Pulse Secure®

WHITE PAPER

Enhancing Application Delivery and Load Balancing on Amazon Web Services with Pulse Secure Virtual Traffic Manager

ALREADY USING AMAZON ELASTIC LOAD BALANCER?

As an abstracted service, Amazon ELB functions well as a basic web service load balancer. But the demands of many modern global businesses require the greater sophistication that only an application delivery controller can offer.

Pulse Secure Virtual Traffic Manager is designed to seamlessly integrate with any application deployed on Amazon Web Services to provide load balancing, user experience optimization, application scalability, and fine-grained application control.

Pulse Secure Virtual Traffic Manager nicely complements Amazon ELB for creating highly reliable cloud deployments.

Cross-Region Load Balancing

Amazon Elastic Load Balancer (Amazon ELB) can balance loads across instances in one or more Availability Zones (AZs) within a single region. The advanced application delivery capabilities available with Pulse Secure Virtual Traffic Manager permit load balancing across more than one region, including AWS GovCloud. Two approaches are possible: managing dispersed Traffic Manager instances as a single cluster, and geographically-aware global server load balancing. Additionally, Traffic Manager cooperates with Amazon EC2 Auto Scaling and can dynamically add or remove application instances as load varies within and across AZs and Regions.

Multi-Site Cluster Management

Within a multi-site configuration (see figure 1), a group of Traffic Manager instances deployed over one or more AZs or regions forms a centrally managed cluster to provide the delivery of application services in a fault-tolerant manner.

Pulse Secure Virtual Traffic Manager service configuration is replicated between all instances in the cluster. As with non-multi-site environments, the majority of the configuration data is shared. However, multi-site management extends this facility by making it possible to set AZ- and region-specific configuration that is active only on the instances marked at that AZ or region. This provides a form of sub-clustering, or local service delivery. The Traffic Manager instances in each region can be configured to run in active/active mode, minimizing potential outages that might result from single points of failure.

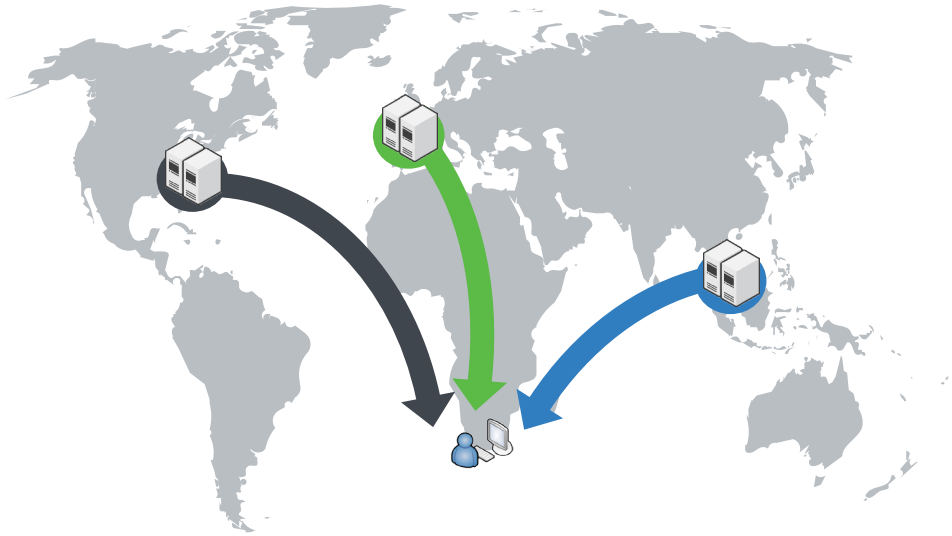

Figure 1: Pulse Secure Virtual Traffic Manager performs local and global load balancing across AWS regions while providing advanced clustering features for simplified management.

Multisite cluster management is commonly used when the services made available in one region are duplicated in a second region. Regardless of their locations, the Traffic Manager instances are managed as a single virtual cluster. The management interface places the instances on an interactive world map, which allows administrators to see a broad overview of resource deployment and also narrow the view to specific Traffic Manager instances, application servers, and users requesting services.

Global Server Load Balancing

Pulse Secure Virtual Traffic Manager can be configured to provide powerful and easy-to-use global server load balancing (GSLB) and failover functionality across multiple regions (see figure 2). It is easy to deploy and gives rich feedback on site performance and traffic distribution.

The primary purpose of GSLB is *business continuity*—to ensure that services are always available, even when one or more service locations become unavailable. A second purpose of GSLB is to *improve customer experience*—to route each user to the best location from a choice of several that are distributed across the globe.

GSLB manages how clients connect to a particular geographical location when a service is hosted in multiple regions:

- **Active-passive**. One location is nominated the active one for each service. The other locations are idle for that service. If the active location becomes unavailable, one of the passive locations becomes active and all clients are directed to it.
- **Active-active**. All locations are used and clients are load-balanced between them based on location performance and proximity.

