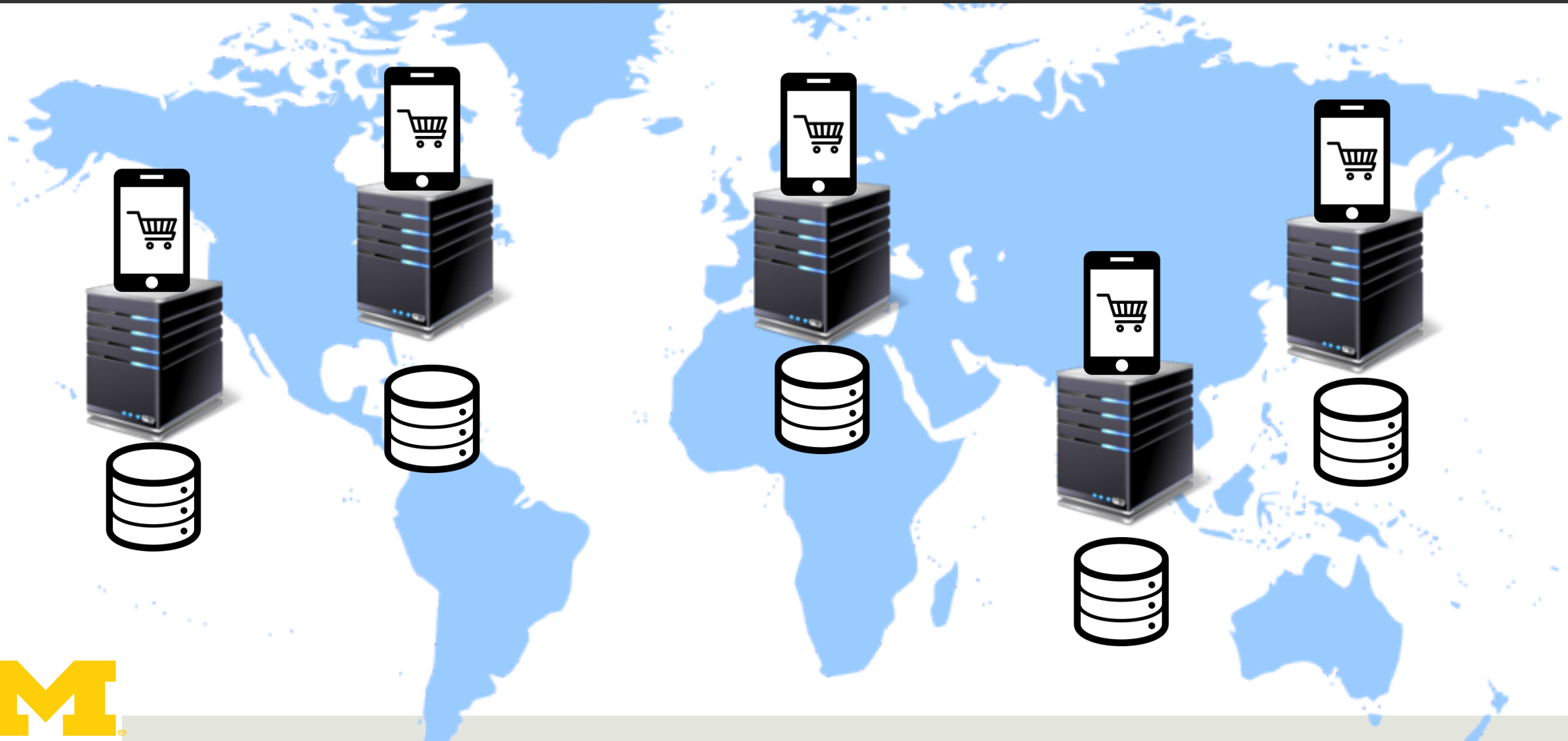
Bolt-On Global Consistency for the Cloud

Zhe Wu, *Edward Wijaya, Muhammed Uluyol, Harsha V. Madhyastha*

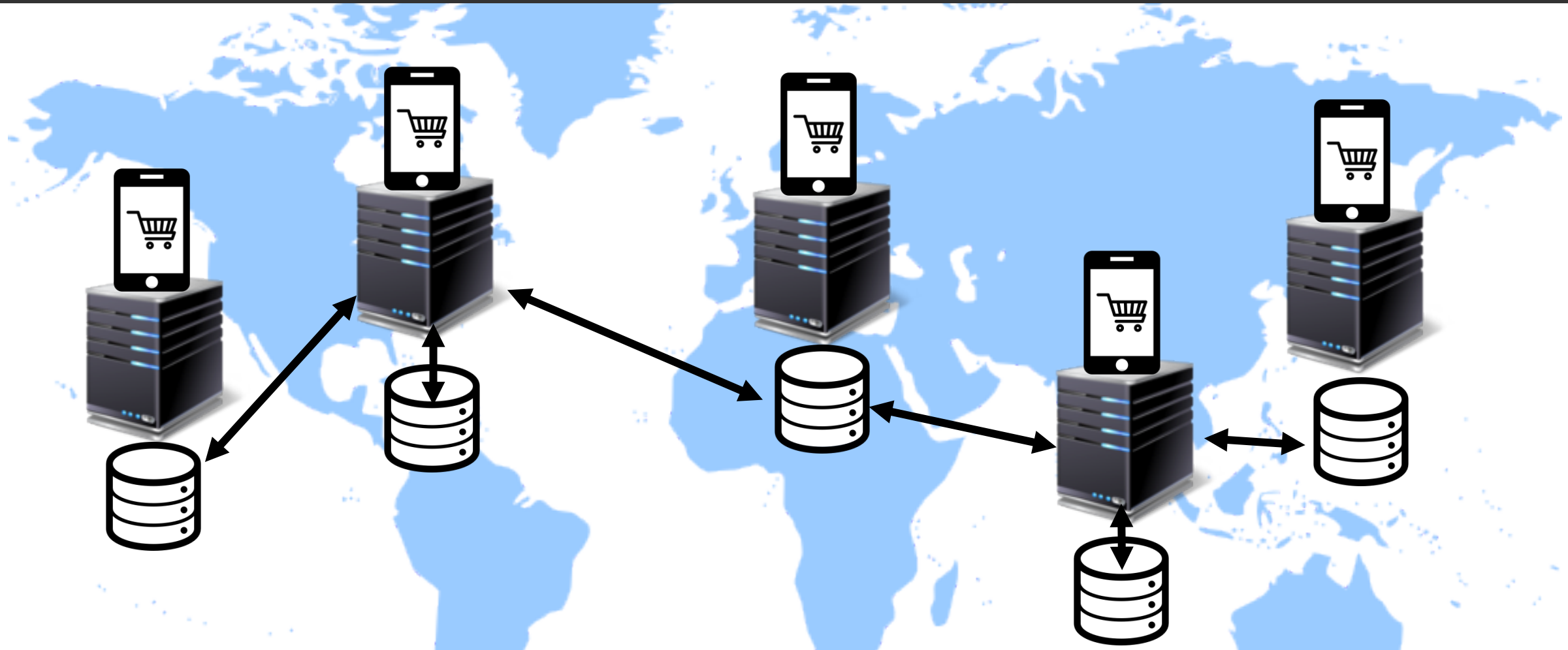
University of Michigan

Geo-distribution for Low Latency

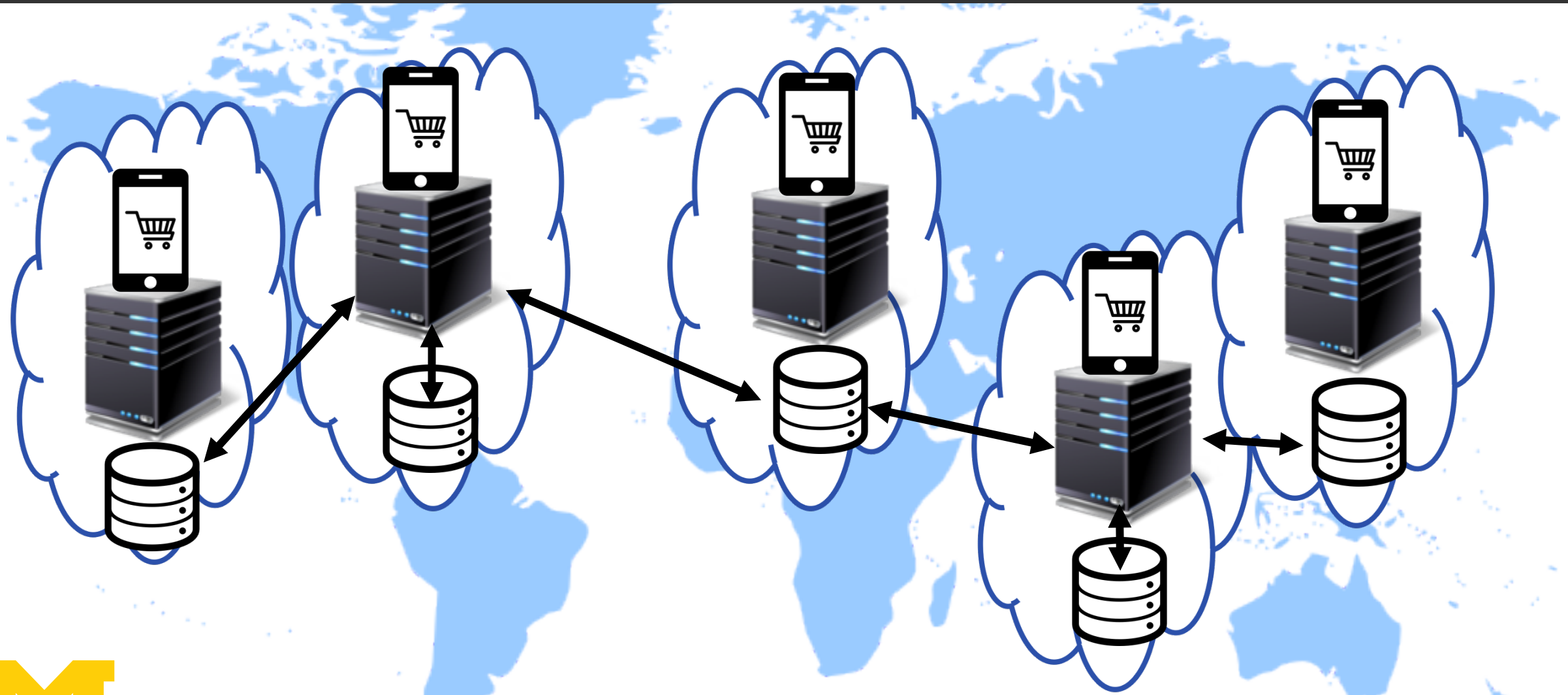
Geo-distribution Requires Data Replication



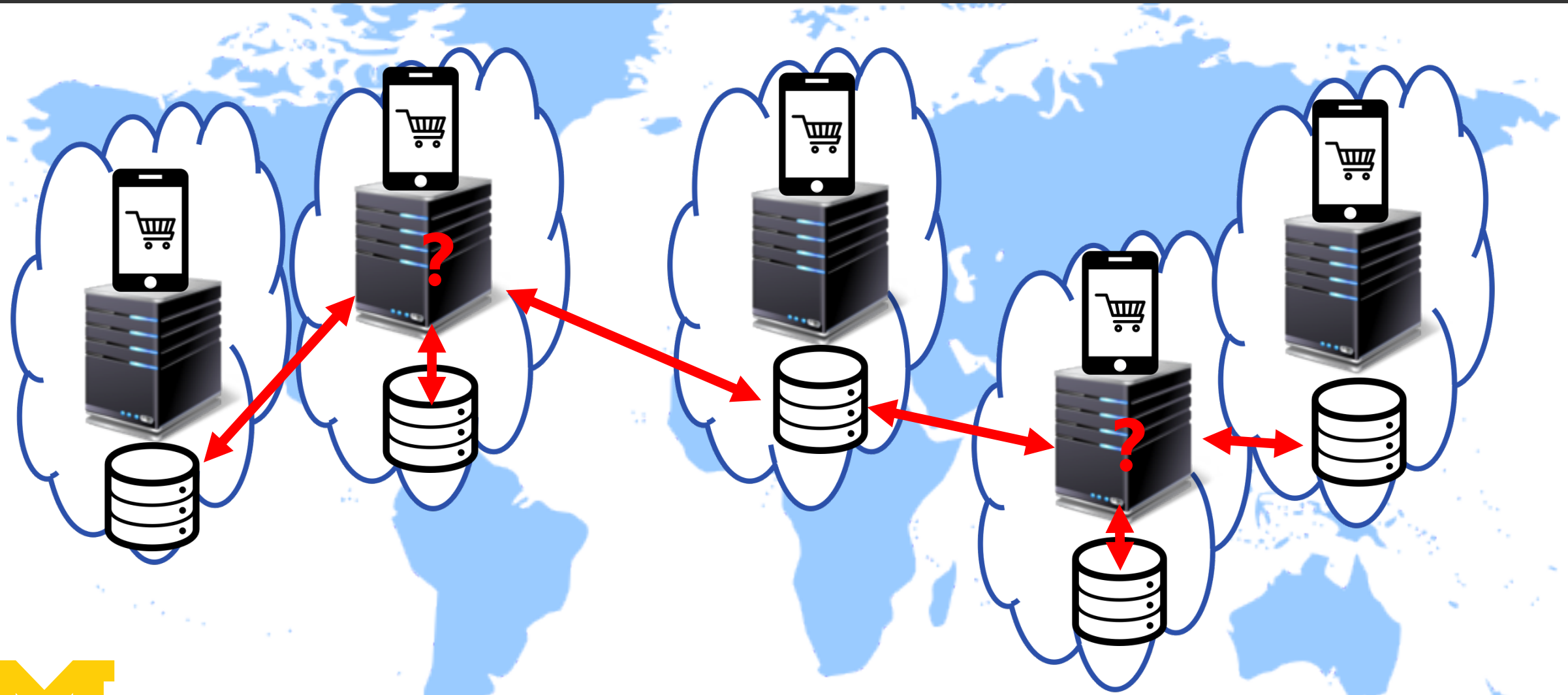
Geo-distribution Requires Data Replication



Cloud Simplifies App Deployment

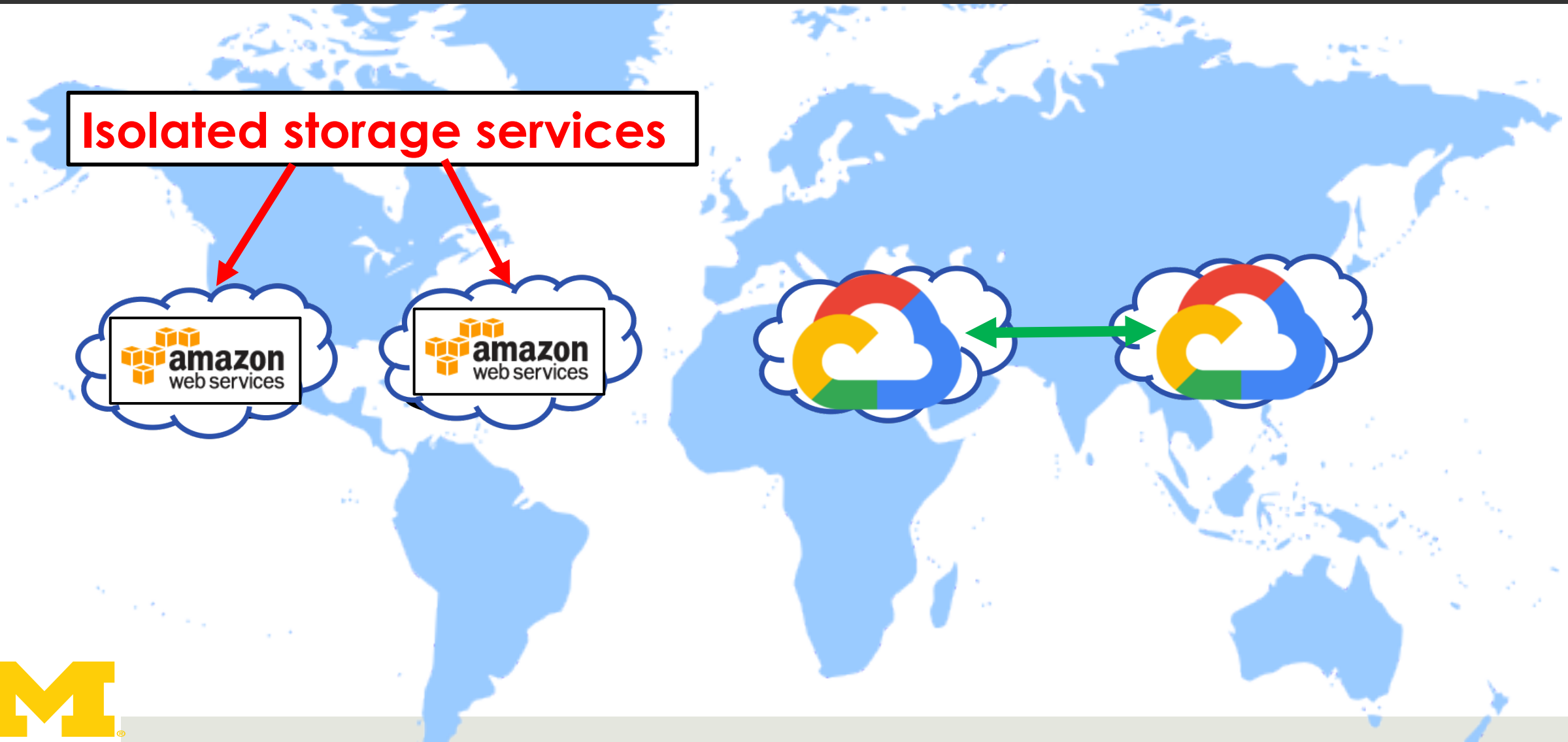


Cloud Simplifies App Deployment

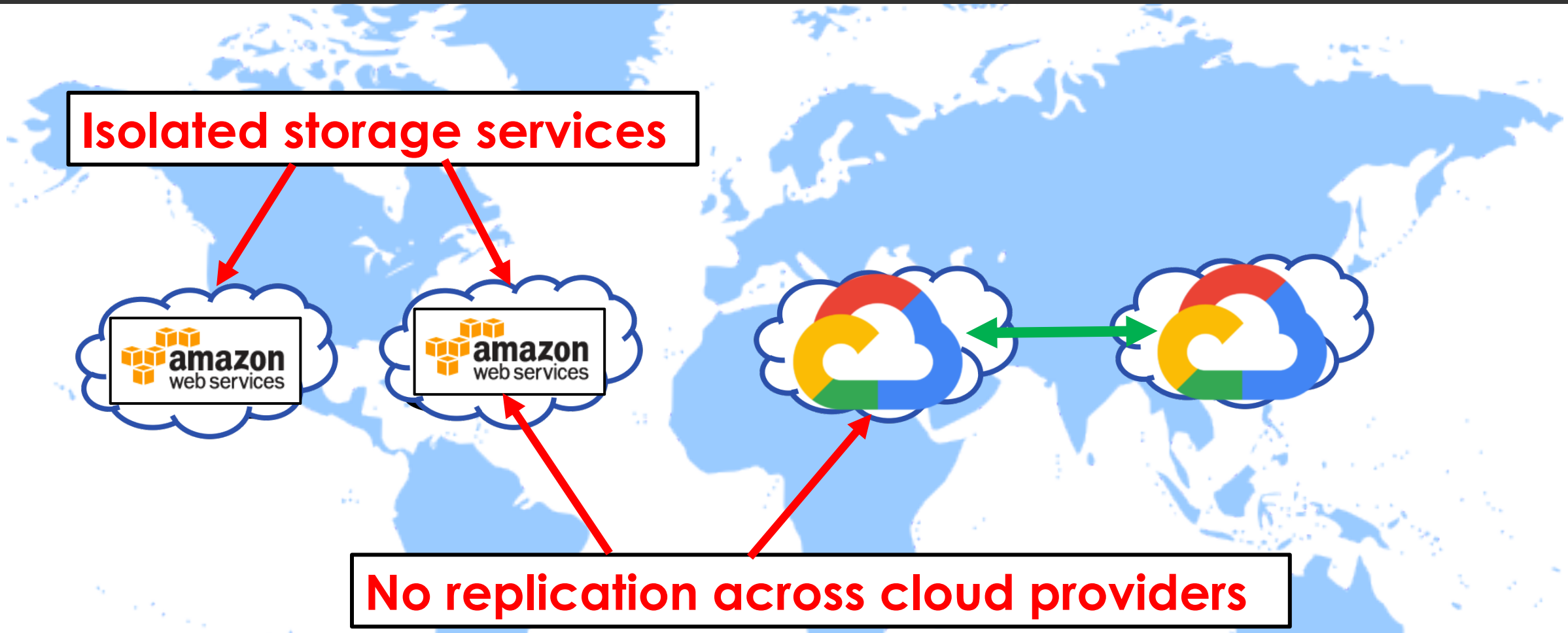


Application Needs to Manage Replication

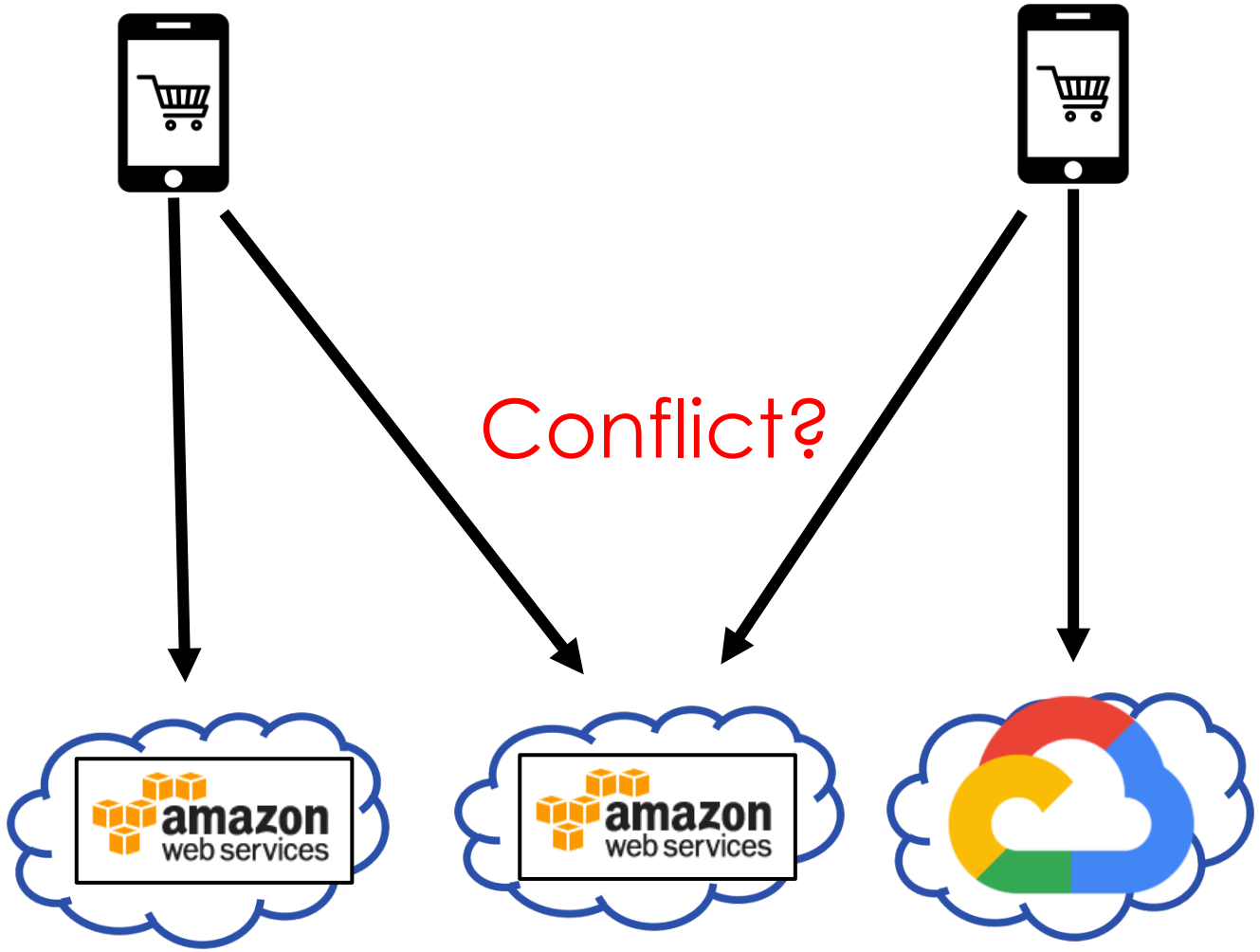
Application Needs to Manage Replication



Application Needs to Manage Replication



Challenges for Data Replication in Cloud



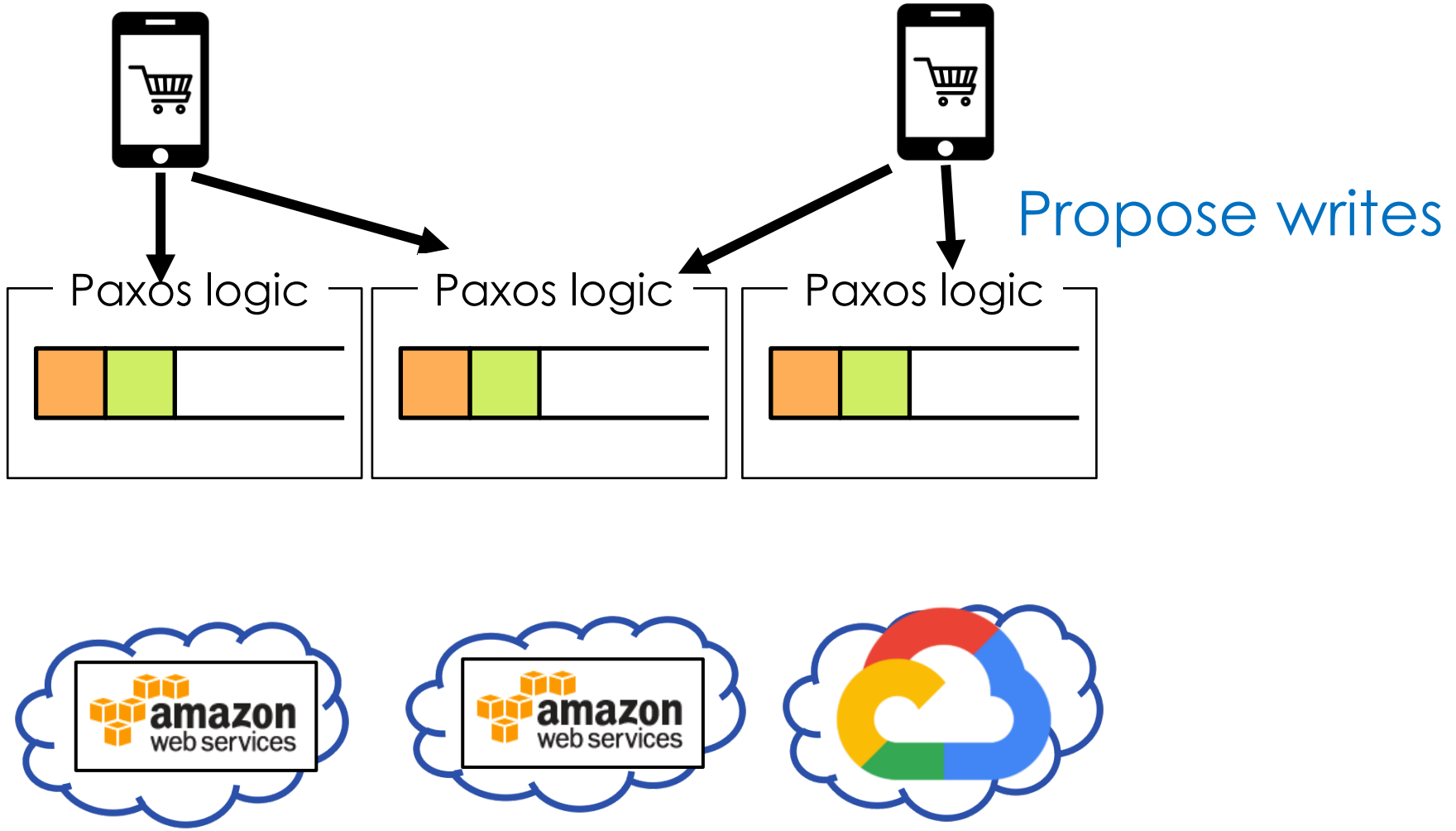
Challenges for Data Replication in Cloud

Paxos

Megastore(CIDR'11)
Spanner(OSDI'12)
MDCC(Eurosys'13)
Tapir(SOSP'15) .....

