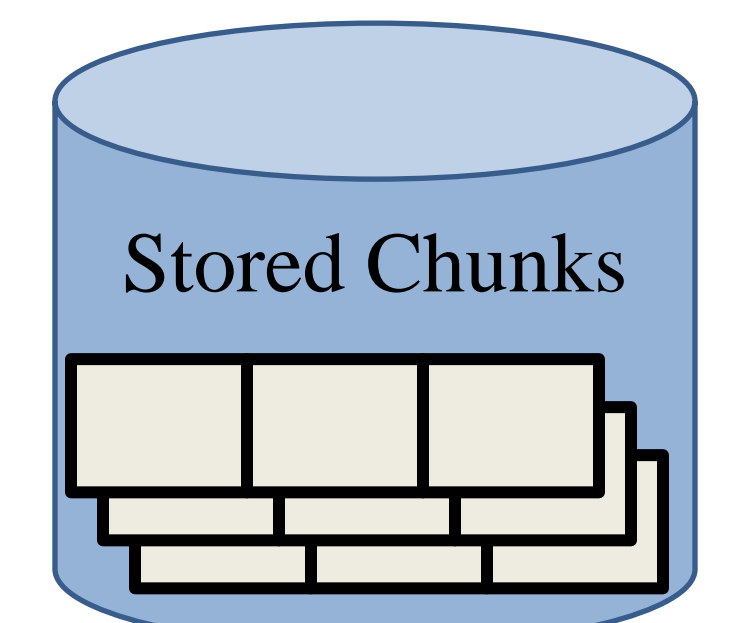


Introduction

- WAN bandwidth often limits replication throughput
- Deduplication and local compression increase replication throughput
- **Delta compression with stream-informed caching adds 2X additional compression**

