Prophecy: Using History for High-Throughput Fault Tolerance

Siddhartha Sen, SOSP 2009

Joint work with: Wyatt Lloyd

and Michael J. Freedman

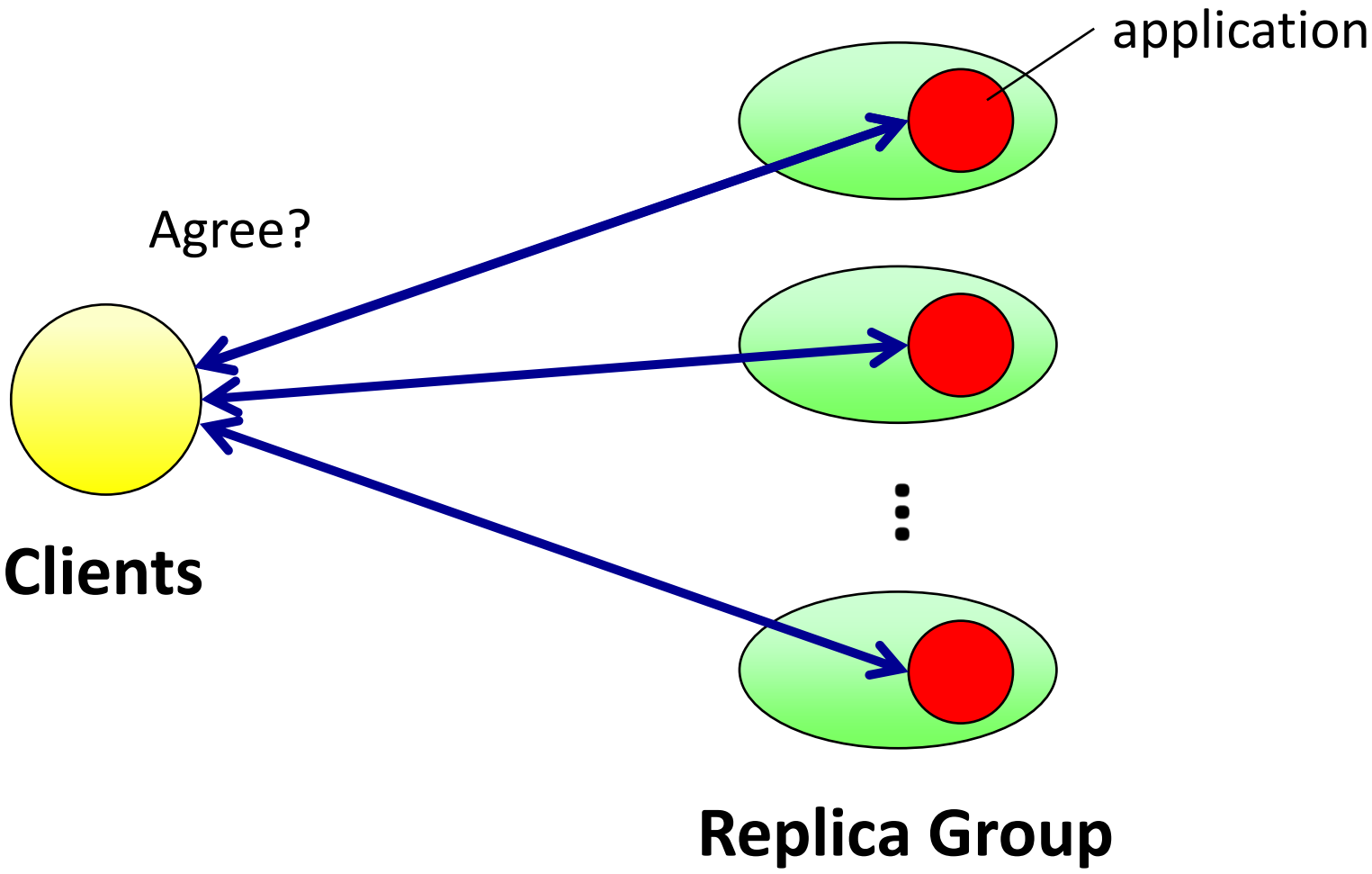
Princeton University

Background & Goals

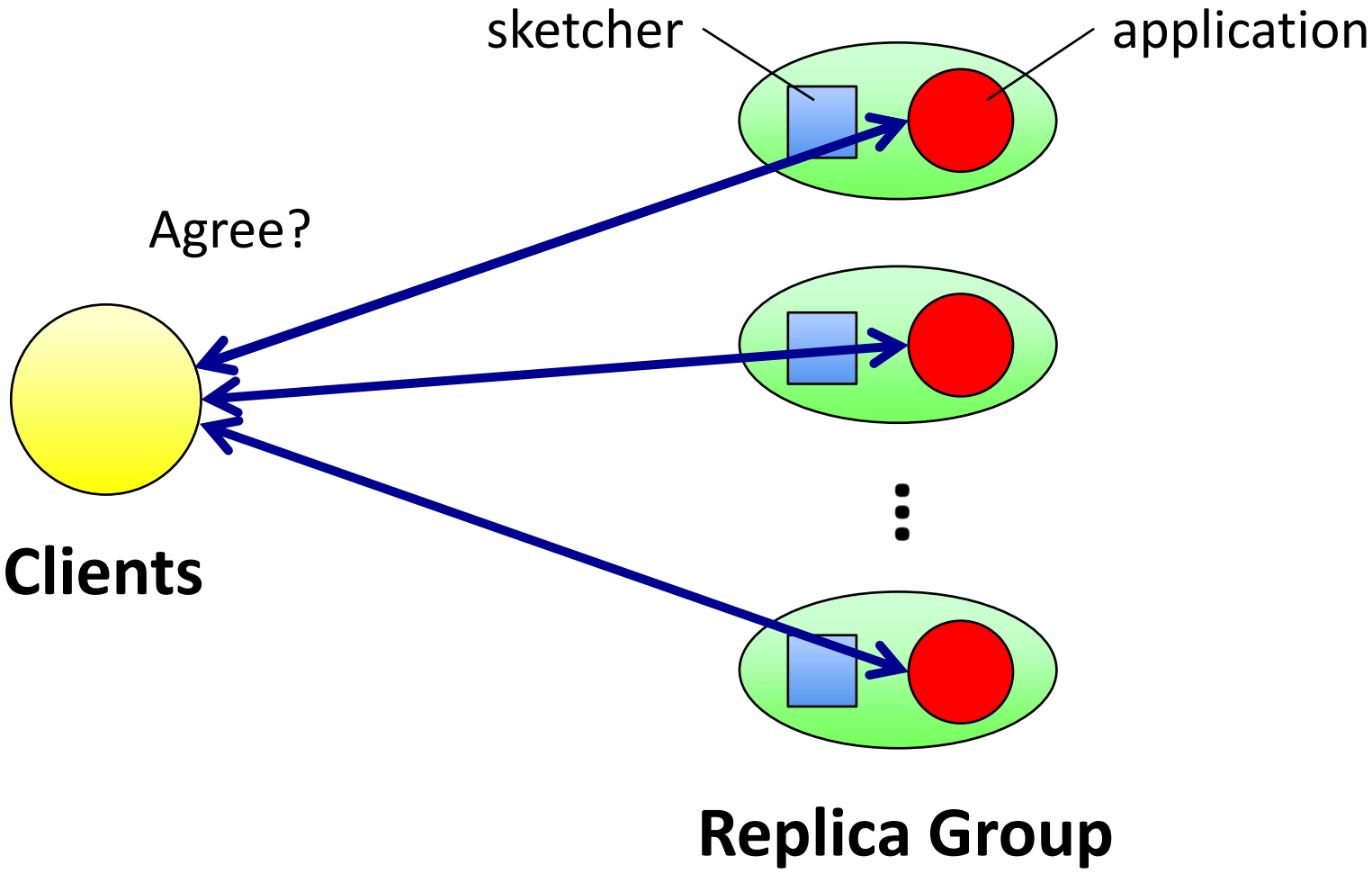
- Replication techniques that tolerate Byzantine (arbitrary) faults have poor throughput:
  - $3f + 1$ replicas needed to tolerate f faulty replicas
  - Every replica participates in every operation