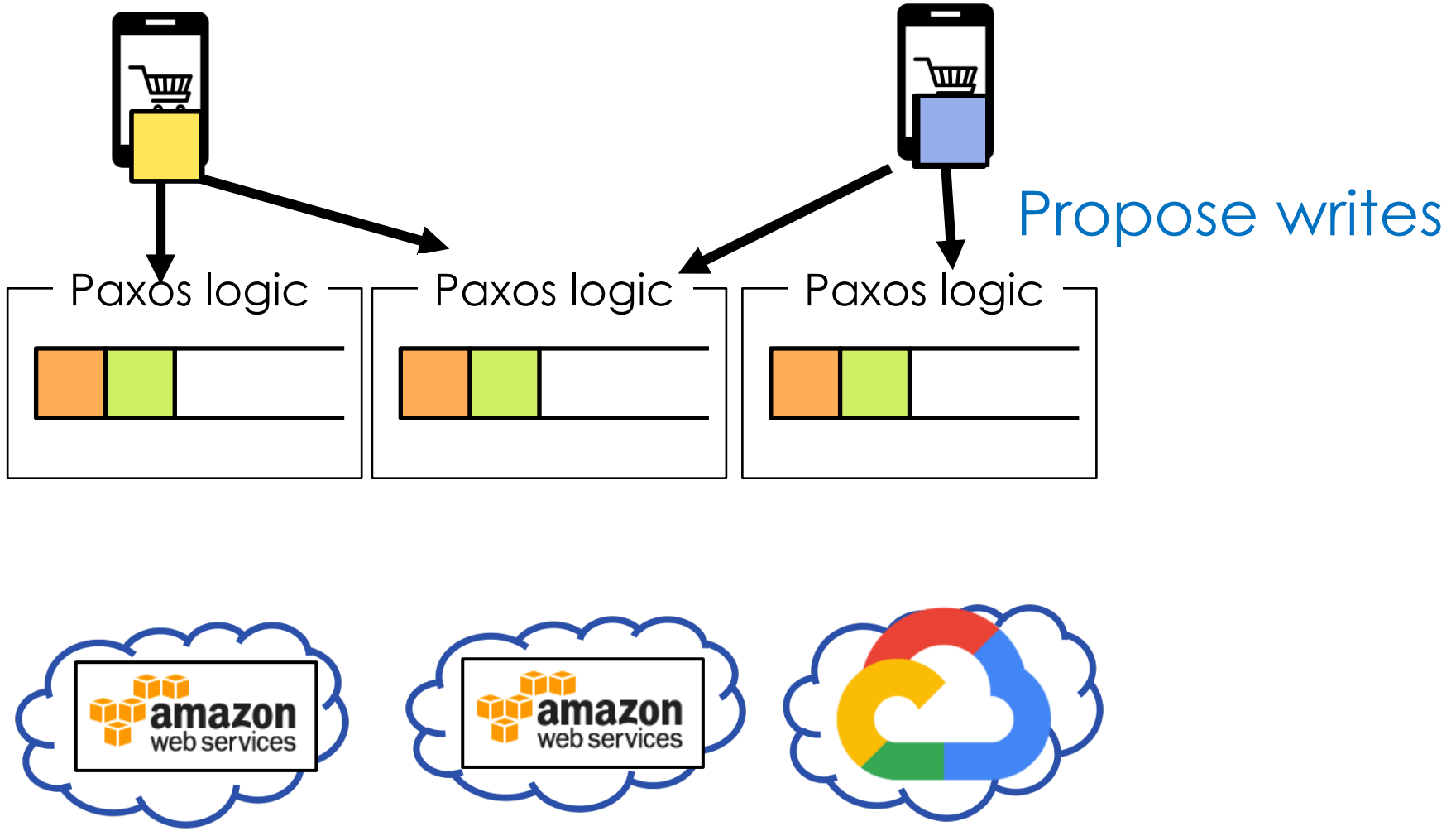
Challenges for Data Replication in Cloud

Paxos



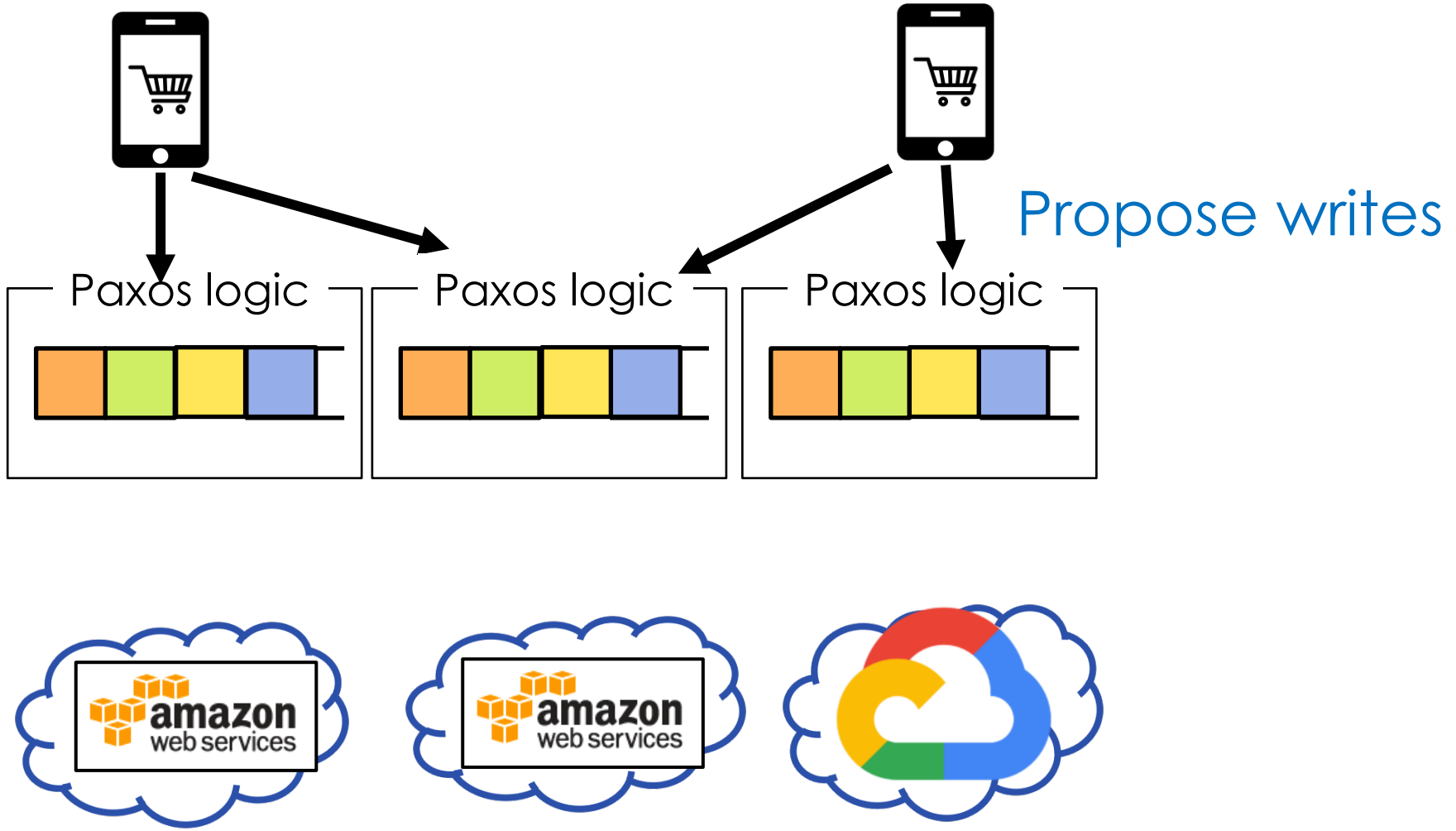
Challenges for Data Replication in Cloud

Paxos



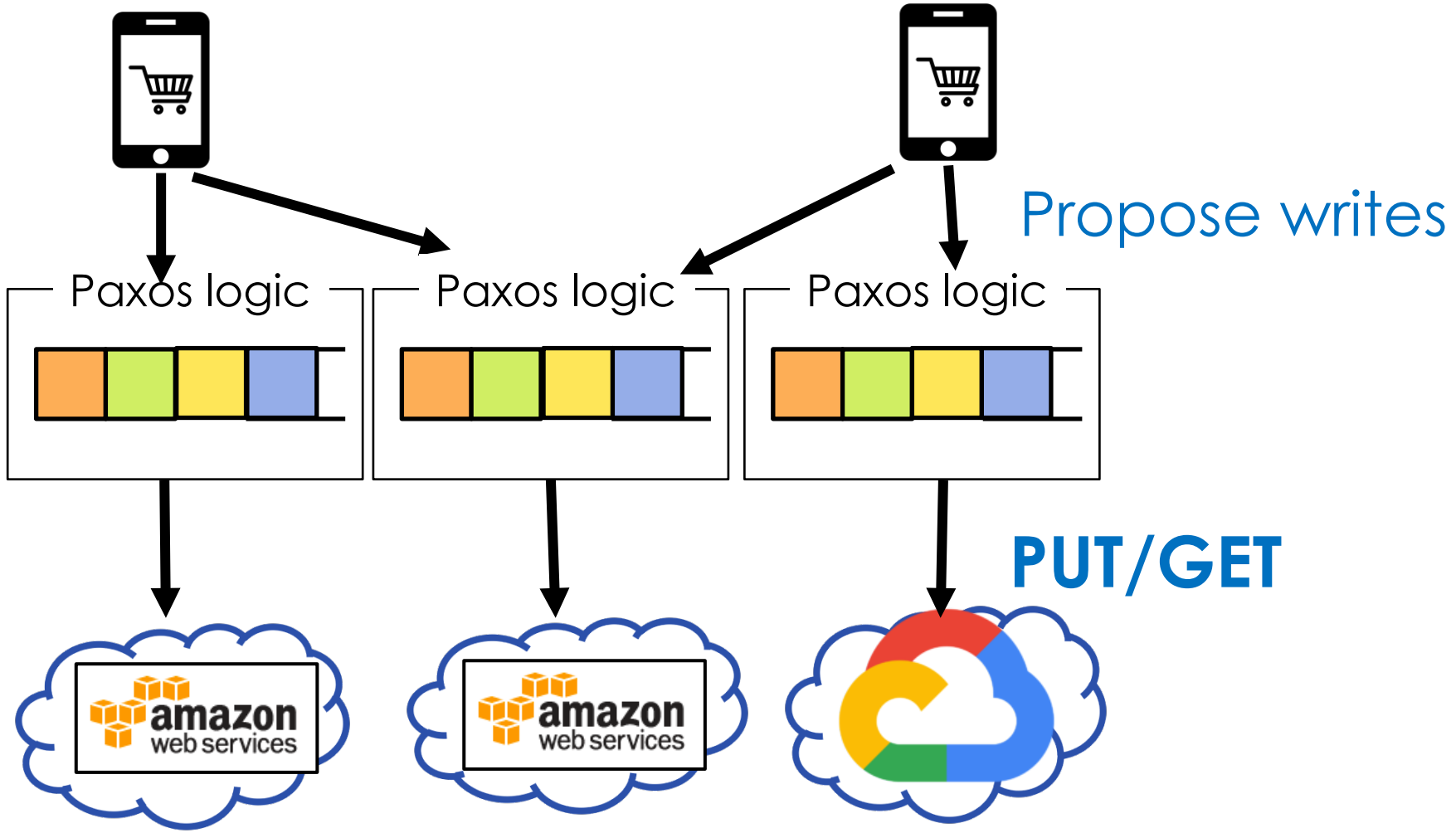
Challenges for Data Replication in Cloud

Paxos



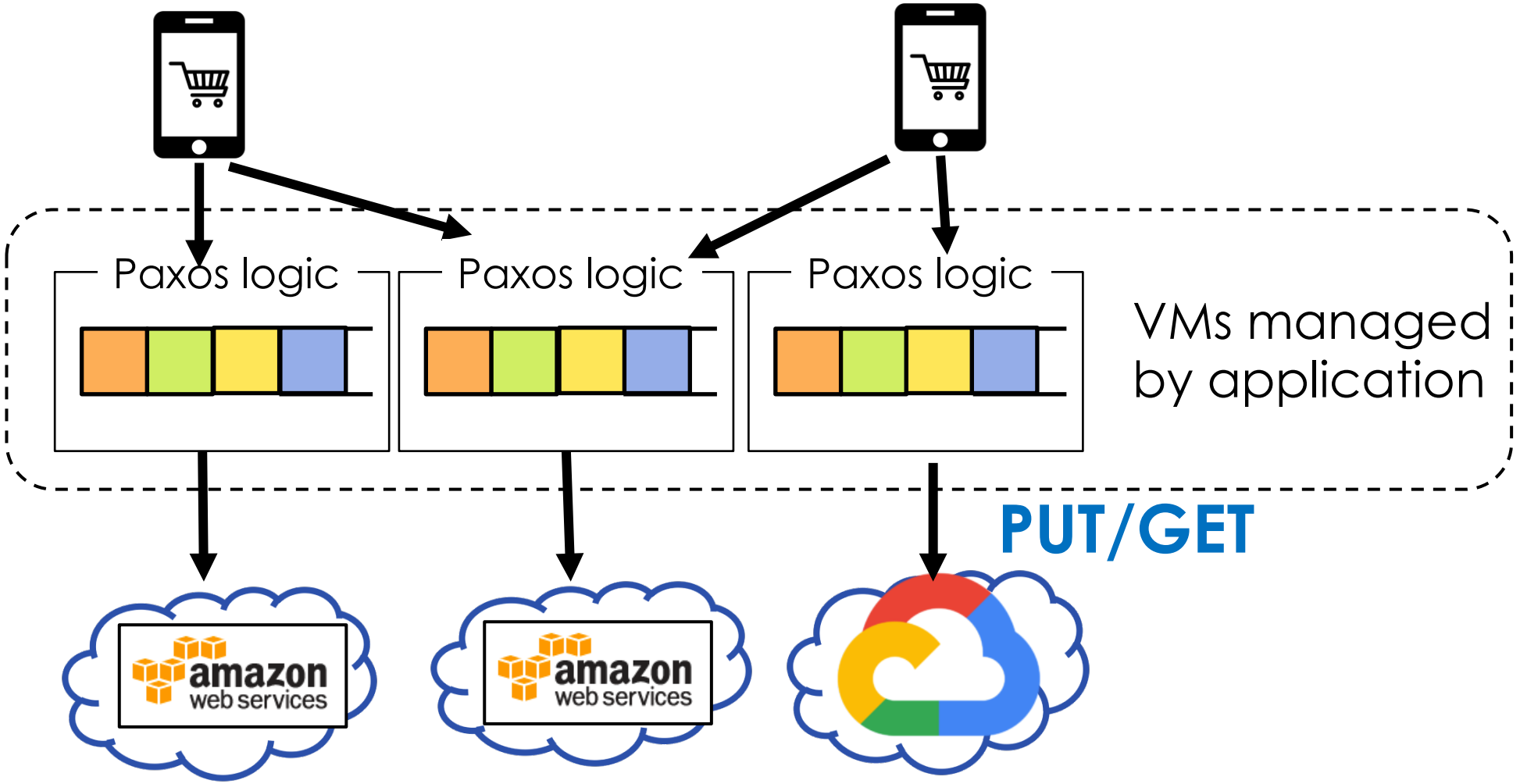
Challenges for Data Replication in Cloud

