Data Replication in Cloud Computing Data Centers With Performance Guarantee

David Ocejo
Email: docejo1@toromail.csudh.edu

Advising Professors: **Dr. Bin Tang**
Dr. Mohsen Beheshti

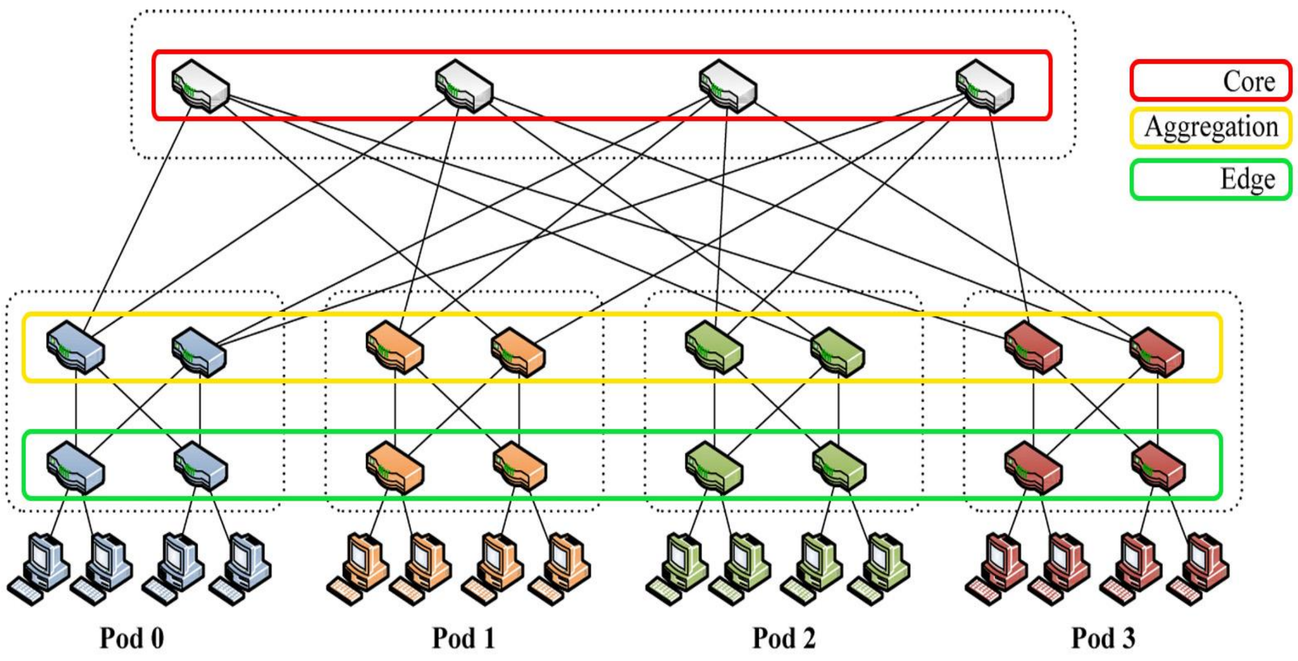
Computer Science Department

California State University DOMINGUEZ HILLS

Abstract

Energy consumption in data centers networks is a growing concern as the size and scope of data center networks increases. Higher energy consumption is concerning for both monetarily and environmentally costs as most of the world's electricity is still generated through the use of fossil fuels which have detrimental greenhouse effects. Efficient data replication across data center networks can reduce network traffic and allow built-in power saving features to take effect, reducing energy consumption. We design a data replication algorithm with performance guarantee. Using FTCloudSim, a cloud computing simulator for fat-tree data center networks, we'll show that our data replication algorithm performs better than other existing heuristic algorithms for data replication in cloud computing. The findings will show that through the designed data replication algorithm with performance guarantee, energy consumption can be decreased, lowering monetary costs and even improving data center performance by reducing network response times, decreasing bandwidth usage and intelligently managing data replicas.

