Ananta: Cloud Scale Load Balancing

Ashutosh Mittal (E-mail: amittal@kth.se)

I. Paper summary

THIS paper introduces *Ananta*, which is a distributed layer-4 load balancer and NAT specifically designed to meet the scale, reliability, tenant isolation and operational requirements of multi-tenant cloud environments. It is developed with the idea to build a scale-out data plane by leveraging existing routing primitives and offloading some heavy packet processing tasks to the host. It is a loosely coupled distributed system comprising three main components Ananta Manager (AM), Multiplexer (Mux) and Host Agent (HA).

The *Multiplexer* (Mux) handles all incoming traffic. It is responsible for receiving traffic for all the configured VIPs from the router and forwarding it to appropriate DIPs. Each instance of Ananta has one or more sets of Muxes called Mux Pool.The *Host Agent* (HA) is present on the host partition of every physical machine that is served by Ananta. It is the key to achieving DSR and SNAT across layer-2 domains. It also enables data plane scale by implementing Fastpath and NAT; and control plane scale by implementing VM health monitoring. The *Ananta Manager* (AM) implements the control plane of Ananta. It exposes an API to configure VIPs for load balancing and SNAT. Based on the VIP Configuration, it configures the Host Agents and Mux Pools and monitors for any changes in DIP health. It is also responsible for health monitoring of Muxes and Hosts. AM achieves high availability using the Paxos distributed consensus protocol.

Ananta has been rigorously tested with wide scale deployment. Over a 100 instances of Ananta had been deployed in the Windows Azure public cloud, serving over 100,000 VIPs when the paper was written. With this implementation experience, many improvements and optimizations had been put in place as well.

II. Significant Contributions

- A key element of Ananta's design is the ability to offload multiplexer functionality down to the host. This design enables greater than 80% of the load balanced traffic to bypass the load balancer and go direct, thereby eliminating throughput bottleneck and reducing latency. This division of data plane scales naturally with the size of the network and introduces minimal bottlenecks along the path.
- Traditional systems provide 1 + 1 redundancy, which can leave the system with no redundancy during times of repair and upgrade. On the other hand Ananta cloud environment provides N+1 redundancy to meet the high uptime requirements. Each instance of Ananta runs multiple replicas that are placed to avoid correlated failures. Three replicas need to be available at any given time to make forward progress. The AM uses Paxos to elect a primary, which is responsible for performing all configuration and state management tasks.
- Ananta can scale upto the 100s of terabit of intra-DC traffic. For this, it uses a technique called Fastpath. The load balancer makes its decision about which DIP a new connection should go to when the first packet of that connection arrives. This information is then sent to the HAs on the source and destination machines so that they can communicate directly. This enables communication at full capacity supported by the underlying network. This change is transparent to both the source and destination VMs.

III. Unresolved Issues

- Ananta Muxes were loaded with moderate to heavy baseline load and a SYN-flood attack using spoofed source IP addresses on one of the VIPs was launched. It took quite long for Ananta to detect an attack as it got hard to distinguish between legitimate and attack traffic. It needs improvement in DoS detection algorithms to overcome this limitation. However, it could detect and isolate an abusive VIP within 120 seconds when it is running under no load.
- There is security concern associated with Fastpath a rogue host could send a redirect message impersonating the Mux and hi-jack traffic. If IP spoofing cannot be prevented, a more dynamic security protocol such as IPSEC need to be deployed. However, Ananta does not evaluate the effect of using security protocols on the latency of the fastpath.
- Upgrading Ananta is a complex process that takes place in three phases in order to maintain backwards compatibility between various components. First, instances of the Ananta Manager updated one at a time. During this phase, AM also adapts its persistent state from previous schema version to the new version. Second, the Muxes are upgraded and third, the Host Agents. During testing phase, false positives of availability drops were shown due to this complex process. More work is required to make it error free.