

**CHENNAI – PONDICHERRY**

**DEADLINE GUARANTEED SERVICE FOR MULTI-TENANT**

**CLOUD STORAGE**

**ABSTRACT**

It is imperative for cloud storage systems to be able to provide deadline guaranteed services according to service level agreements (SLAs) for online services. In spite of many previous works on deadline aware solutions, most of them focus on scheduling work flows or resource reservation in datacenter networks but neglect the server overload problem in cloud storage systems that prevents providing the deadline guaranteed services. In this paper, we introduce a new form of SLAs, which enables each tenant to specify a percentage of its requests it wishes to serve within a specified deadline. We first identify the multiple objectives (i.e., traffic and latency minimization, resource utilization maximization) in developing schemes to satisfy the SLAs. To satisfy the SLAs while achieving the multi-objectives, we propose a Parallel Deadline Guaranteed (PDG) scheme, which schedules data reallocation (through load re-assignment and data replication) using a tree-based bottom-up parallel process. The observation from our model also motivates our deadline strictness clustered data allocation algorithm that maps tenants with the similar SLA strictness into the same server to enhance SLA guarantees. We further enhance PDG in supplying SLA guaranteed services through two algorithms: i) a prioritized data reallocation algorithm that deals with request arrival rate variation, and ii) an adaptive request retransmission algorithm that deals with SLA requirement variation. Our trace-driven experiments on a simulator and Amazon EC2 show the effectiveness of our schemes for guaranteeing the SLAs while achieving the multi-objectives.