Title: Modelling and Resource Scheduling approaches on Cloud Computing

Authors: Dimitrios Dechouniotis and Symeon Papavassiliou

Abstract: Cloud computing is the dominant service delivery paradigm over the last decade. The rapid development of the Internet of Things (IoT)-based applications shifts the cloud model to more decentralized approaches, such as Edge Computing and Fog Computing. The resource allocation in such distributed environments must guarantee both the Quality of Service (QoS) requirements of each application and the optimal utilization of the underlying computing resources. Thus, the static resource allocation policies are not adequate to fulfill these objectives and lead to over or under-provisioning. On the other hand, control-theoretic approaches guarantee important system properties, such as stability. This presentation will present already proposed modeling and resource scheduling methodologies on cloud computing and the future challenges of this interesting research area.

Introduction: Cloud data centers provide vast amount of computing resources and the resource allocation policy is coarse and the providers usually offer predefined resources to the customers. Most of the proposed studies in literature, the performance modeling of cloud-based services is either empirical or derived by simple static models. In order to satisfy the varying workload demand, the resource allocation mechanism is based on instantiation of many service replicas. The modern applications rely on mobile devices and the involving wireless communication makes the workload profile more complex and time-variant. Furthermore, the QoS requirements of these application include ultra-low time constraints and high data throughput. However, edge data centers provide fewer computing resources than the cloud ones. Thus, the performance modeling and the resource allocation of the modern smart application should answer the following challenges:

- What are the most important performance criteria for infrastructure providers and service consumers?
- Which are the appropriate methodologies for modeling smart applications deployed on cloud/edge infrastructure?
- Which are the appropriate scheduling approaches for guaranteeing important system properties such as stability, reachability?

978-3-907144-02-2 