WIRELESS TELECOMMUNICATIONS SYMPOSIUM

Dynamically Adapt Nodes of Data Center by Popularity and Predication

Kuo-chi Fang
Husnu Saner Narman
Ibrahim Hussein Mwinyi
Wook-Sung Yoo

Department of Computer Science
Marshall University, USA

narman@marshall.edu
http://mupfc.marshall.edu/~narman/

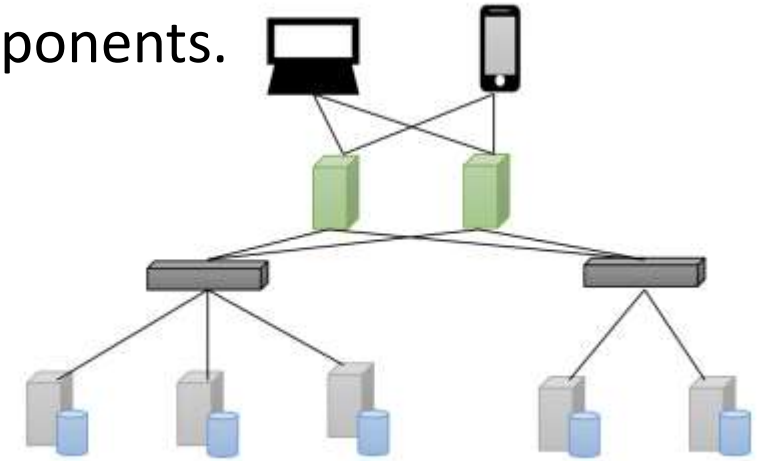
April 2018

Outline

- Introduction
- Proposed System
- Experiment Design and Simulation
- Results
- Conclusion

Introduction

- What is Datacenter?
  - A **data center** is a facility used to house computer systems and associated components.
  - Contains:
    - Server
    - Storage
    - Network (Wire, wireless)



Usage

- What is Datacenter used for?
  - Media searching
  - Data analysis
  - Web application
  - Large-scale network service

Datacenter Example

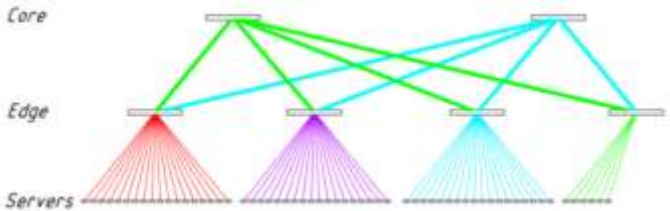
- Facebook
  - 1.13 billion daily active users (September 2016)
  - At least 60,000 servers in its data centers (June 2010)

From : http://www.datacenterknowledge.com/the-facebook-data-center-faq/

Challenges

- Challenges of Datacenter Management
  - Availability and durability
  - Response time
  - Power cost

Previous Work Strategies



- Current solutions
  - Network topologies, such as fat-tree
  - Load Balancing
  - Replication (Replication is also used for Failure Recovery) based on popularity

Problem

- Problem with current solution
  - Network bottleneck (partially solved by replication)
  - Possible Future data popularity unknown

Objective

- To increase data availability in multimedia data centers
- Consider not only the current data popularity but also possibilities for future data popularity
- Consider bandwidth limitation and load balancing.