Paxos



Challenges for Data Replication in Cloud

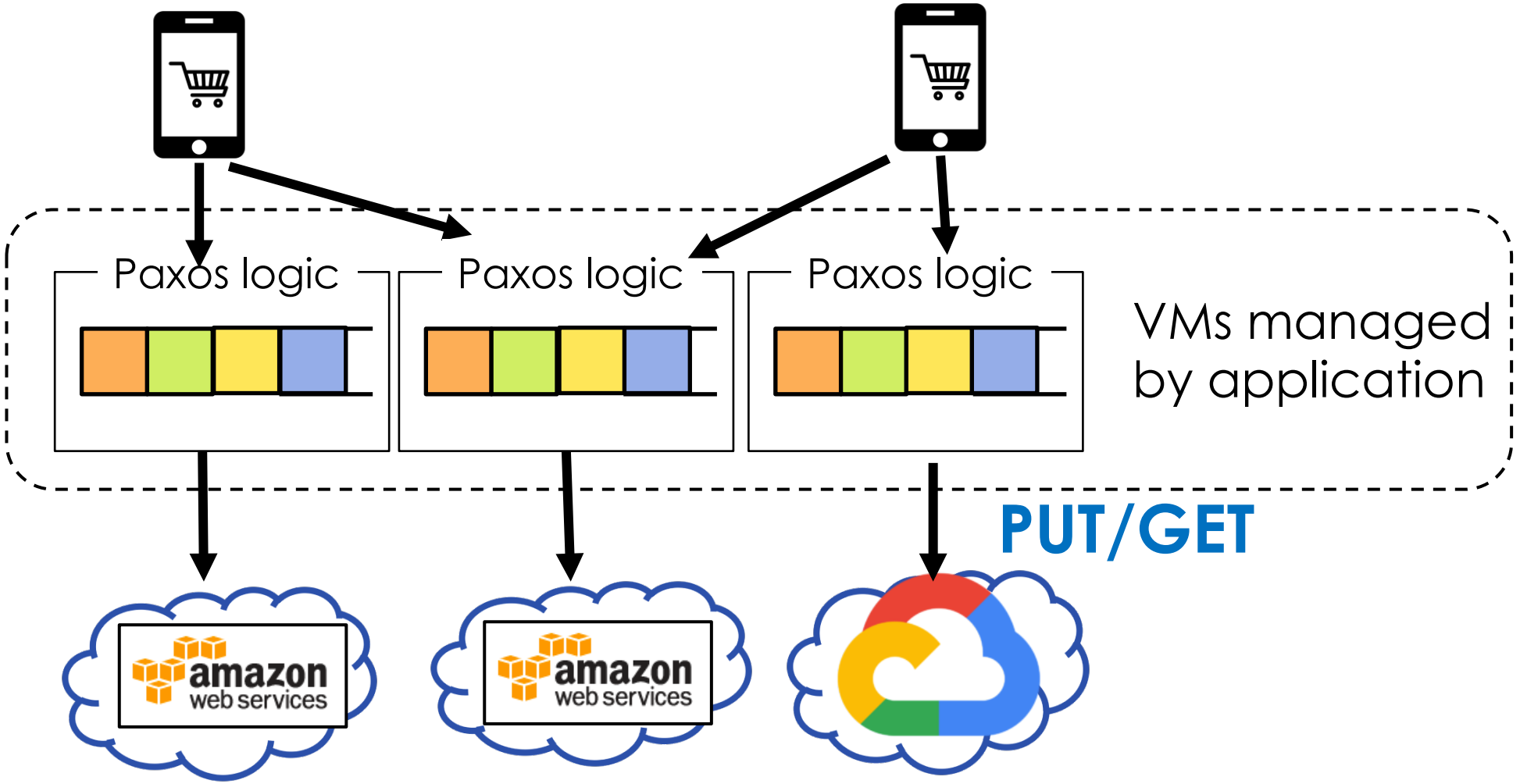
Paxos



Challenges for Data Replication in Cloud

Paxos

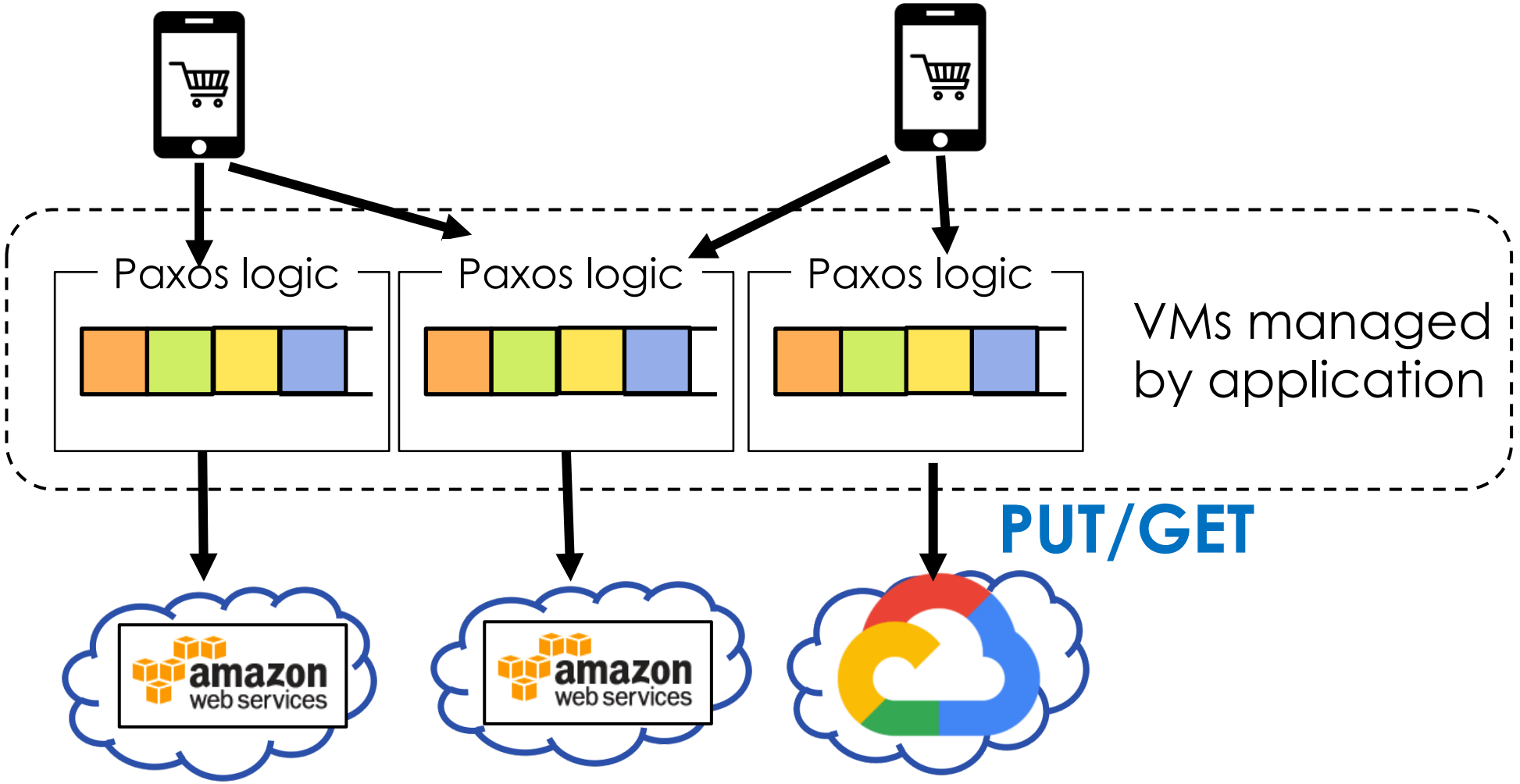
1. High cost



Challenges for Data Replication in Cloud

Paxos

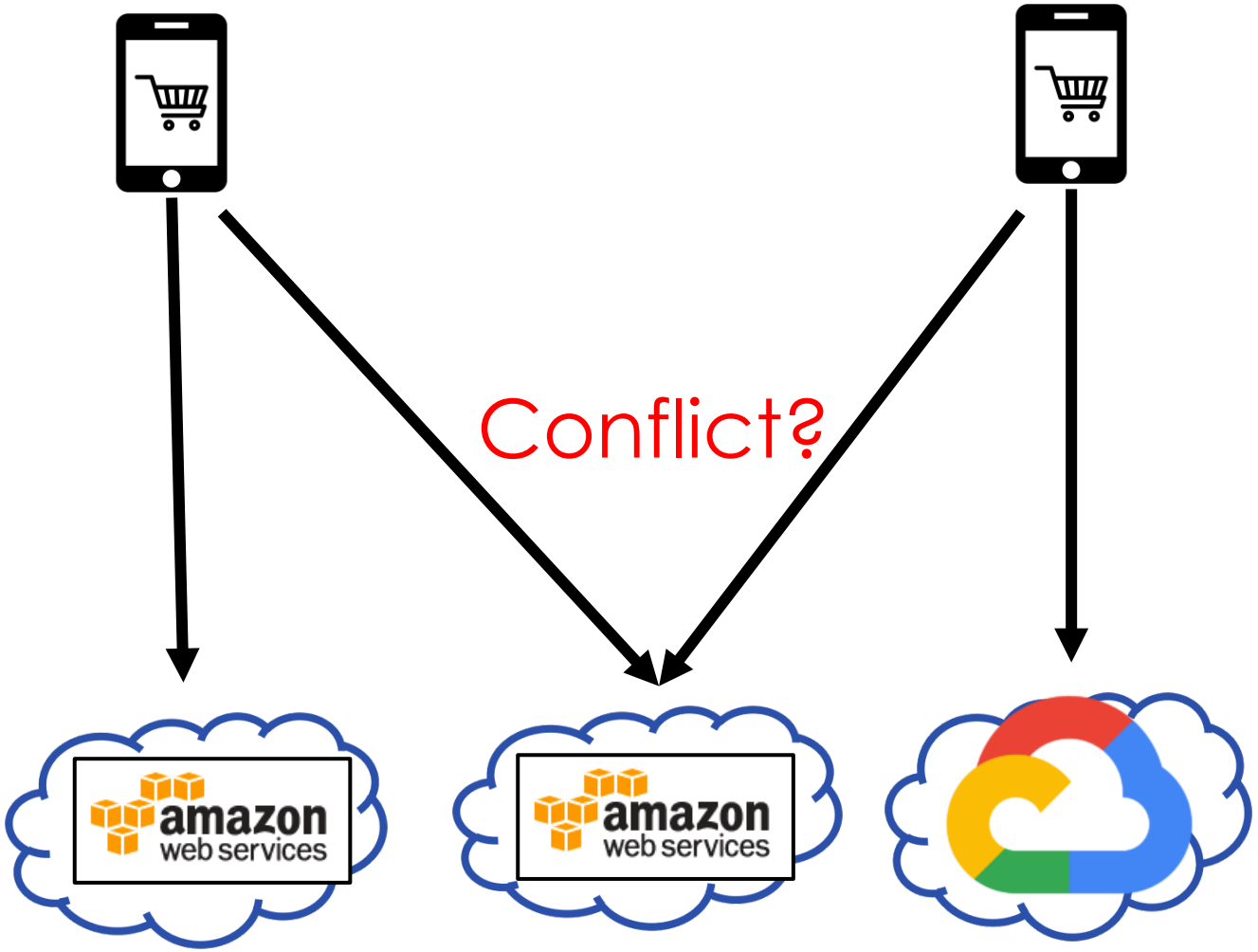
1. High cost
2. Bottleneck



Challenges for Data Replication in Cloud

Paxos

1. High cost
2. Bottleneck



Challenges for Data Replication in Cloud

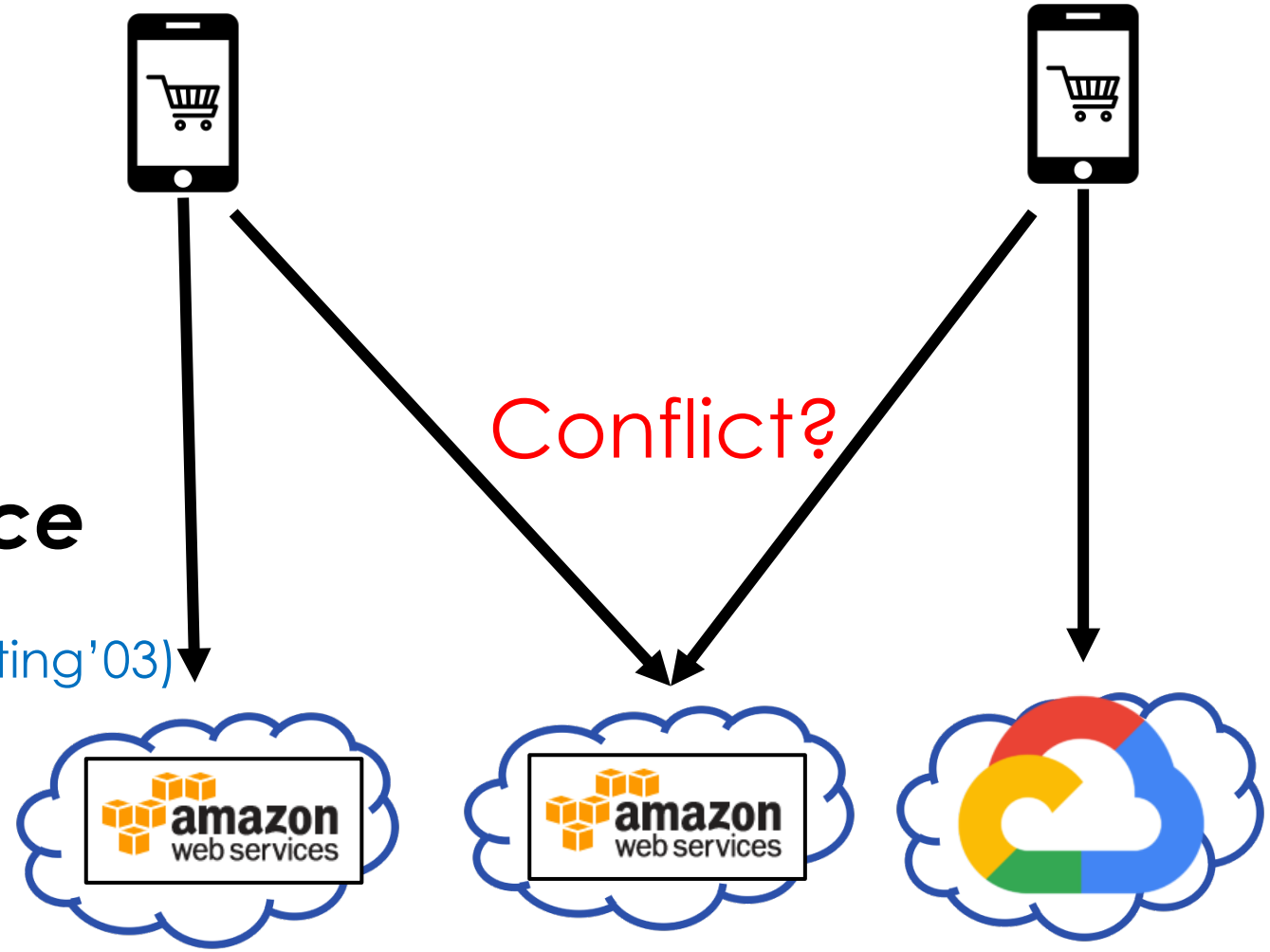
Paxos

1. High cost
2. Bottleneck

Paxos with limited interface

Disk Paxos (Distributed Computing'03)

pPaxos (ATC'15)



Challenges for Data Replication in Cloud

Paxos

1. High cost
2. Bottleneck

DiskPaxos, pPaxos

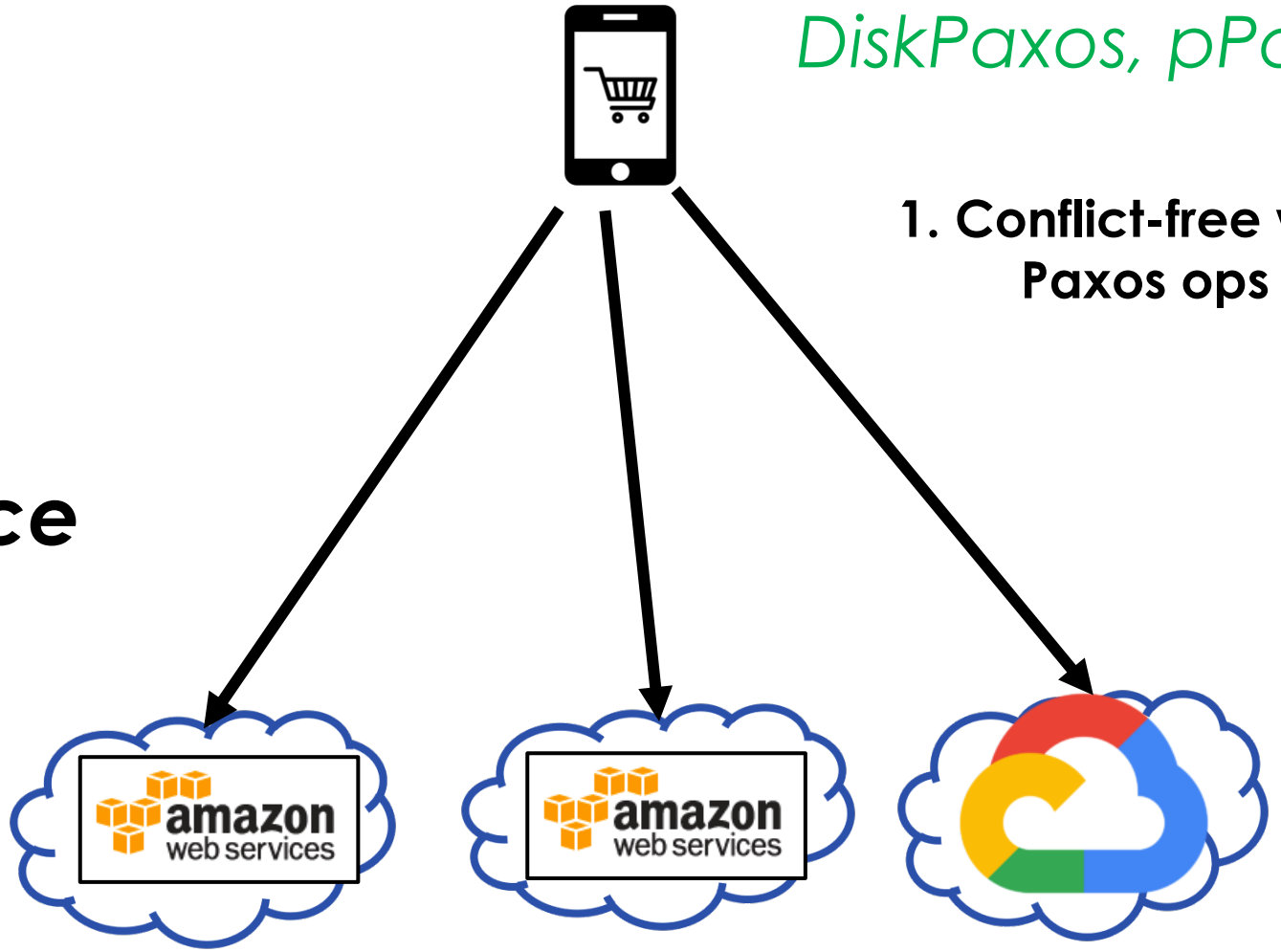
Paxos with limited interface

Challenges for Data Replication in Cloud

Paxos

1. High cost
2. Bottleneck

Paxos with limited interface



DiskPaxos, pPaxos

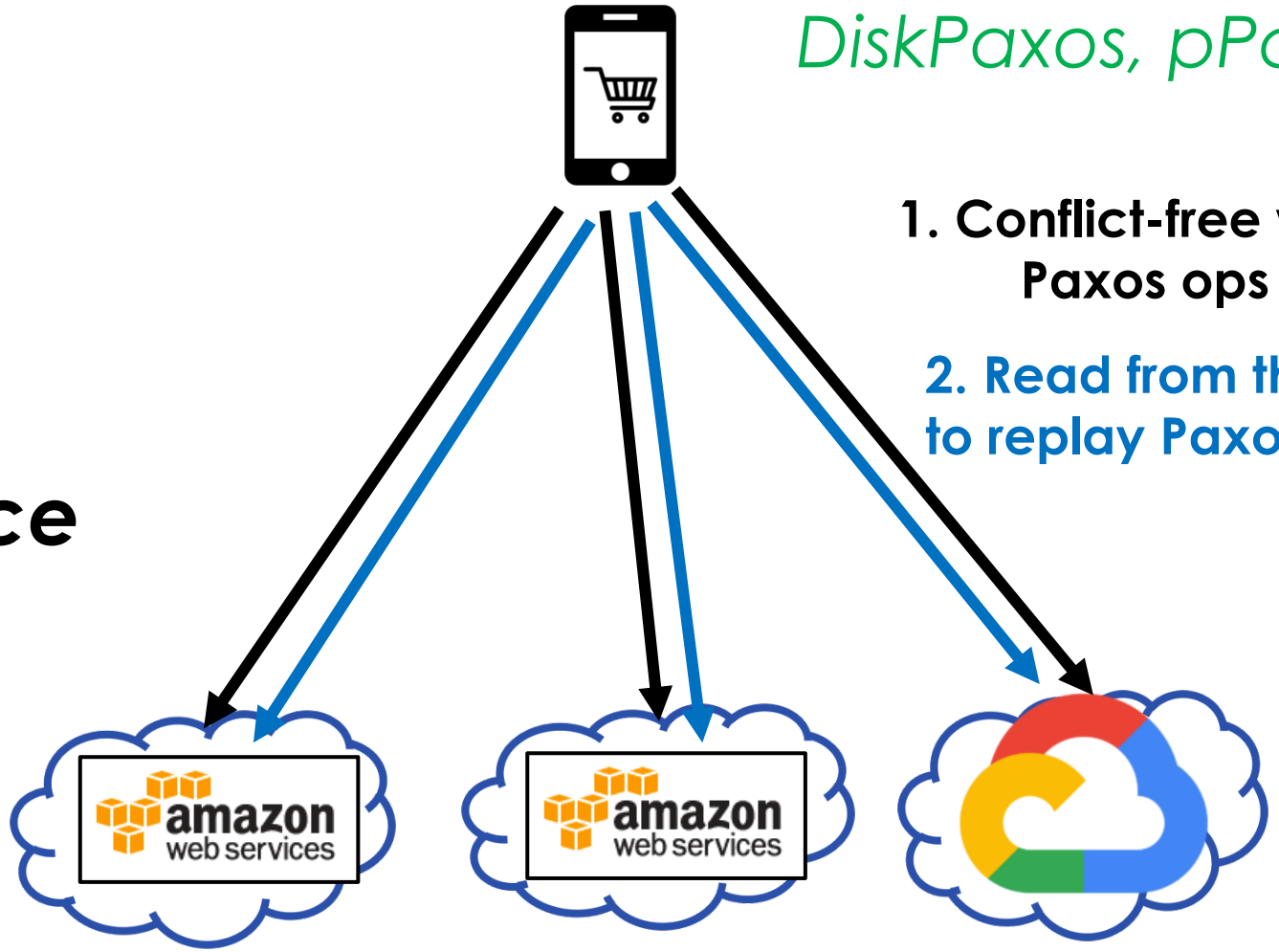
1. Conflict-free write Paxos ops

Challenges for Data Replication in Cloud

Paxos

1. High cost
2. Bottleneck

Paxos with limited interface



DiskPaxos, pPaxos

1. Conflict-free write Paxos ops

2. Read from the log to replay Paxos logic

Challenges for Data Replication in Cloud

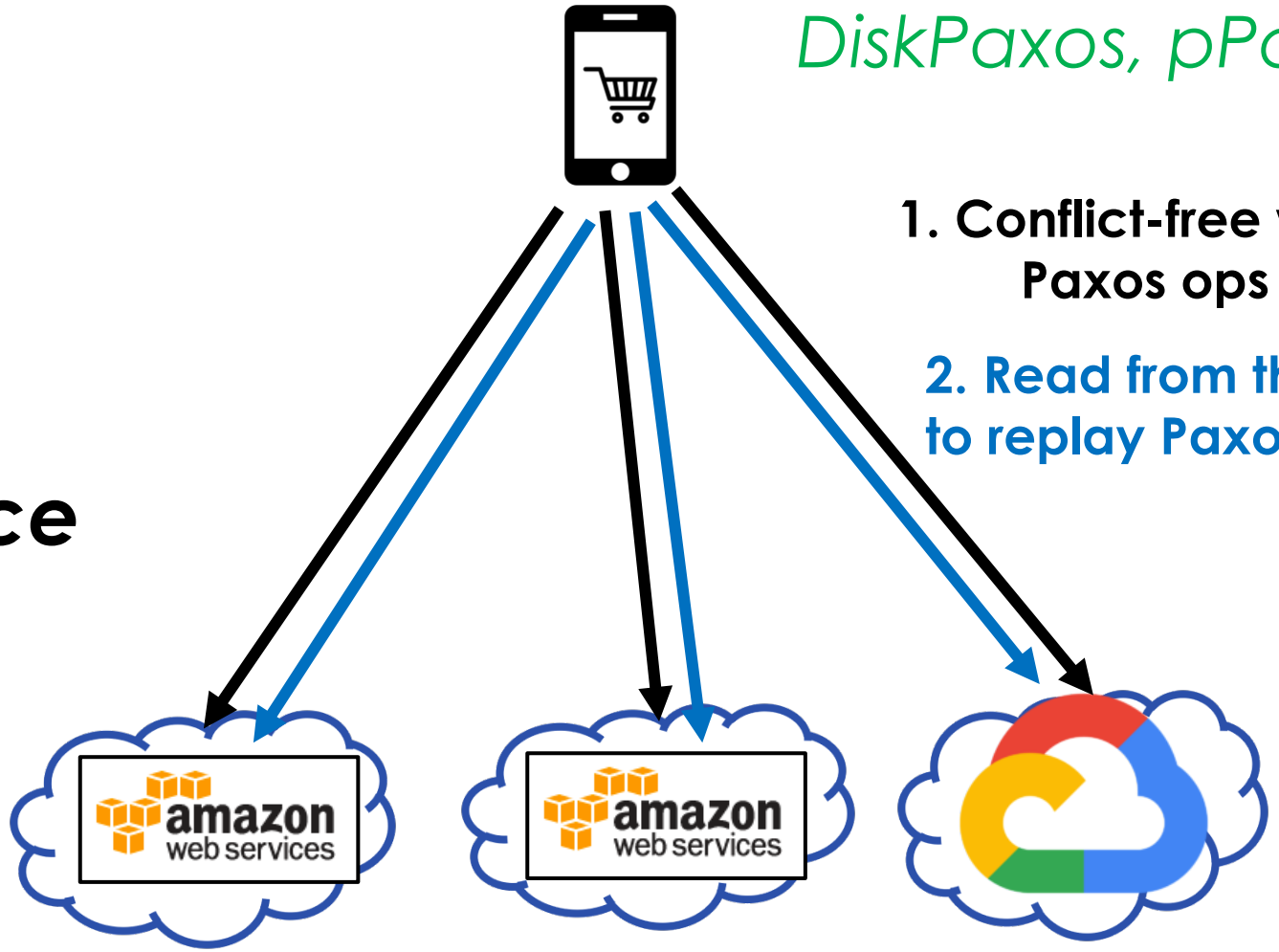
Paxos

1. High cost
2. Bottleneck

Paxos with limited interface

1. High latency

DiskPaxos, pPaxos



1. Conflict-free write Paxos ops

2. Read from the log to replay Paxos logic

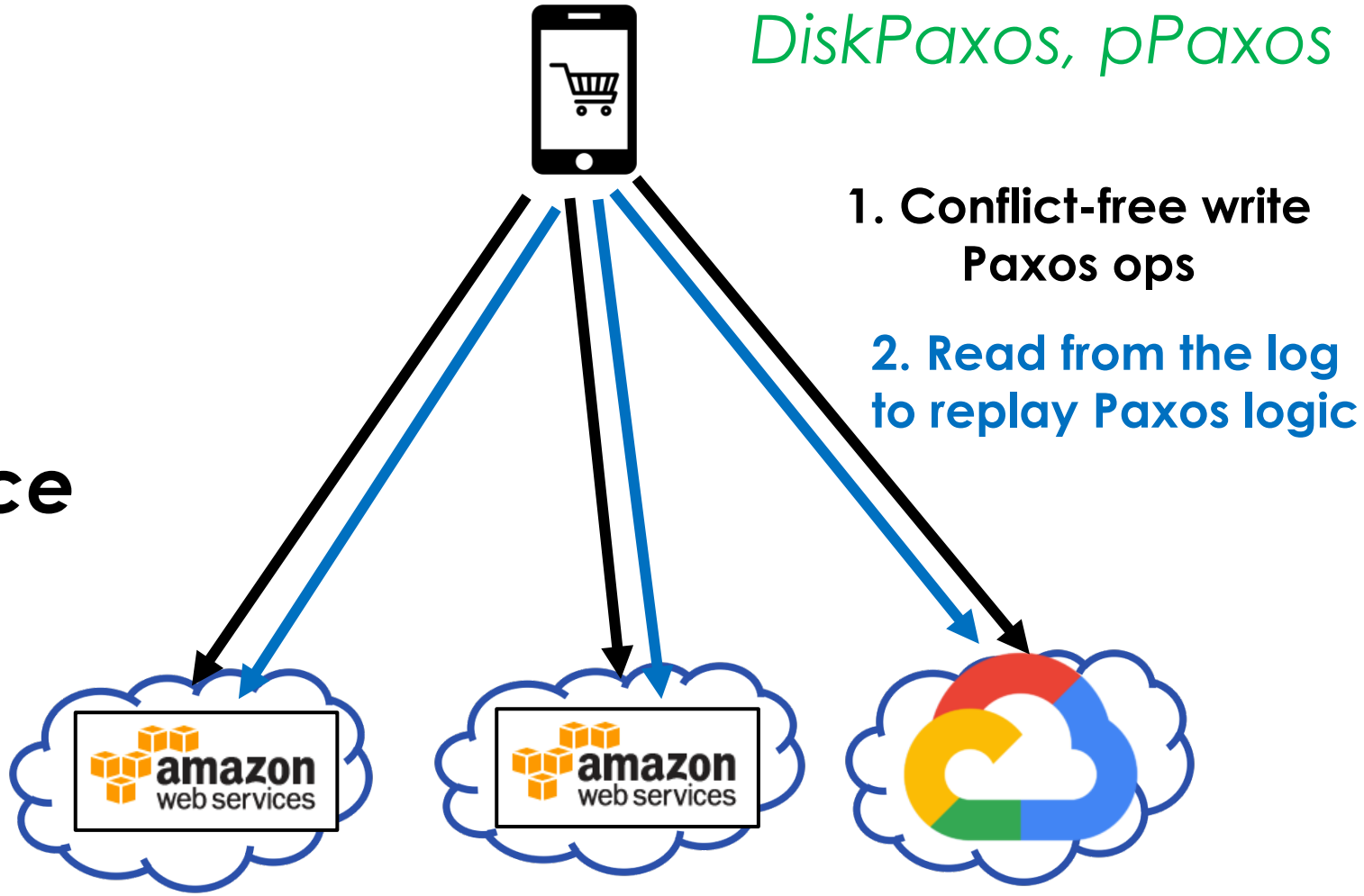
Challenges for Data Replication in Cloud

