High-Performance Disk Imaging with Deduplicated Storage

Raghuveer Pullakandam, Xing Lin,
Mike Hibler, Eric Eide, Robert Ricci
School of Computing, University of Utah

Introduction

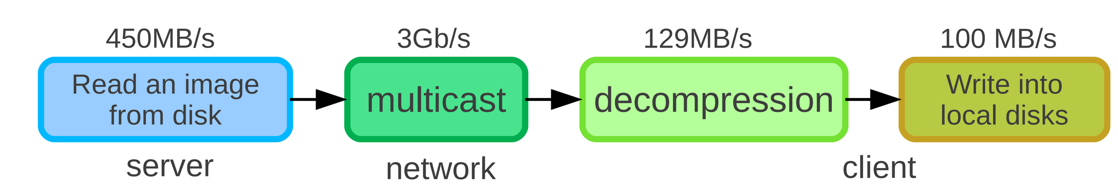
Disk imaging in Emulab:
- Emulab: a 'Cloud' providing machines and a network for running network experiments
- Frisbee: create, distribute and install disk images
  fast(<30s for a 400MB compressed disk image)
  scalable(>80 clients)

Problem: storage consumption for Frisbee disk images is inefficient (lots of duplication)

Goal:
- Reduce storage consumption
- Minimize performance overhead in disk imaging

Approach:
- Efficient deduplication with Venti
- Optimize the image construction overhead introduced with Venti

Effective throughput, assuming 3X compression



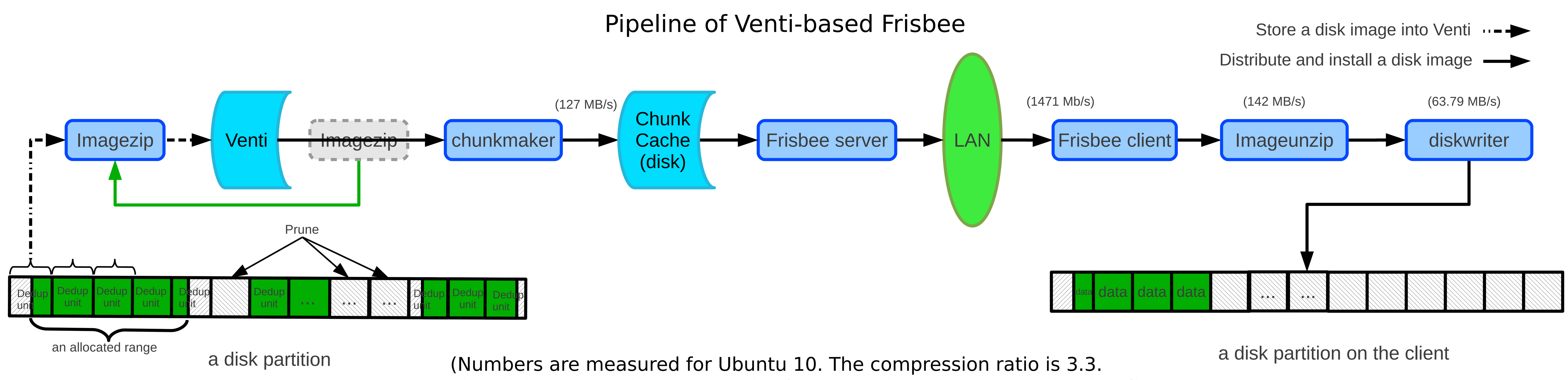
Frisbee pipeline

Frisbee Performance Analysis:
- The bottleneck: disk write speed on the client

Design Implication:
- Image construction based on Venti should be ***fast enough***