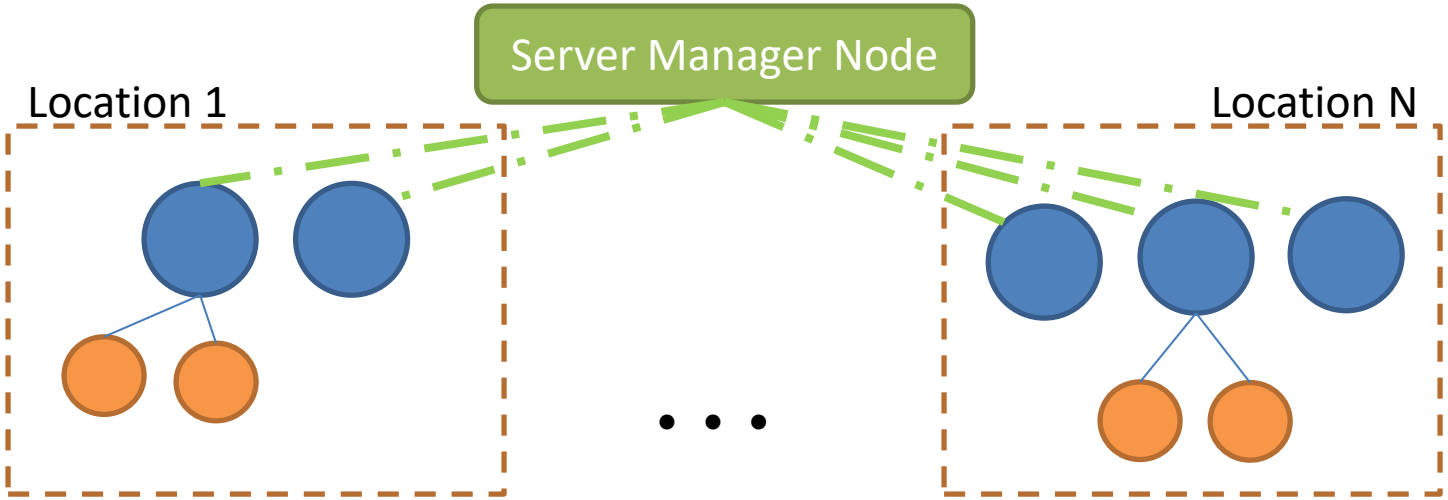
Proposed System

- Management Algorithm
  - Calculate usage of each Node
  - Classify Data based on Nodes
  - Based on the status of data (hot, warm and cold [2])
  - Dynamically active or inactive server node
  - Compress or non-compress replication

[2] Asaf Cidon et al. “Tiered Replication: A Cost-effective Alternative to Full Cluster Geo-replication” *Usenix Annual Technical Conference*. 2015.

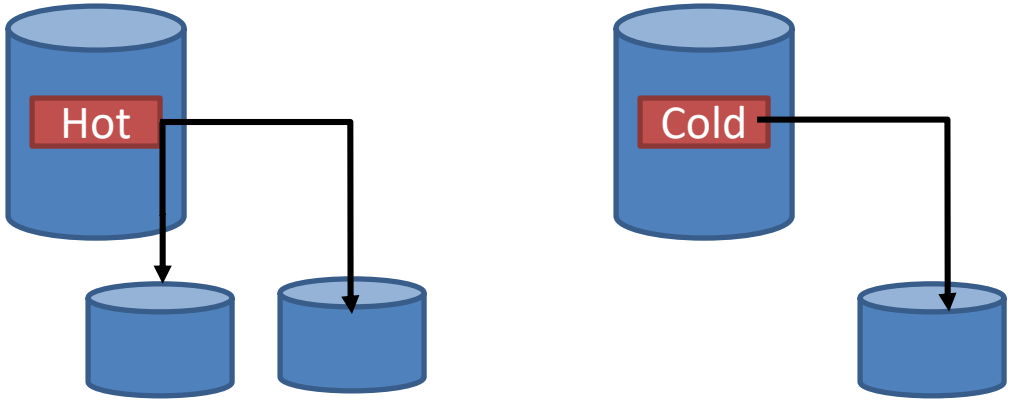
System Overview

- Datacenter Model



Popularity

- Replication for Scalability and Durability [2]
  - System will replica data of node according the rank of data.



Hot data Levels
Cold data Levels

[2] Asaf Cidon et al. “Tiered Replication: A Cost-effective Alternative to Full Cluster Geo-replication” *Usenix Annual Technical Conference*. 2015.

Prediction Metrics

- Ranking method of prediction

$$R = \alpha DA + \beta IA + \left(\gamma \frac{AD}{10} + \eta \frac{DL}{10}\right) + \kappa TU + \varsigma AR$$

Parameters	
DA	The direct access amount of the data object
IA	Indirect access amount of the data object
AD	Application or user diversities of accesses to the data object
DL	Location diversities of accesses to the data object
TU	Duration of the data object existence
AR	The access rates of the related data objects

Proposed System (Cont.)

