Karma: Cost-effective Geo-replicated Cloud Storage with Dynamic Enforcement of Causal Consistency
Tariq Mahmood, Shankaranarayan P. N., Sanjay Rao, T. N. Vijaykumar, and Mithuna Thottethodi
School of Electrical and Computer Engineering, Purdue University

Motivation

The Consistency Spectrum
Replication, asynchronous write propagation create ordering issues
- Weak "eventually consistent" systems
- Widely deployed, but ordering can be confusing
- Strong ordering of all reads and writes across all clients
CAP Theorem ⇒ unavailable on partition

Causal consistency:
Partial order that preserves causality
Not confusing for users
Available under partition

Linearizable ⇒ Causal ⇒ COPS/Eiger/Orbe/Karma

Caching/Write Buffers
Persistence thread-private write buffers enable fast writes
- If a client reads an in-flight object from Ring-1
- Client can access any ring once in-flight objects are stable

Partial Replication
- Decouple rings and DCs
- Rings span multiple DCs
- Each ring contains full replica of dataset
- Availability in wide-area rings guaranteed by causality-preserving dynamic ring binding
- DC level caching used for fast reads of remote objects

Dynamic Ring Binding
- In-flight (Violation possible)
- Stable (No violations possible)
- Karma’s novel mechanism: Dynamic Ring Restrictions (DRR)
- If a client reads an in-flight object from Ring-1
- Temporarily restrict client to read all objects from Ring-1

Karma’s Key Ideas
- Decouple rings and DCs
- Rings span multiple DCs
- Each ring contains full replica of dataset
- Availability in wide-area rings guaranteed by causality-preserving dynamic ring binding
- DC level caching used for fast reads of remote objects

Performance Evaluation (R/W : 95/5)

Experimental Setup
- 64-node testbed on PRObE cluster
- 8 data centers, 8 nodes each
- Amazon AWS emulation using DummyNet

Four Schemes:
- COPS-Ideal
- Karma
- COPS-PR
- Karma-NC

Replication
- Full
- Partial
- Partial
- Partial

Ring binding
- Static
- Dynamic
- Static
- Dynamic

Write buffers
- -
- 
- 
- 

Caching
- -
- 
- 
- 

Karma achieves 43% higher throughput than COPS-PR

No benefit from dynamic ring binding under fault free conditions

Karma-NC performs better than COPS-PR

Importance of Dynamic Ring Binding
- Induce congestion in Europe DC
- All traffic (in and out) is affected
- Table below shows avg. performance hit

Fault Tolerance Analysis
- Backend Server: Chain replication
- Cache Server: Stable state
- Frontend Server: Chain replication
- Rack: Chain replication
- Single AZ: Dynamic binding
- Partition: Dynamic binding

Summary
- First causally-consistent cloud storage system with:
  - Practical, cost effective
  - Stronger availability guarantees
- 43% throughput improvement iso-cost
- Significant reduction in operational and capital expenditures