Paxos

1. High cost
2. Bottleneck

Paxos with limited interface

1. High latency
2. High cost



DiskPaxos, pPaxos

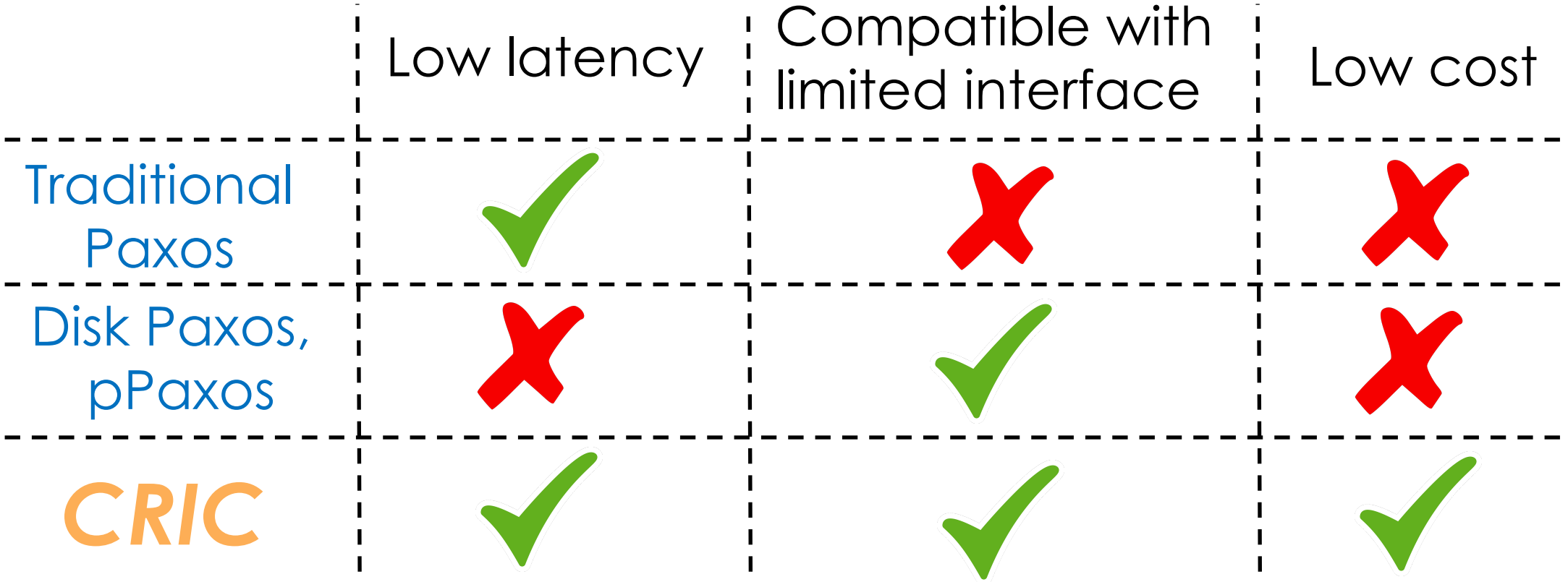
1. Conflict-free write Paxos ops

2. Read from the log to replay Paxos logic

Problems with Existing Solutions

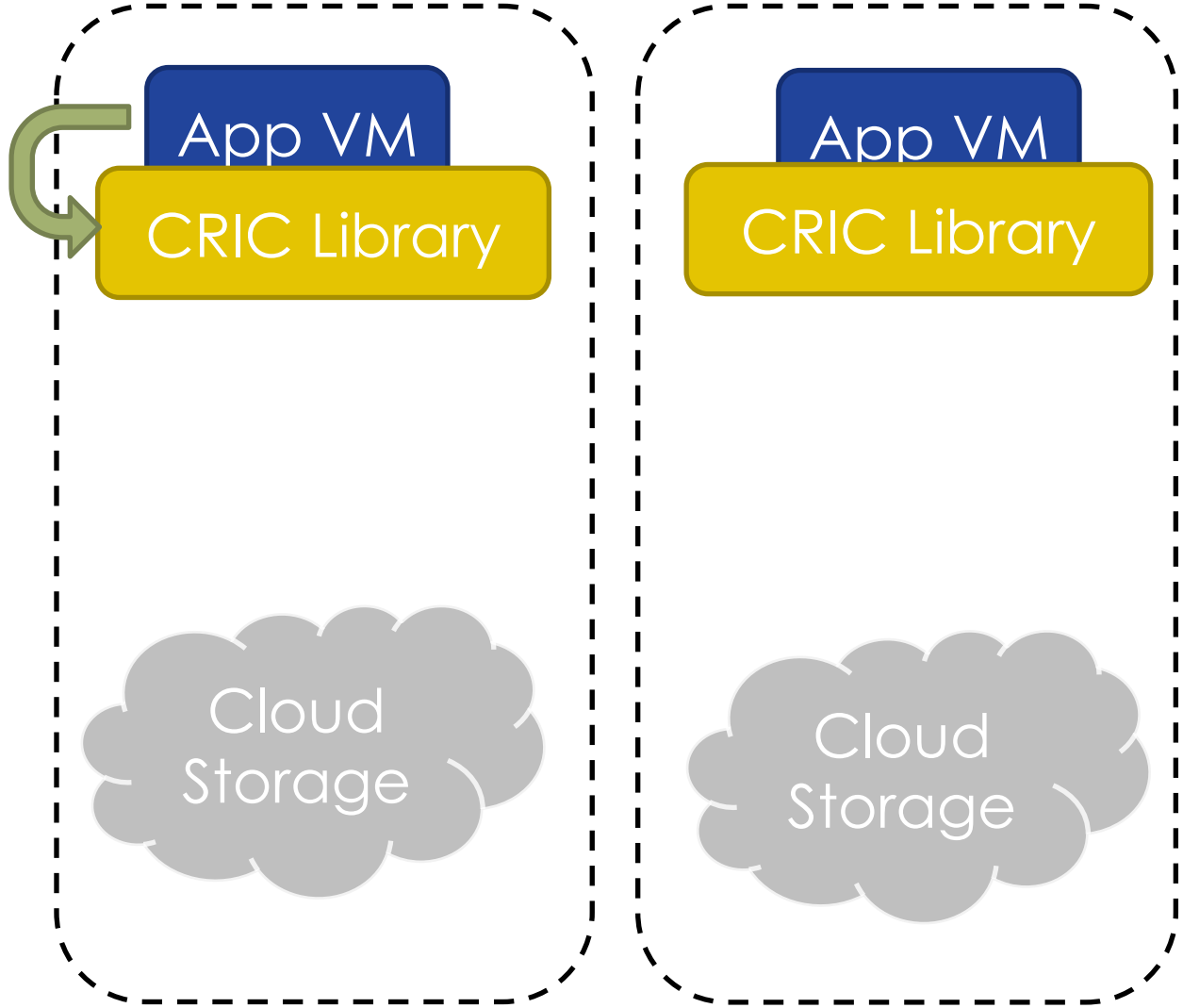
	Low latency	Compatible with limited interface	Low cost
Traditional Paxos	✓	✗	✗
Disk Paxos, pPaxos	✗	✓	✗

Our Solution: Consistent Replication In the Cloud

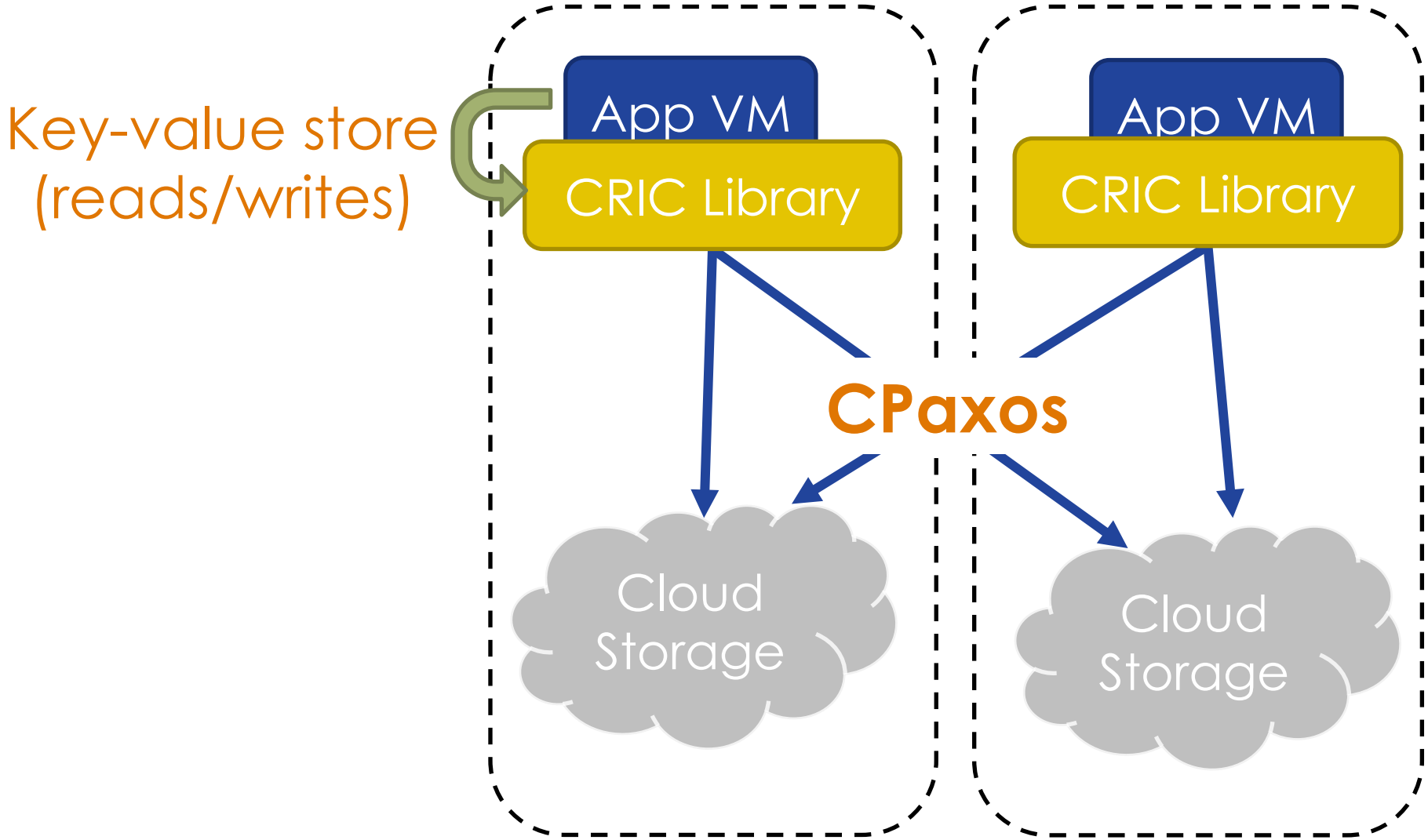


CRIC Overview

Key-value store
(reads/writes)



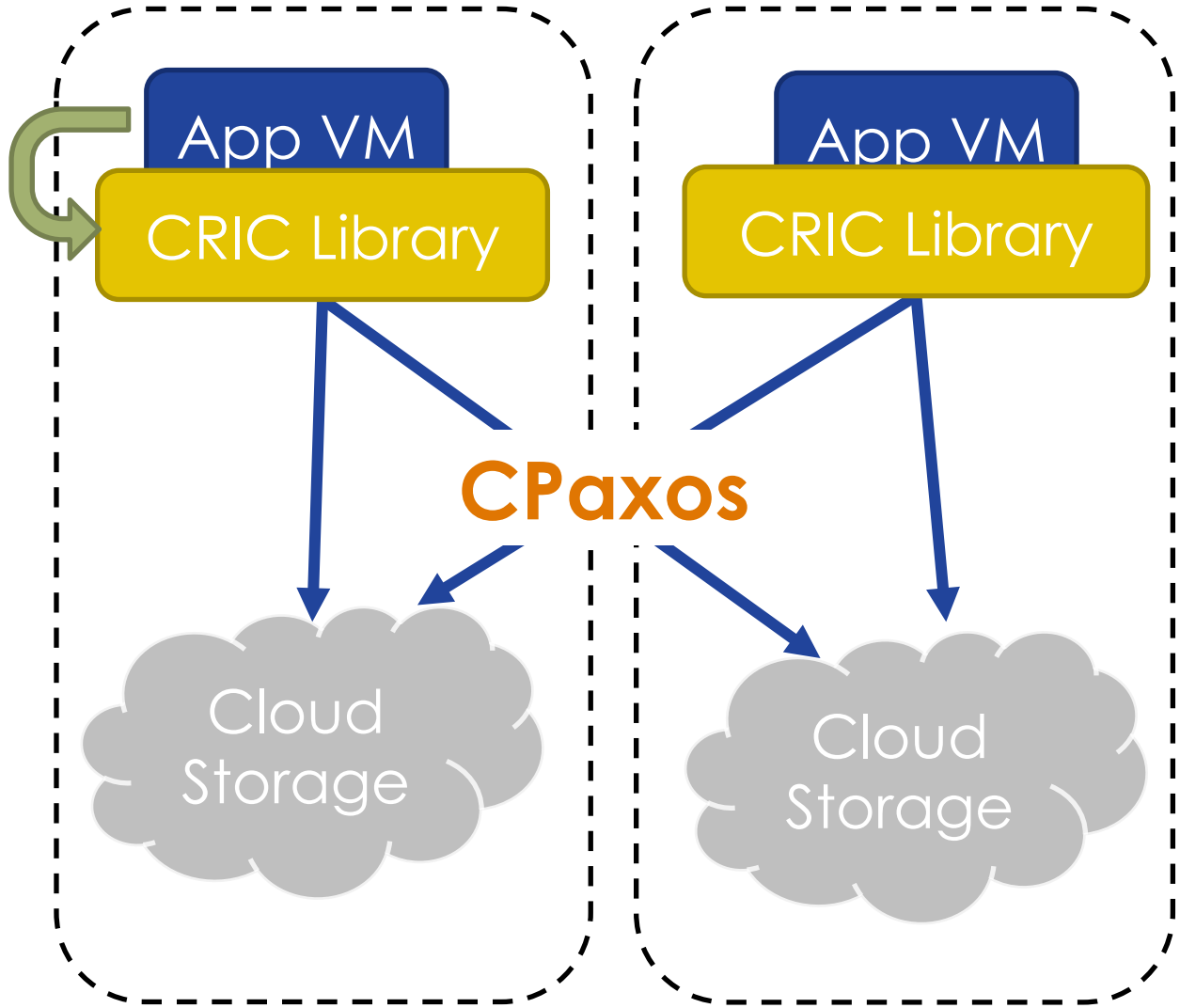
CRIC Overview



CRIC Overview

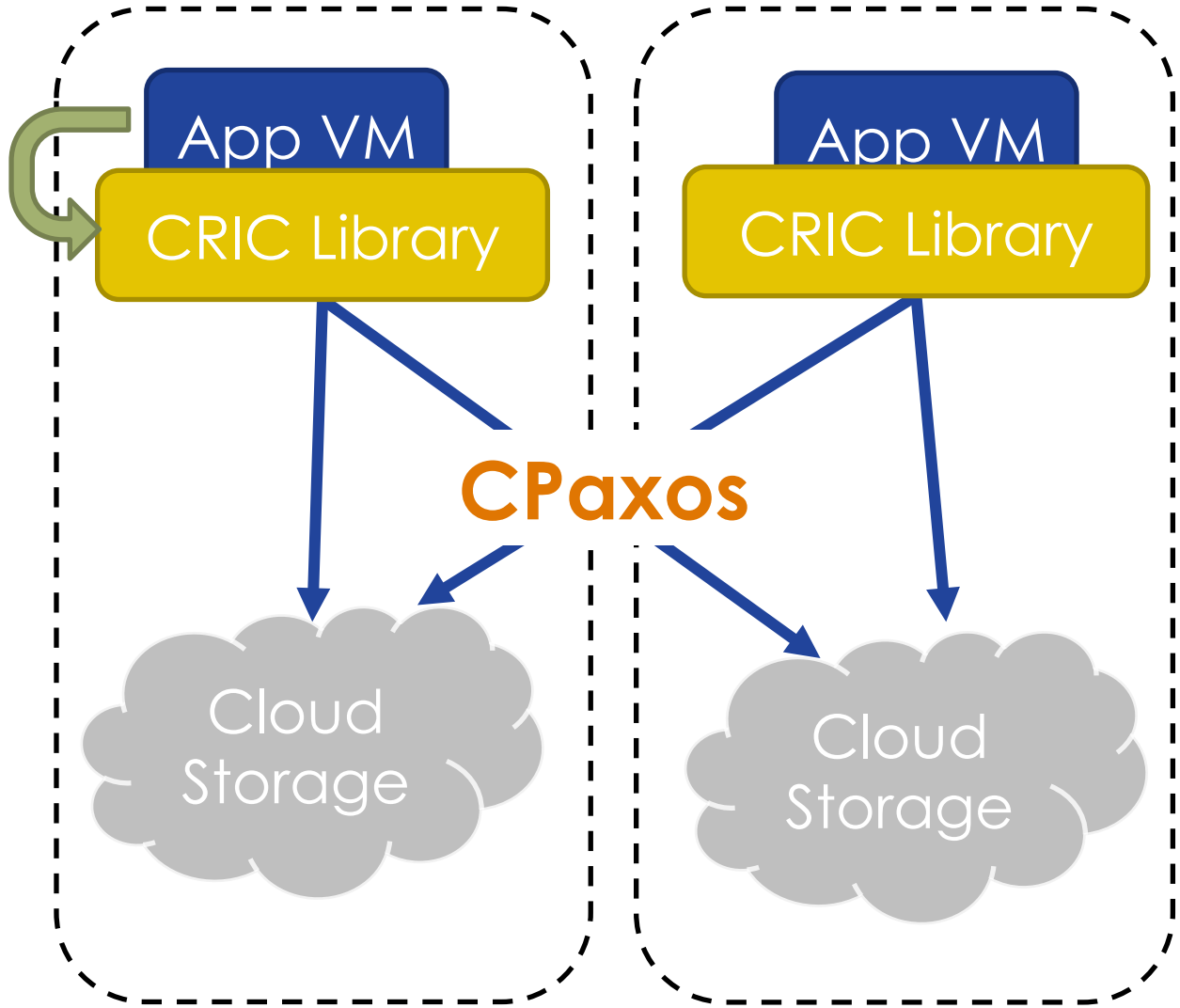
Key-value store (reads/writes)

- ✓ **Apps directly read/write data from/to cloud storage**



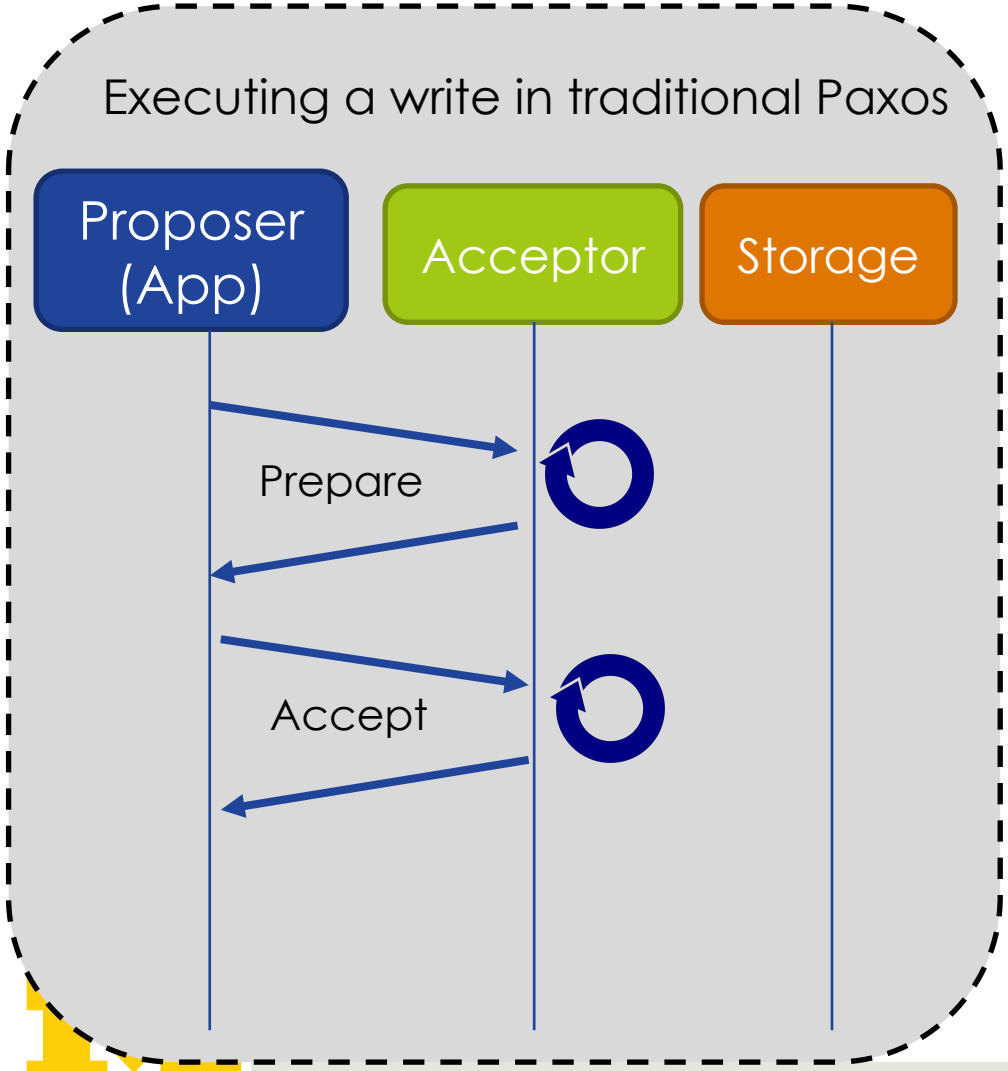
CRIC Overview

Key-value store (reads/writes)

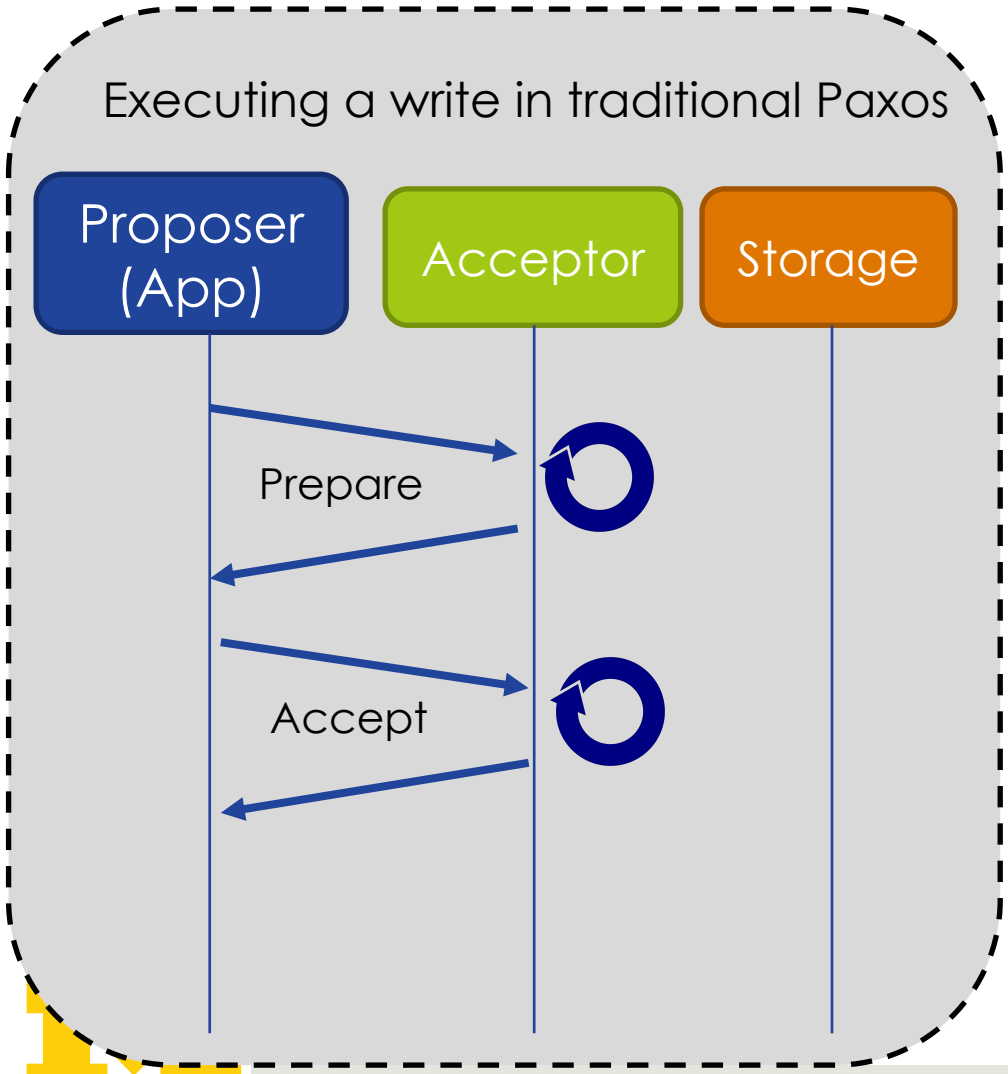


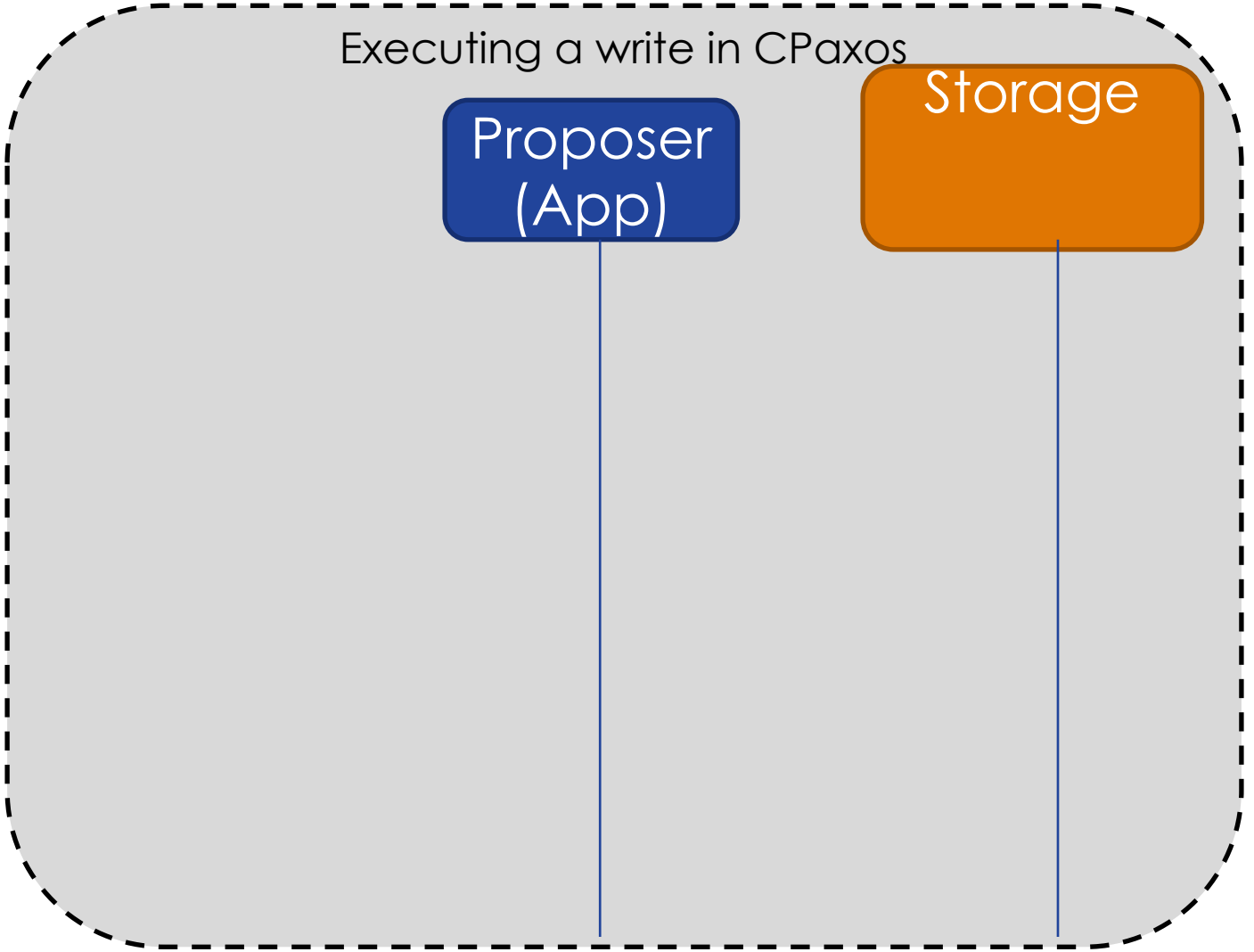
- ✓ **Apps directly read/write data from/to cloud storage**
- ✓ **Low latency (1 RTT)**