- Ranking Technique

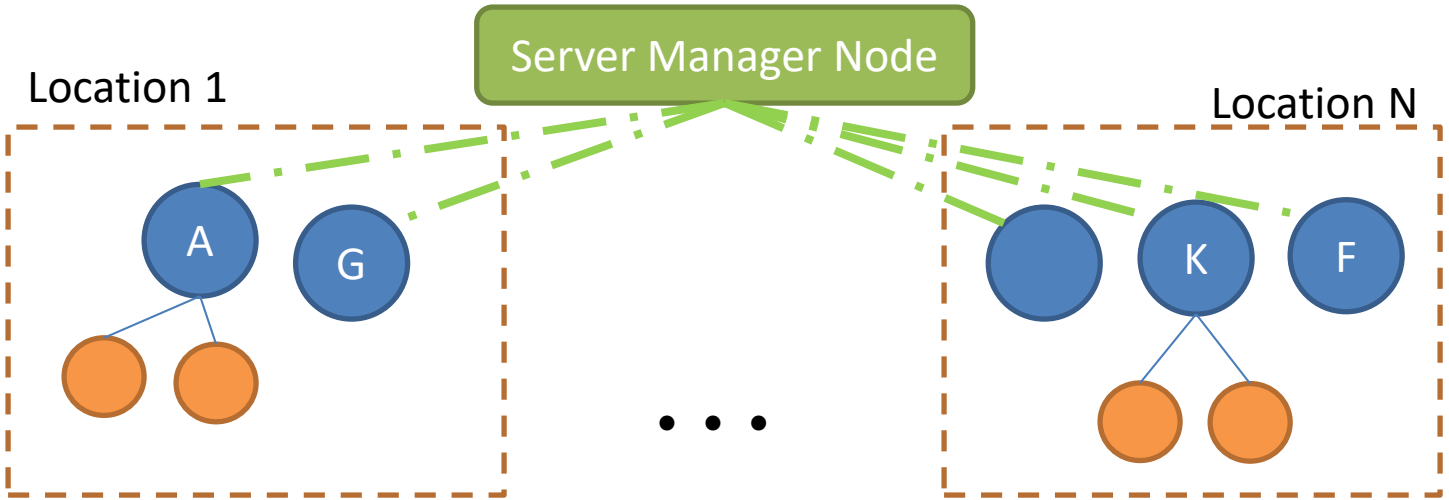
Popularity	
Data Object	Frequency
A	6
G	5
K	1
F	0

Prediction	
Data Object	Frequency
K	9
G	2
A	0
F	0

Ranking table		
Data Object	Frequency	Level
K	10	Hot
G	7	Warm
A	6	Warm
F	0	Cold

Example

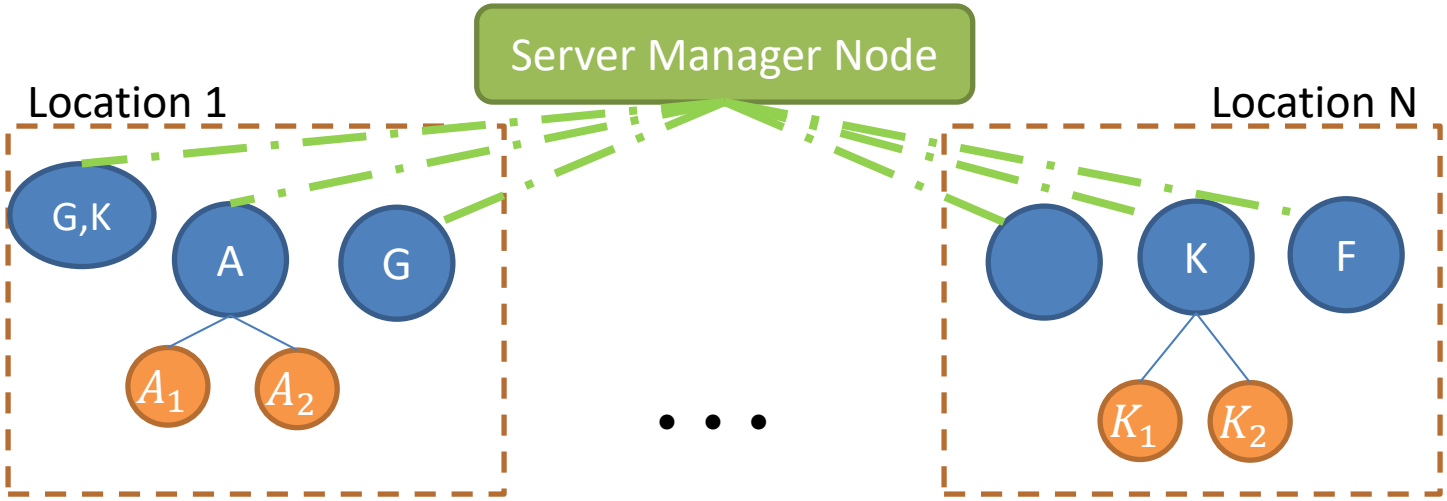
- Before prediction: Data Model



inspired by [1]Ito, Cloud storage with key-value stores over content-centric networking architecture

Example (cont.)