- **Goal**: Leverage properties of Internet services to improve throughput
  - Focus on read-mostly workloads

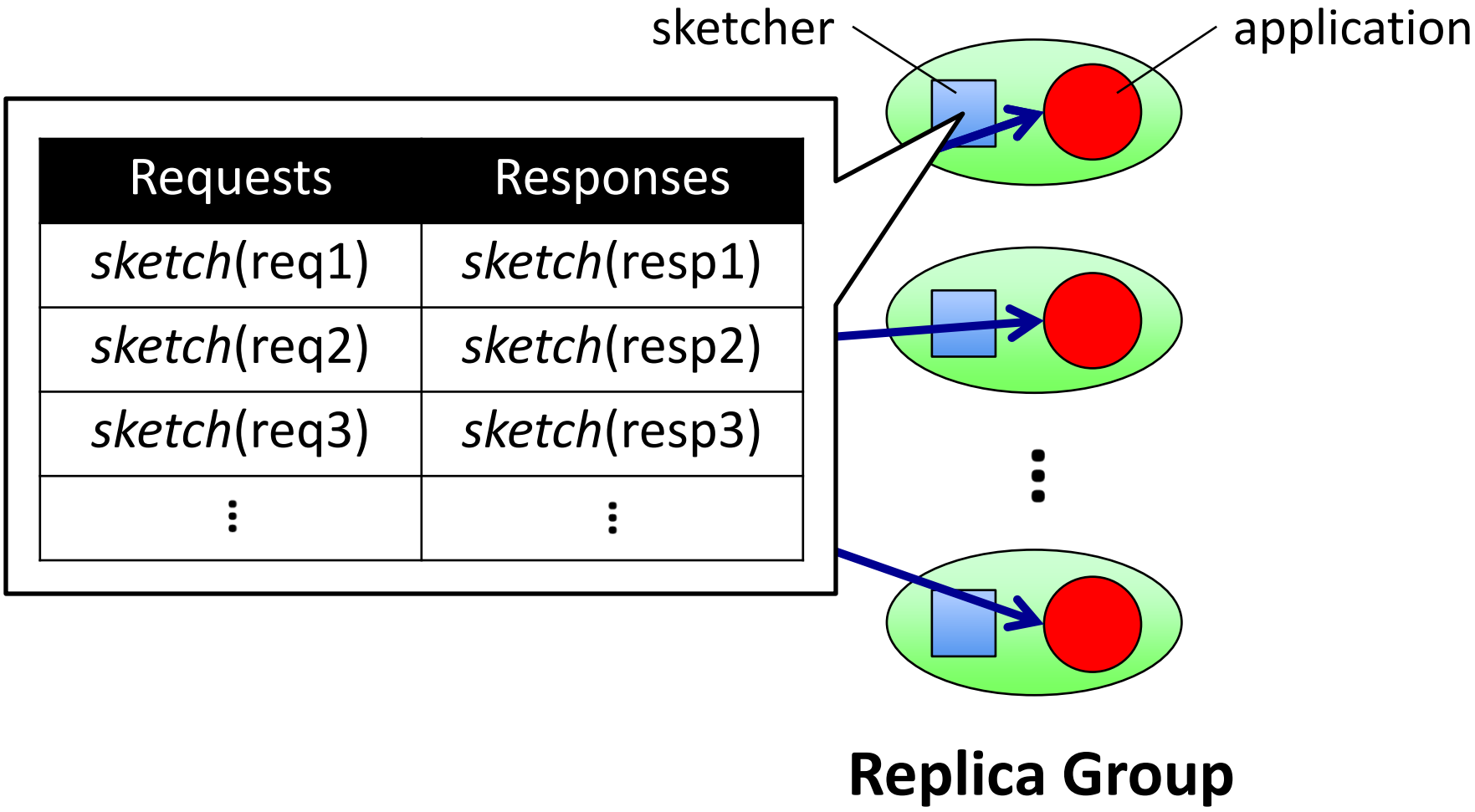
Traditional BFT



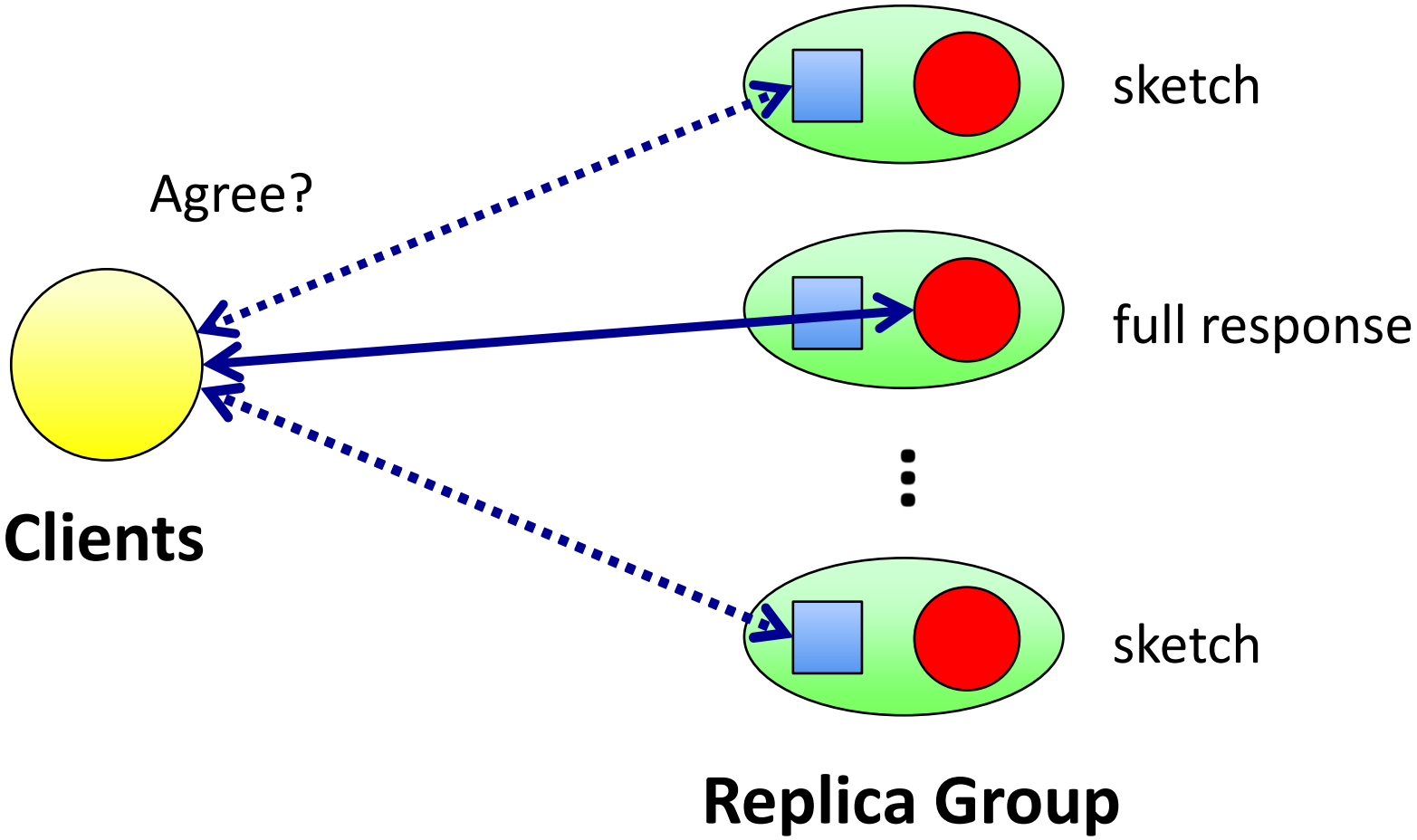
D-Prophecy: A distributed sketcher



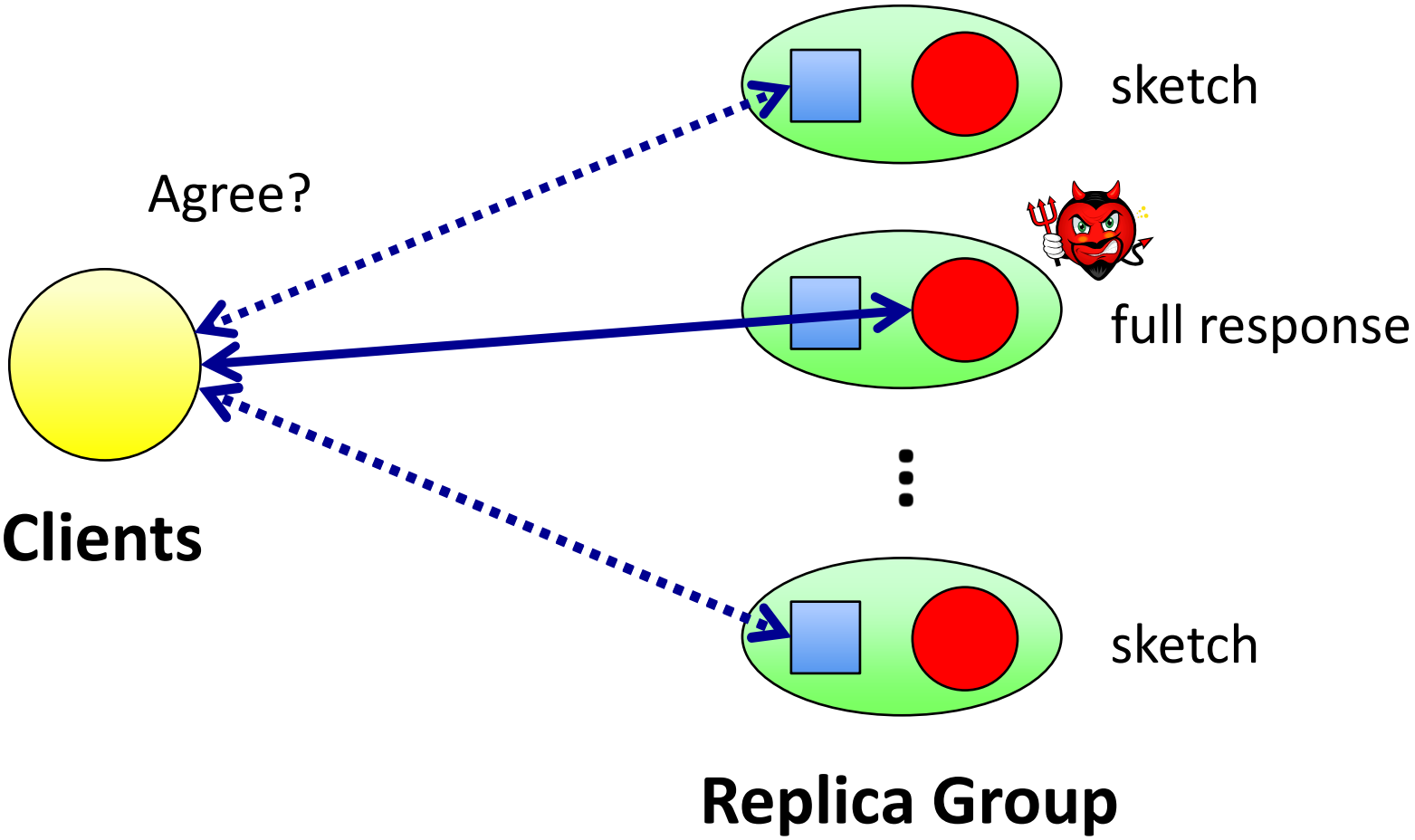
D-Prophecy: A distributed sketcher