System Architecture and Design Detail

Pipeline of Venti-based Frisbee



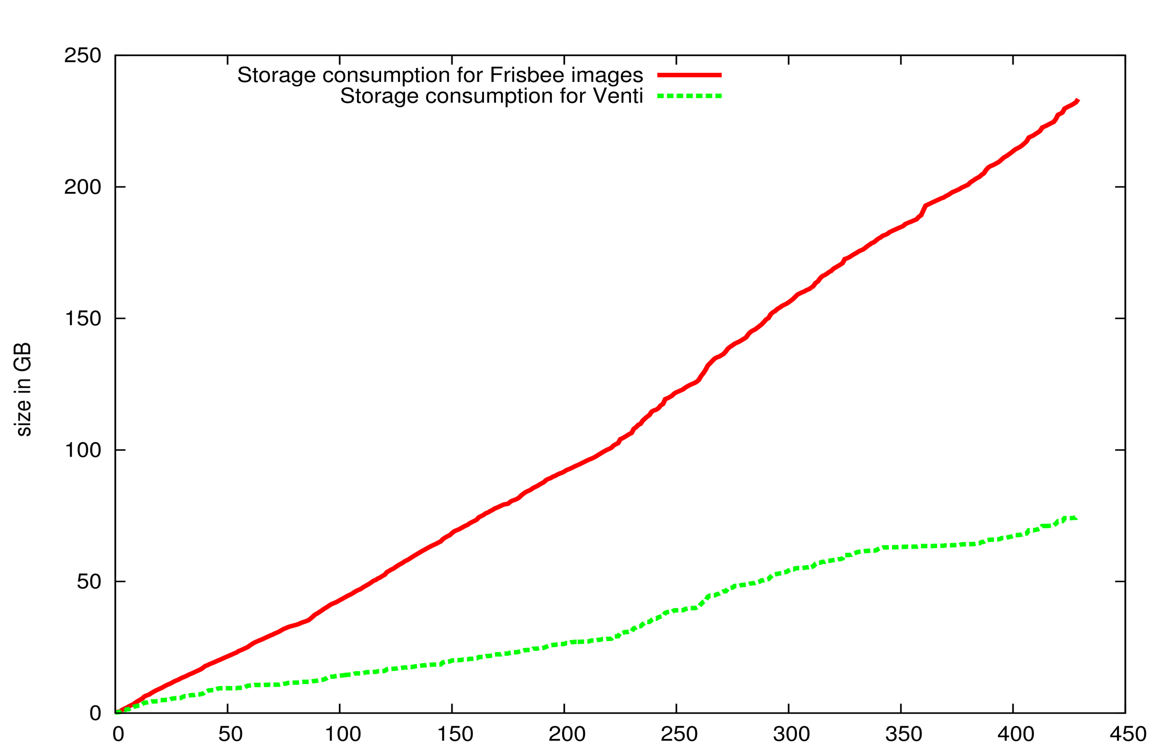
(Numbers are measured for Ubuntu 10. The compression ratio is 3.3. The Frisbee server is set to send only at 500Mb/s. The reason why the client write thoughput is only 64MB/s is that it needs to seek over free sectors.)

1. Efficient Deduplication:
- Frisbee disk images are not ideal for deduplication
- Deduplicate based on uncompressed disk data, not compressed Frisbee images
- Deduplication units are aligned based on disk block offsets

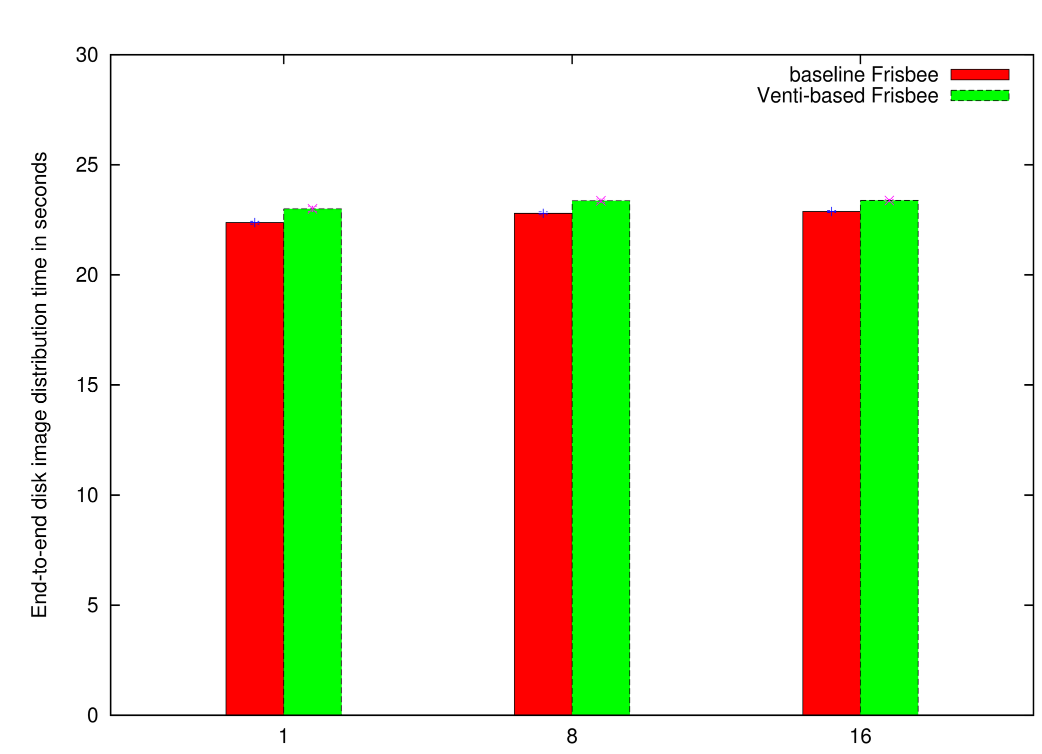
2. Effective Venti Read Throughput
- Venti read rate is about 40MB/s
- Pre-compressed data: average compression ratio of 3.0 gives 120MB/s effective throughput

