

Optimistic Replicated Two-Phase Commit

Irene Zhang, Naveen Kr. Sharma, Adriana Szekeres, Arvind Krishnamurthy, Dan R. K. Ports

Overview

ABSTRACT

Optimistic Replicated Two-Phase Commit (OR-2PC) is a new protocol for distributed transactional storage systems.

Conventional storage systems use:

Paxos/VR **Concurrency Control**
Two-Phase Commit

OR-2PC integrates these into a single protocol, taking advantage of cross-layer optimizations.

BENEFITS

Lower Latency

Commits in a single round-trip.

Higher Throughput

No bottlenecks at the leader.

Same Guarantees

Same transaction model and consistency guarantees.

TECHNIQUES

NEW

Optimistic Ordering Technique

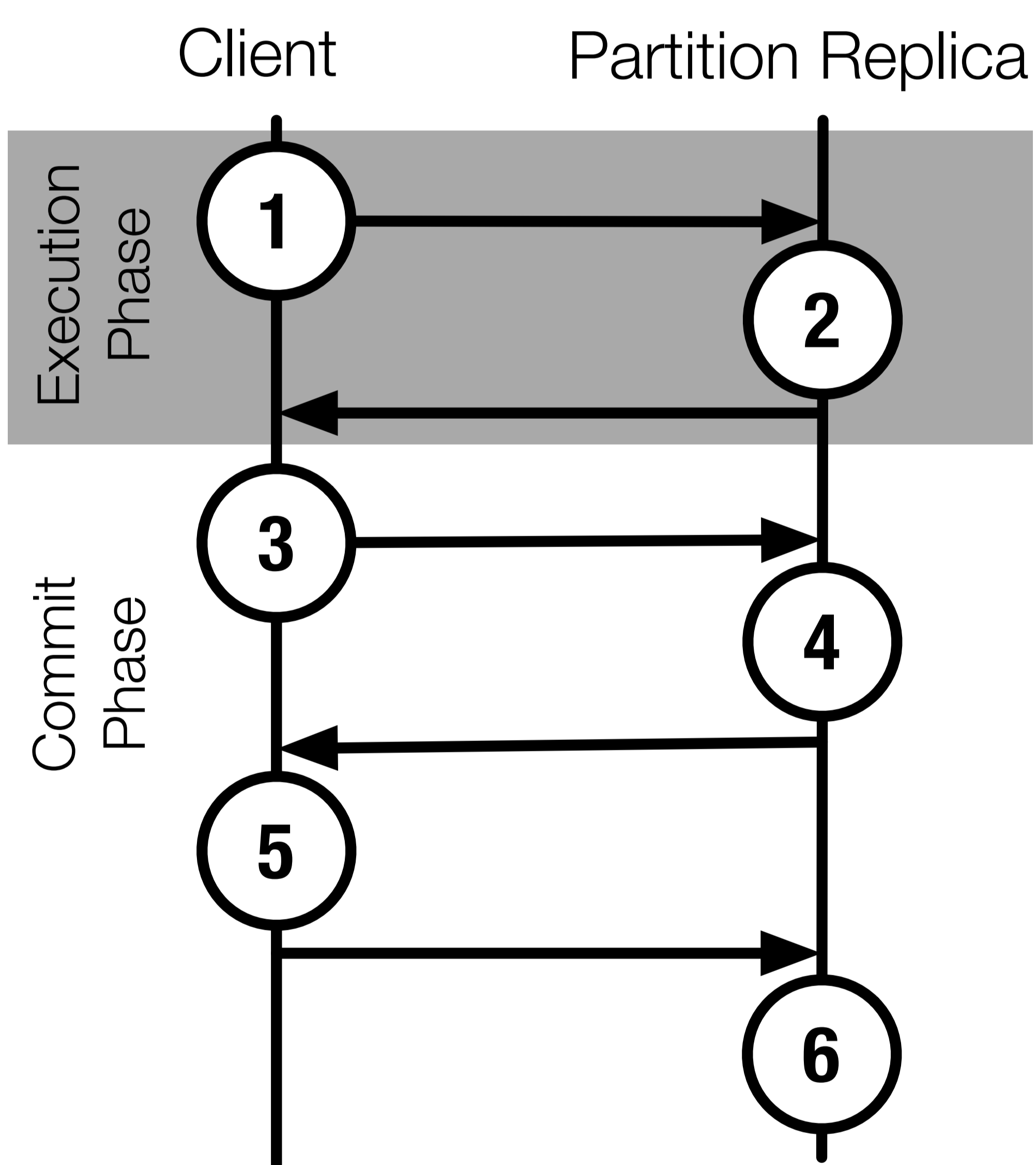
Uses loosely synchronized clocks to pick timestamps.

NEW

Optimistic Concurrency Control Mechanism

Enforces consistency across both transactions and replicas at the same time.

The Protocol



1. Client sends **READ** requests to any replica in the partition responsible for the object and buffers all writes.
2. Partition replica returns its latest version of the object for **READ** requests.
3. Client selects a proposed timestamp for the transaction and sends **PREPARE** with the proposed timestamp to all participant replicas.
4. Participant replicas validate the transaction at the proposed timestamp and respond **PREPARE-OK** if validation succeeds.
5. Once $f+1$ replicas in each participant partition respond **PREPARE-OK**, the client sends **COMMIT** messages.
6. Participant replicas commit the transaction at the proposed timestamp.

