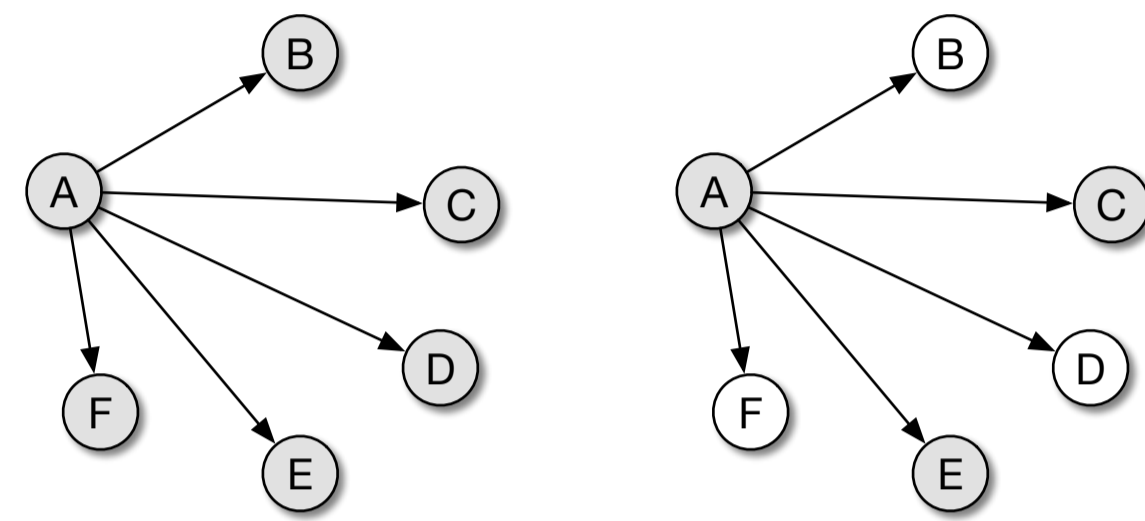


Acorn: Replicate Your Data Where it is Likely to be Used

Hobin Yoon (hobinyoon@gatech.edu) Georgia Institute of Technology

Problem

- **Cost of geo-replicated storage** in global-scale applications keeps increasing as the volume of data and the number of datacenters increase.
- Many **partial geo-replication systems** have been introduced lately to lower the cost [Kadambi '11] [SpanStore '13] [Pileus '13] [Tuba '14]