3. Fast Image Construction
- Store **compressed** data in Venti to remove expensive compression from the critical path and **pre-compute** chunk boundaries
- Exploit pipelining to partially mask chunk construction overhead
- Cache constructed chunks for later client machines

Evaluation



Storage consumption is significantly lower with Venti than with current Frisbee images



End-to-end performance in the Venti-based system is within 3% of the original Frisbee and scaling is unaffected

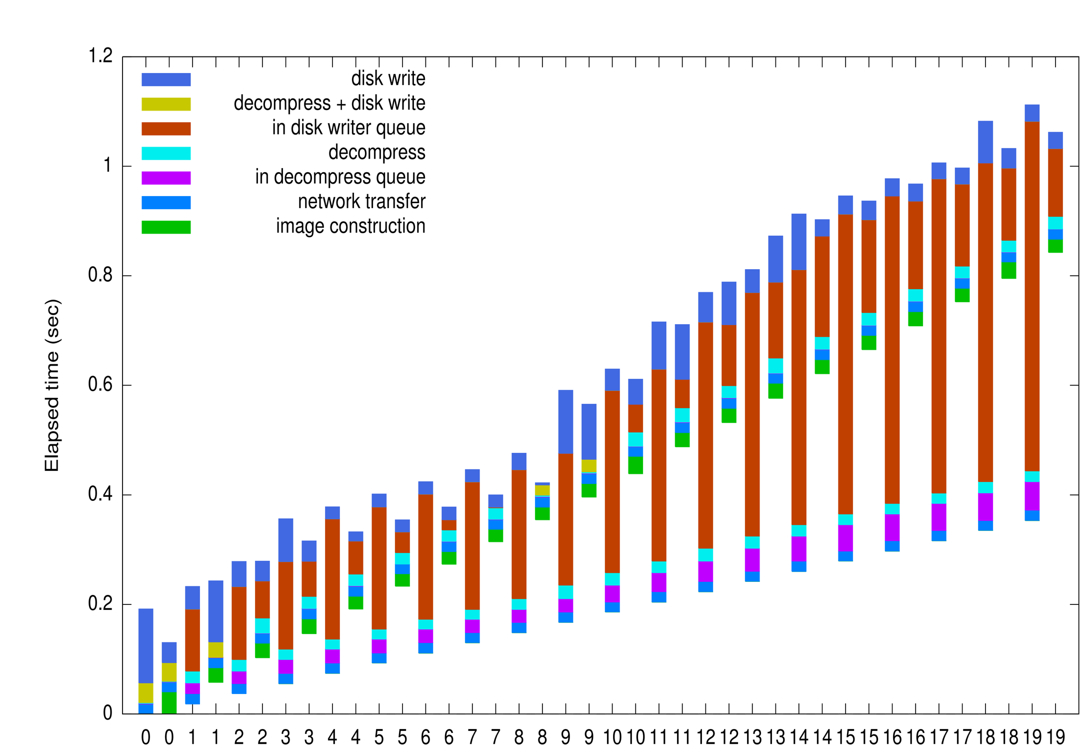


Image construction overhead is masked by disk write speed on the client

This material is based upon work supported by the National Science Foundation under Grant No. 0709427. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.