Evaluation

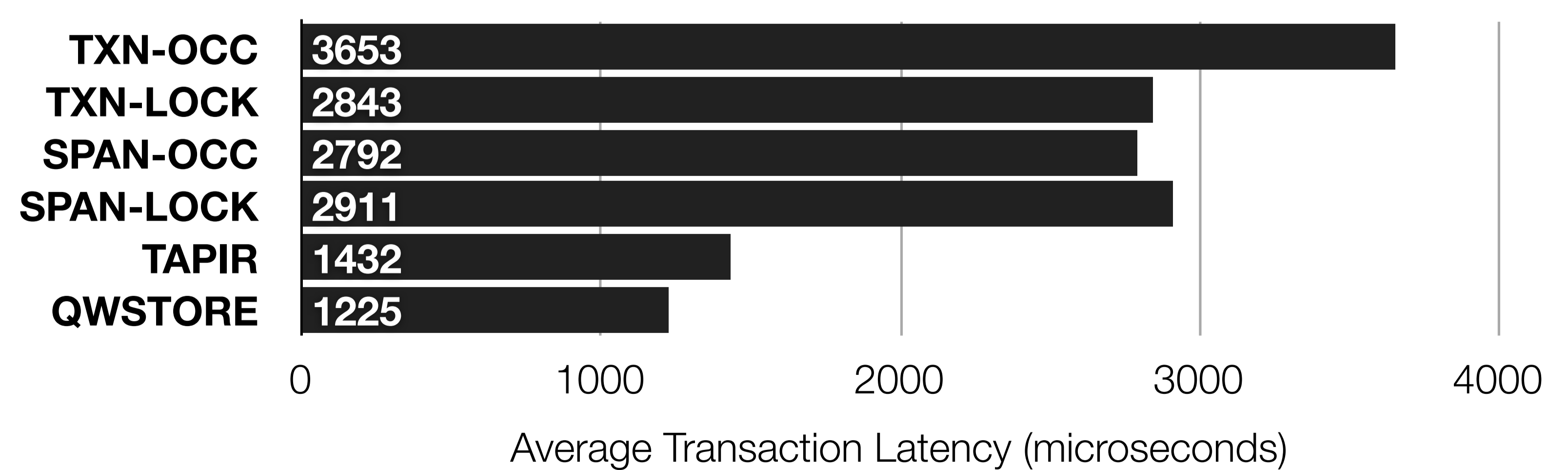
SUMMARY

Compared to conventional systems, OR-2PC provides:

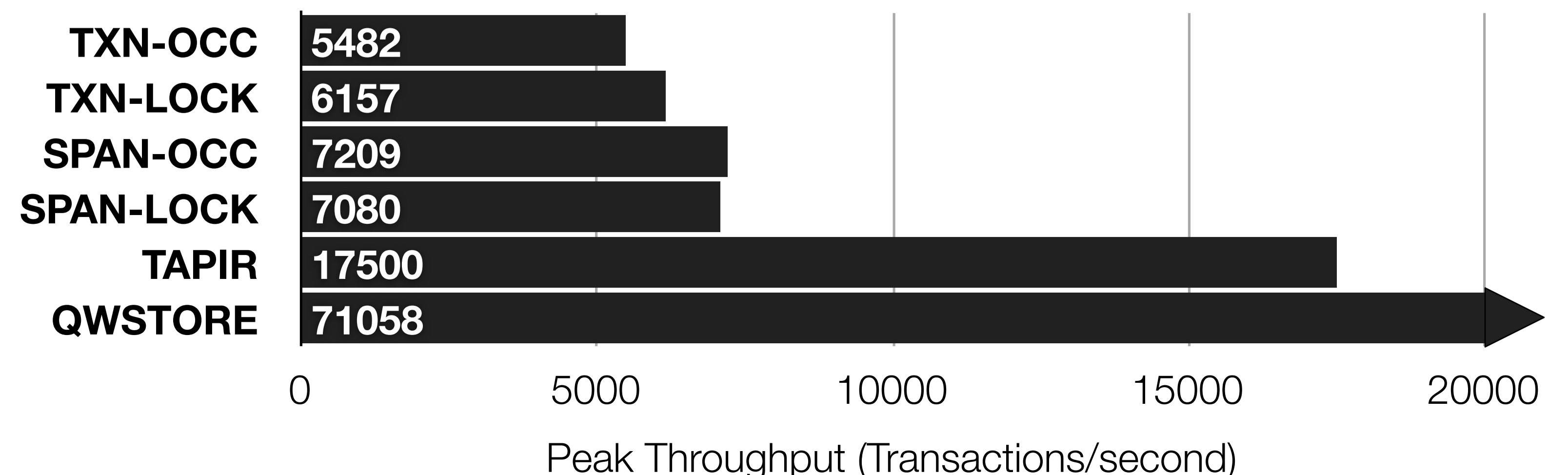
2x improvement in latency

3x improvement in throughput

LATENCY



THROUGHPUT



COMPARISON SYSTEMS

TXN-OCC = 2PC + OCC + VR
TXN-LOCK = 2PC + S2PL + VR
SPAN-OCC = Spanner protocol + OCC
SPAN-LOCK = Spanner protocol (S2PL)
TAPIR = OR-2PC
QWSTORE = Quorum Writes (Dynamo)