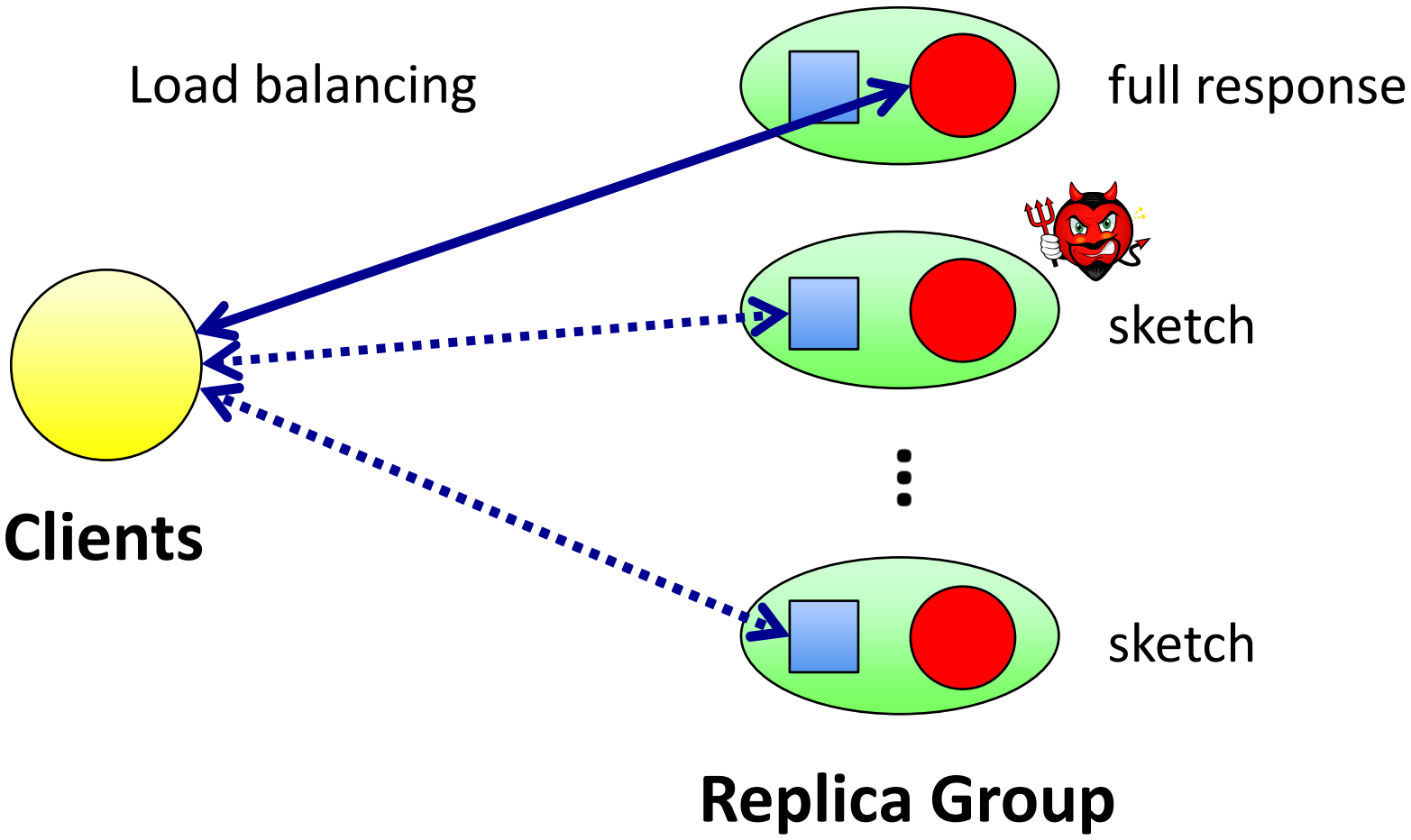
D-Prophecy: A distributed sketcher



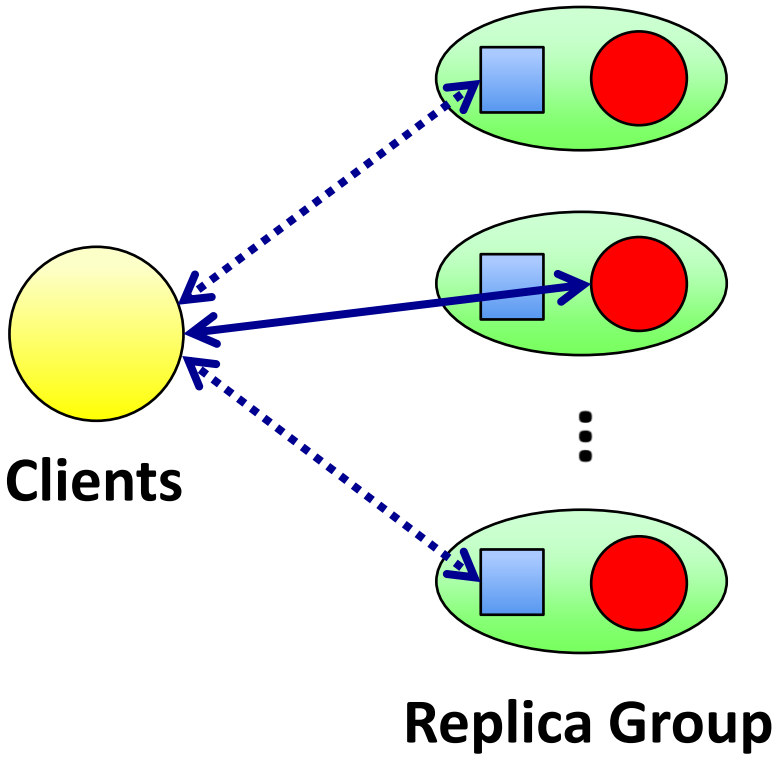
D-Prophecy: A distributed sketcher



D-Prophecy: A distributed sketcher



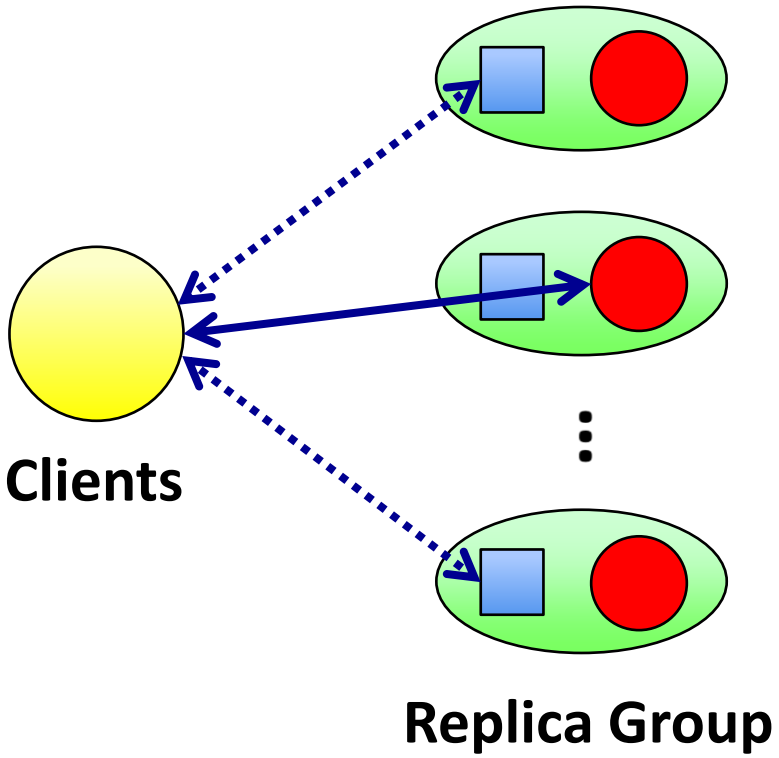
D-Prophecy: A distributed sketcher



Trades off consistency for performance

