

ABSTRACT

Increased traffic that enters the server excessively (overload) causes the server to be less than optimal. To overcome these problems, it is necessary to apply a Load balancing technique. Efforts to improve Web server performance need an algorithm in the Load balancing method whose task is to distribute workloads to many servers. This study uses the Weighted round robin (WRR) algorithm because the WRR algorithm has the advantage of being able to consider server loads based on the specifications of the server equipment used, where this is not found in other algorithms such as round robin. To get the optimal performance load sharing from the WRR 1:1:1, WRR 2:1:1, and WRR 3:1:1 load ratio scenarios, it is necessary to measure Response Time and CPU Utilization. The test was carried out 30 times in each test scenario and the average value was taken. By providing a traffic load of 1,000, 2,000 and 3,000 requests using the H2load benchmark. The results showed that the WRR ratio of 2:1:1 was the most optimal, because it was able to distribute the load evenly to the three web servers used, with an average CPU usage of 1,000-3,000 requests for traffic reaching 71%-79% on Server 1, 47%-56% on server 2 and 48%-56% on server 3. This results in an average Response Time of 223.77ms for 1,000 requests, 233.13ms for 2,000 requests and 235.37ms for 3,000 requests.

Keywords: Load balancing, Weighted round robin, web server