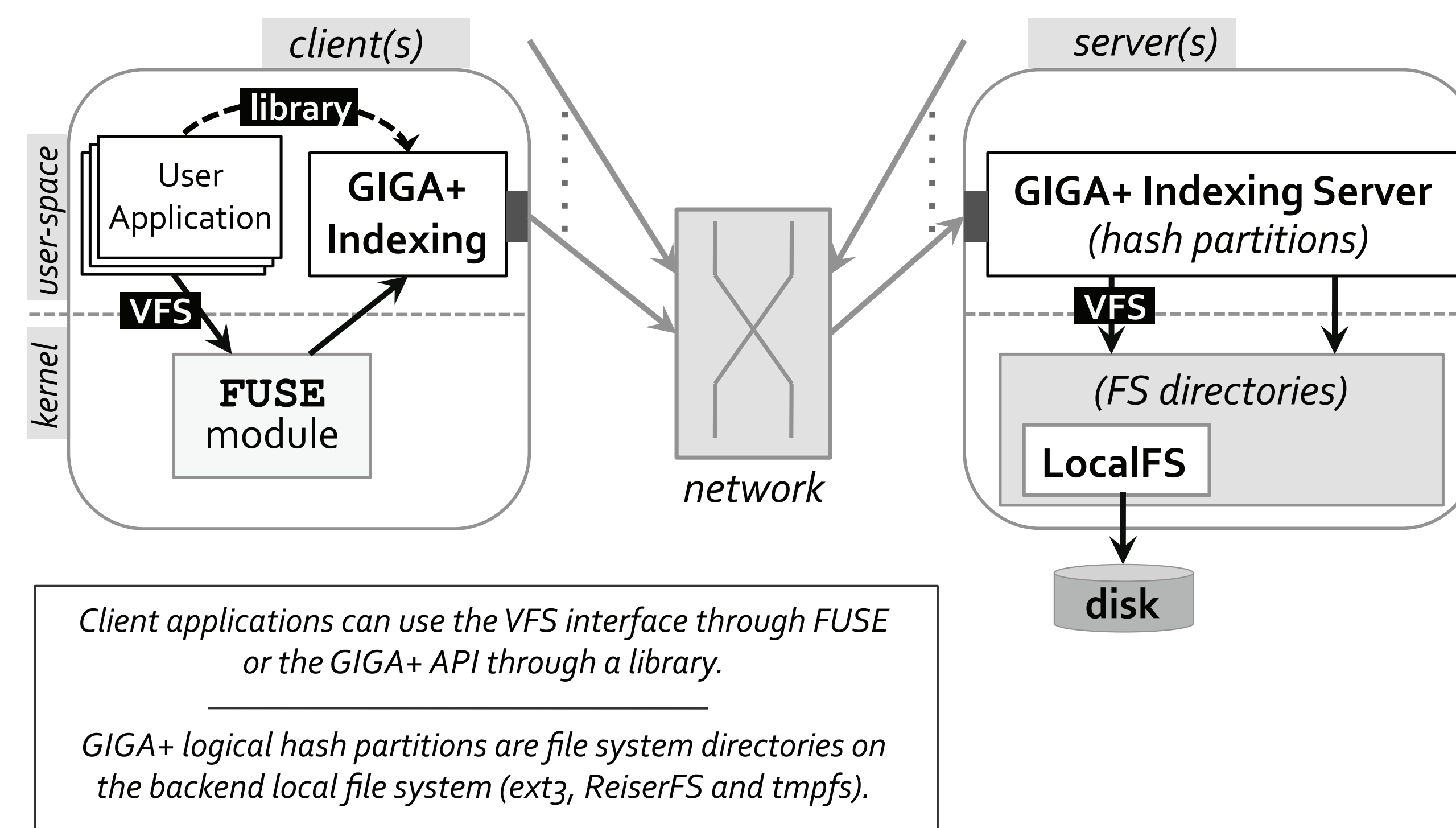


Evaluation of GIGA+ Scalable Directories (using the FUSE user-level prototype)