- Traditional BFT: executes read at every replica, linearizability
- D-Prophecy: in-memory lookup at most, delay-once linearizability

D-Prophecy: A distributed sketcher



Delay-once linearizability

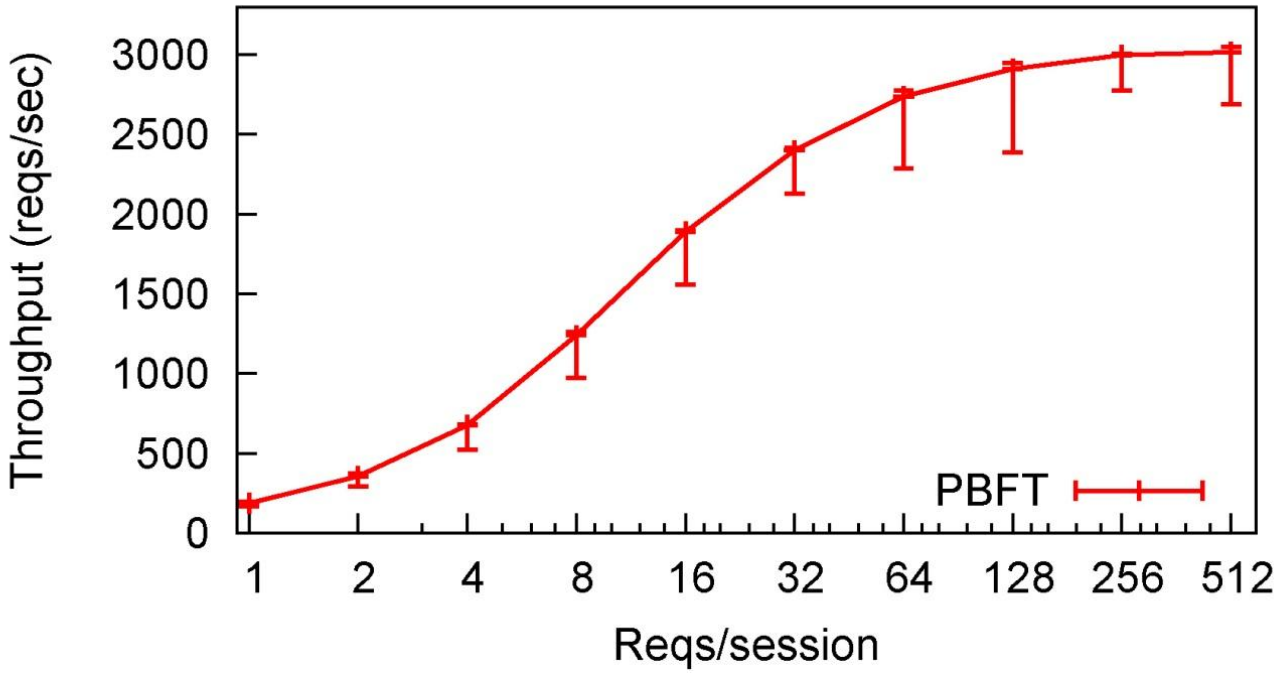
- Faulty replicas can return only stale (not arbitrary) data
- Load balancing limits repeated stale results

