

ABSTRAK

Pertumbuhan penggunaan layanan berbasis internet yang meningkat dalam beberapa tahun terakhir menuntut sistem server memiliki kemampuan distribusi beban yang optimal. Permasalahan utama yang muncul adalah penurunan performa akibat *overload* pada server. Untuk mengatasi hal ini, penelitian ini melakukan perbandingan terhadap dua algoritma *load balancing*, yaitu *Least Connection* dan *Weighted Round Robin* (WRR), yang diimplementasikan menggunakan HAProxy di atas infrastruktur Google Cloud Platform. Pengujian dilakukan dengan menggunakan apache benchmark, dengan variasi jumlah request 1000 hingga 5000 dan jumlah koneksi bersamaan 20 hingga 100. Hasil pengujian menunjukkan bahwa algoritma *Weighted Round Robin* mampu menurunkan *delay* hingga 0,317 ms, sedangkan algoritma *Least Connection* mampu mempertahankan throughput lebih tinggi sebesar 3181,824 KB/s pada beban maksimum. Selain itu, jitter pada algoritma *Weighted Round Robin* tercatat lebih rendah, mencapai hanya 0,001 ms. Penelitian ini memberikan kontribusi berupa analisis performansi berbasis Quality of Service (QoS), yang bisa menjadi acuan dalam pemilihan metode *load balancing* untuk sistem berbasis google cloud platform.

Kata Kunci: *load balancing, least connection, HAProxy, weighted round robin, Google Cloud Platform*