**ABSTRACT**

Cloud computing has been an upcoming trend in the field of computer science for the past few years. Cloud computing entails the distribution of centralised services and dynamic features distributed across several clients or end users over a reliable network. The challenges encountered by cloud computing are many in terms of constant resource provisioning, creation and maintenance of virtual machines, need for migration of virtual machines on a priority basis, scalability of CPU resources. The need for migration and scalability arises from a situation in which the respective virtual machine operating under a server which is part of the cloud is overloaded or underloaded.

It is essential if not Imminent to optimize the processes of migration of virtual machines and scalability. One of the approaches that we have undertaken to perform this optimization, is the use of the load balancing algorithm which dynamically tracks the load and optimizes the process of resource allocation for the concerned virtual machine(s), scales the resources allocated in a dynamic manner and prepares the relevant virtual machine(s) for migration to a new server thus providing a smooth functioning environment. The experimental set-up consists of a suitable number of Xen Cloud Platform enabled inter-connected servers, a network file system which acts as a virtualized hard drive and suitable number of client systems. Based on the experiments conducted, it has been observed that the load balancing algorithm improves the response time of virtual machines, by using the features of scalability and migration of virtual machines.