

In Distributed Cloud Computing, applications are deployed across many data centres at topologically diverse locations to improved network-related quality of service (QoS). As we focus on interactive applications, we minimize the latency between users and an application by allocating Cloud resources nearby the customers. Allocating resources at all locations will result in the best latency but also in the highest expenses. So we need to find an optimal subset of locations which reduces the latency but also the expenses – the facility location problem (FLP). In addition, we consider resource capacity restrictions, as a resource can only serve a limited amount of users.

An FLP can be globally solved. Additionally, we propose a local, distributed heuristic. This heuristic is running within the network and does not depend on a global component. No distributed, local approximations for the capacitated FLP have been proposed so far due to the complexity of the problem. We compared the heuristic with an optimal solution obtained from a mixed integer program for different network topologies. We investigated the influence of different parameters like overall resource utilization or different latency weights.