

1 Evaluation and Performance Measurement

1.1 Does the latency of the application change with and without Docker containers? Did virtualization add any overheads?

- From fig 4 and fig 7 we can see that while both applications were run with a trade request probability of 0.5, the addition of the docker containers consistently added roughly 20ms in latency for each request. This could be because Docker introduces additional access time in order to access the memory of the system, or because the whole CPU is not allocated for Docker use. Additionally there is additional overhead receiving requests as they must be passed from the client to the host computer to the docker container as opposed to simply from the client to the host computer.

1.2 How does the latency of the lookup requests compare to trade? Since trade requests involve all these microservices, while lookup requests only involve two microservices, does it impact the observed latency?

- From fig 2, and fig 3, we can see that in the Docker containerized system, the trade function introduces an additional 5ms of latency. These results are consistent with the native system, which can be seen in fig 5, and fig 6. These results demonstrate that since trade requests involve all three microservices, they consistently introduce an increase of 5ms to the latency as opposed to the latency of lookup requests only.

1.3 How does the latency change as the number of clients change? Does it change for different types of requests?

- From fig 2 and fig 5, we can see that as the number of clients increase in both the docker and native systems, the latency of lookup requests remains relatively the same, or trivially increases by 4ms maximum. From fig 3 and fig 6, we can see that as the number of clients increase in both the docker and native systems, the latency of trade requests seems to increase relatively drastically. This could be because that trade requests involve all three microservices, at which point an increase of the number of clients may require the catalog and order services to introduce locks more frequently, which would in turn increase the latency.

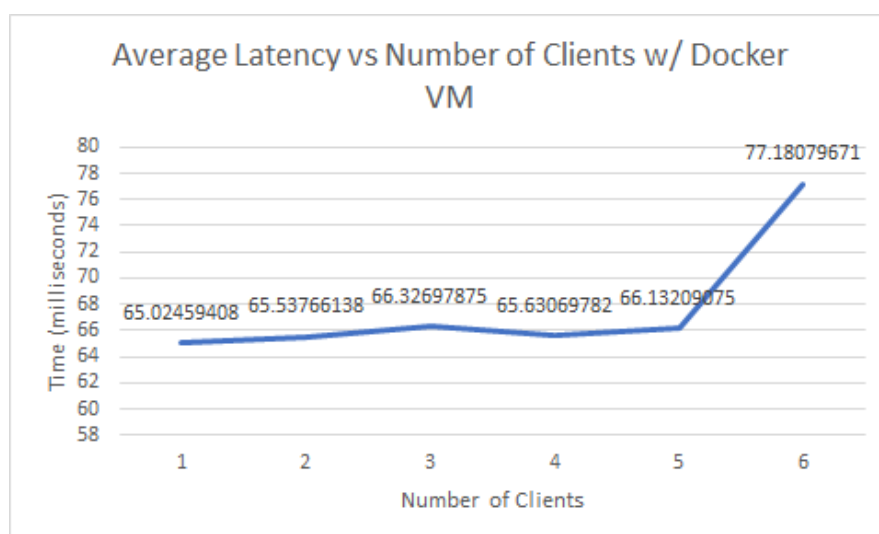


Figure 1: Average Latency of Requests for Docker Server. Trade requests follow lookup requests with a probability of 0.5. The dramatic increase in latency once 6 clients were connected demonstrate the thread per session approach taken with the front-end service. (maxpool size of 5)

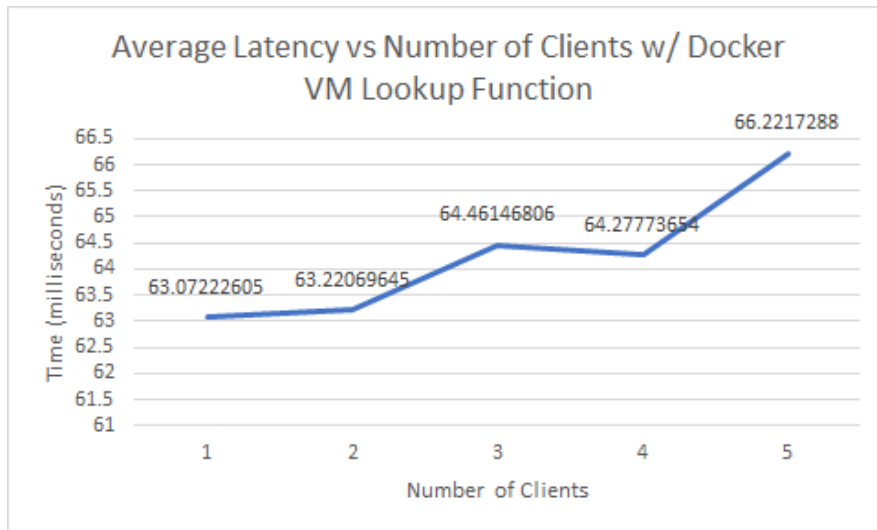


Figure 2: Average Latency of Lookup Requests for Docker Server.