Internet services

- Unmodified clients
- Short-lived sessions

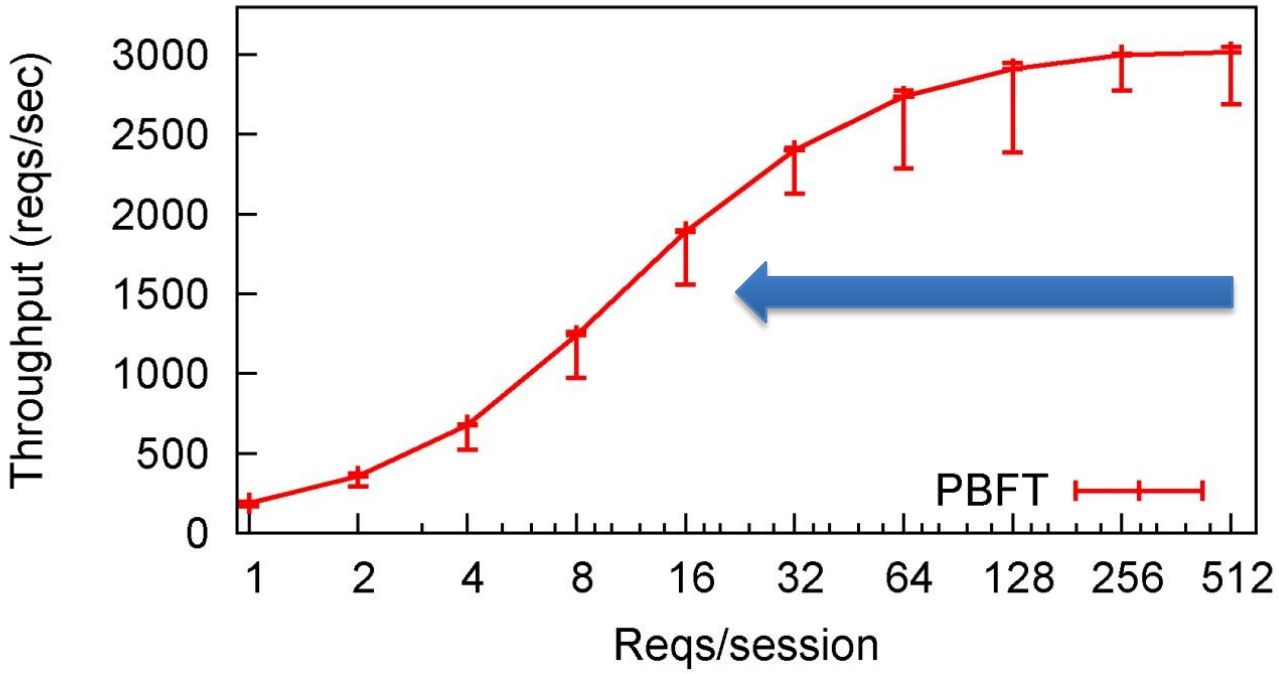
Internet services

- Unmodified clients
- Short-lived sessions

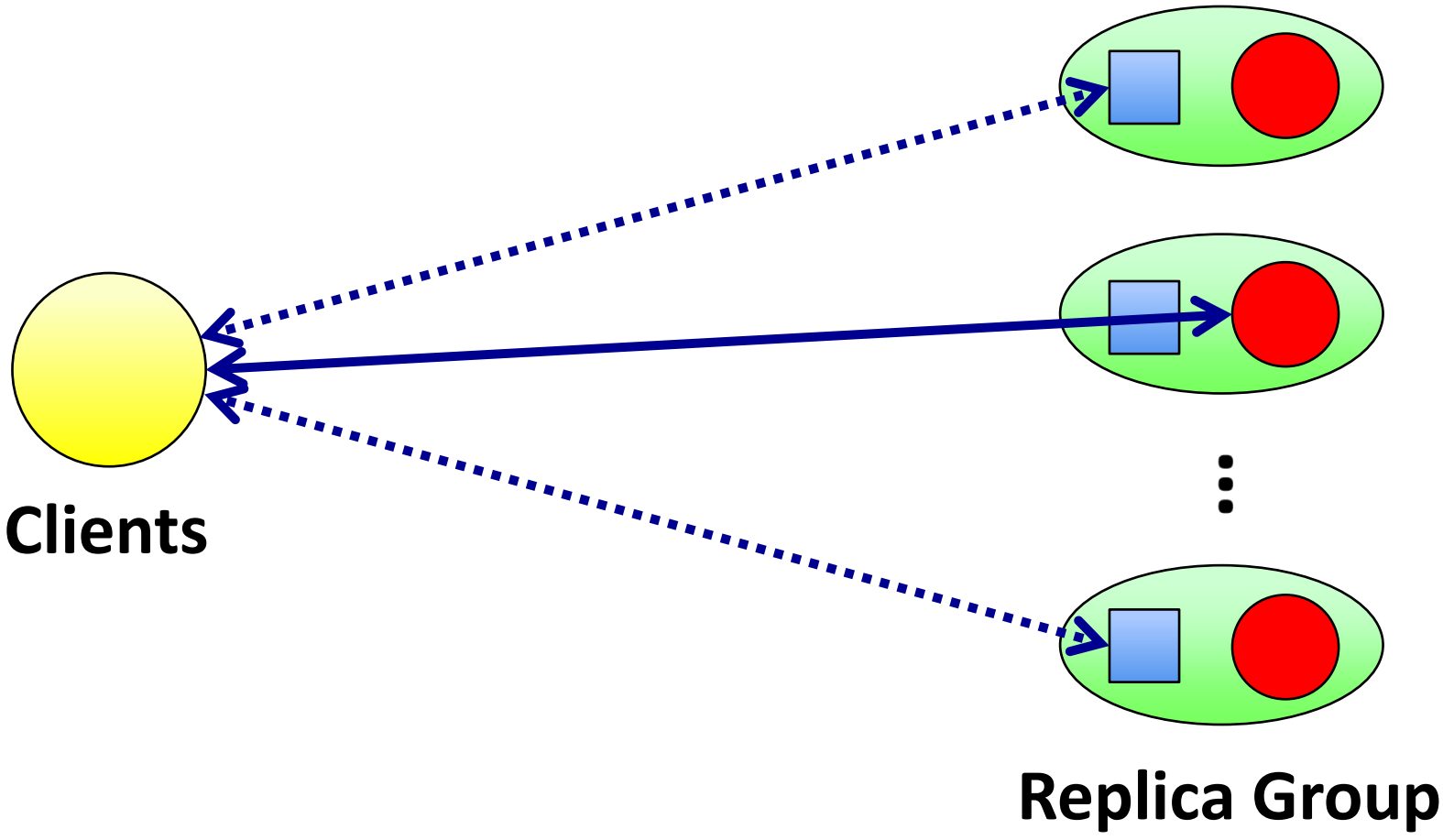


Internet services

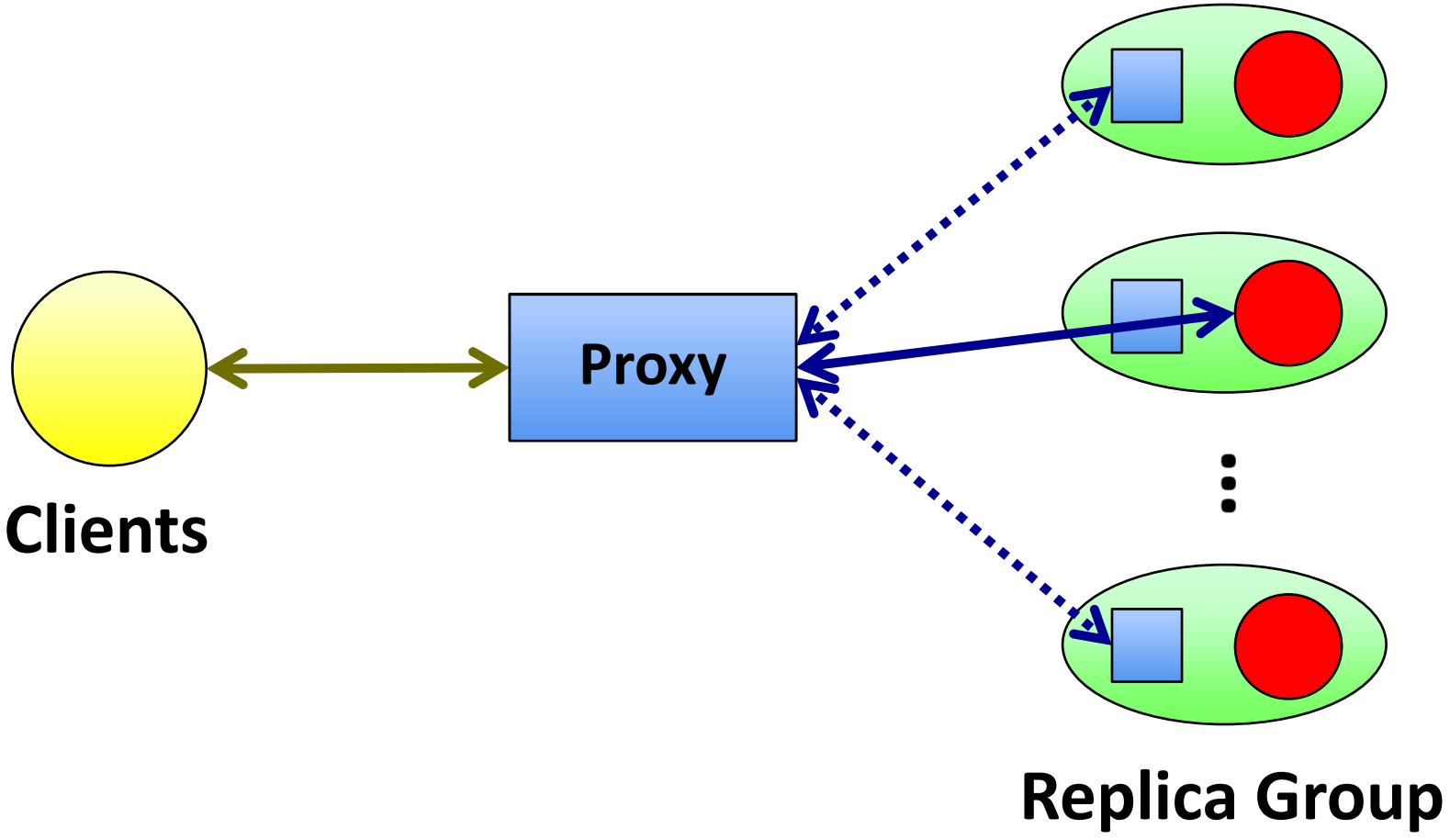
- Unmodified clients
- Short-lived sessions



