

Layer-4 Load Balancer on Programmable Data Plane using IP Hash and Weighted Round Robin

Muchlis Ramadhan Usman, Muhammad Arief Nugroho, Ibnu Asror,

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹muchlisramadhan@students.telkomuniversity.ac.id,

²arif.nugroho@telkomuniversity.ac.id,

³iasror@telkomuniversity.ac.id

Abstrak

Layer-4 Load balancer digunakan di data center untuk mendistribusikan permintaan yang datang dari klien ke beberapa server yang terletak di data center. Ada dua persyaratan untuk membangun penyeimbangan beban Layer-4 yang layak. Yang pertama adalah memastikan keseragaman distribusi permintaan di beberapa server dan afinitas koneksi. IP Hash adalah algoritma load balancing yang umum digunakan untuk mengimplementasikan penyeimbangan beban Layer-4 tetapi dapat menyebabkan ketidakseimbangan beban di beberapa server. Weighted Round-Robin diusulkan untuk mencegah ketidakseimbangan beban di beberapa server. Implementasi load balancing menggunakan programmable data plane daripada menggunakan penyeimbangan beban perangkat keras atau perangkat lunak tambahan. Hasilnya menunjukkan bahwa WRR mampu mengurangi ketidakseimbangan beban di beberapa server dengan mencapai distribusi permintaan yang seragam. WRR mencapai throughput 13% lebih tinggi tetapi memberikan response time 2% lebih tinggi daripada IP Hash dan memiliki lebih sedikit packet loss daripada IP Hash saat menangani permintaan HTTP

Kata kunci: Layer-4, Load Balancer, Connection Affinity, IP Hash, Weighted Round Robin, Data Plane

Abstract

Layer-4 load balancer used in data center to distribute requests coming from client to multiple servers located in data center. There are two requirements to build proper Layer-4 load balancing. The first is ensuring the uniformity of request distribution across multiple servers and connection affinity. IP Hash is a common load balancing algorithm that is used to implement Layer-4 load balancing but can cause load imbalances across multiple servers. Weighted Round-Robin is proposed to prevent load imbalance across multiple servers. The implementation of load balancing is using a programmable data plane rather than additional hardware or software load balancing. The result shows that WRR is able to mitigate load imbalance across multiple servers by achieving uniform request distribution. WRR achieves 13% higher throughput but gives 2% higher response times than IP Hash and has less packet loss than IP Hash when handling HTTP requests.

Keywords: keyword Layer-4, Load Balancer, Connection Affinity, IP Hash, Weighted Round Robin, Data Plane