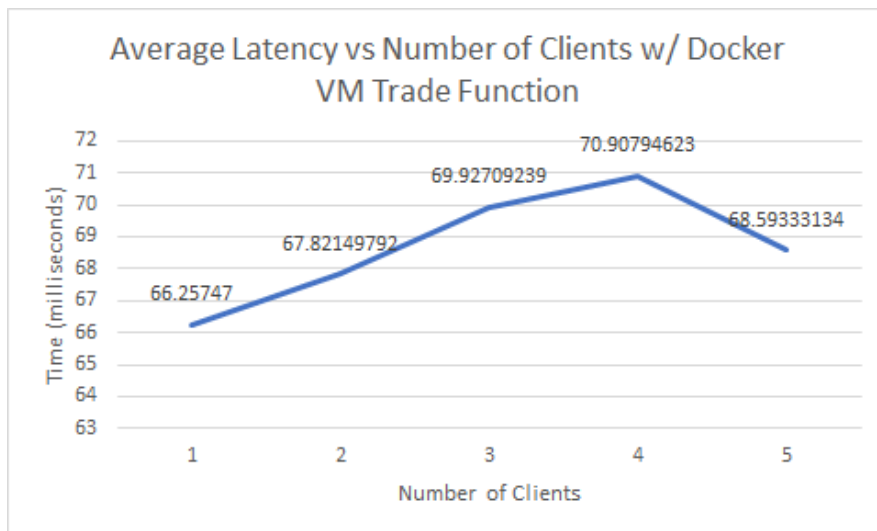


Figure 3: Average Latency of Trade Requests for Docker Server.

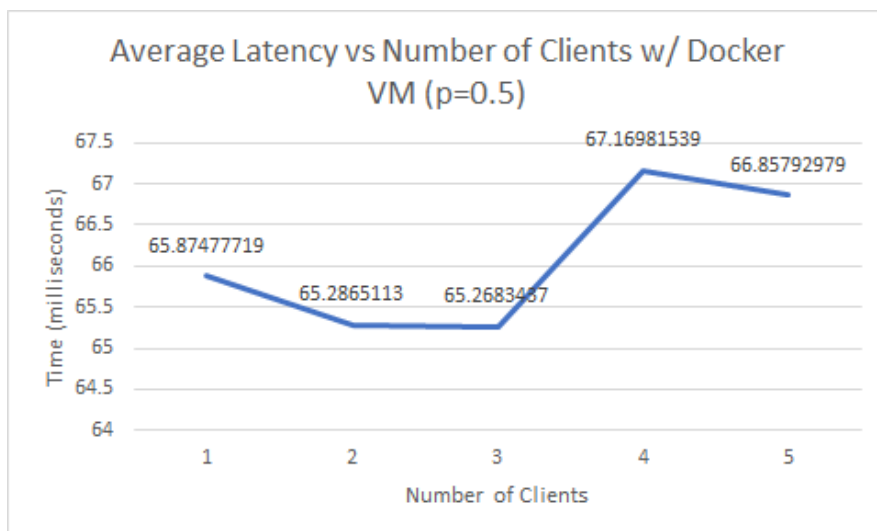


Figure 4: Average Latency of Requests for Docker Server. Trade requests follow lookup requests with a probability of 0.5.