Full and partial geo-replication

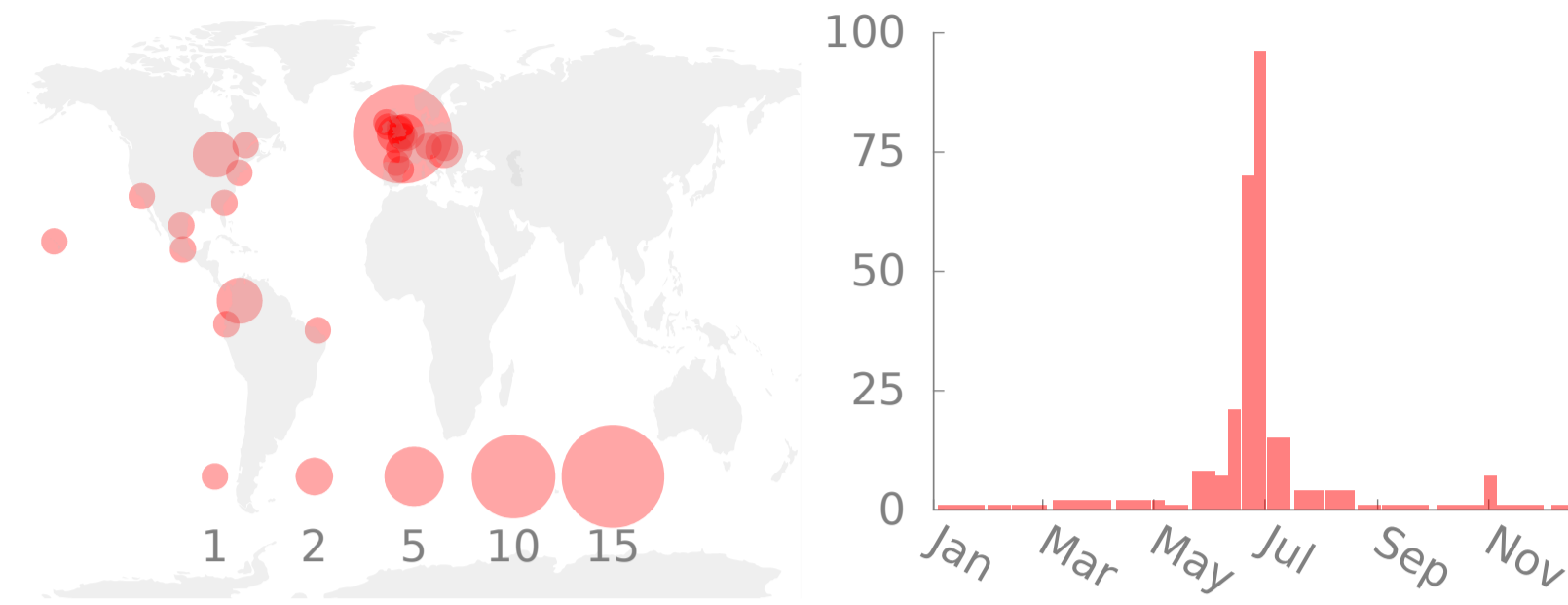
- However, they have **high cost and latency** due to suboptimal replication decisions
 - They use a static record attribute “user” regardless of the application types
 - They use only a single attribute to make replication decisions
 - They use non-continuous replications, leading to suboptimal performance under SLO constraints

We propose **Acorn**, an Atribute-based continuous partial geo-replication system to achieve lower cost and latency than existing systems.

Design Principles

1. Use the right attributes for different types applications

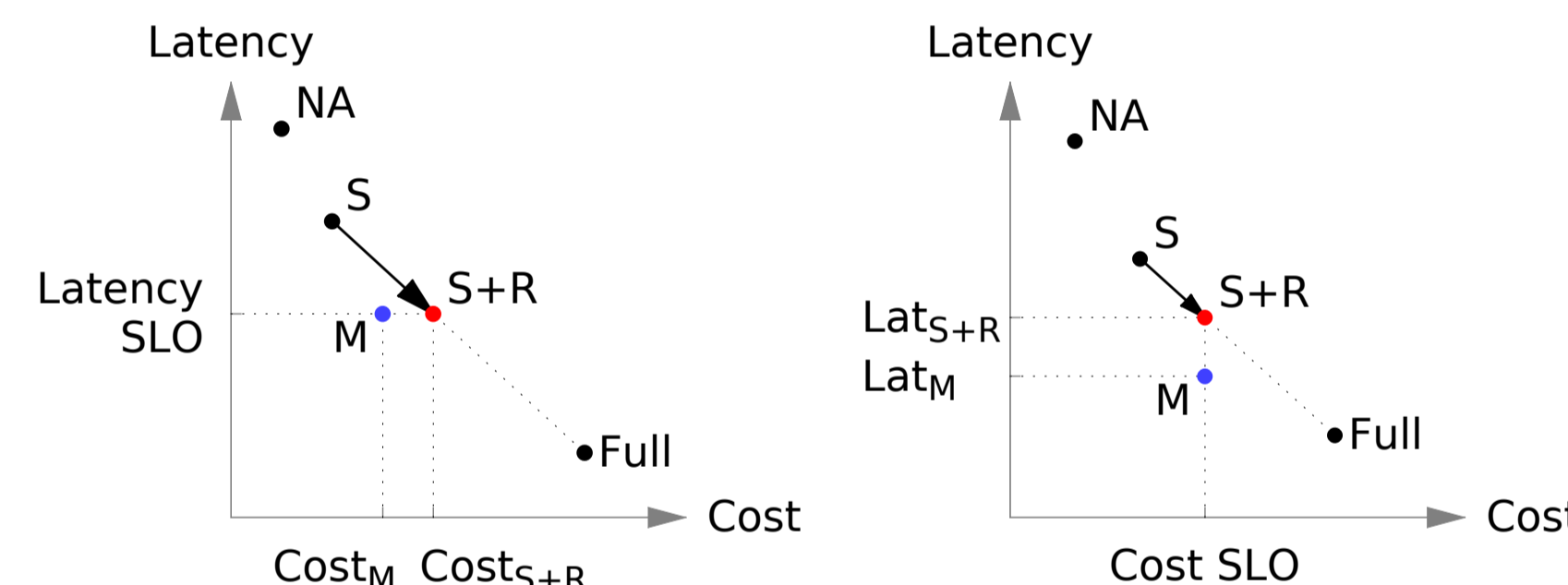
- **Private data-sharing applications**, such as Facebook or Snapchat, objects are accessed mostly through friends, thus “**user**” is the best attribute to monitor and predict future accesses.
- **Public ones**, such as YouTube or Flickr, have diverse sources of accesses. “**Topic**” has stronger geographic and temporal locality than “user”.