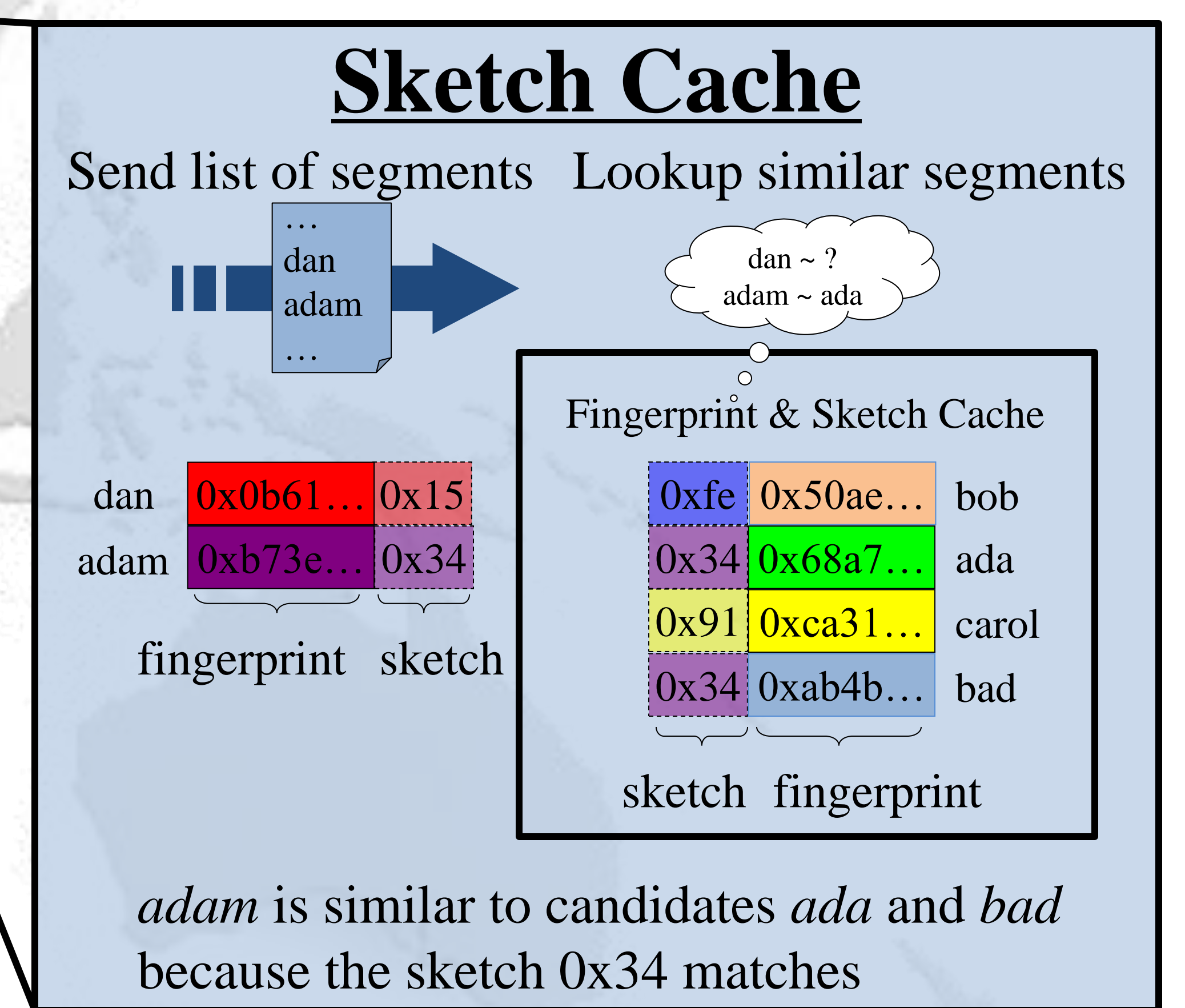
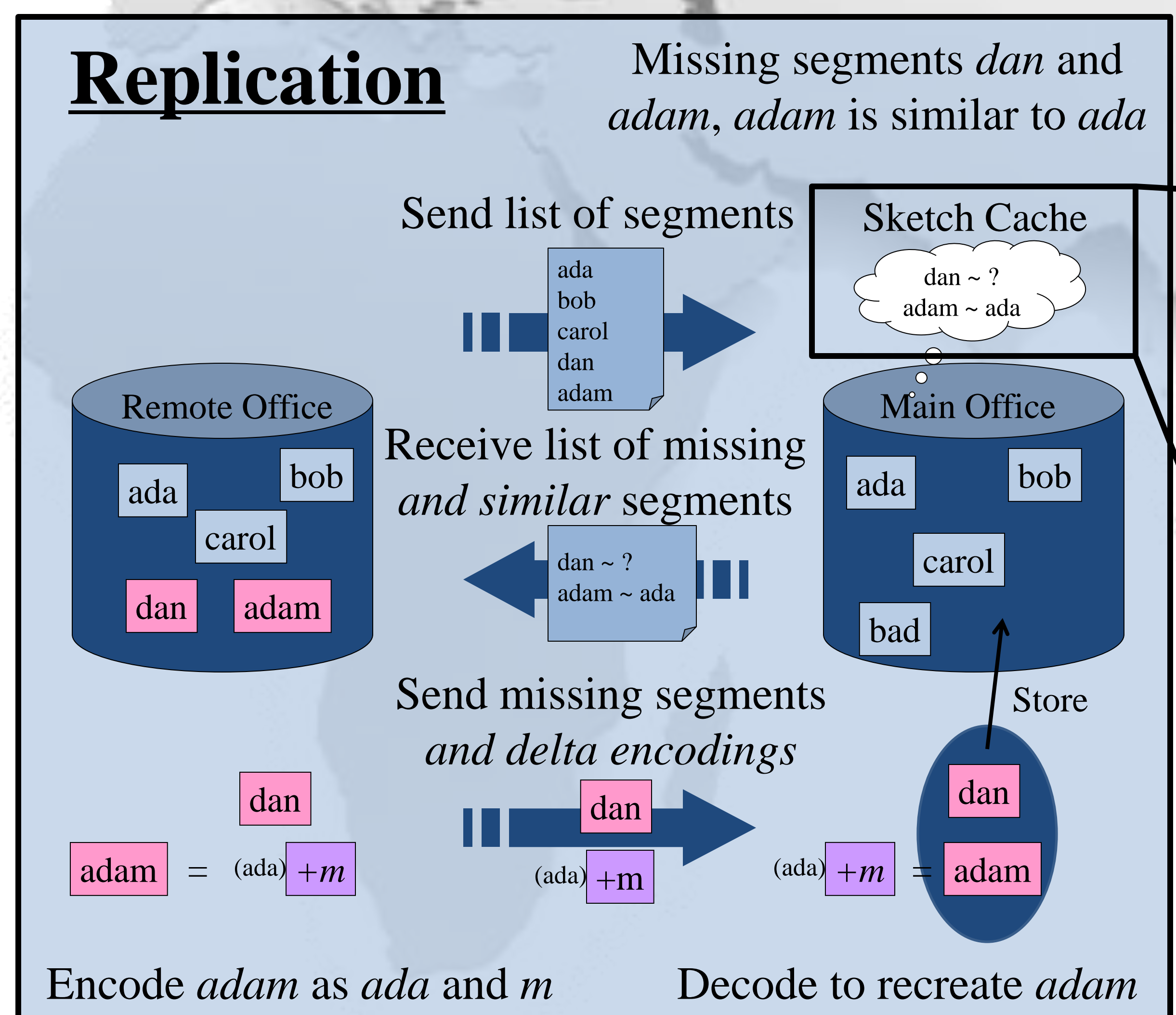