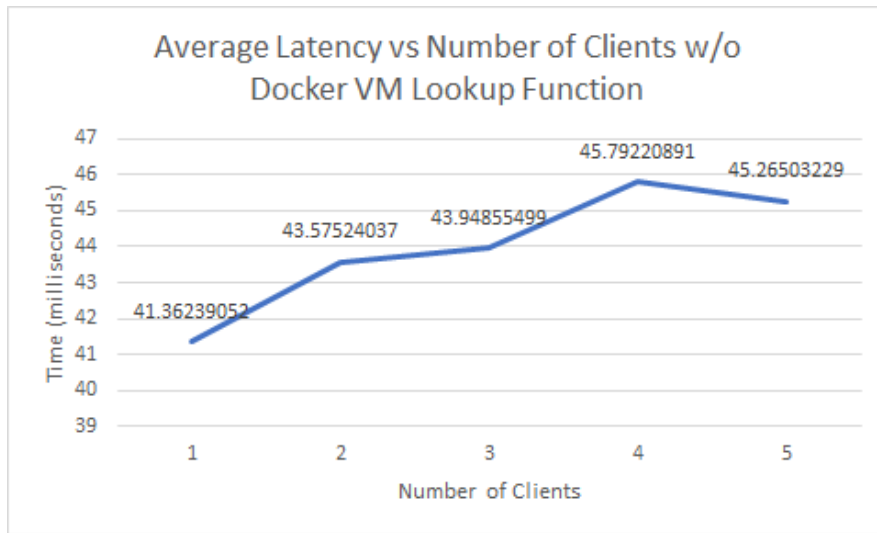


Figure 5: Average Latency of Lookup Requests for Native Server.

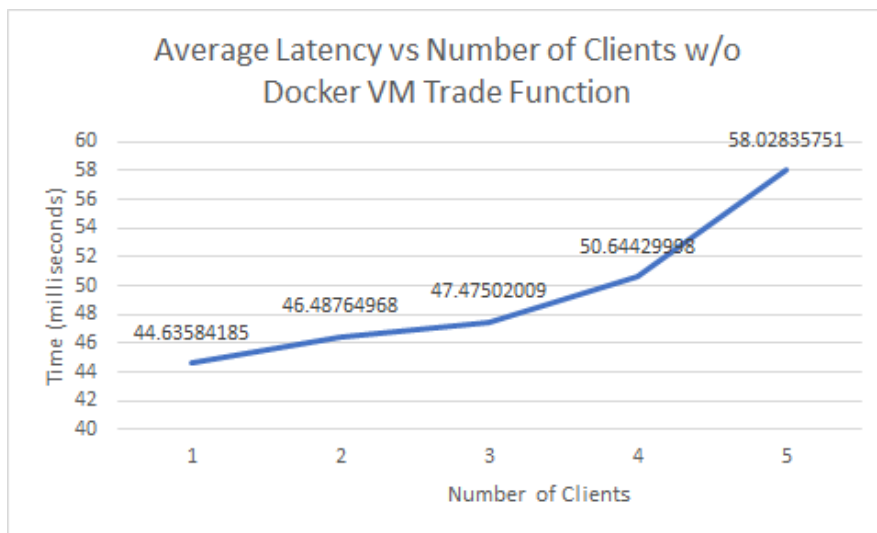


Figure 6: Average Latency of Trade Requests for Native Server.

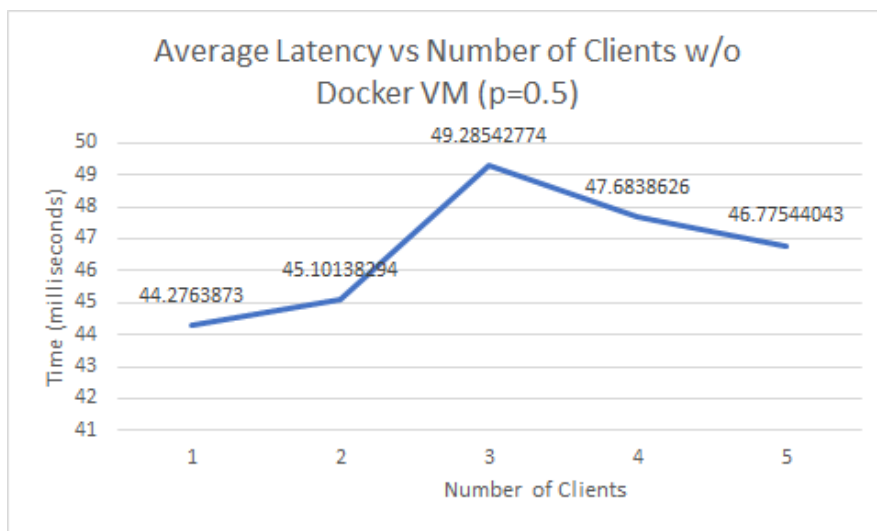


Figure 7: Average Latency of Requests for Native Server. Trade requests follow lookup requests with a probability of 0.5.