

# Replication Degree Customization for High Availability

Ming Zhong  
Google

Kai Shen    Joel Seiferas  
University of Rochester

## Problem Context

- Replication degree tradeoff between availability and space
- Skewed data popularity distributions
  - intuitive to highly replicate popular objects
  - **goal**: improve availability under certain space constraint
- Should we worry about space constraint today?
  - **decentralized wide-area systems** - high machine failure/inaccessibility rate demand high-degree replication
    - 0.265 failure rate from a Planetlab machine accessibility trace
  - **centrally-managed local-area clusters** - often requiring data in-memory and relatively high memory cost

4/10/2008

EuroSys 2008

2

## Basic Analytical Result

- Problem formulation:
  - $p$  is machine failure probability  $\Rightarrow$  unavailability of an object with  $k$  replicas is  $\text{pow}(p,k)$
  - a system with  $n$  objects: object  $i$  popularity is  $r_i$ , size is  $s_i$
  - find object replication degrees  $k_1, k_2, \dots, k_n$  to minimize expected unavailability  $\sum_{1 \leq i \leq n} r_i \cdot \text{pow}(p, k_i)$
  - subject to space constraint  $\sum_{1 \leq i \leq n} s_i \cdot k_i \leq K$
- Result:
  - Lagrangian function:  $\sum_{1 \leq i \leq n} r_i \cdot \text{pow}(p, k_i) + \lambda \cdot (\sum_{1 \leq i \leq n} s_i \cdot k_i - K)$
  - optimization is reached when the function's partial derivatives on  $k_i$ 's and  $\lambda$  are all zero
  - therefore  $k_i = C + \log_{1/p}(r_i/s_i)$ ,  $C$  is a constant

4/10/2008

EuroSys 2008

3

## Challenges in Systems Context

- Basic analytical result:
  - optimal object replication degree  $k_i = C + \log_{1/p}(r_i/s_i)$
  - $p$  is machine failure probability;  $r_i$  is object popularity;  $s_i$  is object size
- Systems issues:
  - **complex system models**: nonuniform and correlated machine failure rates, multi-object operations

4/10/2008

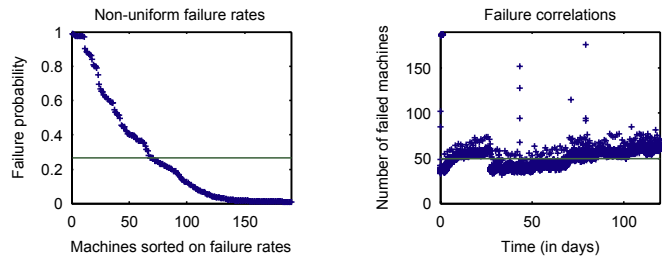
EuroSys 2008

4

realistic workload behaviors: skewness and

## Non-uniform/correlated Failures

- Planetlab machine failures in a four-month trace



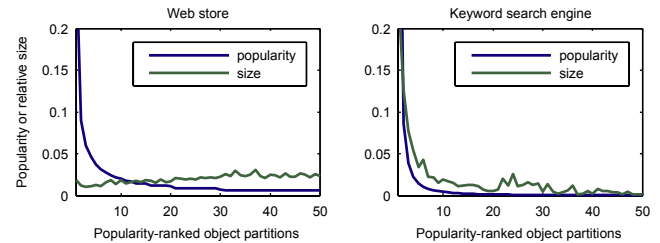
4/10/2008

EuroSys 2008

5

## Object Popularity/Size Skewness

- Popularity to size skewness correlation
- Trace-driven examination:



- Most popular objects are not necessarily subject to most replication

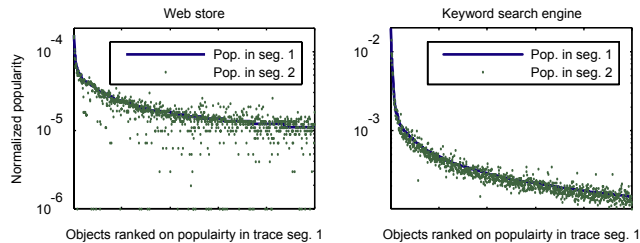
4/10/2008

EuroSys 2008

6

## Object Popularity Stability

- Stable object popularity is important for learning the object popularity and for low adjustment overhead
- Illustration of stability across month-long trace segments:



- Not designed to handle flash crowds

4/10/2008

EuroSys 2008

7

## Dynamic Maintenance Overhead

- System adaptation may require dynamic maintenance
  - object popularity may change over time
- Low maintenance overhead due to
  - stable object popularities
  - stable replica assignment in the analytical result
    - object replication degree  $k_i = C + \log_{1/p}(r_i/s_i)$
  - coarse-grain adaptation
    - e.g., support only two replication degrees - low/high

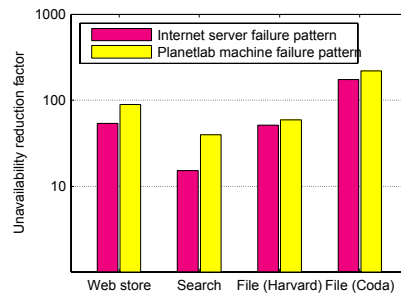
4/10/2008

EuroSys 2008

8

## Trace-driven Evaluation – Availability Improvement

- Availability improvement on four real application traces and two machine failure patterns
- Compared to uniform replication under same space constraint:



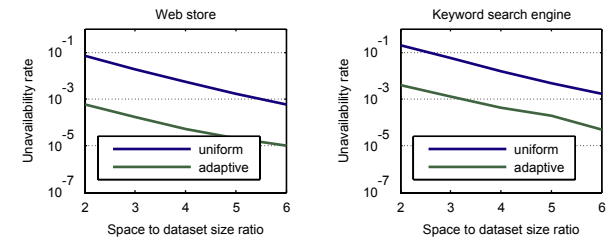
4/10/2008

EuroSys 2008

9

## Trace-driven Evaluation – Changing Space Constraint

- Results on the Planetlab machine failure pattern



- Availability improvement is independent of space limit
- High replication needed for some decentralized wide-area systems

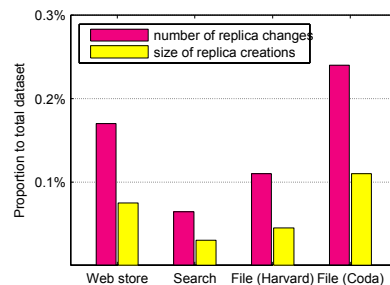
4/10/2008

EuroSys 2008

10

## Trace-driven Evaluation – Dynamic Maintenance Overhead

- Overhead of weekly changes on replication degree
  - number of replica creations/deletions
  - size of replica creations



4/10/2008

EuroSys 2008

11

## Conclusion

- Results
  - analytical result:** optimal replication when object replication degree is linear to  $\log(\text{popularity}/\text{size})$
  - address systems issues in complex system models, realistic workload behaviors, and maintenance overhead
- Big picture:** skewed & stable data distributions motivate per-object adaptation in distributed system management
  - adapt replication degree for high availability [[this paper](#)]
  - adapt Bloom filter hash number for low false-positive rate [[other result](#)]
  - adapt co-placement of correlated data objects for fast multi-object operations [[other result](#)]

4/10/2008

EuroSys 2008

12