Fat-Tree Data Center Topology



K-ary Fat-Tree: Three-Layer Topology

- Each pod consists of $(k/2)^2$ servers and 2 layers of k/2 k-port switches
- Each edge switch connects to k/2 servers and k/2 aggregation switches
- Each aggregation switch connects to k/2 edge and k/2 core switches
- $(k/2)^2$ core switches: each connects to k pods

Virtual Machine (VM) Replication

In the fat-tree topology based data center, initially, VMs are randomly located in different Physical Machines (PMs). There are a sequence of jobs that are submitted, each requesting some of the VMs. The objective of the problem is to replicate the VMs onto PMs, in order to satisfy the requests of the jobs, while incurring minimum energy consumption in the VM replication process.

Algorithm: Our centralized VM replication algorithm is a greedy algorithm, taking place in rounds. In each round, we choose a VM to replicate on a PM such that the reduction of the energy consumption is the most in that round. The algorithm stops when all the storage spaces of all the PM are occupied.

Future Work

- Implement GUI for FTCloudSim simulation
- Implement modular Data Center topology generation
- Implement FTCloudSim data management protocols

Data Center Costs

Data distribution in data center networks can be seen to have tangible effects on energy costs. Fig. 1 [1] shows the lower energy consumption resulting from data more closely located to client nodes as well as the increasing costs of communication with larger data sizes. Fig. 2 displays the rising VM migration costs as data center network size increases (more servers across a larger network).

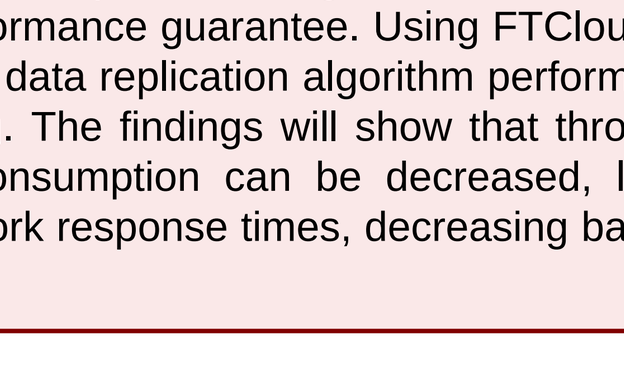


Fig. 1: Energy Consumption of Servers

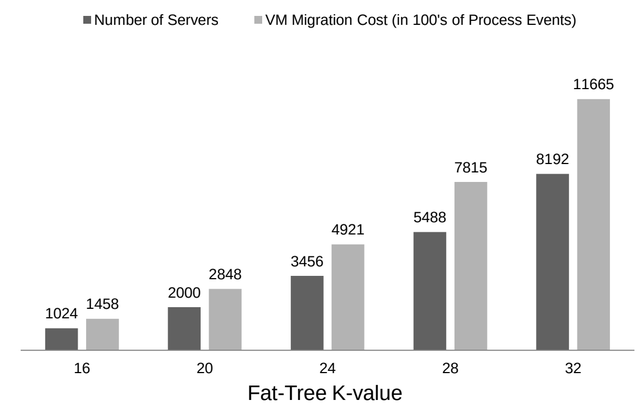


Fig. 2: Data Migration Cost

References

[1] Boru, Dejene., Kliazovich, Dimitry. “Energy-Efficient Data Replication in Cloud Computing Data Centers” Globecom 2013 Workshop

[2] Wang, Shangguang., Zhou, Ao. “FTCloudSim: A Simulation Tool for Cloud Service Reliability Enhancement Mechanisms” Beijing University of Posts and Telecommunications

[3] Dharma Teja Nukarapu, Bin Tang, Liqiang Wang, and Shiyong Lu. “Data Replication in Data Intensive Scientific Applications With Performance Guarantee” IEEE Transactions on Parallel and Distributed Systems, Volume 22, Number 8, 1299 - 1306, August 2011.

[4] Calheiros, Rodrigo N., Ranjan, Rajiv., Beloglazov, Anton., De Rose, Cesar., Buyya, Rajkumar. “CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms” Wiley Online Library. 24 Aug 2010.

Presentation of this work has been supported in part by NSF Grant CNS-1042341.