COM301A Written Assignment I

Introduction and Context

Cloud computing is an emerging information technology (IT) paradigm. It enables us to use a network of remote servers (hosted on the Internet) to store, manage and process data, rather than a local server or a personal computer. There are several cloud services: IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service) [1]. The paper talks only about IaaS. It is the most basic category of cloud services, in which we rent servers, virtual machines (VMs) and operating systems from a cloud provider. In IaaS datacenters, in order to place a VM on a physical host, we use migration techniques. Migration allows us to transfer data from a source host, say, A to a destination host, say, B. Host B processes this data and sends it back to host A. There are two migration techniques: Pre-copy and Post-copy. Both the techniques are equally efficient. The existing migration techniques keep the data images of VMs in the host's memory, during migration, in order to reduce the amount of data transfer. However, with the increase in migrations, the size of data images stored on the hosts increases and this can lead to memory starvation. This paper focuses on the efficient memory management of the source host. It proposes a new technique that performs better than the existing techniques.

The Proposed Technique

During VM migration, the existing techniques keep the data image for that VM in the source host based on a probability factor. This factor decides the possibility of updating the data image in future. A low factor indicates a less probability of update and vice versa. The authors exploit this to their use in the following ways:

1. There is no need of data with least or zero probability of update on another host because it cannot be reused. Therefore, we do not store this data.

2. The image need not include the data that has strong probability of update in a near future.

The authors analyze the data using standard probability protocols, keeping in mind above two pre-conditions. Consequently, the approach helps in significantly reducing the data to be sent back to the source host.

Observations

The proposed techniques demonstrates the availability of 33% (approx.) more memory at the source host. This means if a physical machine could host 10 VMs using the existing techniques, it can host around 3 more using the proposed technique.

Conclusion

The proposed technique optimizes the memory requirement at the remote hosts. This means a single physical machine can host a larger number of VMs. Popularity of cloud computing increases day-by-day and so does the demand for the cloud resources. A better utilization of

cloud resources means lesser investment and an increase in the financial gains for the users of cloud computing.

References

[1] What is Cloud Computing? https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/

[2] Link to the Paper: http://www.sciencedirect.com/science/article/pii/S187705091501902X