Internet services

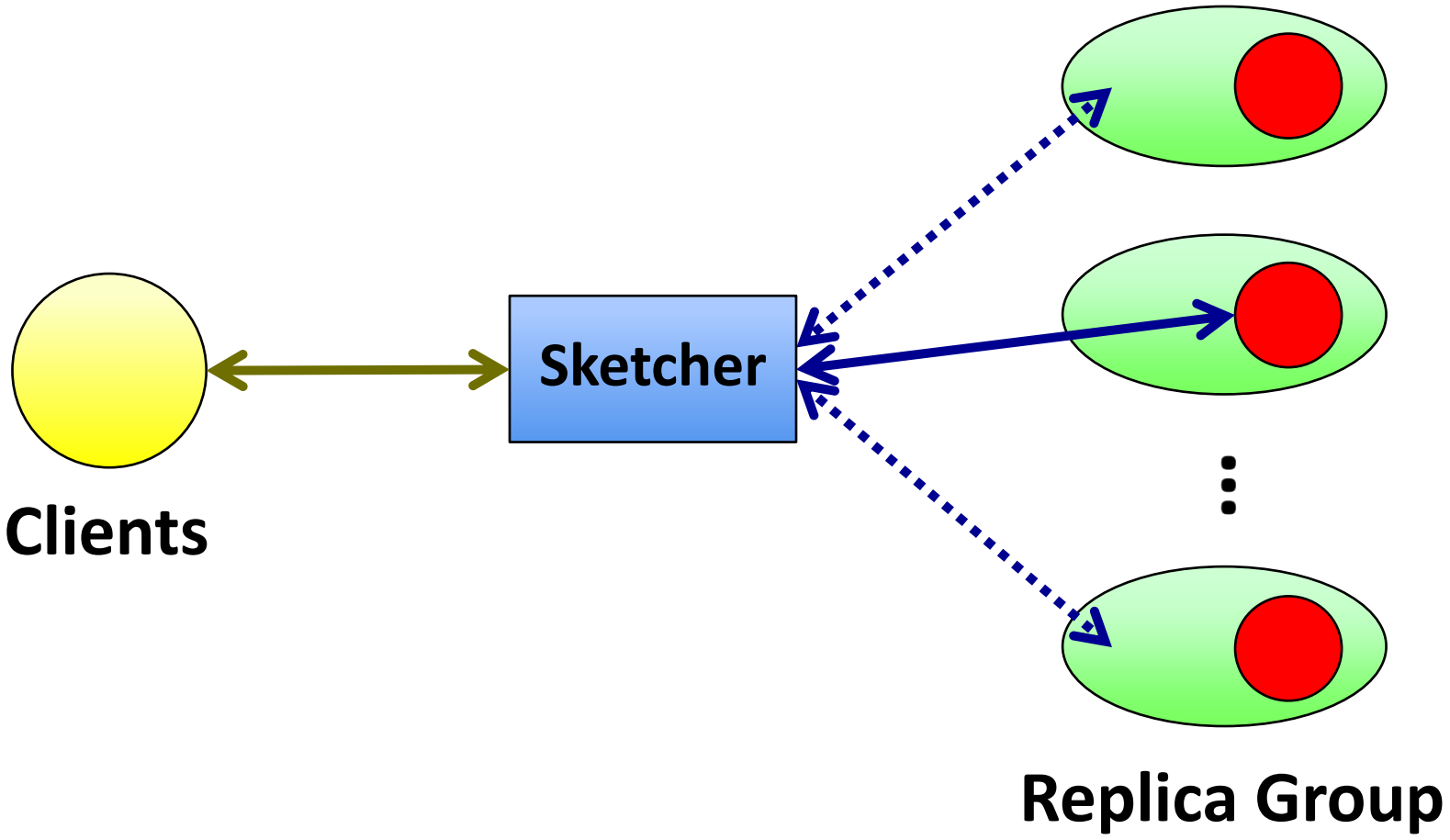


Internet services



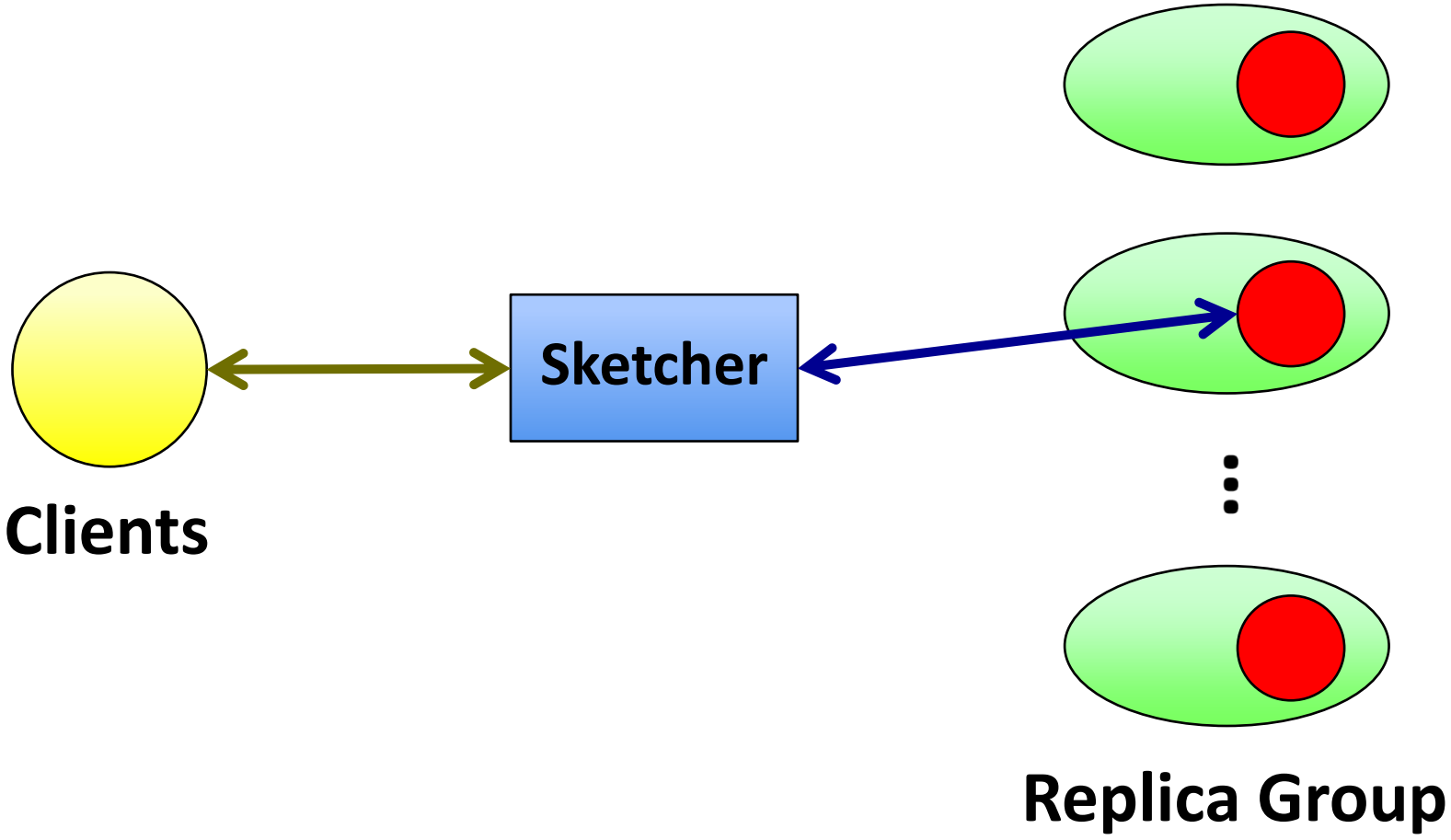
Internet services

Consolidate sketch tables



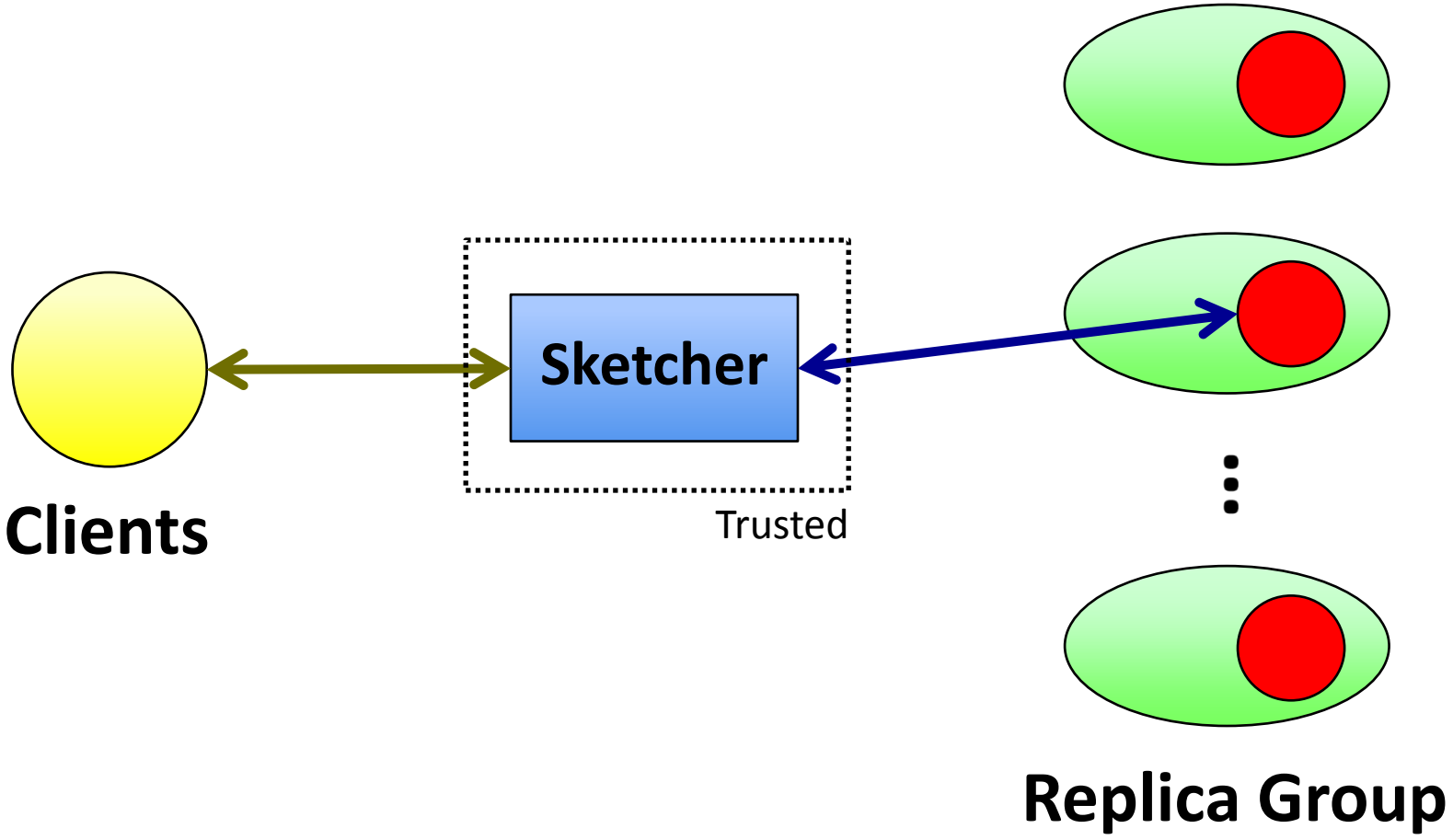
Internet services

Consolidate sketch tables



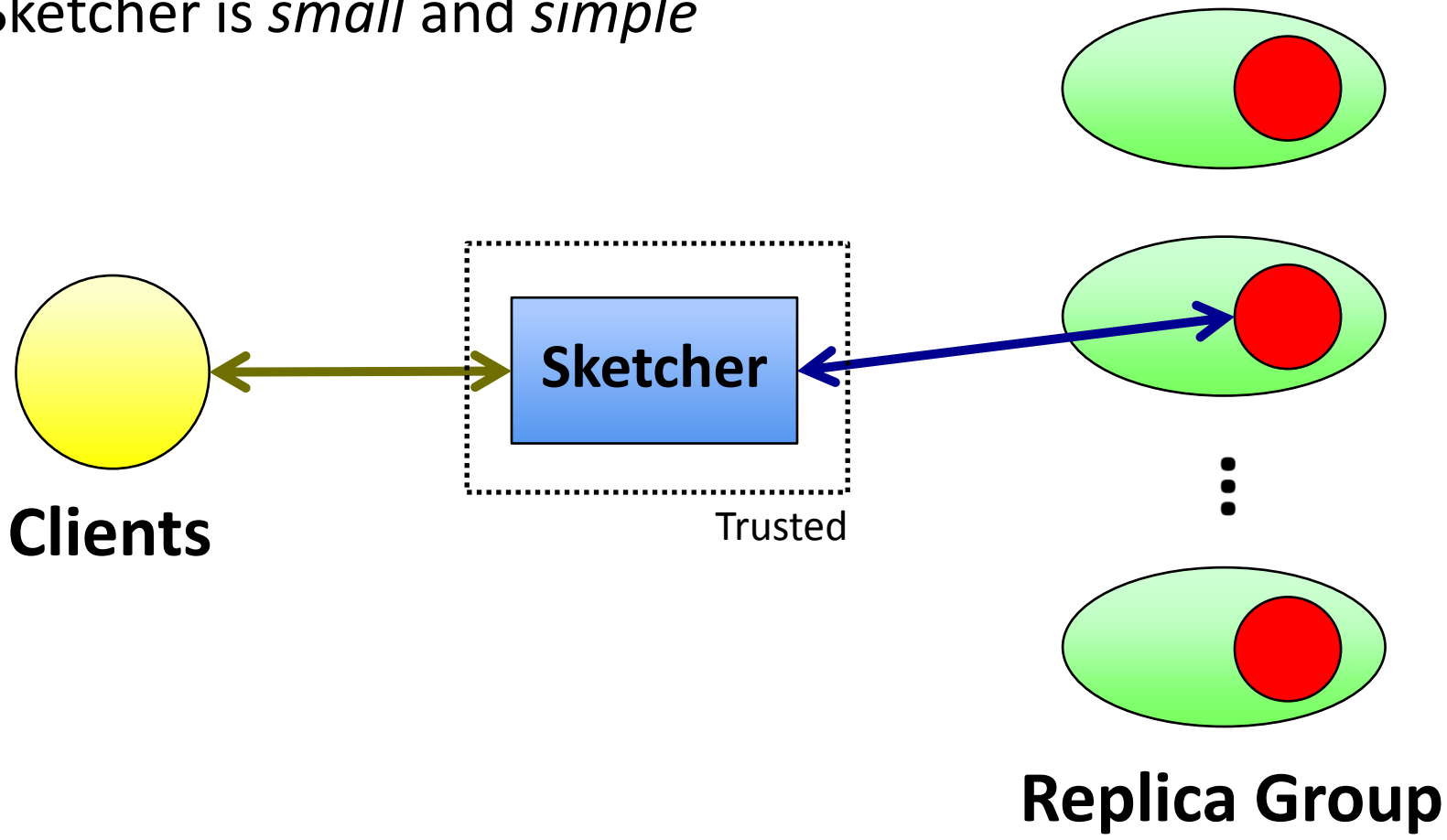
