Project Proposal

By Noah Picarelli-Kombert and Brian Gregg

Envoy is a collection of microservices designed to simplify large-scale SaaS systems. One such use case is active health checking on L3/L4 load balancer implementations. However, these checks are limited to responses indicating only ‘healthy’ or ‘unhealthy’ hosts and do not take into consideration the specific state of the cluster in question. This may become an issue as clusters in varying states of degraded health may be treated the same by the load balancer, even as one may be more capable of accepting traffic. This simplistic approach to health checking may reduce the efficacy of the load balancer.

We propose an extension to L3/L4 health checking that will expand the range of response types that the load balancer can accept. With this modified approach, a host could communicate a range of values for given health metrics including CPU usage and disk utilization. A load balancer could then avoid routing traffic in ways that could result in data loss. To start, we aim to implement an extension making use of disk utilization as a metric. Should time permit, we plan to include more metrics in our extension.

To make use of this new health check data, we will also modify the L3/L4 load balancer to be able to route traffic based on the relative degrees of health indicated by the host. For example, if a host indicates 5% disk space remaining, then the load balancer may temporarily route traffic away to avoid potential data loss. This is in contrast to the current functionality of the load balancer and health checks, wherein that same host might still be considered “healthy” and would receive that problematic traffic.