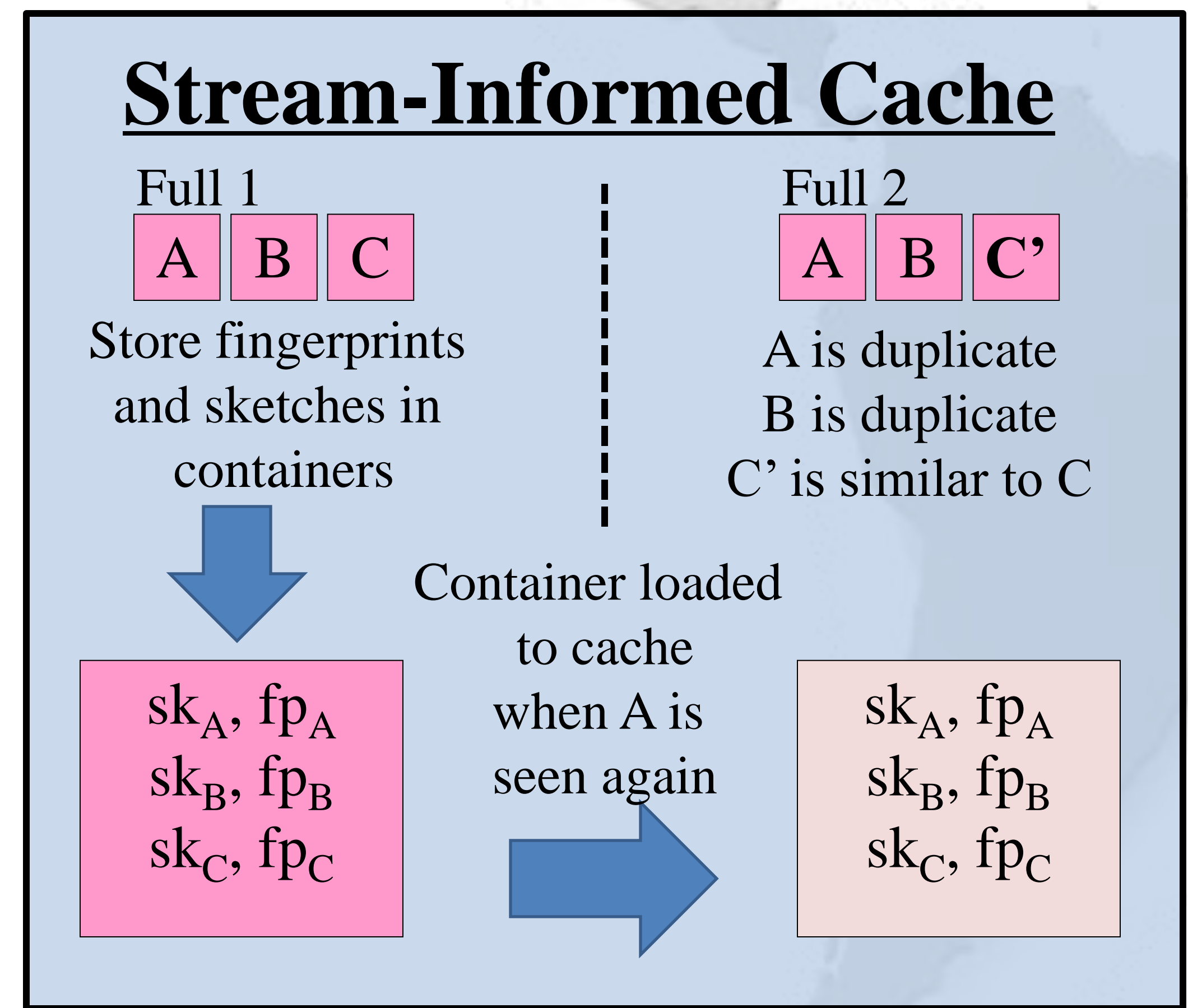