CPaxos In Action

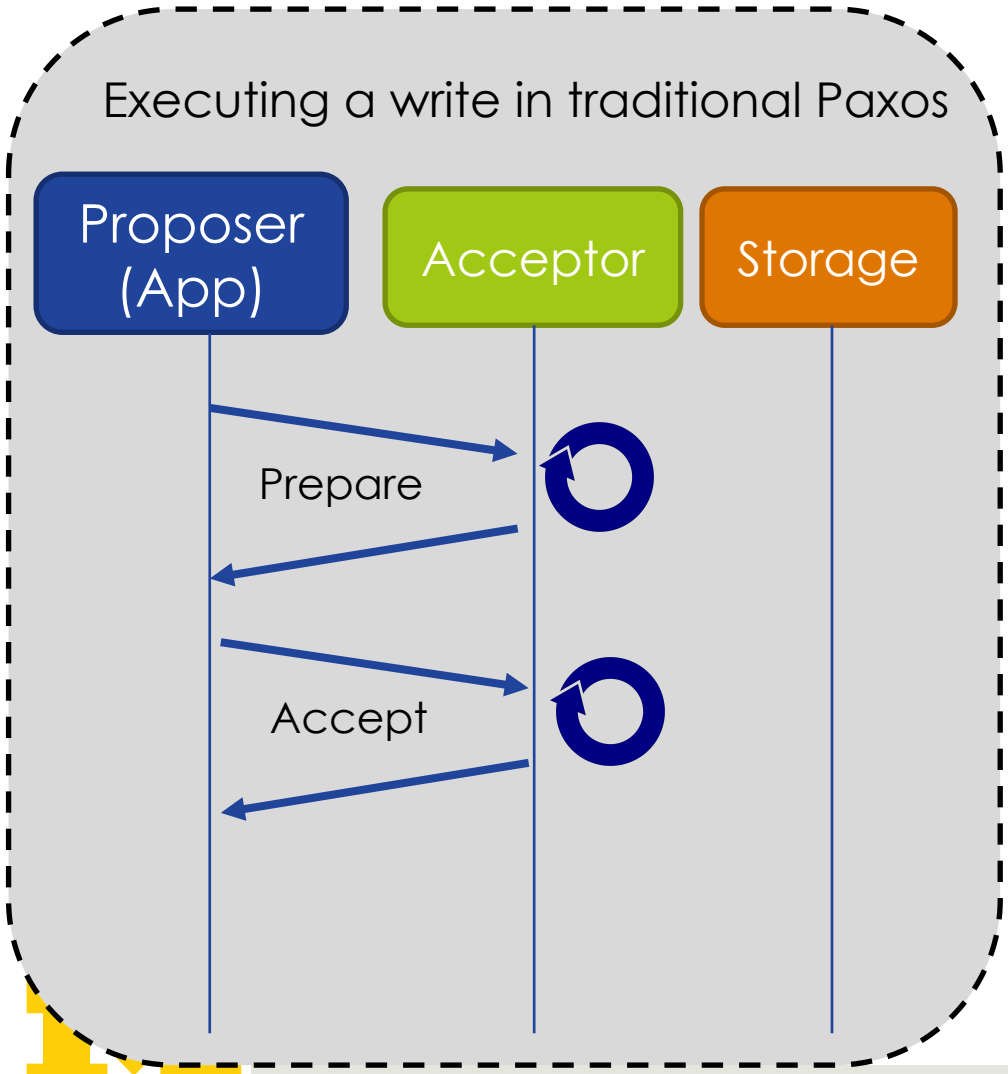


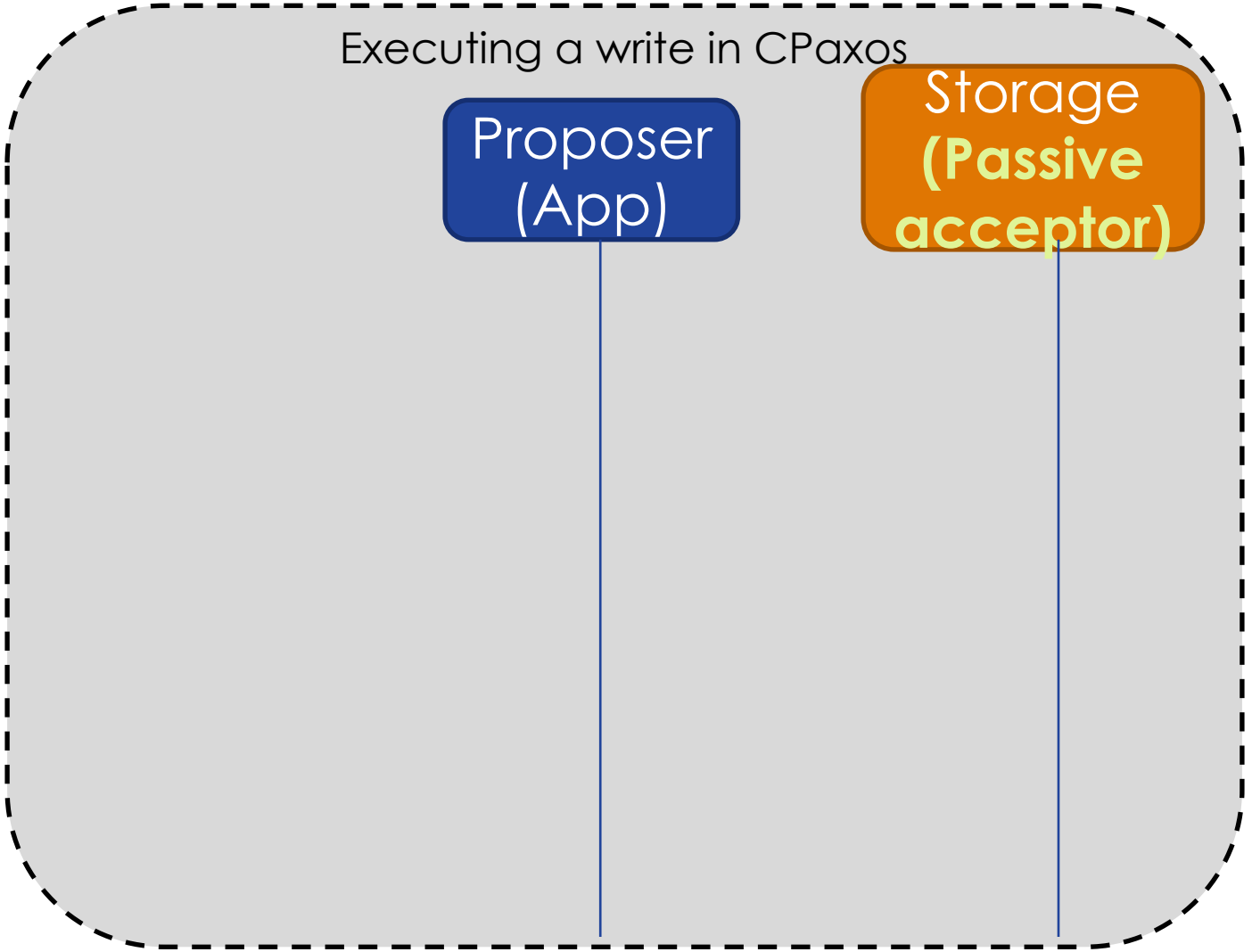
CPaxos In Action



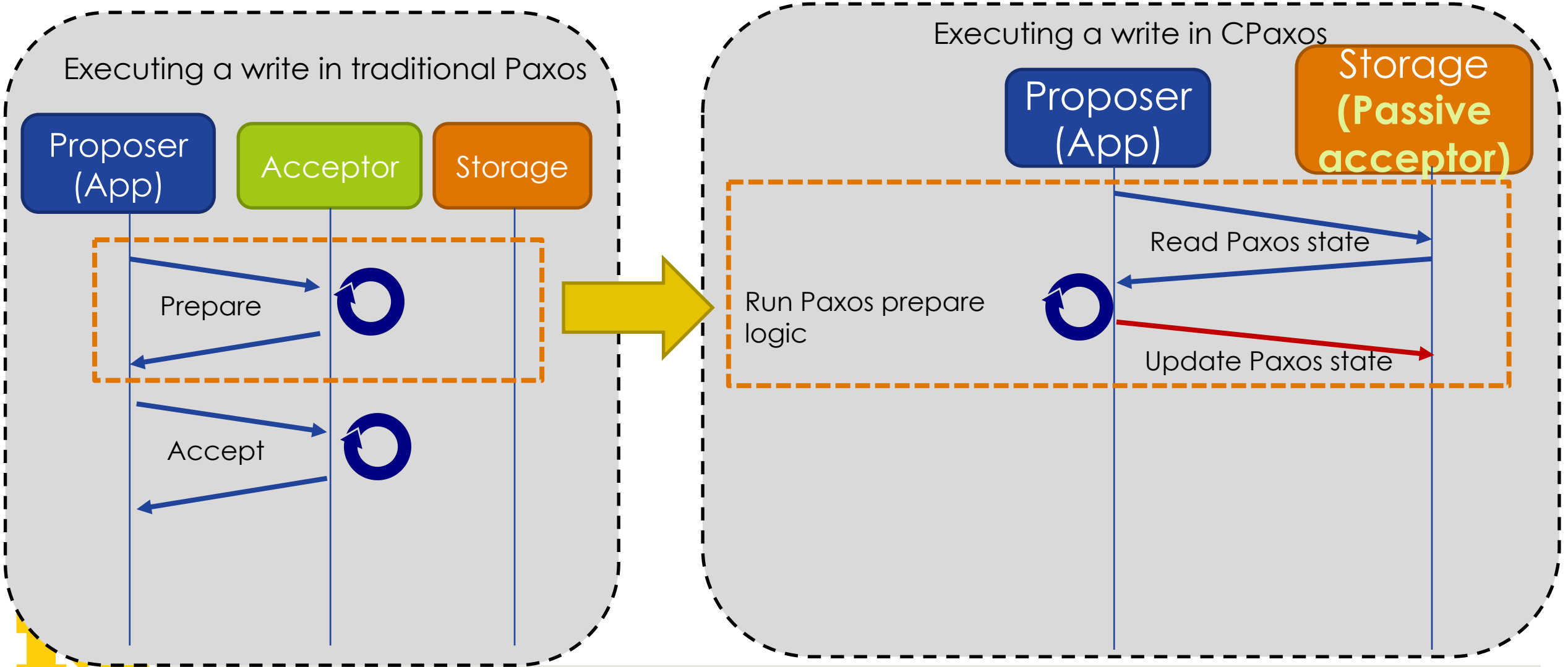


CPaxos In Action

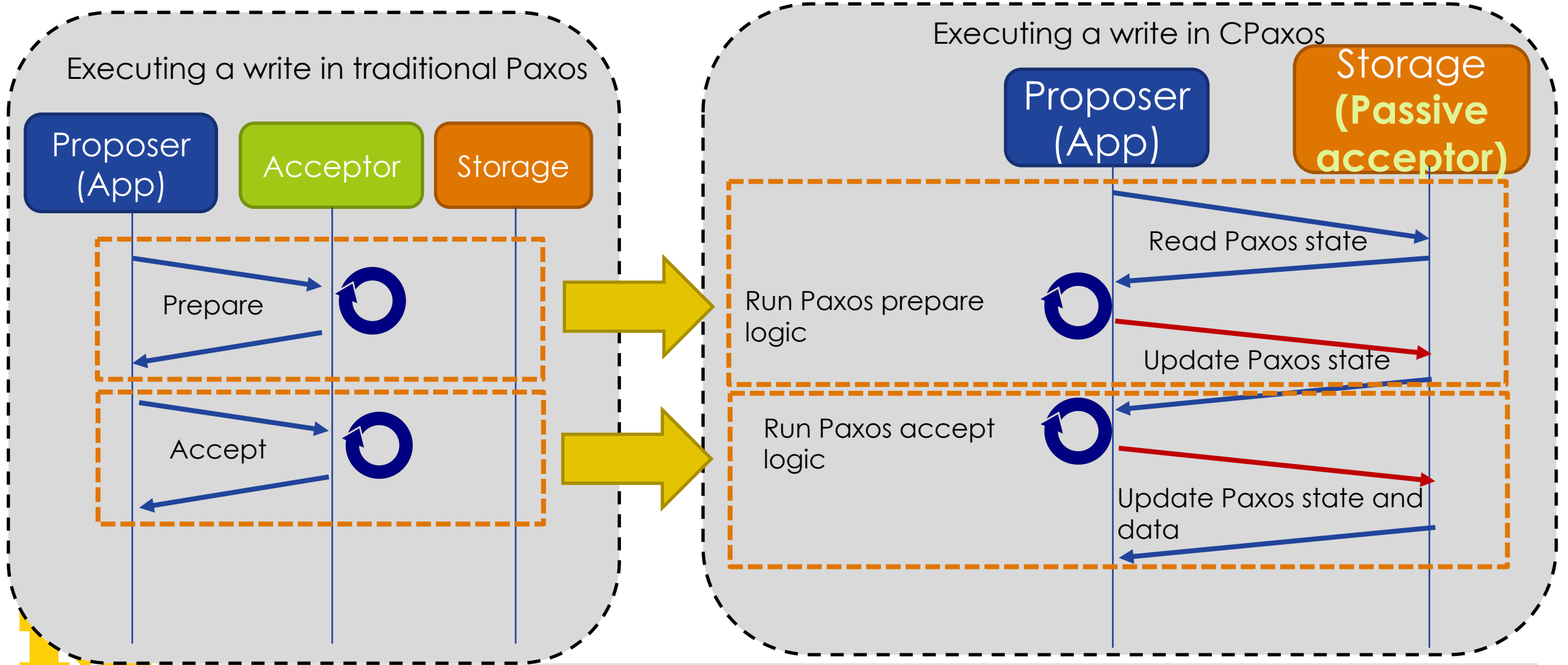




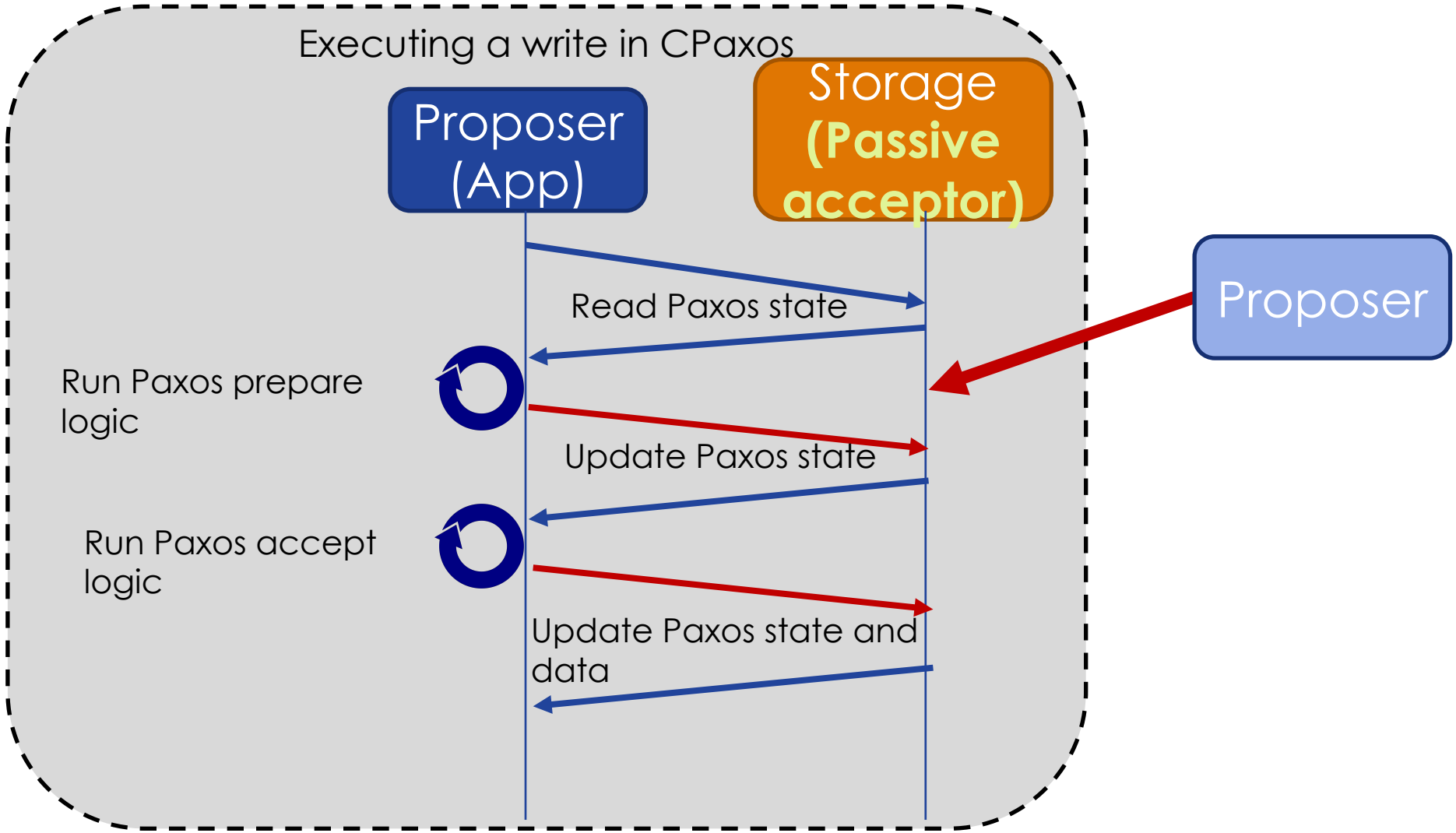
CPaxos In Action



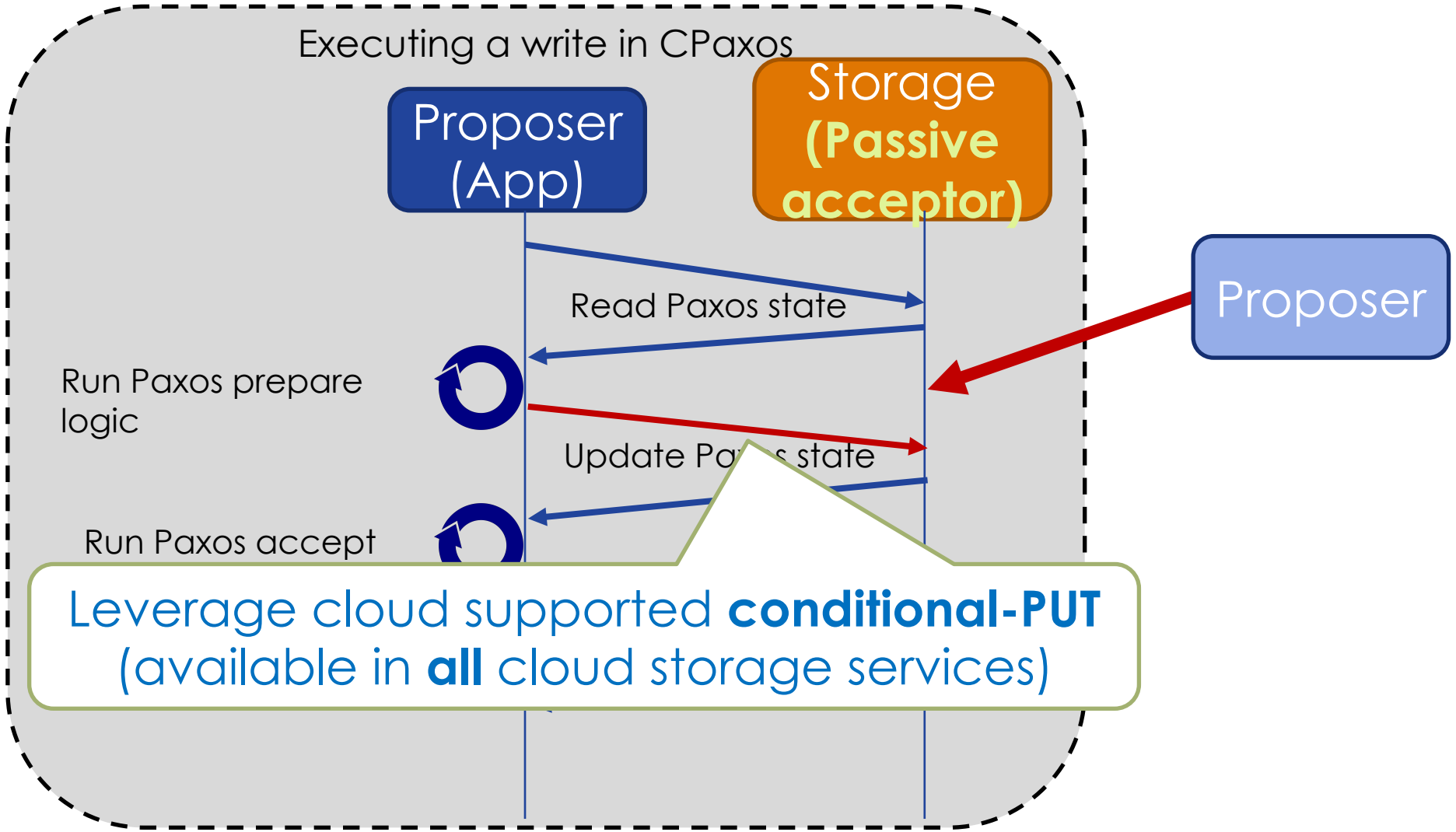
CPaxos In Action



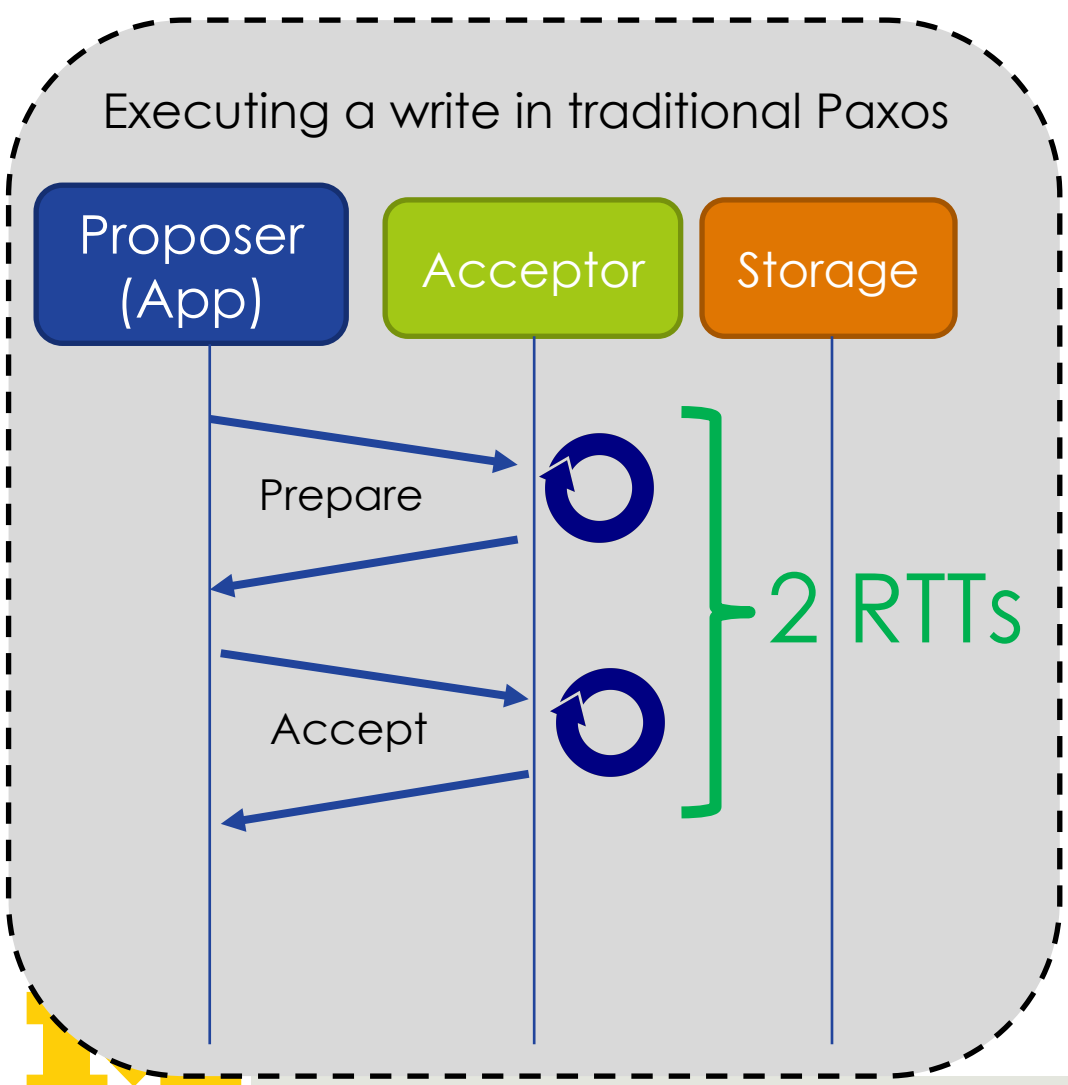
CPaxos In Action

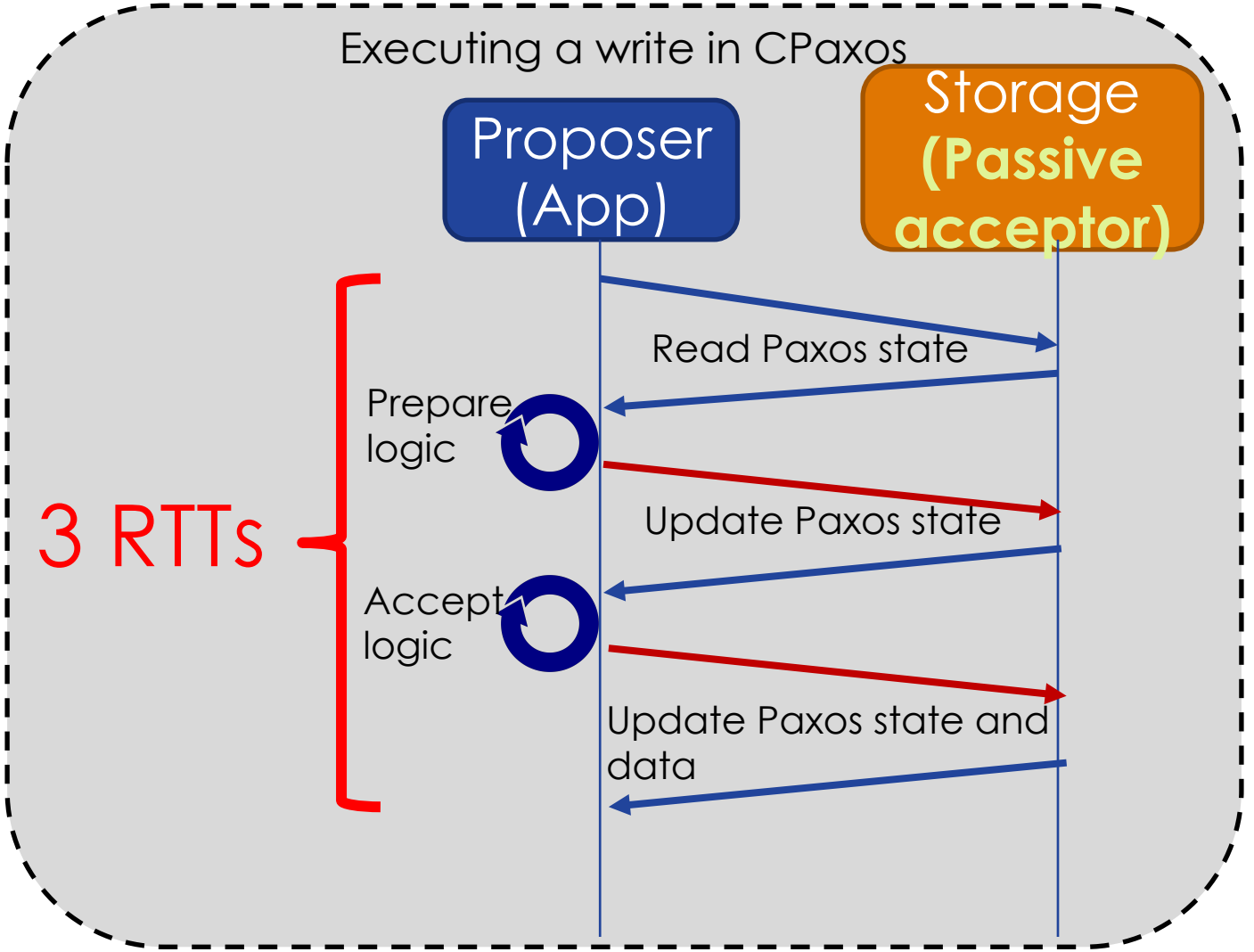


CPaxos In Action

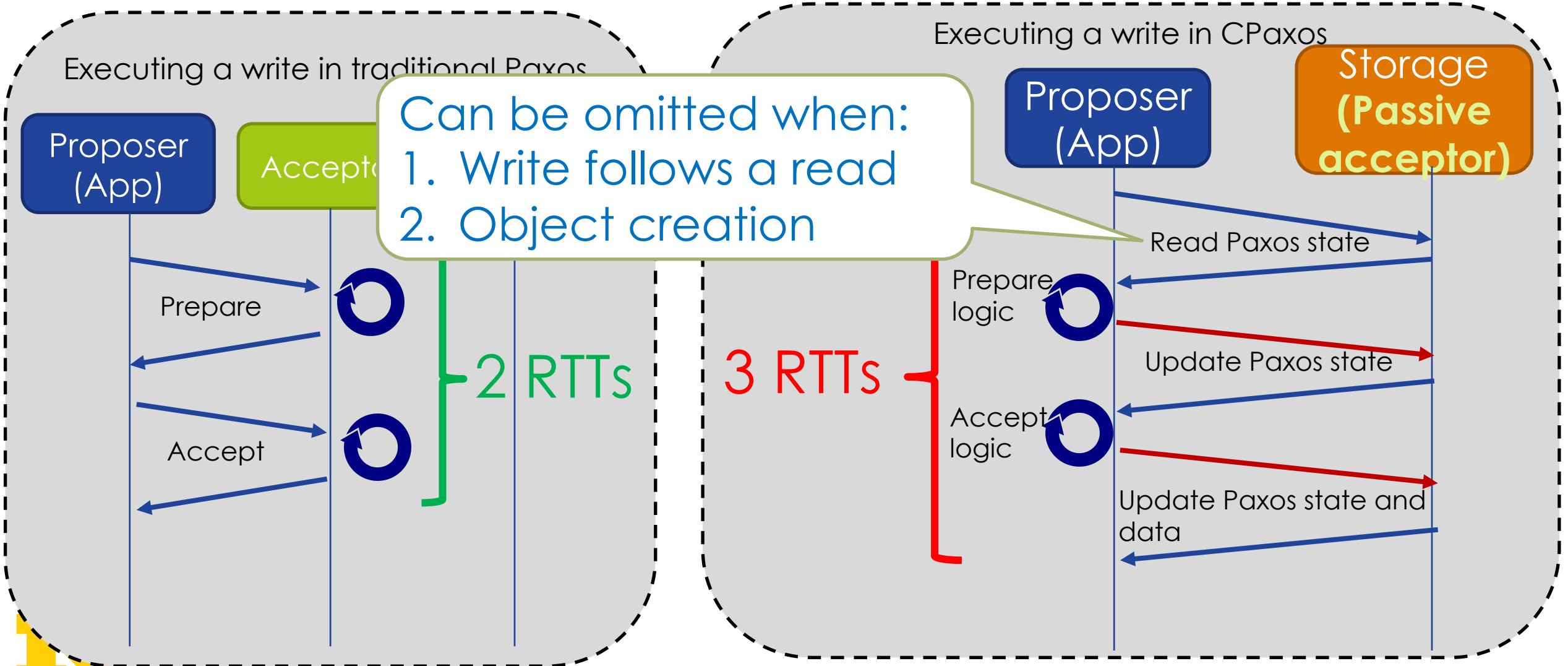


CPaxos In Action

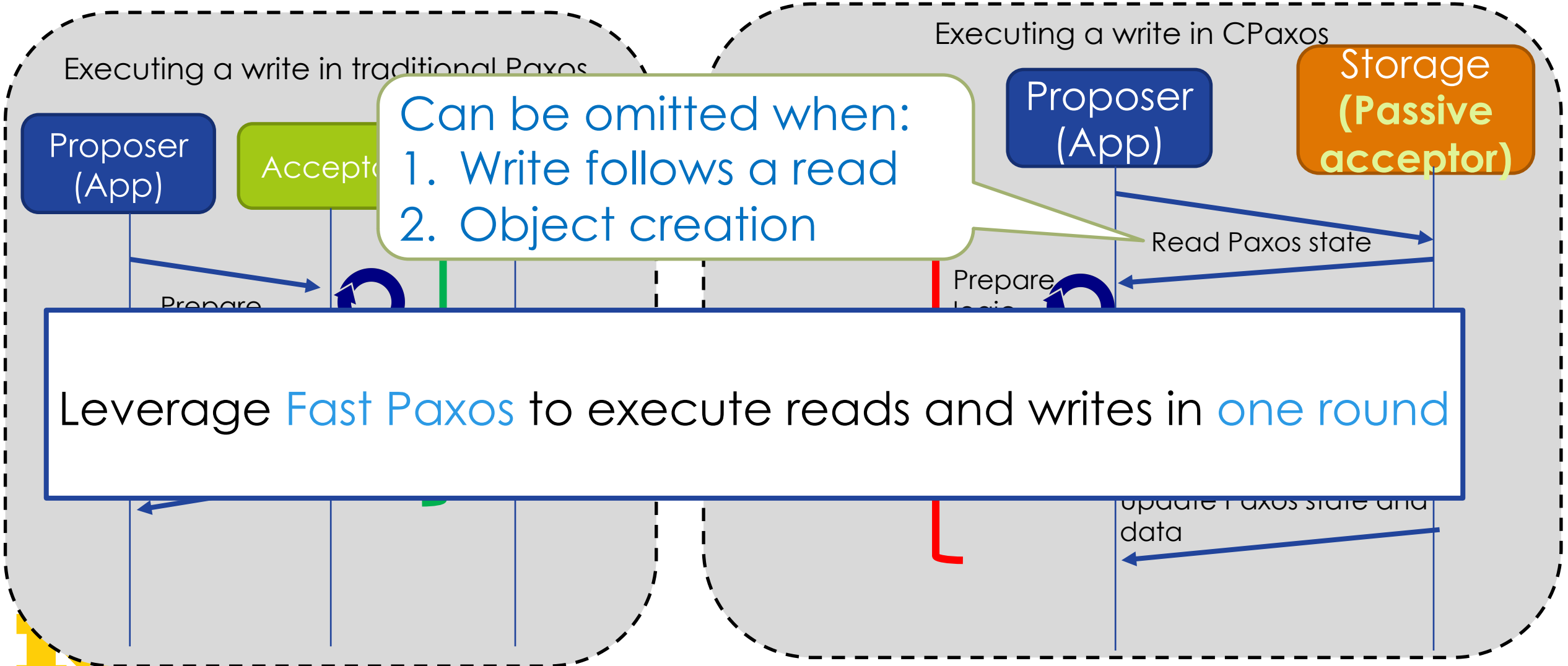




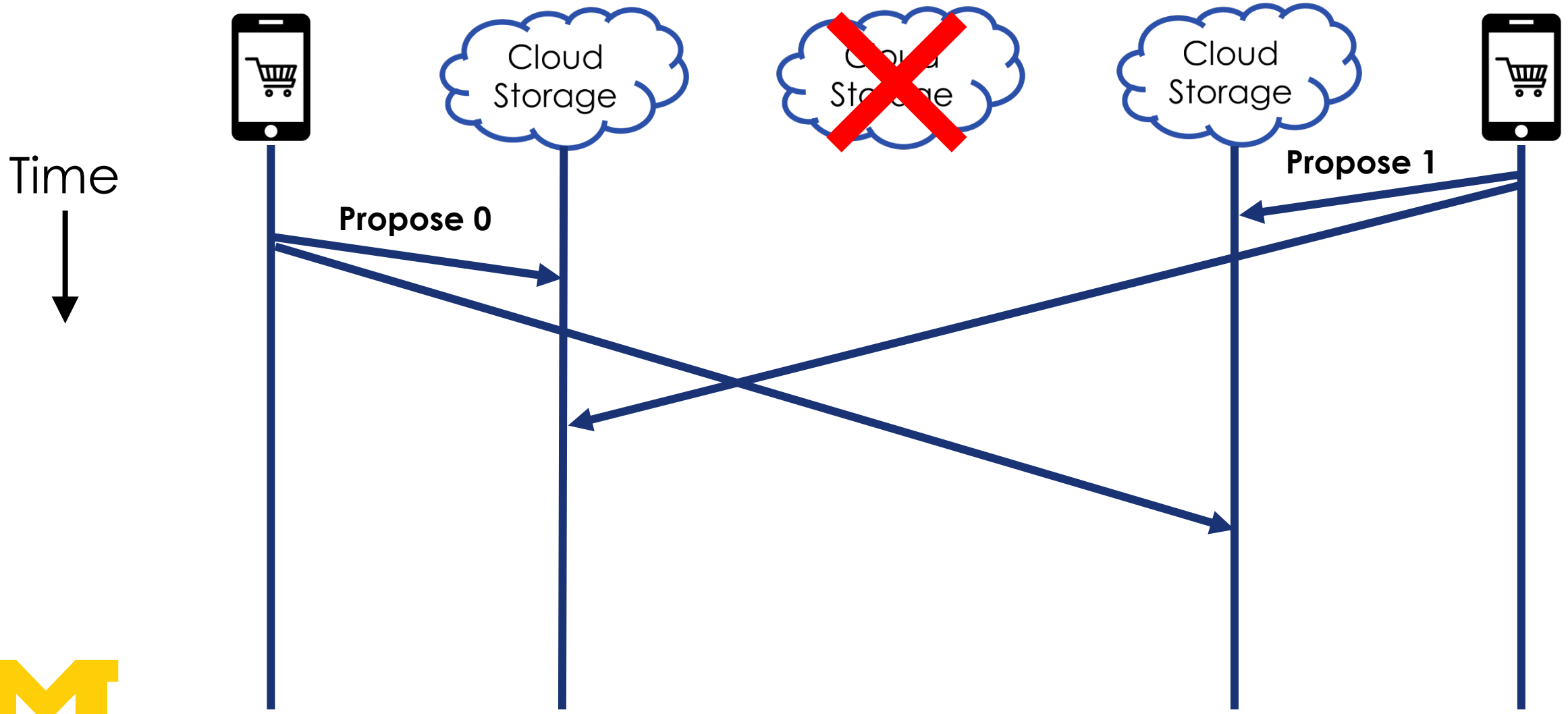
CPaxos In Action



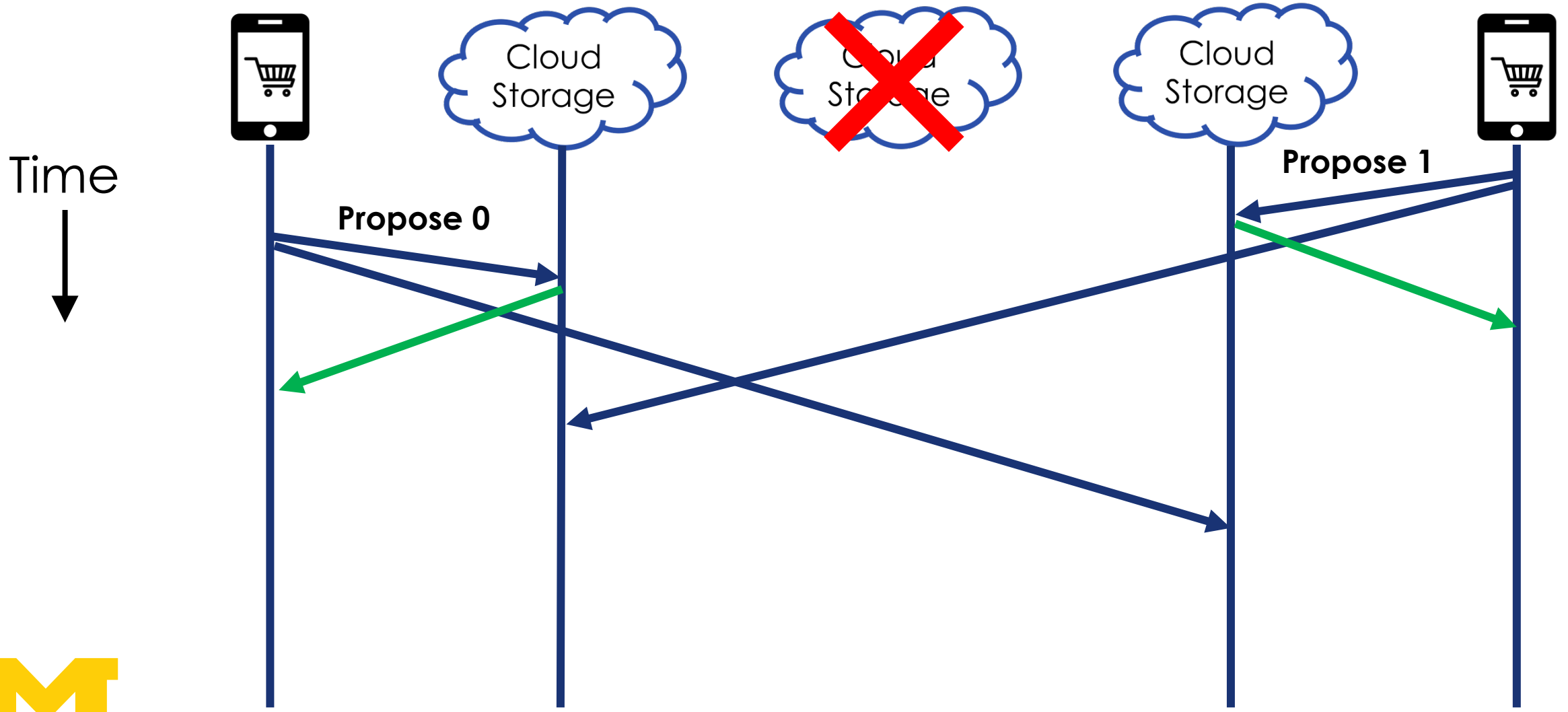
CPaxos In Action



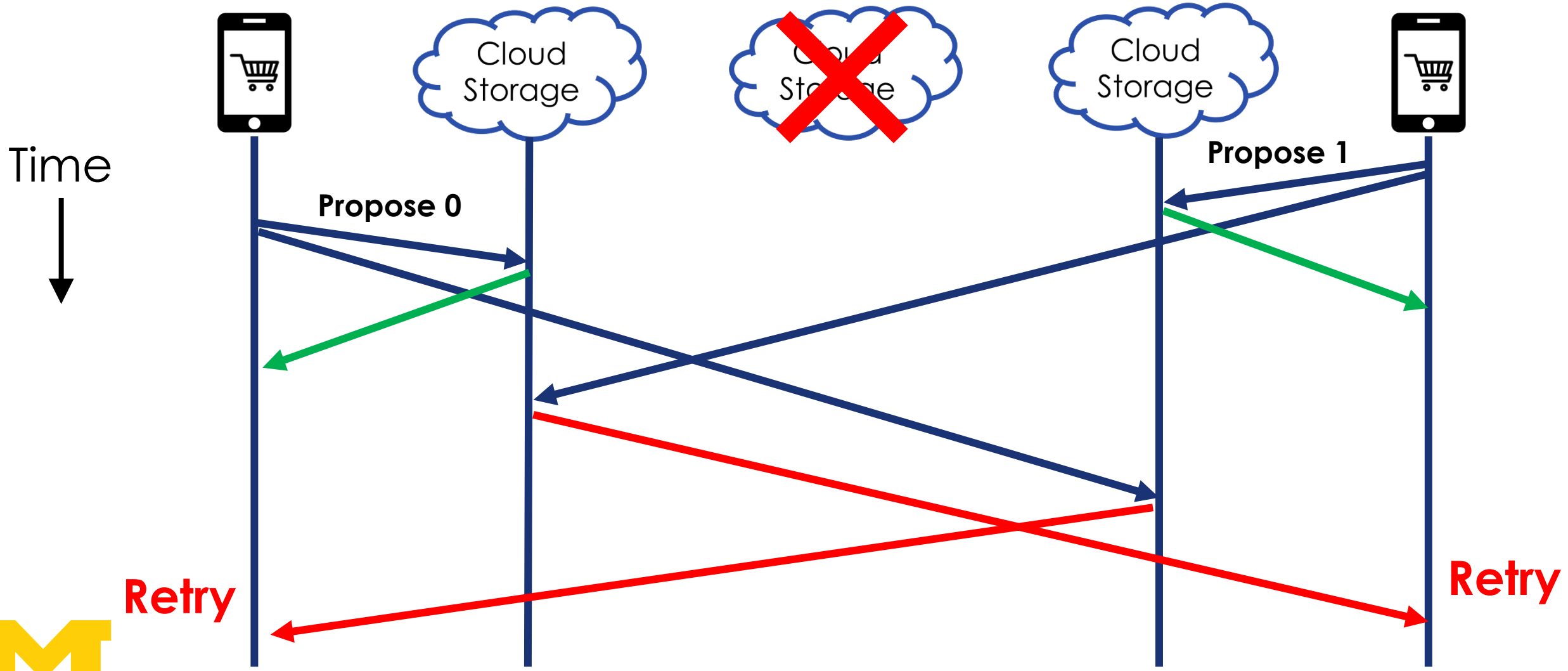
Tradeoff: High Latency under Conflict



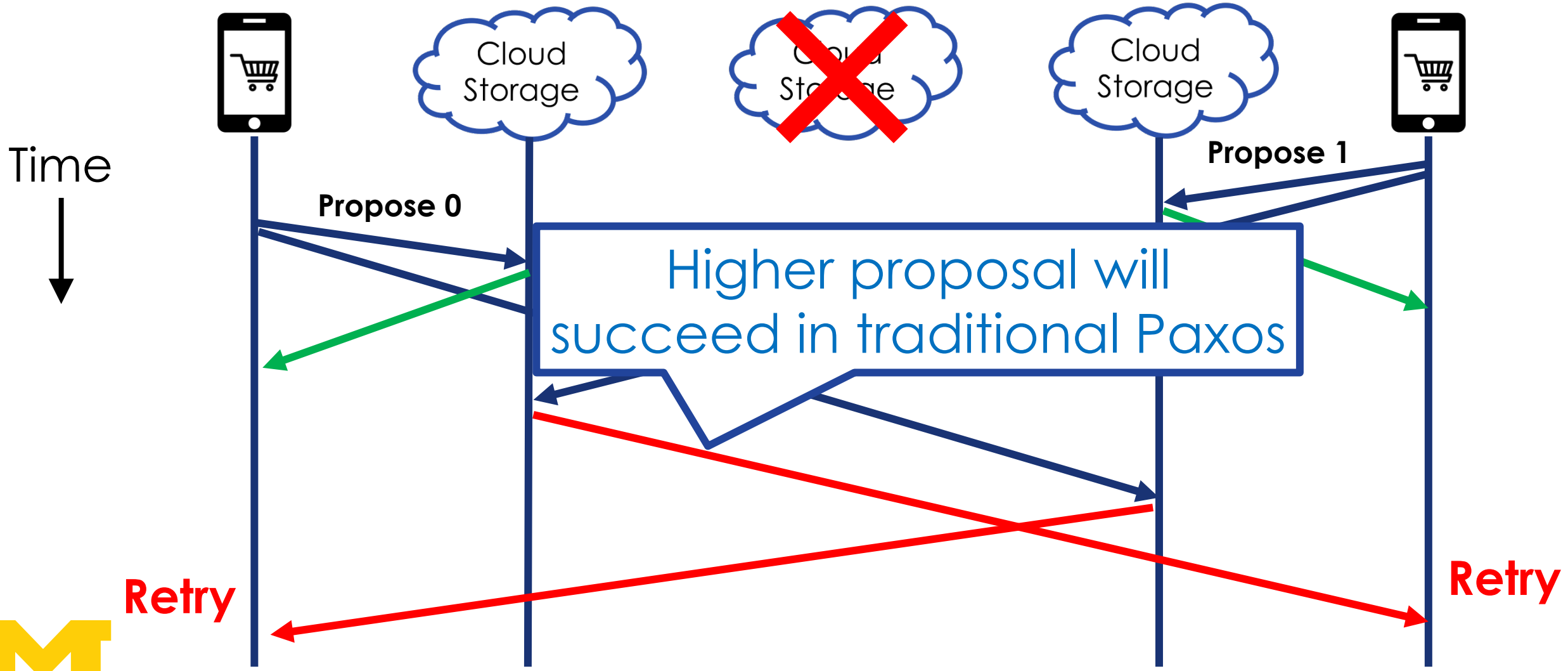
