Towards Deadline Guaranteed Cloud Storage Services

Guoxin Liu †, Haiying Shen †, and Lei Yu ‡

Presenter: Guoxin Liu
Ph.D.

† Department of Electrical and Computer Engineering,
Clemson University, Clemson, USA
‡ Computer Science
Georgia Institute of Technology, Atlanta, USA

Outline

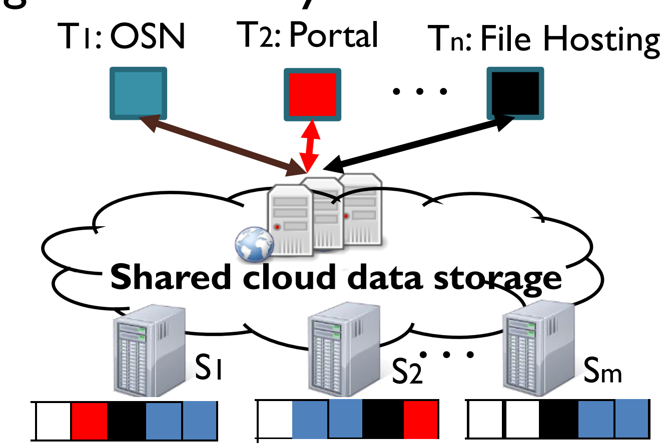
- Introduction
- Related work
- DGCloud Design
- Evaluation
- Conclusion

Introduction

- Cloud storage
  - Tenant incentive
    - Capital investment saving
    - Avoid management and maintenance
    - Pay-as-you-go
  - Provider profits
    - Maximize benefits
      - Resource multiplexing between more tenants
    - Minimize cost
      - Workload consolidation (energy cost)

Introduction

- Cloud storage
  - Challenge for providers
    - Tenant satisfaction
      - Unpredictable service latency with significant variation
      - Tenant income is inversely proportional to latency
        - Amazon portal: increasing page presentation by 100ms degrades sales by 1%.



Introduction

- Solution
  - Avoid server overloaded
    - Allocate workload among servers
- Challenge
  - How to allocate data: non-trivial
    - Satisfy Service level objective (SLO) [1] (e.g. 99% requests within 100ms) baked into cloud storage services
    - Heterogeneous environment
      - Workload among data
      - Service capability among servers
    - Energy saving

[1] C. Wilson, H. Ballani, T. Karagiannis, and A. Rowstron. Better Never than Late: Meeting Deadlines in Datacenter Networks. In Proc. of SIGCOMM, 2011.

Outline

- Introduction
- Related work
- DGCloud design
- Evaluation
- Conclusion

Related work

- Deadline-aware networks
  - Workflow schedule
    - Urgent flow scheduled first
      - Bandwidth allocation
      - Prioritizing workflows
      - Data cache
  - Data operation
    - Allocate close to data requester
    - Proactively replicate
      - Physical closeness
      - Data availability
  - Topology optimization
    - Maximize bandwidth among any pair of servers
- Problem
  - None of them achieve multiple goals as DGCloud in cloud storage

Outline

- Introduction
- Related work
- DGCloud design
- Evaluation
- Conclusion

DLCloud Design

- DGCloud: Deadline Guaranteed Cloud storage service
  - System model: based on M/M/1, derive the upper bound of request arrival rate on each server to satisfy the SLOs of all tenants
  - Cooperative algorithms
    - A load balancing algorithm to satisfy SLO
    - A workload consolidation algorithm to maximize utilization
    - A data placement optimization algorithm to minimize the transmission cost

System Model

- Cloud storage system
  - Different service capacity
  - Different storage capacity
- Tenant
  - Different tenant SLOs
  - Variations of request rates among data
  - Data availability requirement
- Multiple goals
  - Tenant side: satisfy SLO
  - Provider side
    - Minimize network load while data reallocation
    - Maximize system utilization to reduce active servers
- Time complexity
    - NP-hard

Acceptable Request Arrival Rate

- M/M/1 queueing system for Server n
  - Long tail distribution of service latency [2]
  - Poisson process of request arrivals [3]
  - Derive cumulative distribution of service latency $F_{S_n}(t)$
- Acceptable request arrival rate for Tenant k on Server n
  - For a specific SLO$< t, F_{S_n} >$, we get λ'_{S_n,T_k}
- Acceptable request arrival rate for Server n
  - $\lambda'_{S_n} = \min\{\lambda'_{S_n,T_k}\}$

[2] D. Zats, T. Das, P. Mohan, D. Borthakur, and R. Katz. DeTail: Reducing the Flow Completion Time Tail in Datacenter Networks. In Proc. of SIGCOMM, 2012.
[3] D. Wu, Y. Liu, and K. W. Ross. Modeling and Analysis of Multichannel P2P Live Video Systems. TON, 2010.

Load Balancing Scheme

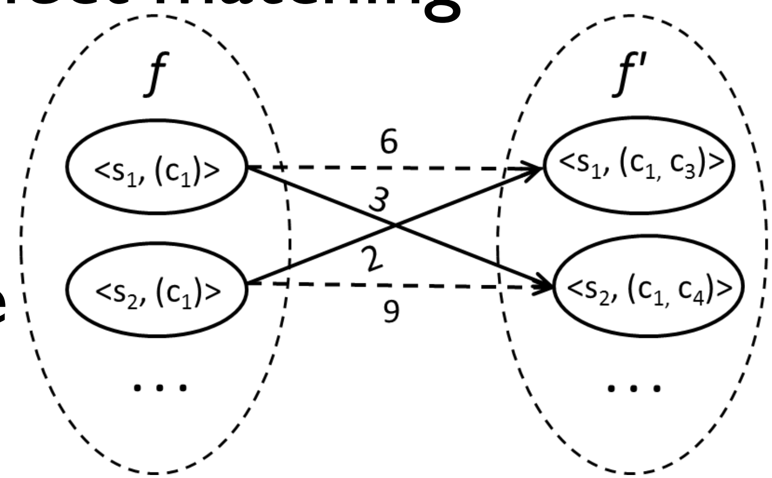
- Request Redirection Algorithm
  - Underloaded servers: $\lambda_{S_n} < \lambda'_{S_n}$
  - Overloaded servers: $\lambda_{S_n} > \lambda'_{S_n}$
  - Loop from most overloaded servers to redirect workload between to other replicas in underloaded servers
- New Replica Allocation Algorithm
  - From most overloaded servers, we created a replica for data partitions with largest workload to most underloaded servers, and then redirect maximum workload to avoid congestion
- Triger: in last measurement period, the worst SLO performance is below a threshold