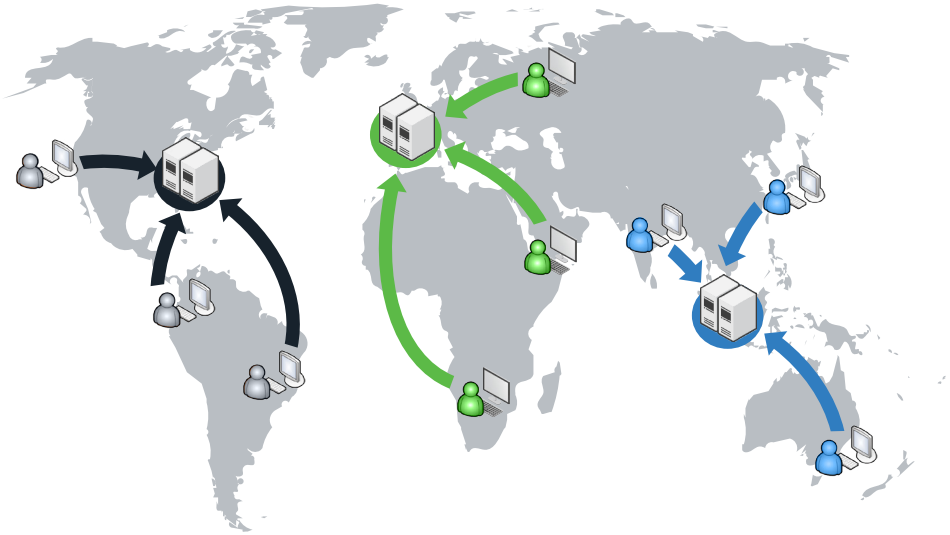

Figure 2: Global server load balancing improves user experience with geographic awareness of services and users.

To ensure that each client receives the best possible level of service, factors that influence the decision are completely customizable, and include location performance, client proximity, and resource demand and availability. Each instance of the service is available from a different IP address. So that every client can use the same fully-qualified domain name to access the service, DNS servers are configured to return the IP addresses of all of the locations hosting a service. Pulse Secure vADC acts as a DNS proxy: it rewrites the round-robin DNS response to ensure that a client is directed to the most appropriate location. Pulse Secure vADC monitors the performance and availability of each location to inform the load-balancing decisions it makes.

HTTP Connection Multiplexing

Pulse Secure Virtual Traffic Manager manages client-side and server-side connections independently, re-using keepalive connections on the server side whenever possible to reduce the number of established and new TCP connections to the server. Connection multiplexing (see figure 3) pools a large number of incoming connection requests from individual clients into a small number of connections to the application servers. Multiplexing also keeps the connections to the servers open for longer periods. These techniques minimize the number of concurrent connections the servers need to handle, and brings big performance and capacity gains.

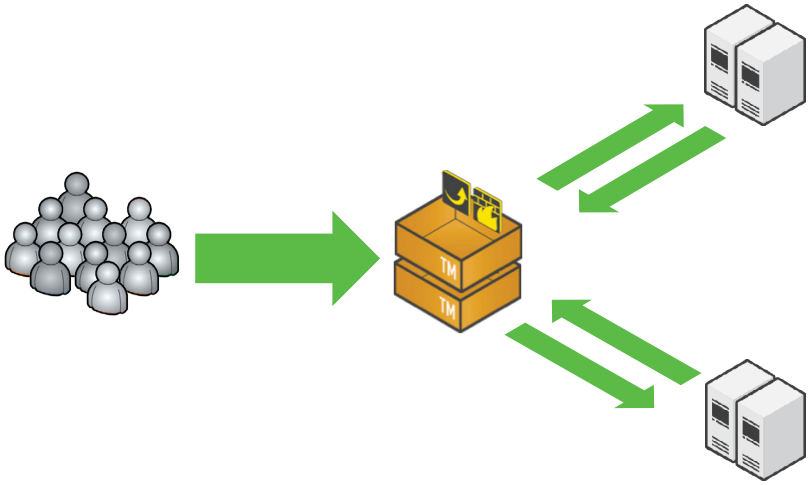

Figure 3: Connection multiplexing consolidates many incoming client connections to reduce load on application instances.

By reducing the overhead required for managing individual client connections Traffic Manager enables servers to respond more rapidly to a higher volume of incoming requests. An existing application server instance immediately benefits from the ability to handle a much greater number of simultaneous incoming requests than before.

Unlike some other ADCs, Pulse Secure Virtual Traffic Manager implements this feature transparently. No special configuration or scripting rules are required to benefit from the performance gains that come from connection multiplexing. Developers can concentrate their efforts on rules that improve application delivery while allowing the ADC to automatically optimize performance.