Swapnil Patil and Garth Gibson

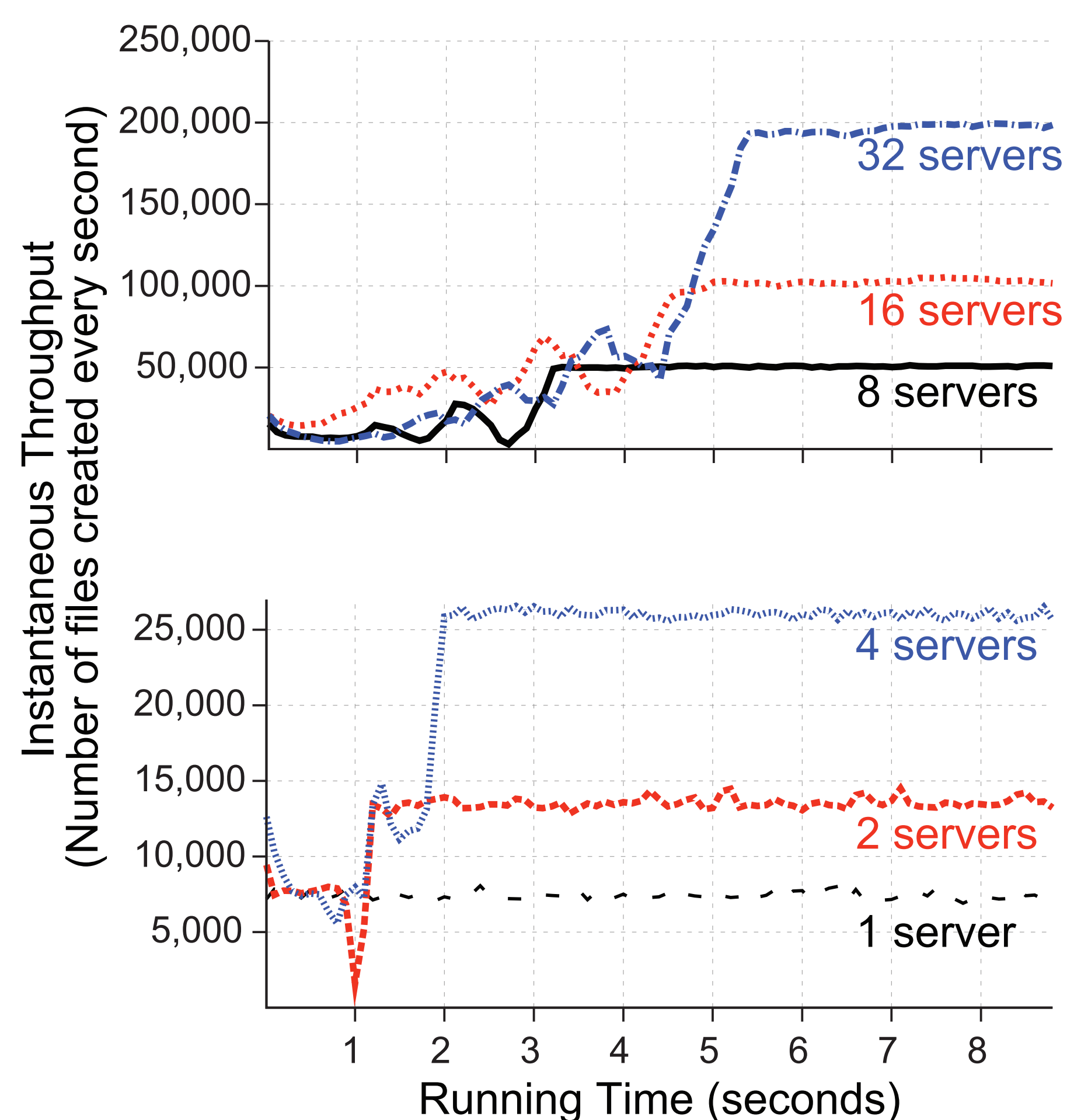
Methodology

- Concurrent create workload on the OpenCloud cluster
- Dual quad-core 2.8 GHz Xeons with 16GB RAM and 10GigE NICs
- Benchmark based on the "mdtest" framework
- 400,000 files created per server
- Two aspects of evaluation
- Scalability of indexing (by storing data in Linux tmpfs)
- Directory performance (by storing data on ext3 and Reiser)



