Week 28

In a practical implementation, the functionality attributed to the service management and orchestration can follow a different modularization path.

a) Service Discovery: It is quite possible that end user devices are specialized for a set of functionalities and request different types of services from the network. However, the user and the application that runs on the mobile device do not have any prior knowledge of the available services on the Edge Computing system. For instance, there should be an environment where a client application may generate a service request by specifying the necessary computation power and the required storage area. Then, a broker mechanism discovers the service and servers that are able to satisfy the aforementioned requirements.

SDN environment leverages the similar functionality where the Service Discovery module maintains the necessary information about the existence and location of the service con- tent available on the Edge Computing facilities. The module contains a mapping table which links the service name with the corresponding server locations. This table is updated frequently since each edge server informs the Service Discovery application about the provided services periodically and new service registrations.

b) Service Commissioning & Migration: This module is responsible for commissioning a service onto a Cloudlet. Since each Cloudlet has only a finite computational capacity which is already preoccupied with a set of services, it may not be possible to just initiate a service on the most nearby Cloudlet. This module decides where to commission the service based on various performance factors such as server utilization and network conditions. Also, whenever a service has low utilization, this management module can decide to migrate the virtual machine (VM) that hosts this service to another Cloudlet on the network assuming other conditions are met. At this point, the space required for the deployment of the new service will be provided which contributes to an increase in the overall performance.

c) Performance Tuning & Optimization: Performance of an Edge Computing system as experienced by the users should always be at the maximum attainable level. This requirement comes from the fact that most of the end-users utilize the Edge Computing facilities due to real-time requirements or related performance criteria. Performance Tuning & Optimization module make use of SDN and OpenFlow capabilities to monitor the service utilization levels and flow sizes to manage the load on various edge servers, and hence the overall throughput of the system is maximized.

There are proposals that utilize the functionalities of SDN and OpenFlow to carry out the load balancing procedures among network resources, computation resources, or both of them. In fact, a load balancer should take both aspects into consideration for optimizing the overall performance of the system. The joint optimization problem can be solved through either a single northbound application that is able to assemble the loads on network and computation resources, or separate inter-operable applications.

d) User Handover: Due to mobility, an edge device can leave the coverage of a Cloudlet and enter the proximity of another one. Without any handover management involved, the user starts over with service discovery and service commissioning procedures whenever the active Cloudlet is changed. This would not only cause performance disruptions as experienced by the end-user but also brings in inefficiencies for the operations of the Edge Computing infrastructure. User Handover module can use related techniques to forecast the next coverage domain and can employ various methods to supply service continuity to the user such as flow redirection or instant live migration of the relevant virtualized resource through SDN .