Workload Consolidation

- Energy cost efficiency
  - Server utilization $\rho_{S_n} = \lambda_{S_n}/\mu_{S_n}$
  - Minimize energy consumption $\equiv$ Minimize server usages $\equiv$ Maximize all server utilization
- Procedure
  - Sort servers in an ascending order of λ_{S_n}
  - From the server at the beginning of the list
    - Shift its workload to all other servers
      - Success: deactivate this server
      - Failed: Stop workload consolidation process

Optimal Data Replication

- Goal: Minimize data transfer during data reallocation
- Observation: replicate cost to convert S_n in f to S_n in f' may not be optimal
- Minimum-weight perfect matching
  - Minimize the total weight of a perfect matching in a bipartite graph



Outline

- Introduction
- Related work
- DGCloud design
- Evaluation
- Conclusion

Evaluation

- Experimental settings
  - 30000 data servers
    - [6TB, 12TB, 24TB] storage capacity
    - [80,100] service capacity
    - Fat-tree with three layers
  - 1000 ~ 6000 tenants
    - [100ms, 200ms] Deadline
    - 5% maximum allowed percentage of requests beyond deadline
    - [100, 900] data partitions with request arrival rate follows distribution in [4]

[4] CTH Trace. http://www.cs.sandia.gov/Scalable IO/SNL_Trace_Data/, 2009.

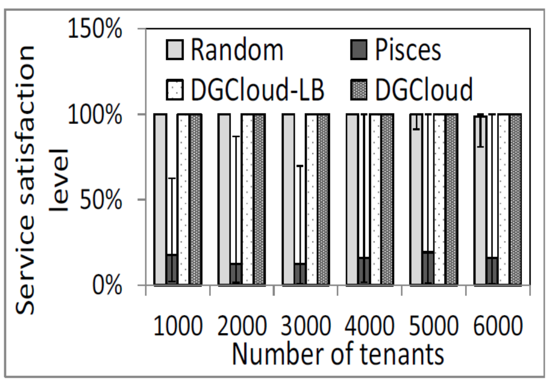
Evaluation

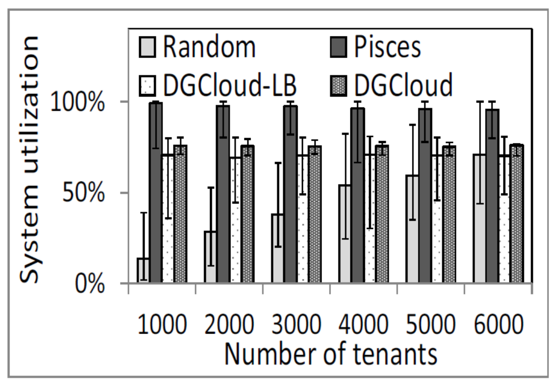
- Comparison methods
  - Deadline guarantee periodically
    - Random: randomly place data among servers
    - Pisces[5]: storage capacity aware data first fit
    - DGCloud-LB: deadline aware first fit
- Important metrics
  - SLO satisfaction level: actual percentage of requests within deadline/required percentage over the desired percentage $\in [0,1]$
  - System utilization: average server utilization of all active servers in the system

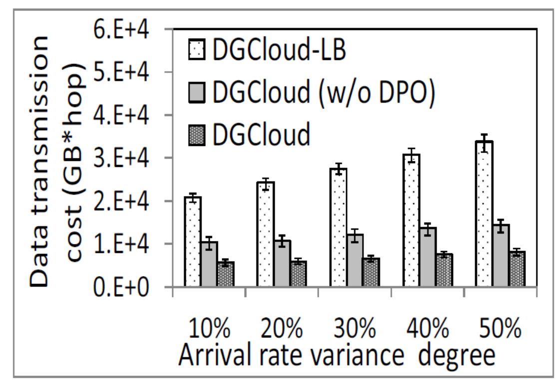
5] D. Shue and M. J. Freedman. Performance Isolation and Fairness for Multi-Tenant Cloud Storage. In Proc. of OSDI, 2012.

Evaluation

- SLO guarantee
  - Ensure SLO
- System utilization
  - Highest system utilization among all systems ensuring SLO
- Data transmission cost
  - Optimal data placement saves transmission cost







Outline

- Introduction
- Related work
- DGCloud design
- Evaluation
- Conclusion

Conclusion

- DGCloud: Deadline-Guaranteed Cloud service while achieving high system resource utilization
  - Mathematically derive the extra load that a server needs to move out to meet the SLOs of all tenants
  - A deadline-aware load balancing scheme to ensure SLOs
  - Work consolidation to maximize the system resource Utilization
  - Data placement optimization to minimize the transmission cost

- Future work
  - Dynamically redirect requests and replicate data to ensure SLO under a request burst

Thank you!

Questions & Comments?

Guoxin Liu

guoxinl@g.clemson.edu

Electrical and Computer Engineering

Clemson University