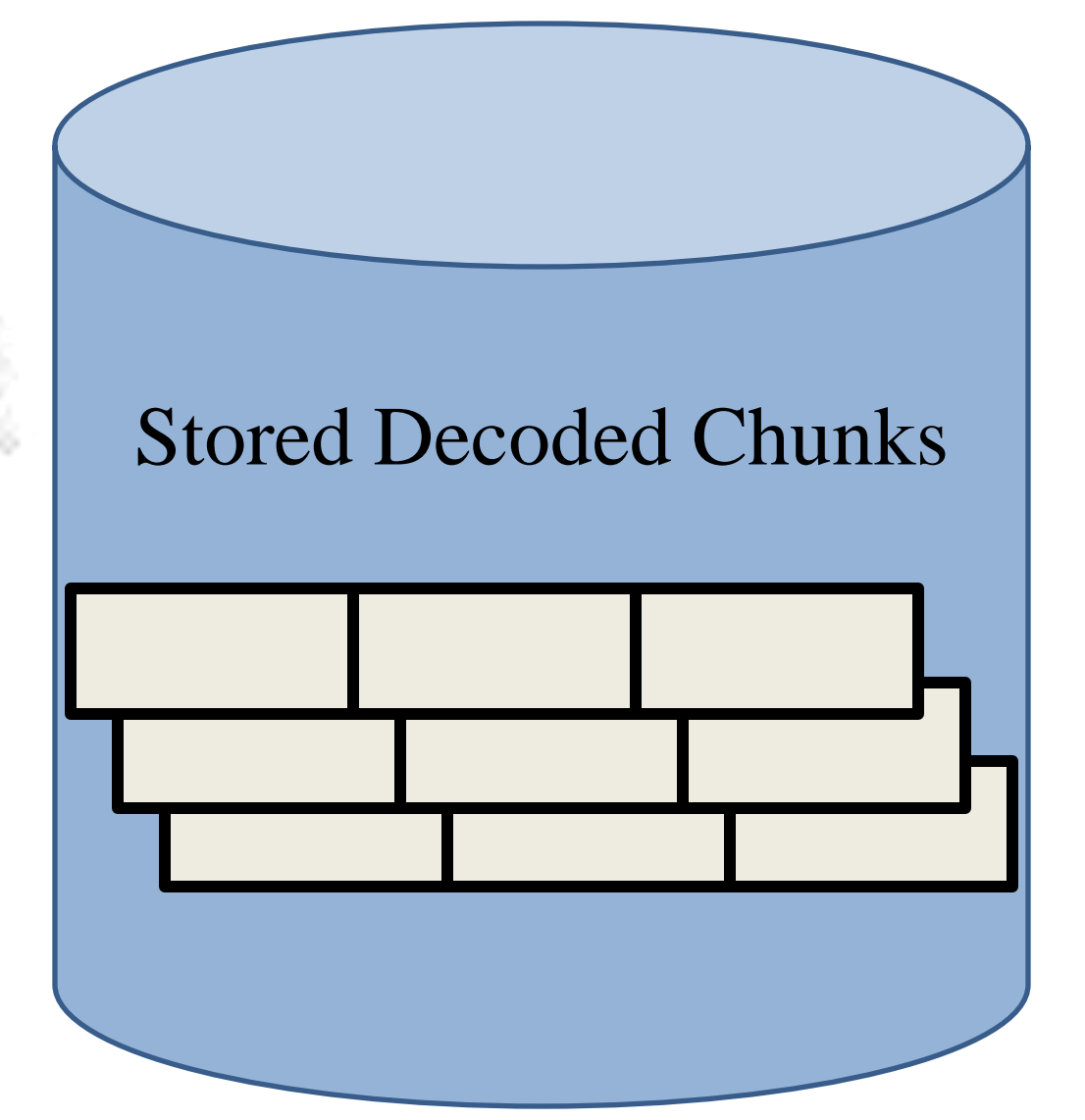
Remote Office Backup System