Examples of geographic and temporal locality. (Left) Access locations of YouTube videos with the topic “Wimbledon” from Jan. 1st to June 7th, 2014. (Right) Accesses of the same videos in 2014 by week.

2. Use multiple attributes

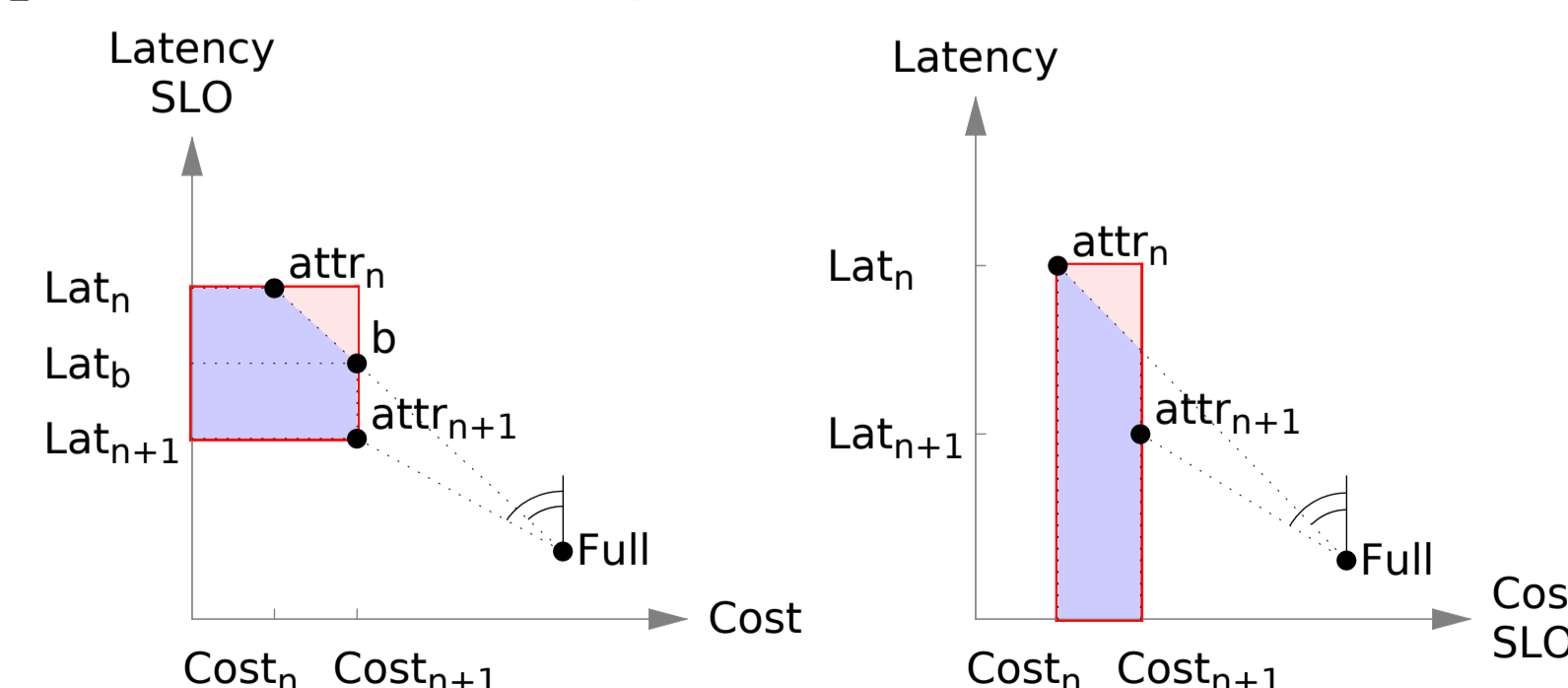
- Using more attributes makes you make better educated guesses where to place replicas.
- E.g., replicate YouTube videos to Atlanta, GA when either they have a topic “tennis”¹ or they are uploaded by “John Isner”²