Tradeoff: High Latency under Conflict



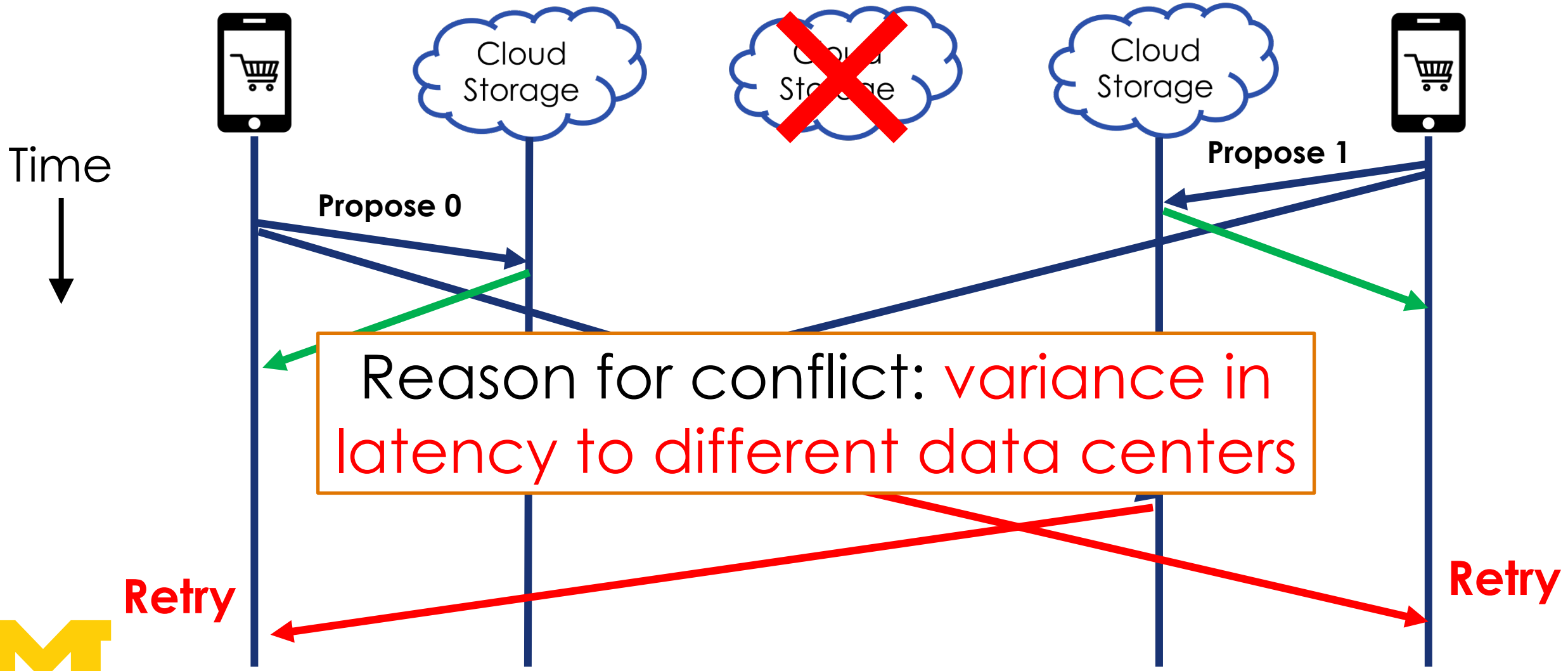
Tradeoff: High Latency under Conflict



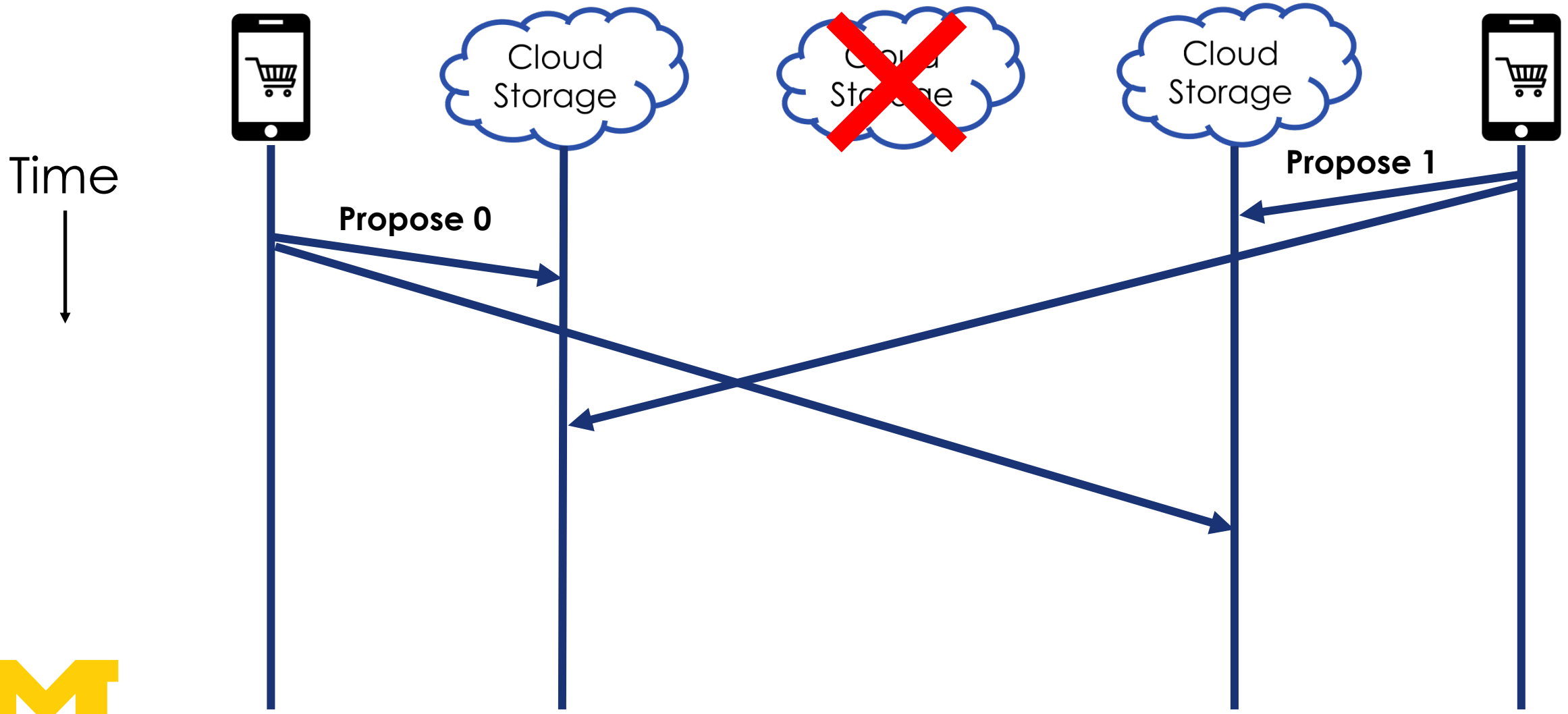
Tradeoff: High Latency under Conflict



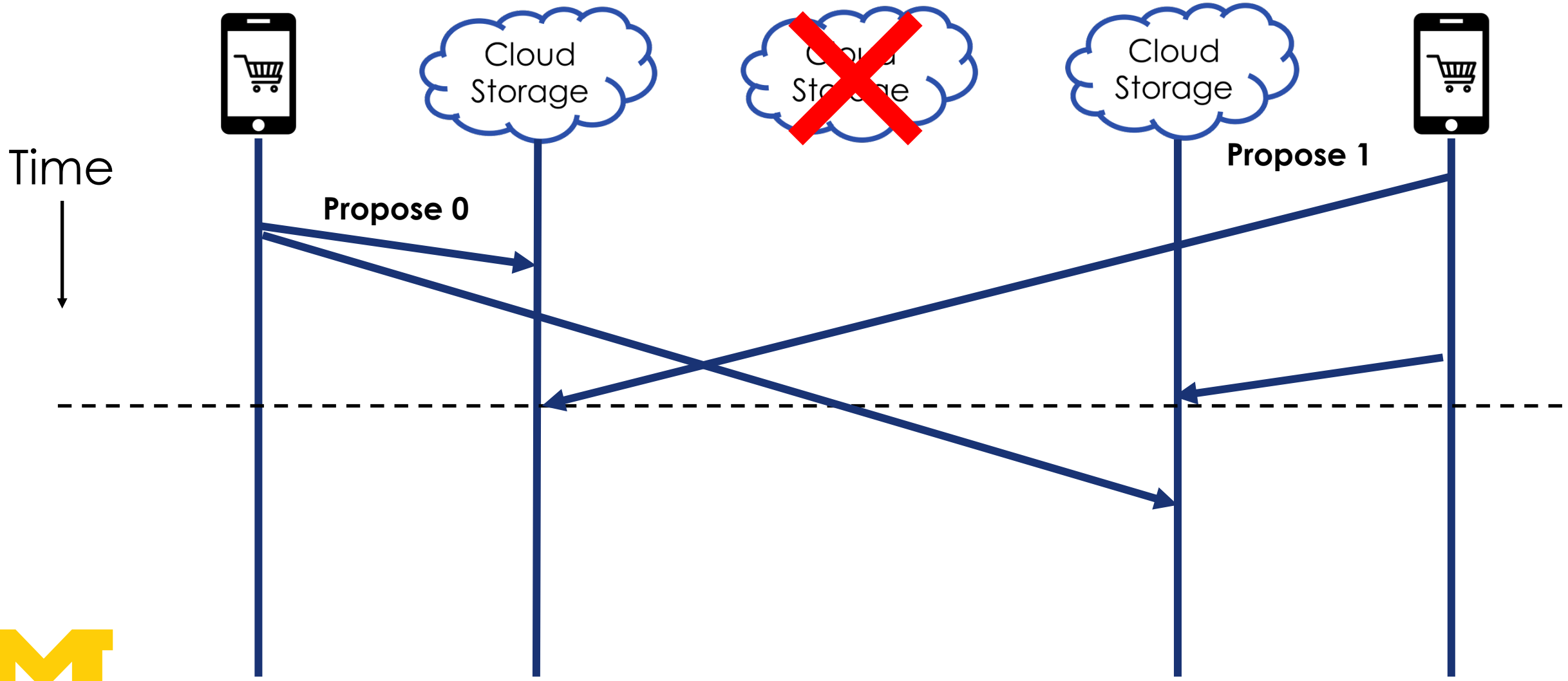
Tradeoff: High Latency under Conflict



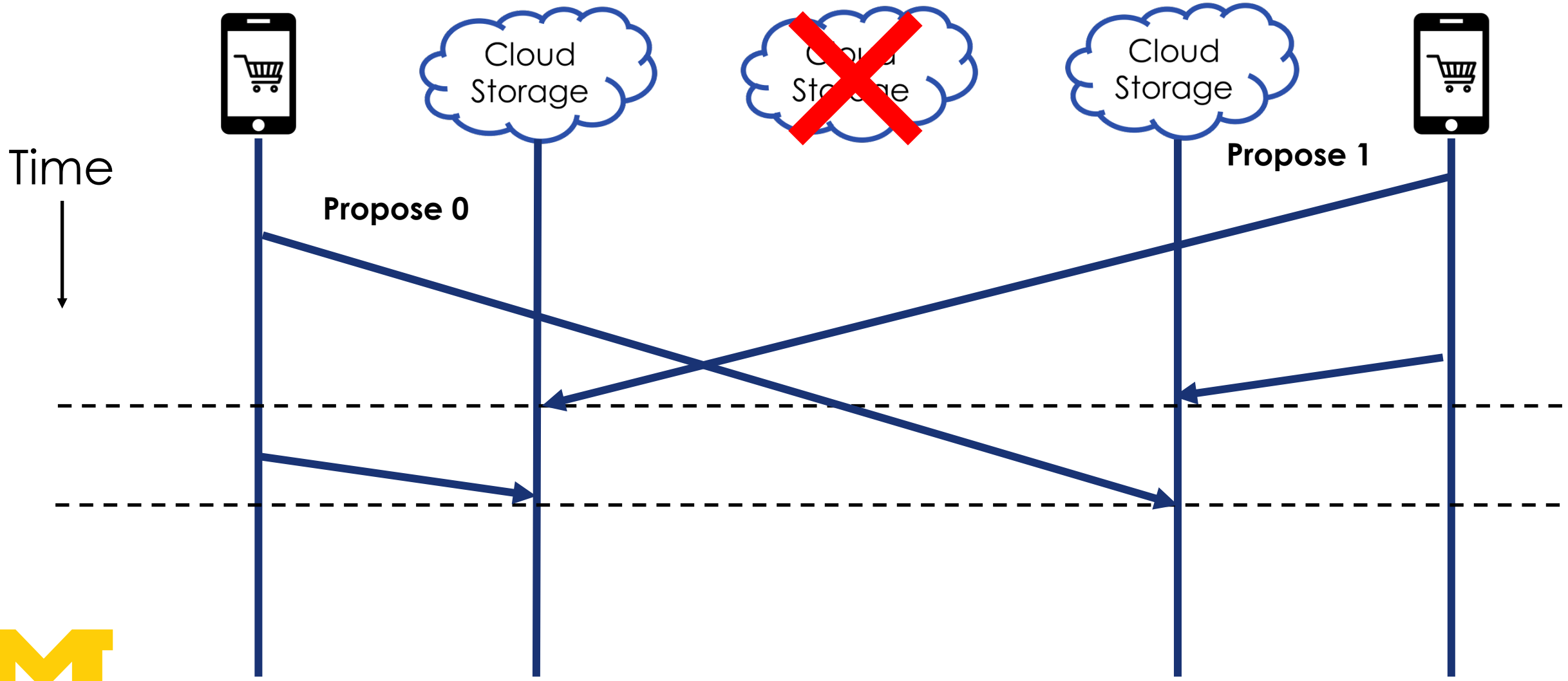
Optimization: Staggered Requests



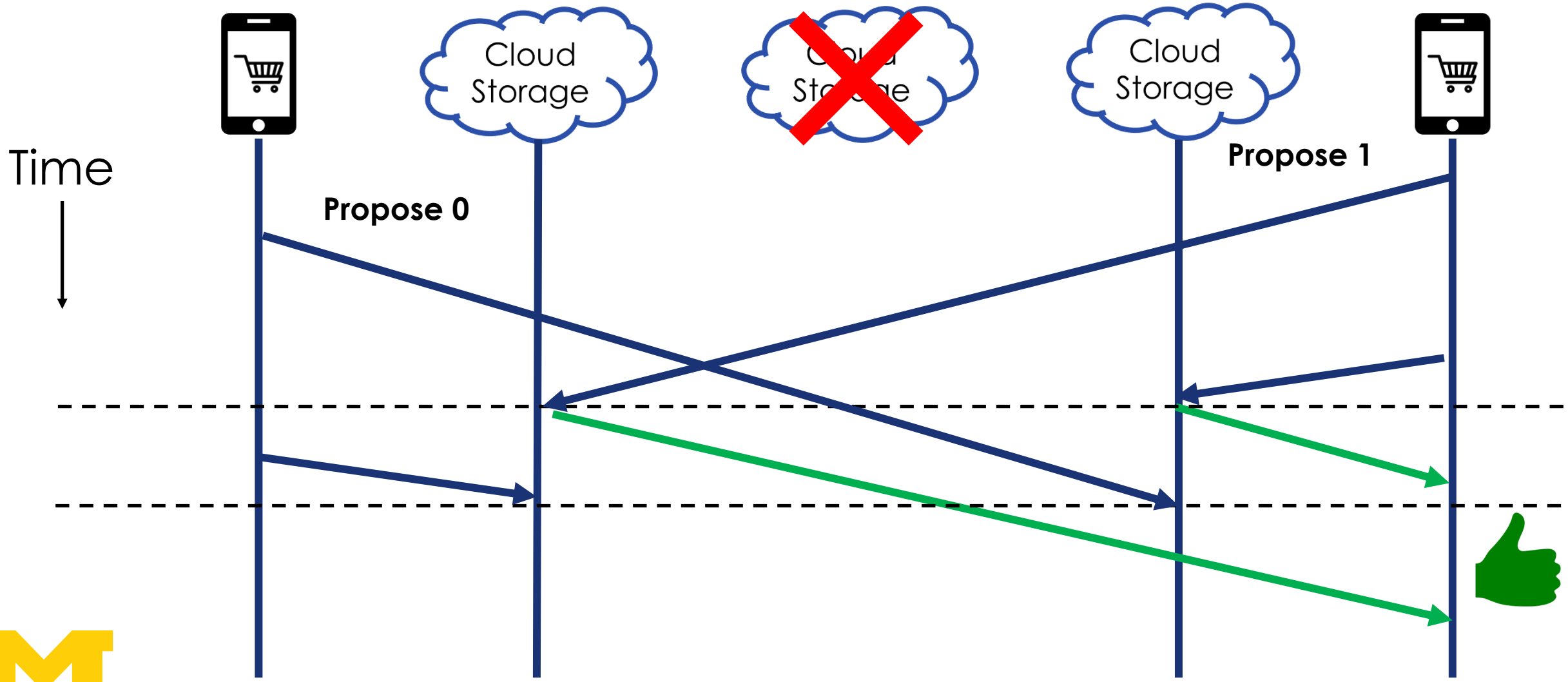
Optimization: Staggered Requests



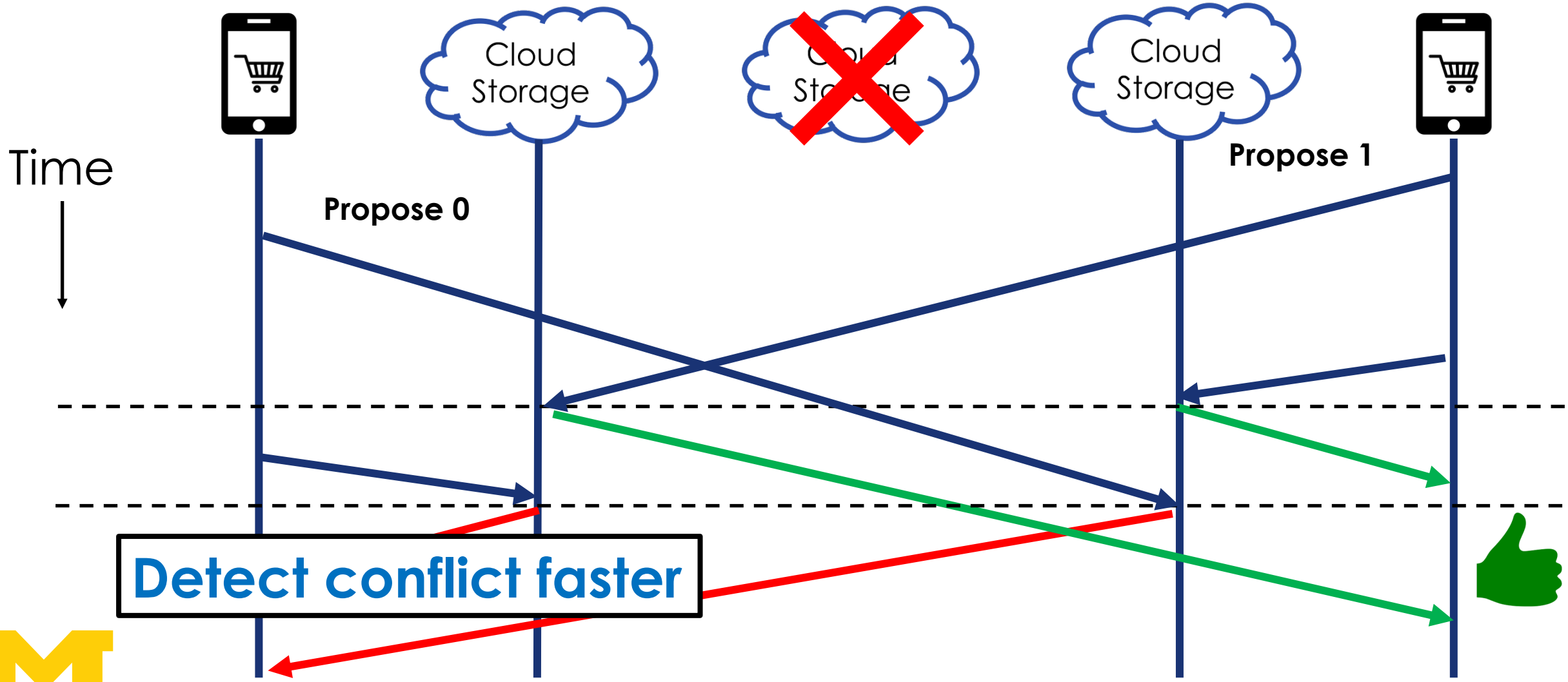
Optimization: Staggered Requests



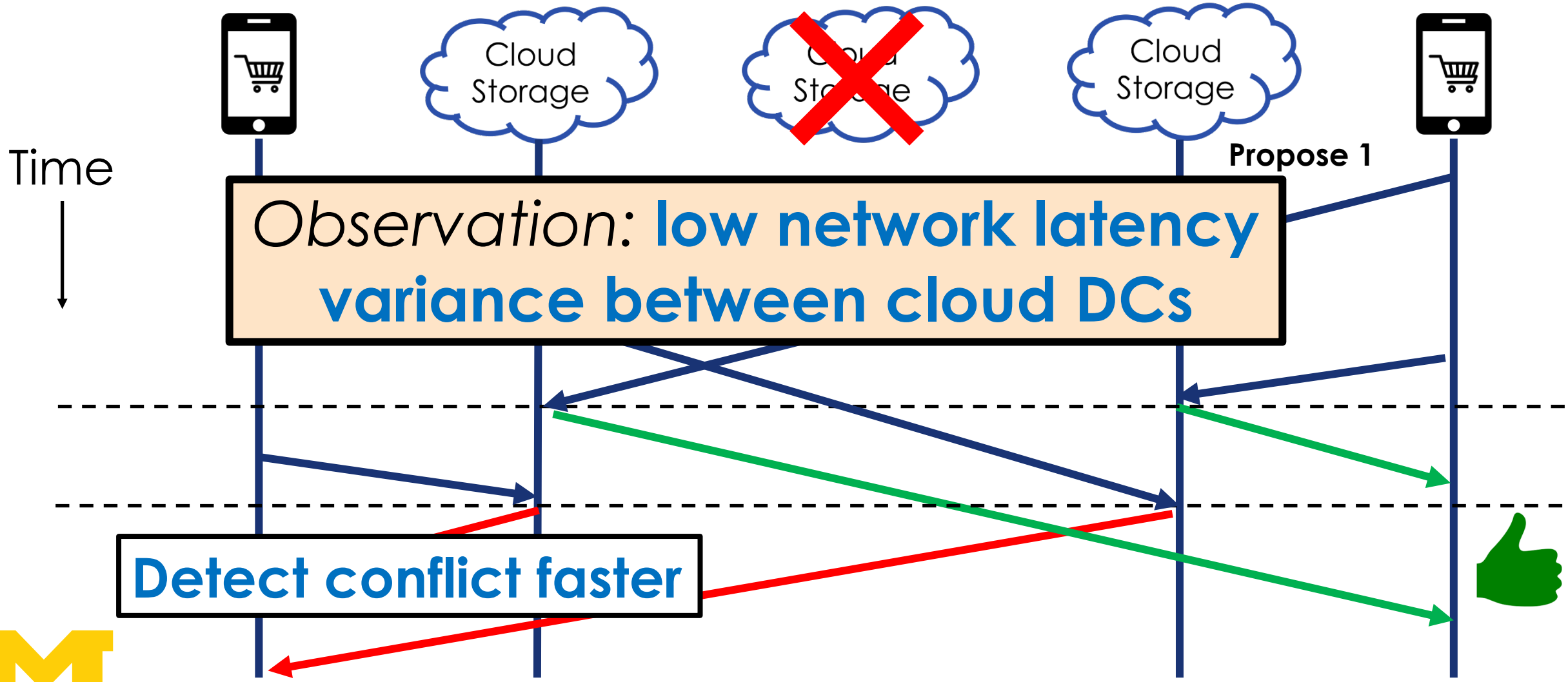
Optimization: Staggered Requests



Optimization: Staggered Requests



Optimization: Staggered Requests



CRIC Optimizations

- Reduce **latency under conflict**
  - Staggered Requests
- Reduce **reader-write-back**
  - Asynchronous commit notification
- Reduce **storage and data transfer cost**