Low bandwidth link



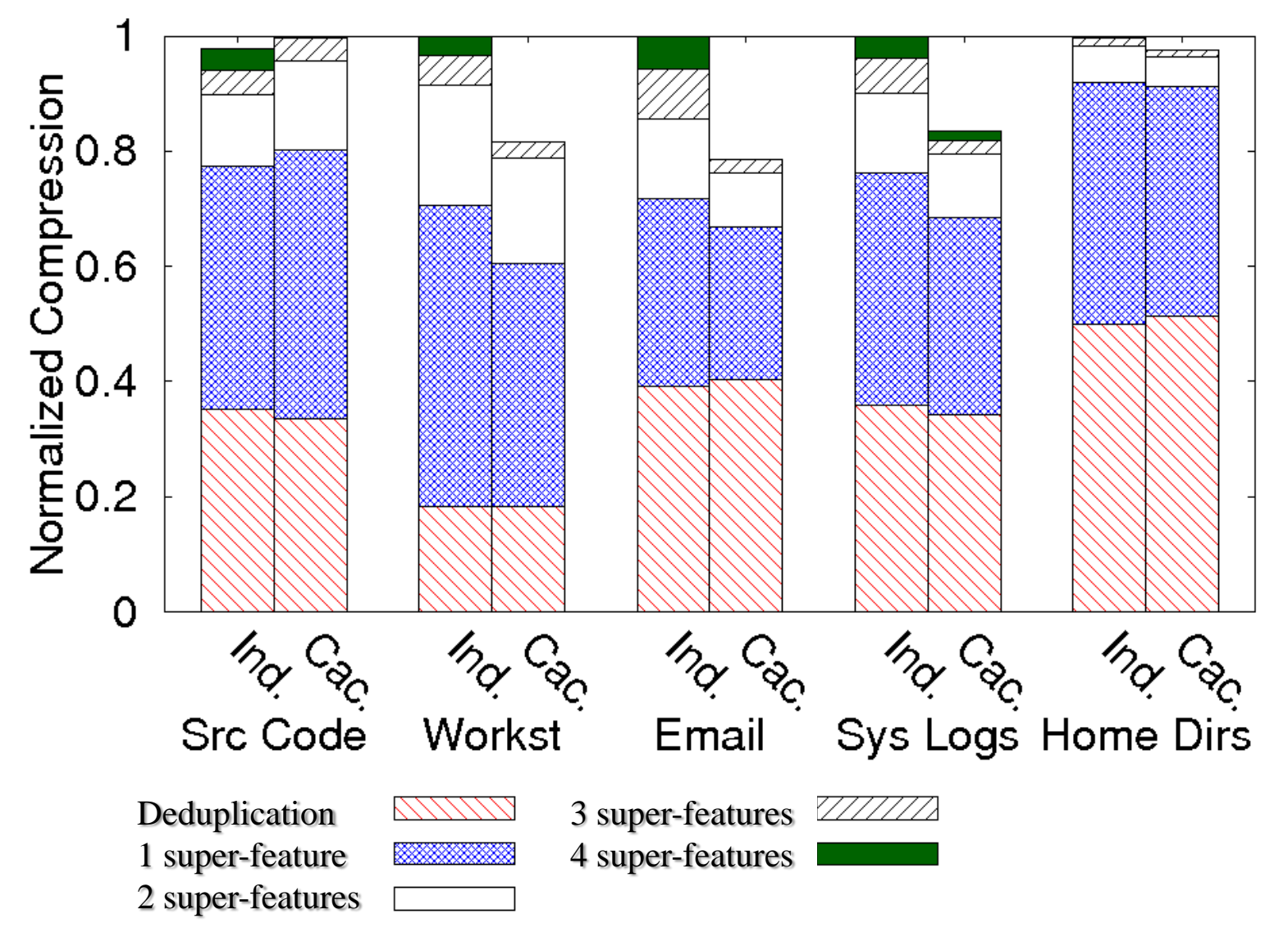
Central Office Backup System



Replication protocol simplified

For non-duplicate chunks, similar matches are found by checking for a sketch match in a cache