Incremental, Scale-out Growth

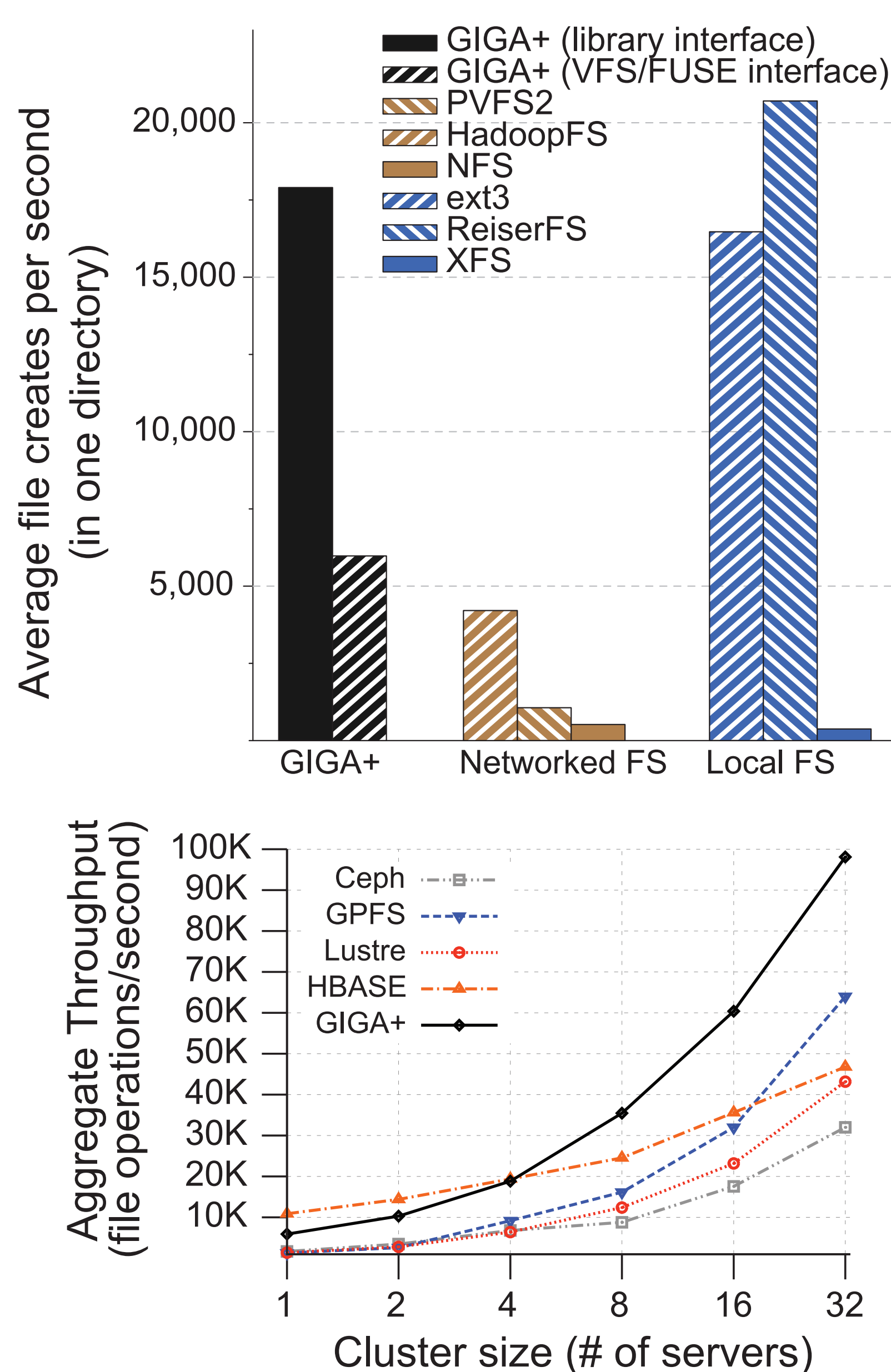
GIGA+ throughput scales linearly after partitions are distributed on all servers