- After prediction: Data Model (manage nodes based on location)



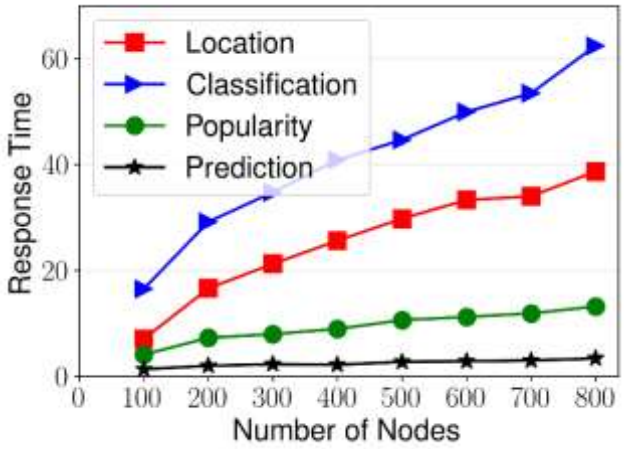
inspired by [1]Ito, Cloud storage with key-value stores over content-centric networking architecture

Experiment Design and Simulation

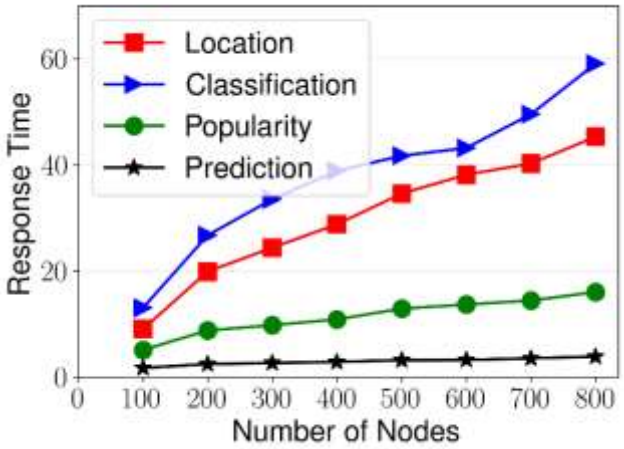
- Input
  - Searching tags with user location information
  - Prediction after finding the
- Assumption
  - Homogeneous system
  - Media files with tag information
- Initial System Setup
  - Allocated data to different nodes based on location

Simulation Result

- Case 1 (Numbers of nodes with randomly requesting 200 times 50/100 different data objects from servers.)