Index vs. Stream-Informed Cache



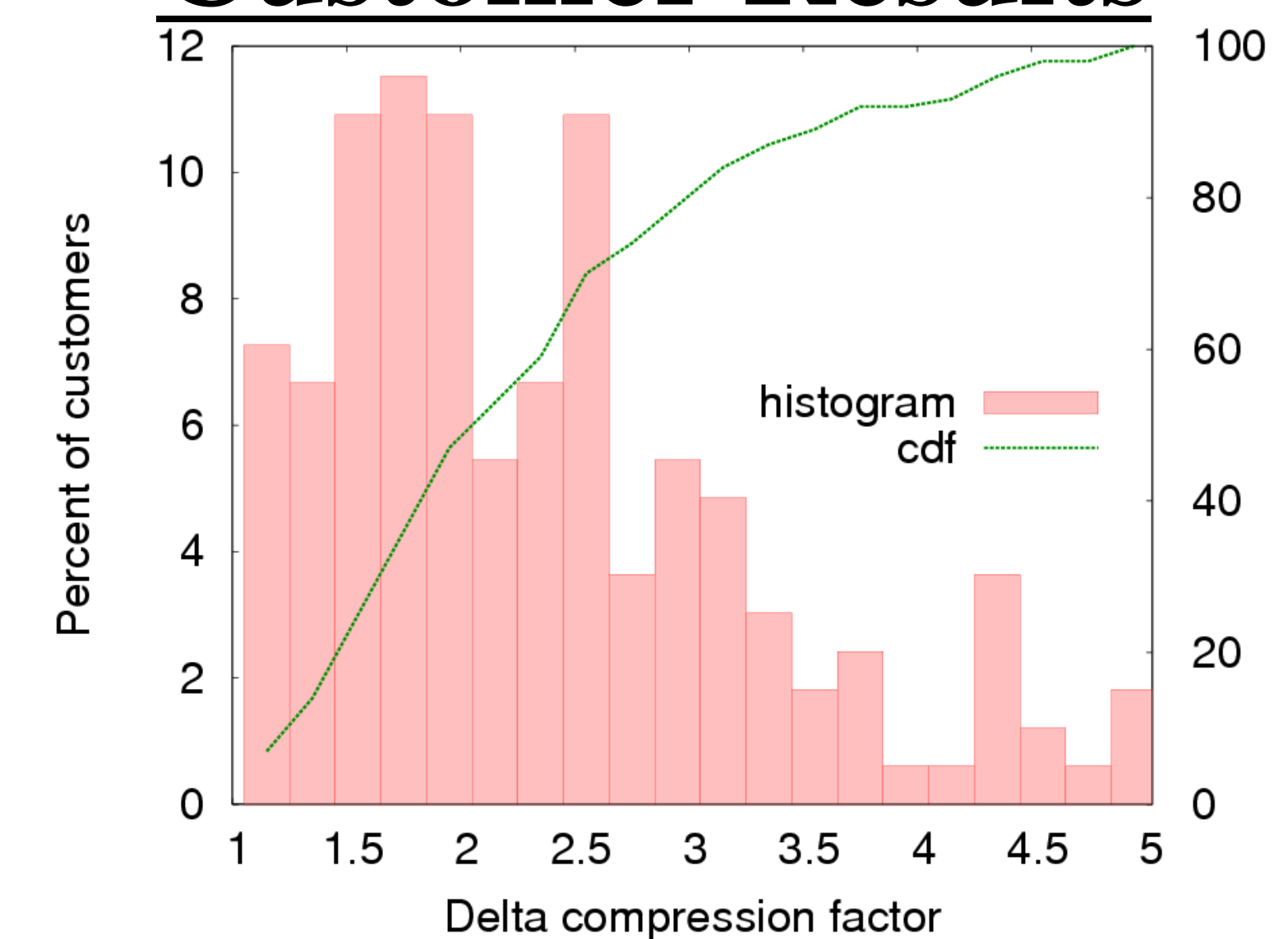
A stream-informed sketch cache (Cac.) has nearly as much compression as a full index (Ind.), and using two super-features offers more compression than an index with one.

Compression Results

Dataset	TB	Months	Dedup.	Choose GZ or combination of delta with GZ		
				GZ	Delta w/ GZ	Delta Improv.
Source Code	4.6	6	24.9X	7.2X	14.9X	2.1X
Workstations	4.9	6	5.7X	2.8X	8.8X	3.1X
Email	5.2	7	6.9X	3.1X	5.8X	1.9X
System Logs	5.4	4	57.9X	4.6X	10.2X	2.2X
Home Dirs.	12.9	3	31.7X	3.1X	5.5X	1.8X

Delta with GZ adds 2X more compression after deduplication than GZ alone. Compression factors are after an initial seeding period of one week.

Customer Results



The median customer experienced 2X delta compression after deduplication.

Conclusion

- Delta locality closely matches deduplication locality for backup workloads
- Stream-informed delta compression is effective with a small cache
- Product allows customers to replicate and protect twice as much data across a WAN