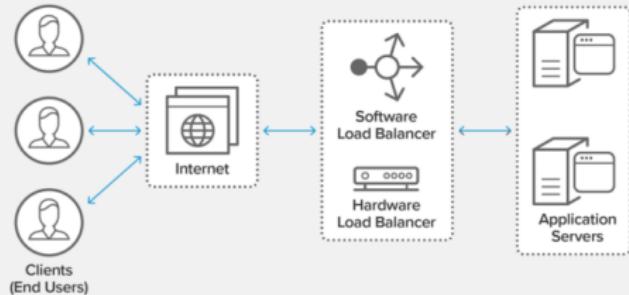


Load Balancer



Load Balancers

- ▶ Load balancers are the bridge between the servers and the network.
- ▶ Load balancers understand many higher-layer protocols, so they can communicate with servers intelligently.
- ▶ Load balancers also understand network protocols, so they can integrate with networks effectively.

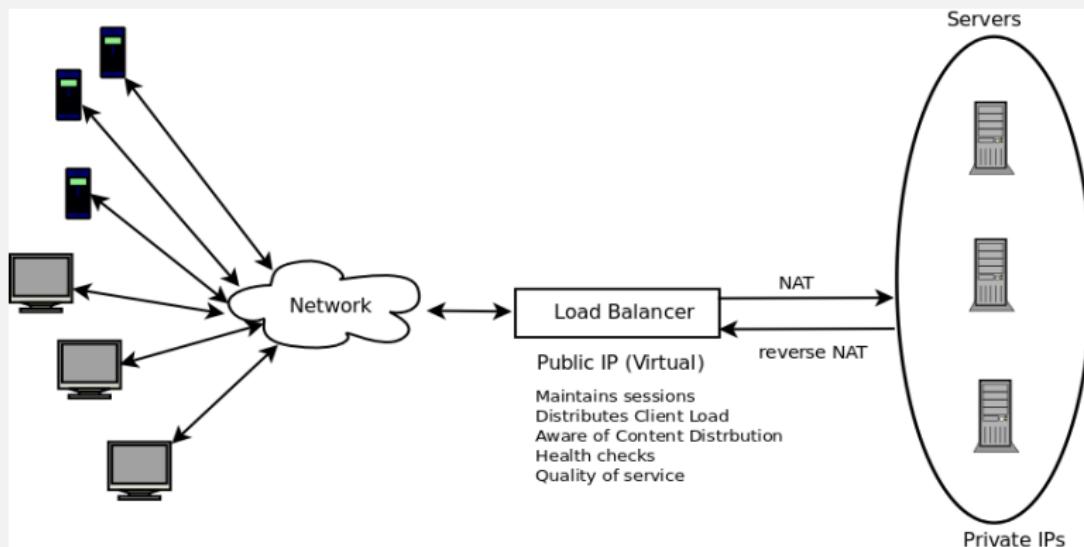


Load Balancer Functionality

- ▶ What does a load balancer do?
 - ▶ Distributes client requests efficiently across multiple servers
 - ▶ Ensures high availability and reliability by sending requests only to servers that are online
 - ▶ Provides the flexibility to add or remove servers as demand dictates
 - ▶ Improves security by protecting against denial-of-service attacks
- ▶ Types of load balancers
 - ▶ Software load-balancers
 - ▶ Hardware load-balancers
 - ▶ Switches with extended functionality



Load Balancers: The Big Picture



Load Distribution Methods (1)

- ▶ **Stateless load balancing:** The load balancer uses some algorithm to distribute all incoming traffic to available servers but does not keep track of any individual session.
 - ▶ Simple hashing based on source IP (we can also include source port for better distribution). Or we can use Hash Buckets for a two-tier distribution method, which is better if a server goes down as only those packets get redistributed to other servers.
- ▶ **Stateful load balancing:** The load balancer keeps track of state information for every session and makes load balancing decisions for each session.
 - ▶ A **session** is identified by the (**source IP**, **destination IP**, **source port**, **destination port**). Easier to identify for TCP than for UDP (why?)
 - ▶ Keeps a session table. We also need an idle timer to remove entries so the table doesn't fill up.

Load Distribution Methods (2)

- ▶ Round robin
- ▶ Least connections
- ▶ Weighted distribution
- ▶ Response time (in-band monitoring versus out-of-band monitoring)
- ▶ Server probes (that run on servers)
- ▶ Server load thresholds

Scalability Options

- ▶ Direct server return
- ▶ Use two load balancers: either in Active-Standby or in Active-Active configurations. We can also duplicate routers and switches to get even higher bandwidth
- ▶ Global server load balancing.
 - ▶ Use standard DNS that allows multiple addresses for the same host address
 - ▶ Use HTTP Redirect
 - ▶ Make the load balancer be the authoritative DNS server
 - ▶ Make the load balancer be the forward DNS proxy server

References

- ▶ *Load Balancing Servers, Firewalls, and Caches.* Chandra Kopparapu. Wiley.