Scale and Performance

GIGA+ scales to a peak throughput of 98,000 empty file creates per second

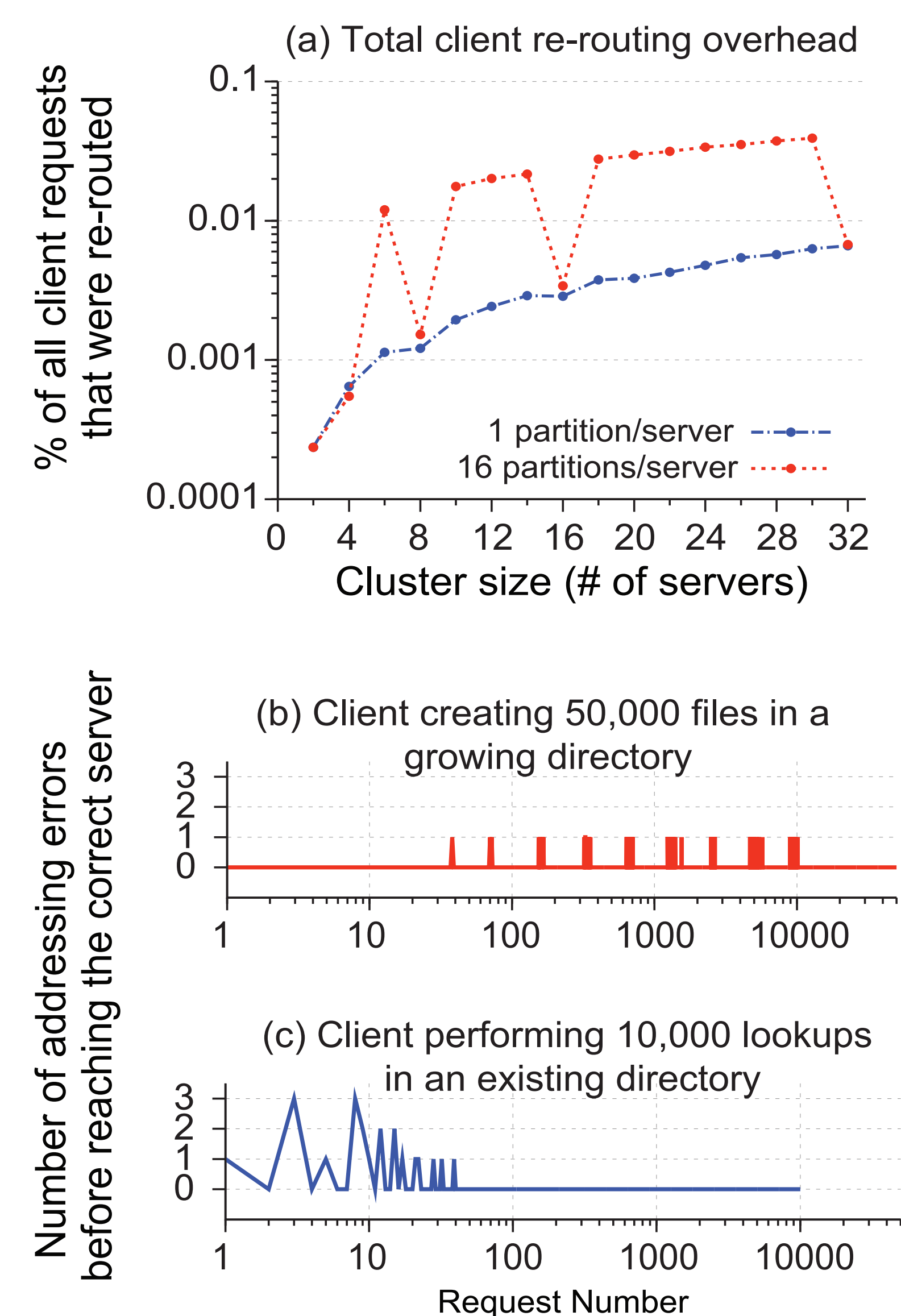
- Outperforms many scalable file systems