  - Separates data and Paxos log
  - Aggressive garbage collection in Accept phase
  - Store data digest in Paxos log

CRIC Optimizations

- Reduce **latency under conflict**
  - Staggered Requests
- Reduce **reader-write-back**
  - Asynchronous commit notification
- Reduce **storage and data transfer cost**

Cost-effective

Only one version of the data is stored in each replica data center

Evaluation

- Deploy CRIC in **5 Azure data centers** and run **YCSB workload**
- Comparison systems:
  - active acceptor **Fast Paxos**
  - passive acceptor **pPaxos**

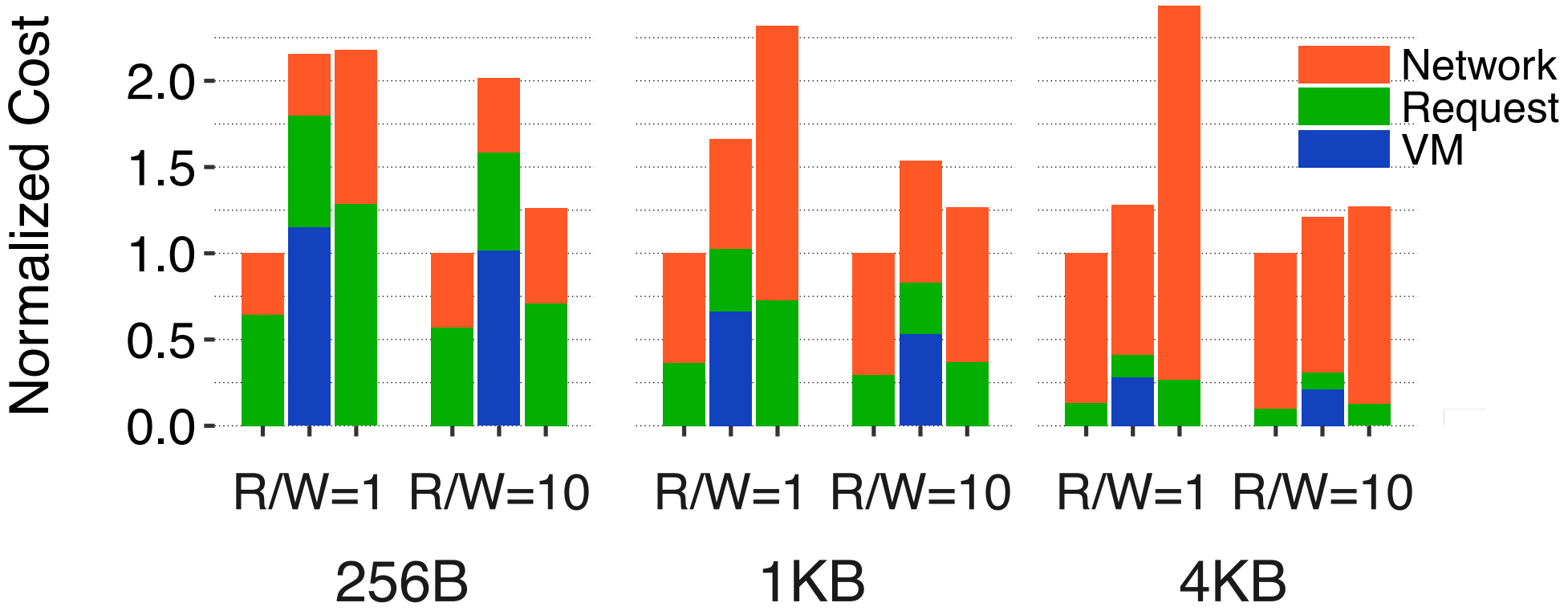
Evaluation

- Deploy CRIC in **5 Azure data centers** and run **YCSB workload**
- Comparison systems:
  - active acceptor **Fast Paxos**
  - passive acceptor **pPaxos**
- How does CRIC compare with respect to cost and performance?

