netidee STIPENDIEN

Reliability of Edge Offloading

Interim Report | Call 18 | Scholarship ID 6851

Content

1 Introduction

The document is an interim report for the research progress of PhD dissertation "Reliability in Edge Offloading". In the next two sections, main milestones and research plan updates will be presented.

2 Status

2.1 Milestone 1 - <Literature Research and Initial Idea Development>

We identified the research problem and its scope we want to tackle. The problem is offloading on resourceful but unreliable edge servers where failures can postpone or prevent offloading. Failures can cancel the main offloading benefit which enables real-time execution of resource-intensive mobile applications on resource-limited mobile devices. Edge servers can be unreliable due to a lack of supporting systems and hardware redundancy. Additional factors can also contribute to unreliable edge servers and prolong offloading time. Volatile workloads generated by multiple mobile devices can make an estimation of application response time uncertain. Combining it with device mobility, where the device moves between different clusters of edge servers, makes offloading even more dynamic and challenging. We want to devise a solution that can cope with the aforementioned challenges of failure-prone edge servers, device mobility, and volatile workloads. The main goal is to achieve high application performance by offloading applications from resource-limited mobile devices to resourceful edge servers.

There were no deviations from the original plan.

2.2 Milestone 2 - <Modeling>

We developed a simulated environment which consists of infrastructure and workload. The simulated environment should exhibit typical edge characteristics as realistically as possible under which our future solution will be evaluated. To simulate multi-server edge networks and multi-device workloads, we employed a formal framework based on queueing network theory. We find it suitable for our edge offloading scenarios because queueing networking can model real-world processes that exhibit (i) *random* arrival of workload, (ii) waiting times if service is busy due to *limited* capacity, and (iii) service capable of *processing* arrived workload. Edge offloading exhibits the aforementioned characteristics due to (i) *random* workload generation by multiple mobile devices, (ii) *limited* resource capacities, and (iii) edge servers capable of *processing* resource-intensive applications. Each edge server can be represented as a distinct queue and interconnected into a queueing network to represent real edge networks. The main goal of edge queueing network model is not just to simulate real-world edge offloading networks and its workload, but also to estimate accurately application response time under multi-device workload settings.

There were no deviations from the original plan.

2.3 Milestone 3 - <Simulation>

Besides application response time, energy consumption of mobile devices is also a critical objective, especially for battery-powered ones. Conserving battery supplies while offloading should be taken into consideration. Furthermore, offloading on edge servers that are commonly owned by resource providers can impose resource utilization costs. Combining application response time, battery lifetime, and resource utilization cost objectives into a unified optimization problem makes offloading not an easy problem to solve. Therefore, we formally defined offloading as a constraint optimization problem that is composed of objective functions, constraints, and variables. Application response time, battery lifetime, and resource utilization cost are represented as objective functions, constraints correspond to edge resource capacities and application response time deadlines, and variables correspond to edge resource factors such as CPU cores, RAM capacity, etc. To solve a constraint optimization offloading problem, we employed the satisfiability modulo theory (SMT). It is a formal method that can provide a formal guarantee of the feasibility of the solution which is important in real-time environments where timely responses are vital as in edge offloading. The constraint optimization offloading problem is encoded into SMT formulas and is solved by SMT solver which is a tool that automatically solves SMT formulas. The output of solving SMT formulas are offloading decisions that optimize objective functions (application response time, battery lifetime, and resource utilization cost) and respect constraints (resource capacities and application timing deadlines). We integrated constraint optimization offloading problem with SMT solver into edge offloading simulator together with edge queueing model of infrastructure and workload from previous milestone.

There were no deviations from original plan.

3 Summary of Plan Update

In general, there were no deviations from the original plan. For future activities, I added a single milestone change. Milestone 5, originally called “Refining simulation models” with description “Fine-tuned simulation models”, is changed into “Simulation” with description “Reliability simulation model”. It is more precise naming of the ongoing activity then previous sheet entry. Technically, it does not deviate from original plan.