Weak Consistency Overhead

Less than 0.01% client requests are rerouted due to stale mapping state

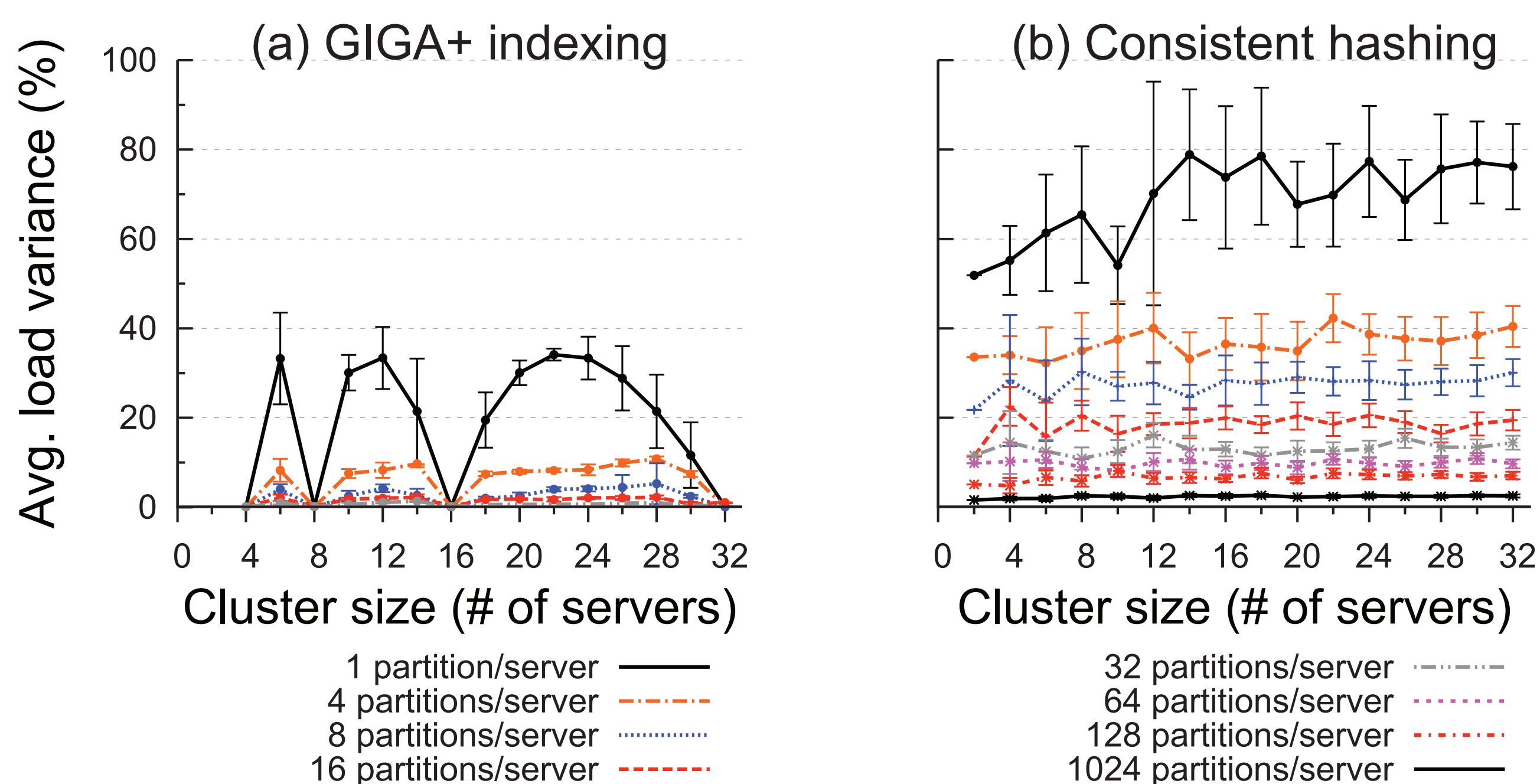
- Request re-routing stops after clients have learnt about all partitions in the system



Load-balancing Effectiveness

Multiple partitions per server improves GIGA+ load balancing (for non power-of-two servers)

- Requires 10-100X less re-partitioning than consistent hashing



Interaction with Local File Systems

Two design choices affect the performance of GIGA+ directories

- On-disk directory representation
- Metadata handling techniques such as journaling