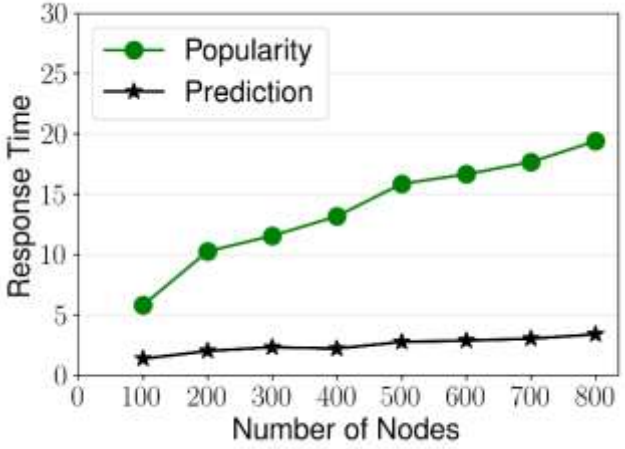
Data objects : 50



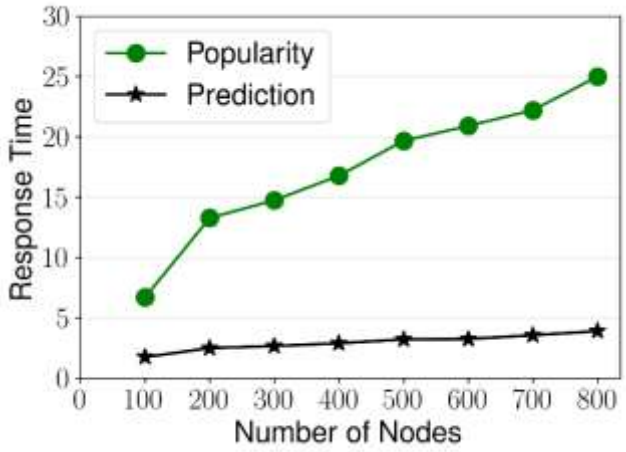
Data objects : 100

Simulation Result(cont.)

- Case 2 (Numbers of nodes with requesting 200 times 50/100 different data objects from servers based on popularity/prediction.)



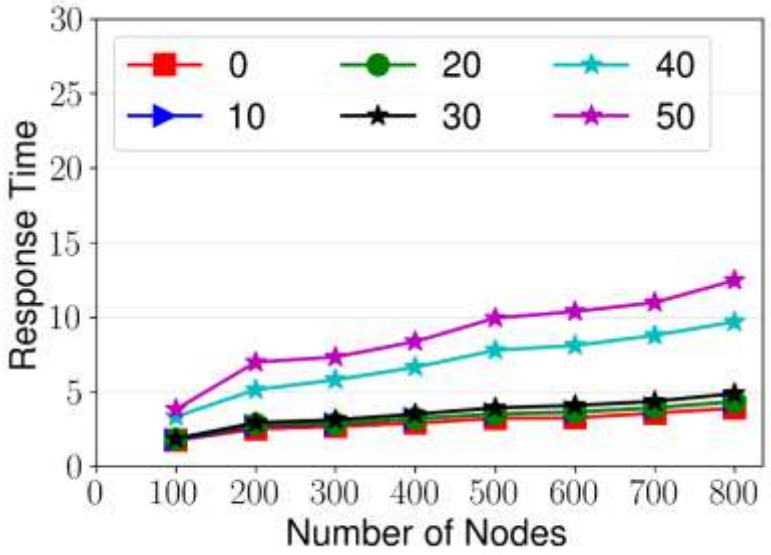
Data objects : 50



Data objects : 100

Simulation Result(cont.)

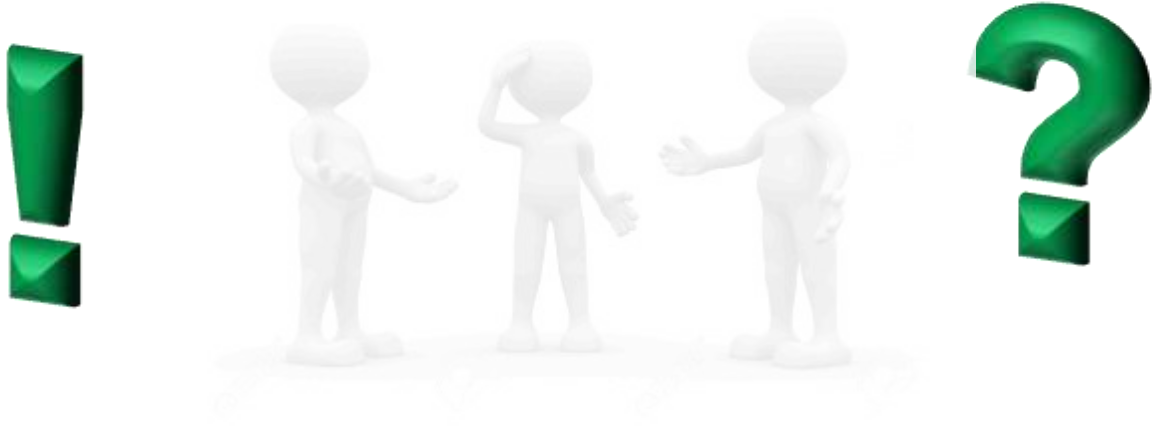
- Case 3
  - error rate with prediction



Data objects : 100

Conclusion and Future Plan

- Prediction-based replication increase availability.
- Real data is used to compare the proposed to three different Popularity, classification and location.
- Need more experiment on cost such as storage.
- Investigating the relation of data for improving the performance.



Thank You

narman@marshall.edu

http://mupfc.marshall.edu/~narman/