(Left) Cost reduction of multiple attribute over single attribute under latency SLO. (Right) Latency reduction of multiple attributes.

3. Use continuous replications

- Acorn adds **extra random geo-replicas**, to achieve the optimal cost or latency under SLO constraints.



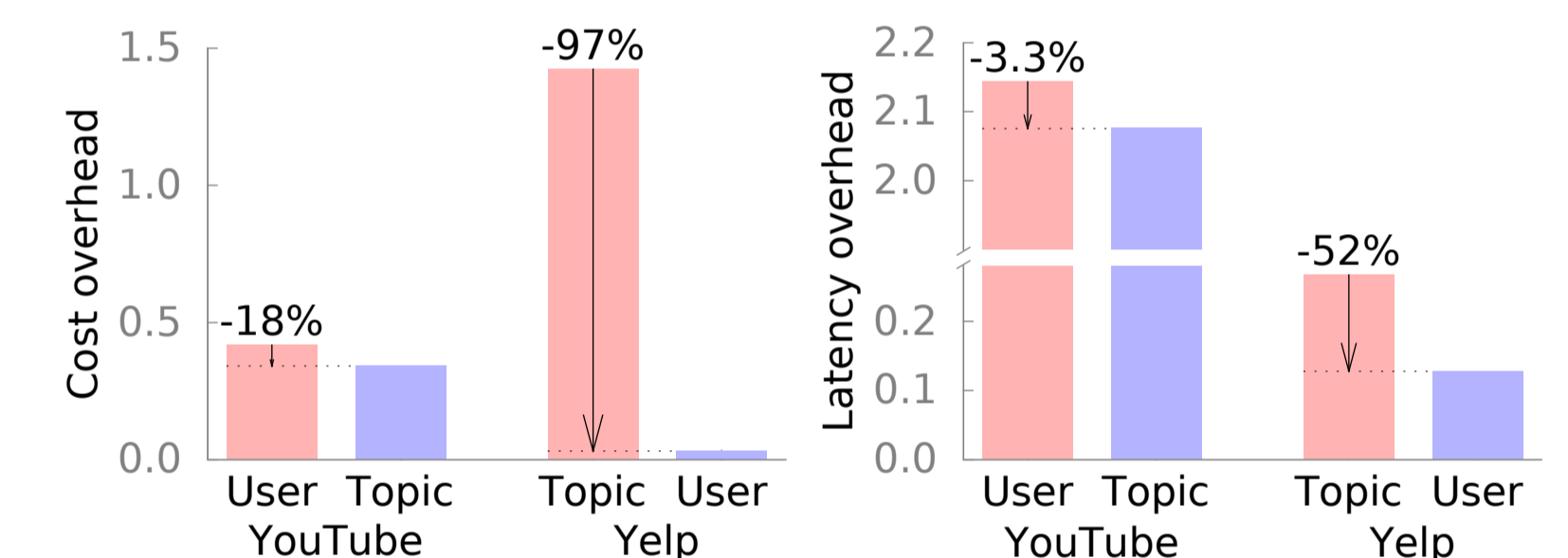
(Left) Cost reduction of continuous replications over non-continuous ones (Right) Latency reduction