Prophecy: A trusted proxy

Sketcher must be fail-stop, but…

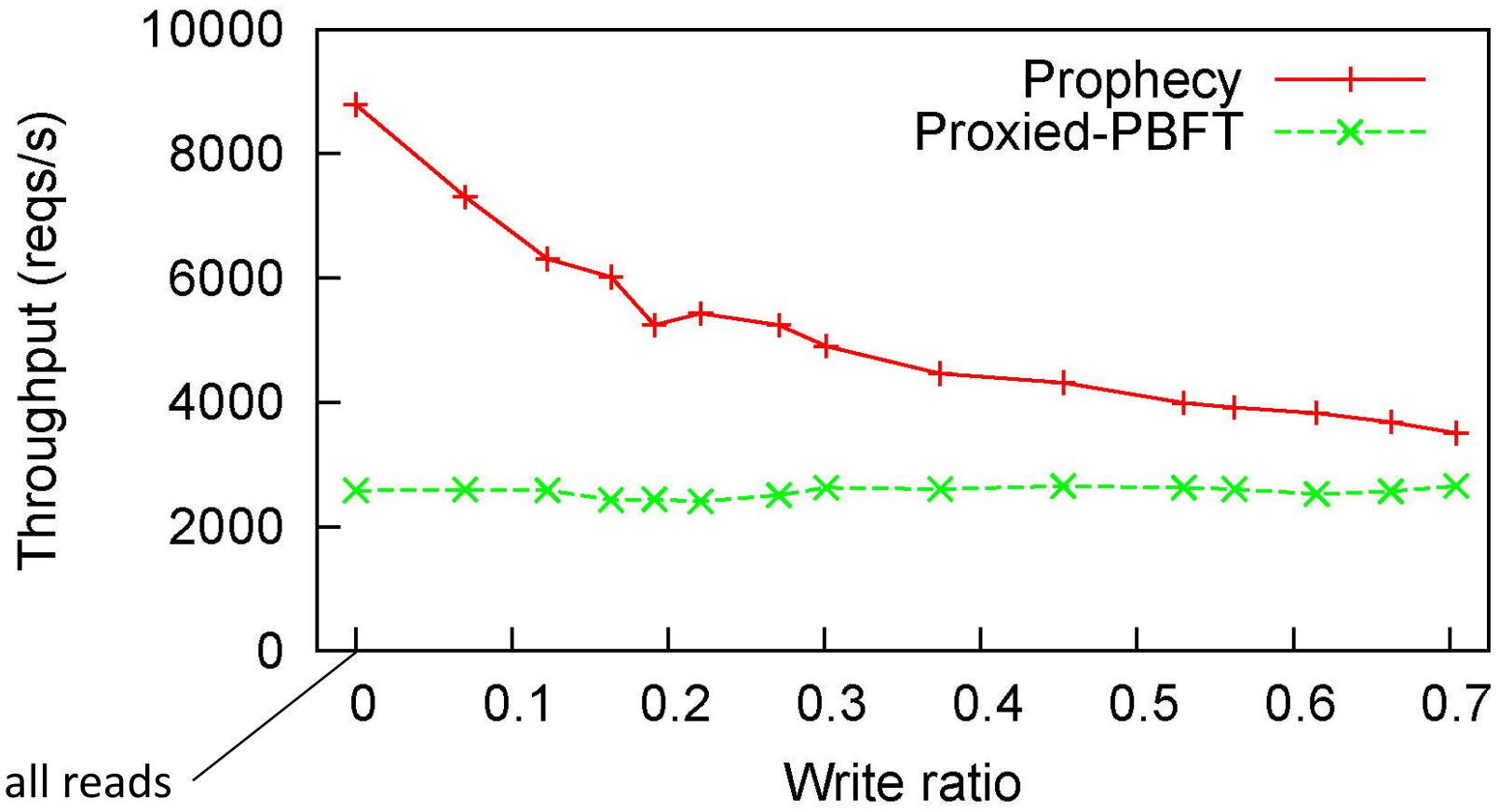


Prophecy: A trusted proxy

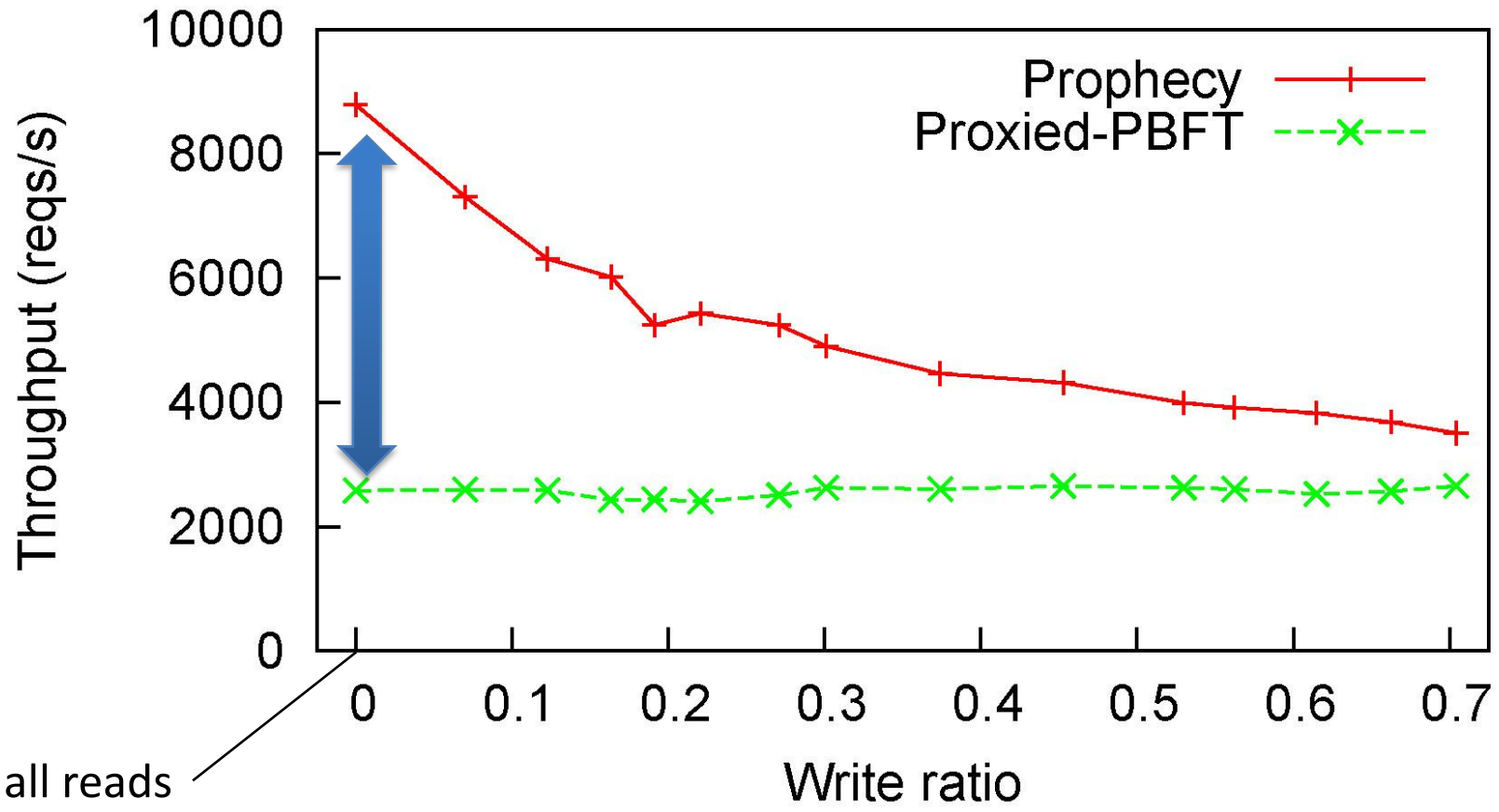
Already trust middleboxes for availability
Sketcher is *small* and *simple*



Prophecy: A trusted proxy



Prophecy: A trusted proxy



Summary

- Prophecy's performance on reads approaches that of unreplicated service
- Relaxes consistency to delay-once semantics

Summary

- Prophecy's performance on reads approaches that of unreplicated service
- Relaxes consistency to delay-once semantics
  - Not specific to BFT!
  - Can apply to Paxos, quorums, etc.