Client-Side Performance Management

Bandwidth Management

Pulse Secure Virtual Traffic Manager can limit the bandwidth used by inbound or outbound traffic. Normally, network bandwidth is provided at the highest rate possible for all connections. This may result in uneven use of an application, possibly with too much bandwidth used by secondary services at the expense of more critical services. Bandwidth management controls this imbalance explicitly. For example, a 20 Mbit/s network connection that is over-utilized by anonymous browsers would negatively affect the responsiveness of logged in users. To ensure better performance for logged in users, a rule can limit outbound bandwidth to

2 Mbit/s for all connections that lack an authentication cookie. Bandwidth limits are automatically shared and enforced across all Traffic Manager instances in a cluster. Depending on the load on each instance, individual instances take different proportions of the total limit and unused bandwidth is equitably allocated across the cluster depending on the need of each instance.

Request Rate Shaping

Individual users may dominate the use of a service to the detriment of other users. A back-end application infrastructure with limited scalability can be easily overwhelmed when too many requests are given to it. Request rate shaping restricts the rate at which certain activities can occur. Per-second and per-minute limits can be specified on a wide range of events, with very fine-grained control over how events are identified. Some examples include:

- **Rate-shaping** individual Web spiders, stopping them from overwhelming a Web site. Each Web spider, from each remote IP address, can be given maximum request rates.
- **Throttling** individual connections, or groups of connections from the same client, so that each connection is limited to a maximum number of transactions per second.
- **Globally rate-shaping** the numbers of connections per second that are forwarded to an application.
- **Restricting** each user to a limited number of login attempts per minute to thwart dictionary-based login attacks.

Active-Active Load Balancing

By providing active-active failover support for advanced load balancing on Amazon Web Services, Pulse Secure Virtual Traffic Manager gives 100 percent increase in ROI, because all ADCs can now actively serve traffic without having un-used ADC instances in an active-passive configuration.

Active-active gives enterprises massive scalability and reliability at the traffic-management layer with the ability to cluster multiple active members and support multiple traffic IPs in multiple traffic IP groups, which results in all traffic managers actively processing traffic.

Integrated Application Security

Amazon EC2 instances are protected by basic security groups, which define the permitted sources of inbound traffic and permitted destination ports. Both Pulse Secure Virtual Traffic Manager and Pulse Secure Virtual Web Application Firewall enhance the protection offered by security groups, giving more protection from attack.

Because the Traffic Manager inspects all incoming and outgoing traffic at the application level, TrafficScript rules could be used to minimize data leakage. For example, numeric patterns that may appear to be credit card numbers, social insurance numbers, or other personally identifying information can be removed from server responses. Such approaches are useful if it isn't possible to modify applications to eliminate unwanted behavior. The Traffic Manager also protects applications from basic misuse and attacks such as denials of service (DoS) and distributed denials of service (DDoS).

A web application firewall provides more sophisticated protection by detecting and blocking attacks at the application layer. Using baseline rulesets from OWASP and other sources, the Web Application Firewall scans all traffic for known threats and unknown anomalies. The rulesets receive frequent updates as attackers continually evolve their techniques.

Distributed Deployment for Scalability and Performance

To minimize potential performance impacts, the Web Application Firewall can be deployed in a distributed configuration. Enforcer modules, running on application servers, automatically pass allowed traffic. Suspect traffic is sent to a farm of Decider nodes for validation. If the traffic is valid, Enforcer nodes will subsequently permit the traffic. If the traffic is determined to be malicious, Enforcer nodes will block it. Rules can be run in "shadow" mode, for testing and modifying rules using production traffic without affecting availability. Once the rules are operating as expected, they can be fully enabled. The Pulse Secure Virtual Web Application Firewall helps organizations meet PCI-DSS 6.6 security requirements.

Additional Capabilities

In addition to the capabilities described above, Pulse Secure Virtual Traffic Manager on Amazon Web Services provides traffic management features that are not available with traditional load balancers, including:

- Automatic application auto scaling for backend servers
- Advanced session persistence of non-HTTP/HTTPS applications
- Client IP address transparency
- Application availability and latency monitoring
- Content-based routing decisions
- Incoming and outgoing request/response traffic manipulation
- Balancing of TCP traffic on any port
- Instance draining (to prepare it for clean removal)
- Web content caching and Web traffic compression (when necessary)

Combined with the robust and reliable Amazon Web Services cloud computing infrastructure, Pulse Secure Virtual Traffic Manager allows organizations of all sizes to quickly attain global reach and scale.

To learn more, please visit www.pulsesecure.net/vadc

Corporate and Sales Headquarters
Pulse Secure LLC
2700 Zanker Rd. Suite 200
San Jose, CA 95134
www.pulsesecure.net

Pulse Secure®