Implementation

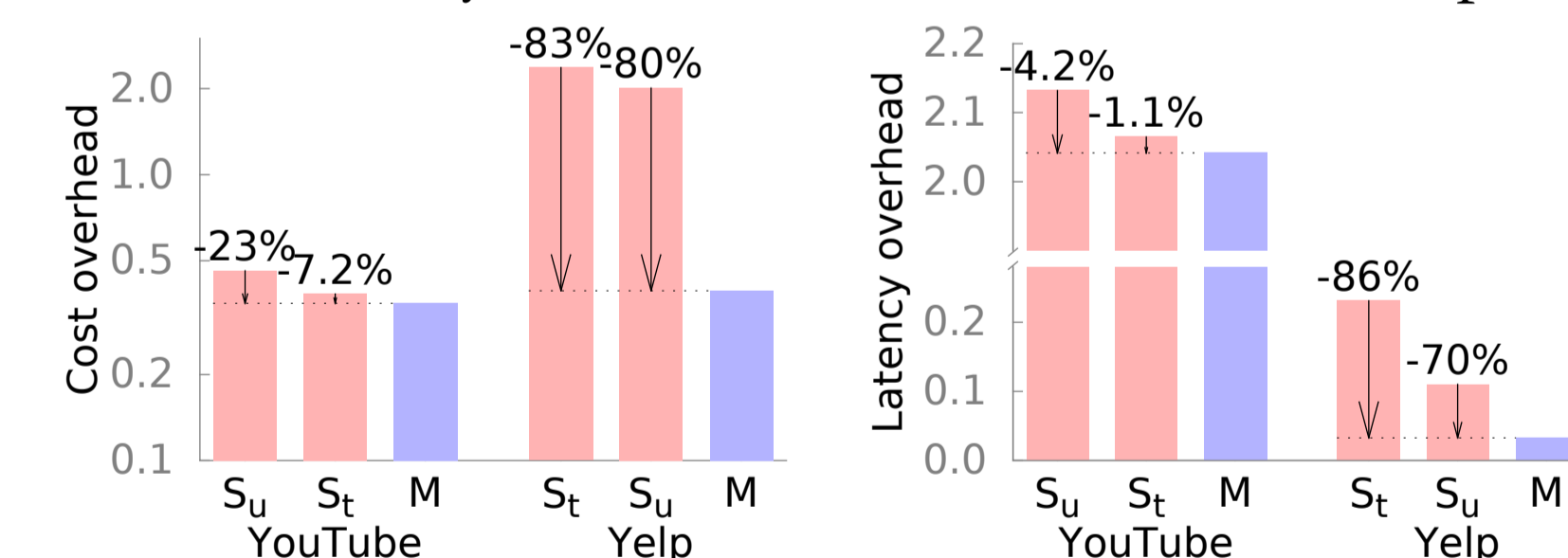
- Modified Apache Cassandra
- write makes selective geo-replicas.
- read (a) first reads the object in local datacenter, and when misses, fetches it from a remote datacenter. (b) updates attribute popularity monitor.
- Popularity metadata synchronizer
- Object location metadata

Preliminary Results

- Evaluation with a public and a private data-sharing application using workload of YouTube and Yelp on 8 and 5 simulated datacenters
- Cost and latency reduction with the right attribute



- Cost and latency reduction with multi-attribute replication



- 40.62% cost reduction with continuous replications
- Overall, Acorn achieves up to 54.28% and 89.91% cost overhead reduction and 43.12% and 91.69% latency overhead reduction for the public and private data-sharing applications

¹ Atlanta has the highest number of USTA members per capita in the US.
² Isner played for University of Georgia and is a top-ranked tennis player in the US as of Aug. 2015.