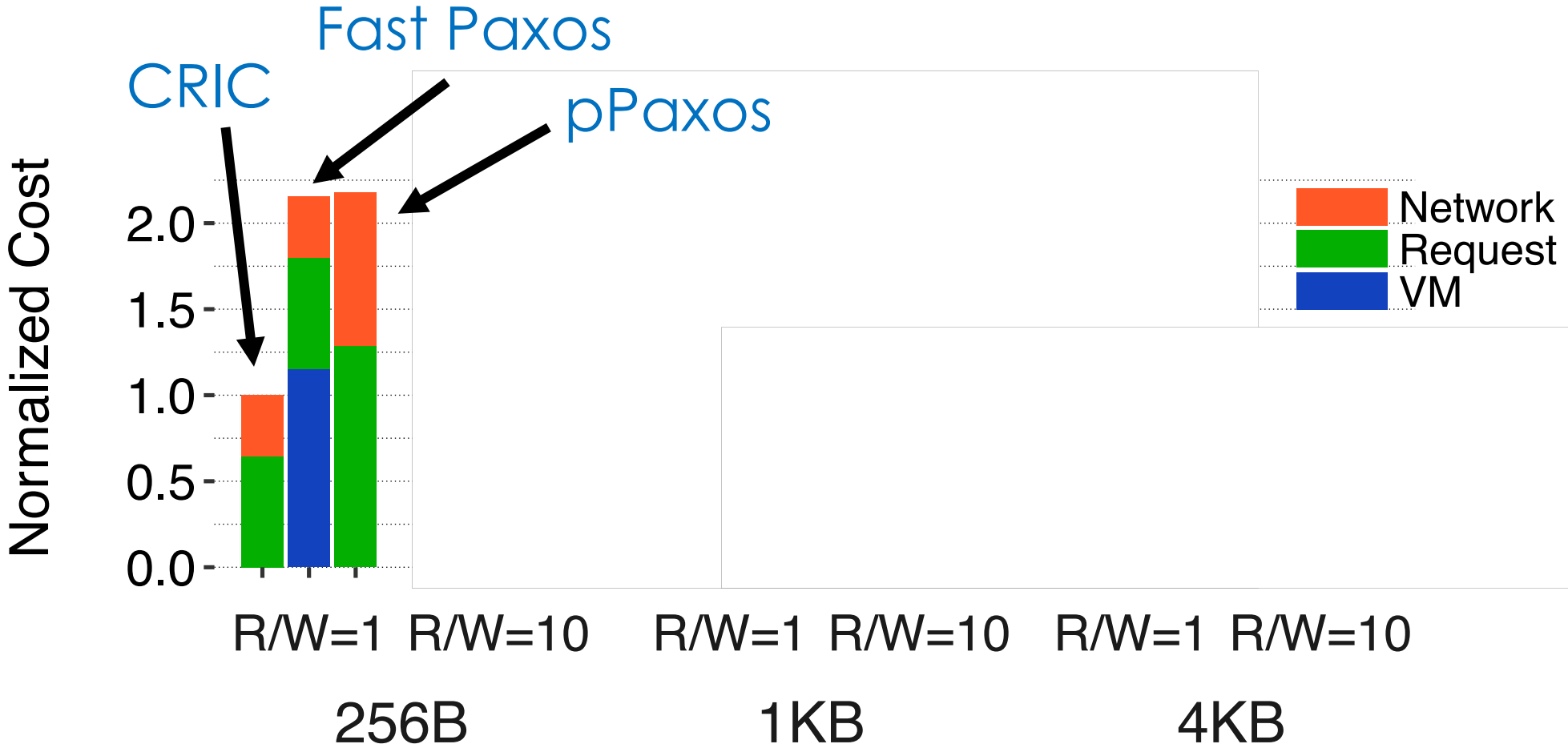
Evaluation

- Deploy CRIC in **5 Azure data centers** and run **YCSB workload**
- Comparison systems:
  - active acceptor **Fast Paxos**
  - passive acceptor **pPaxos**
- How does CRIC compare with respect to cost and performance?
- How effective are staggered requests?

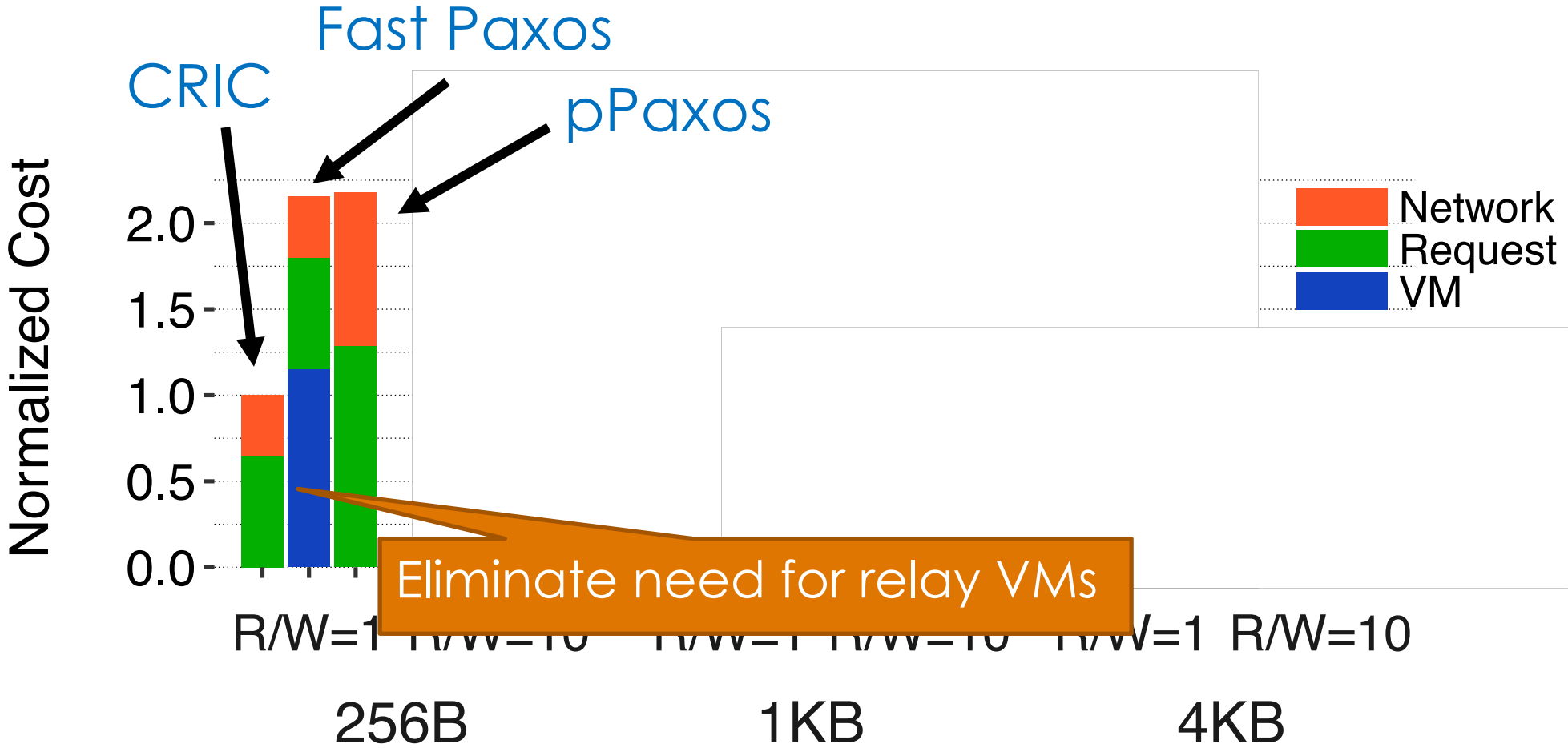
CRIC Enables Low Cost



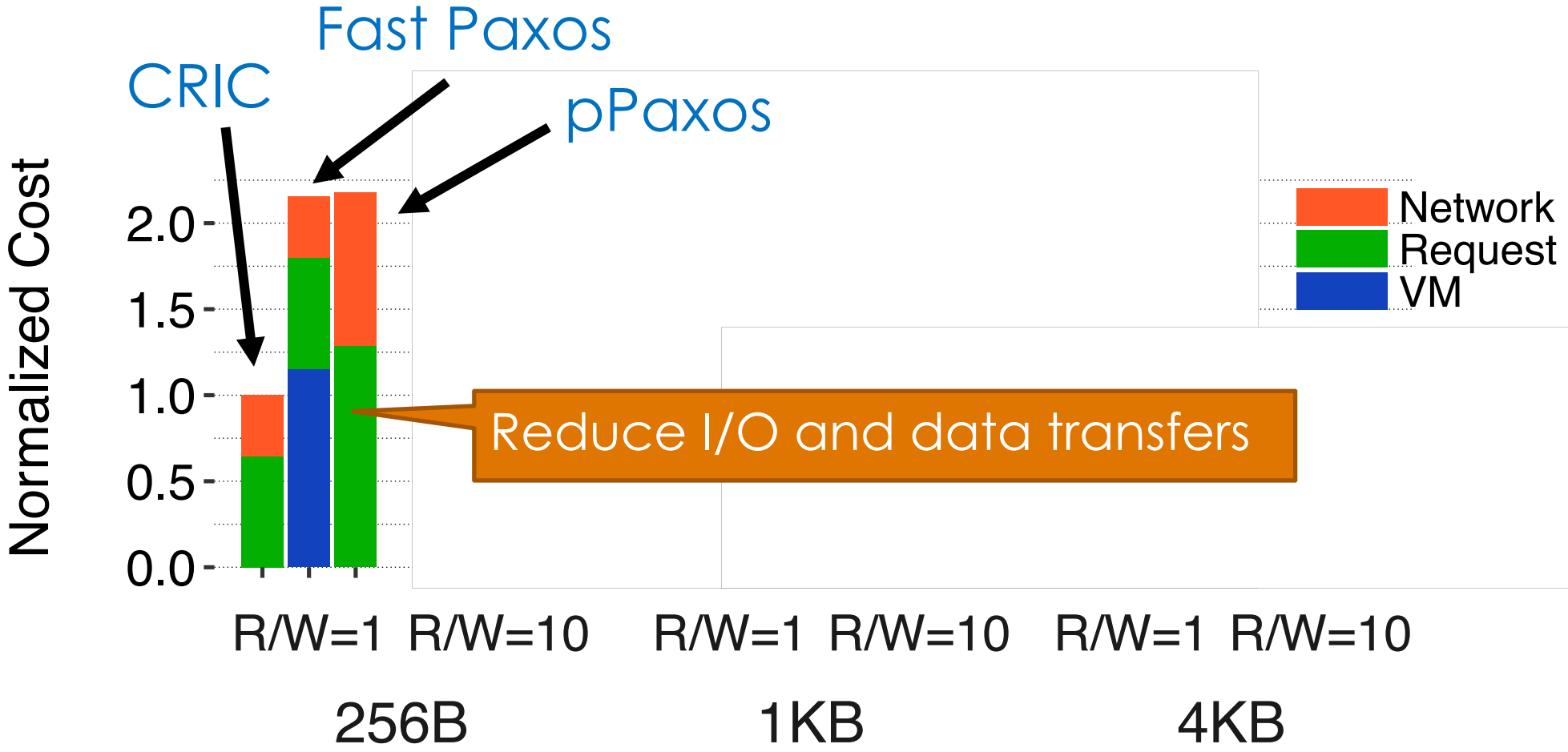
CRIC Enables Low Cost



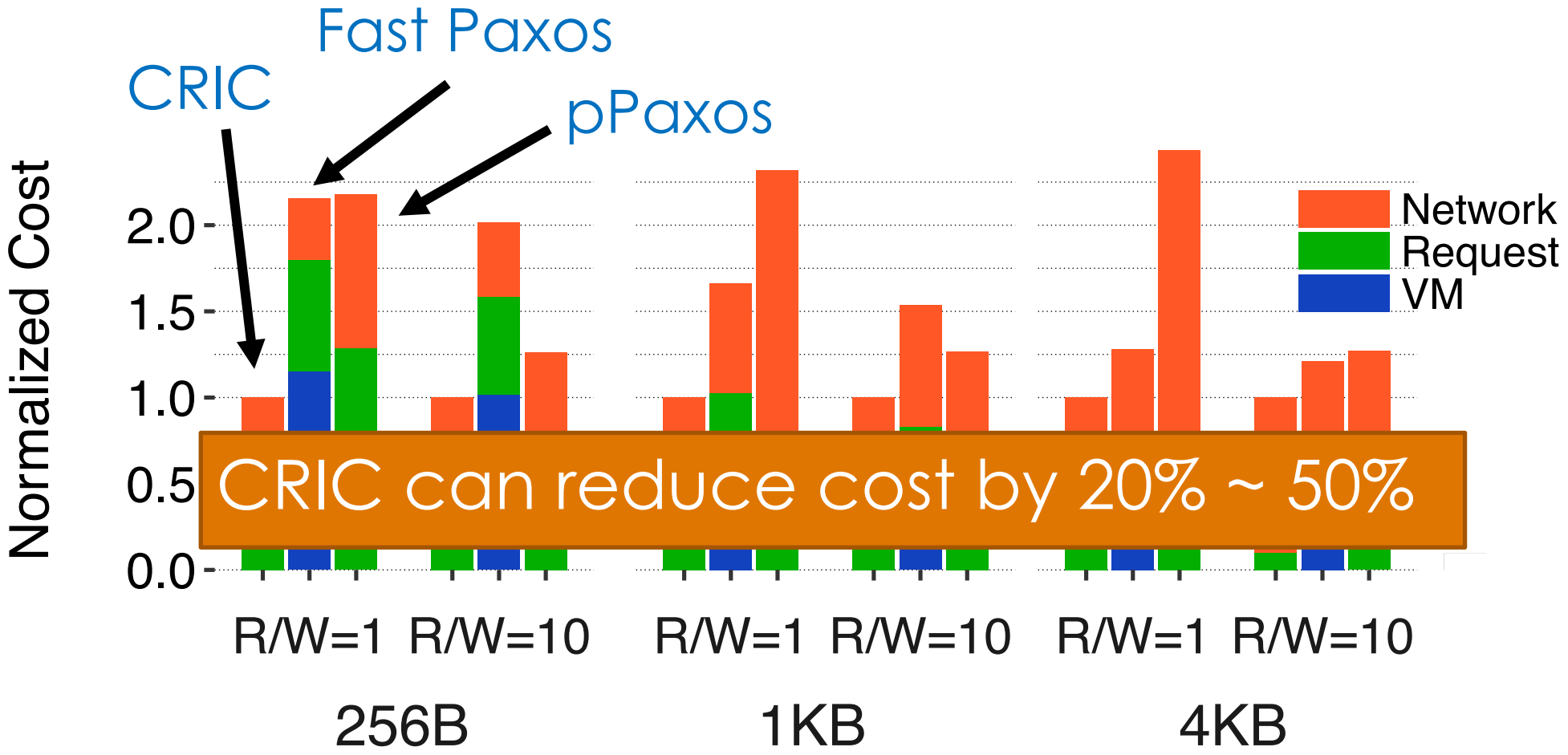
CRIC Enables Low Cost



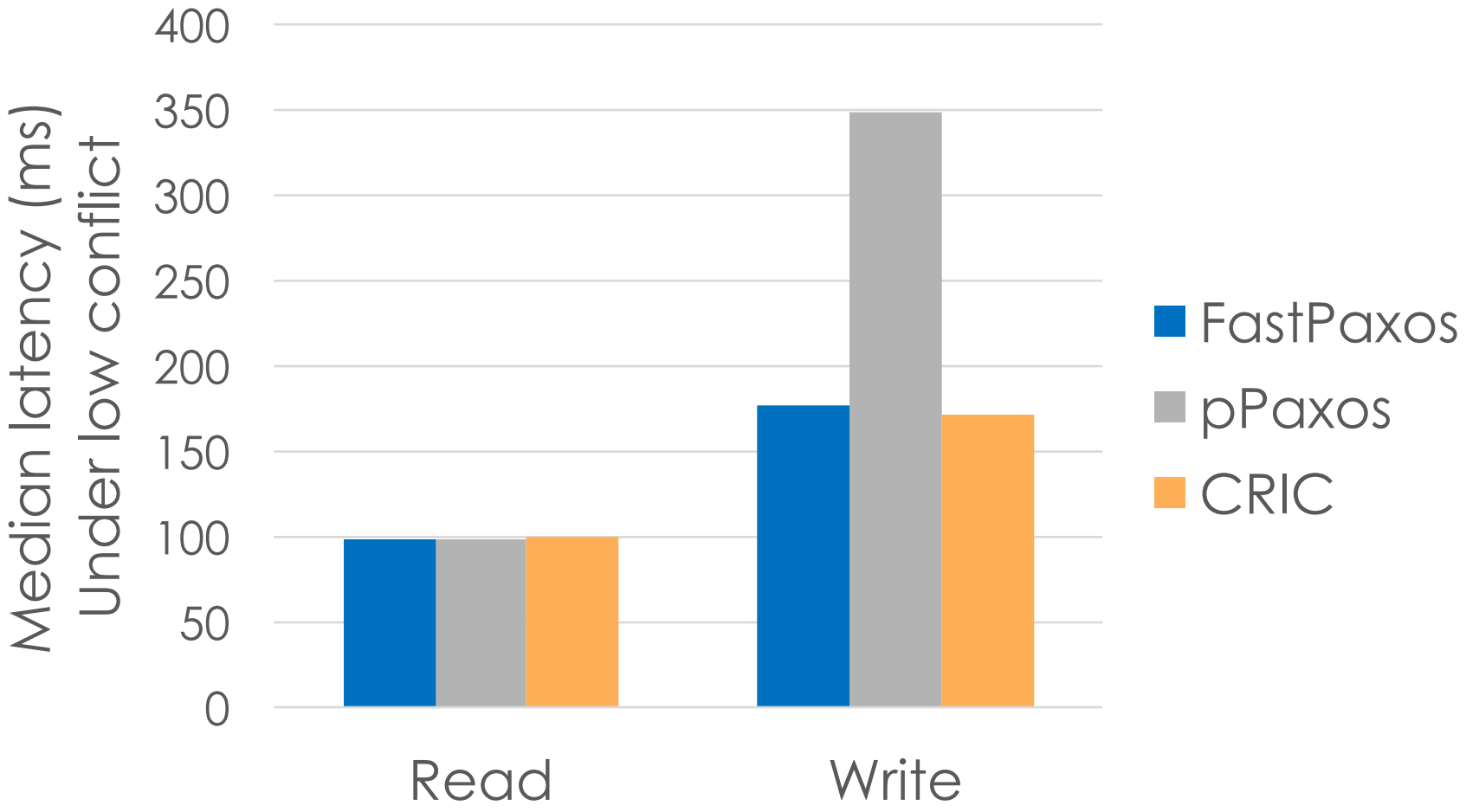
CRIC Enables Low Cost



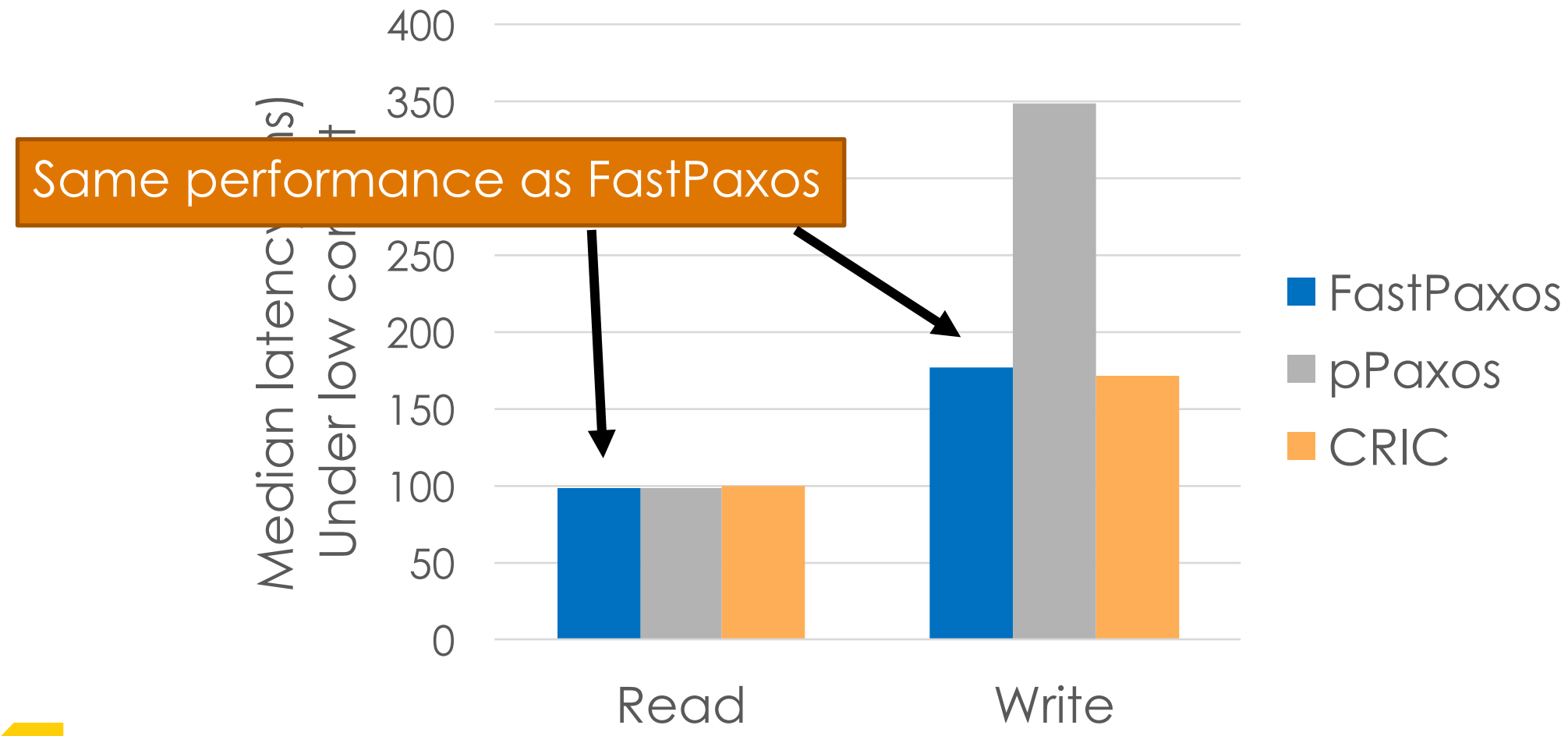
CRIC Enables Low Cost



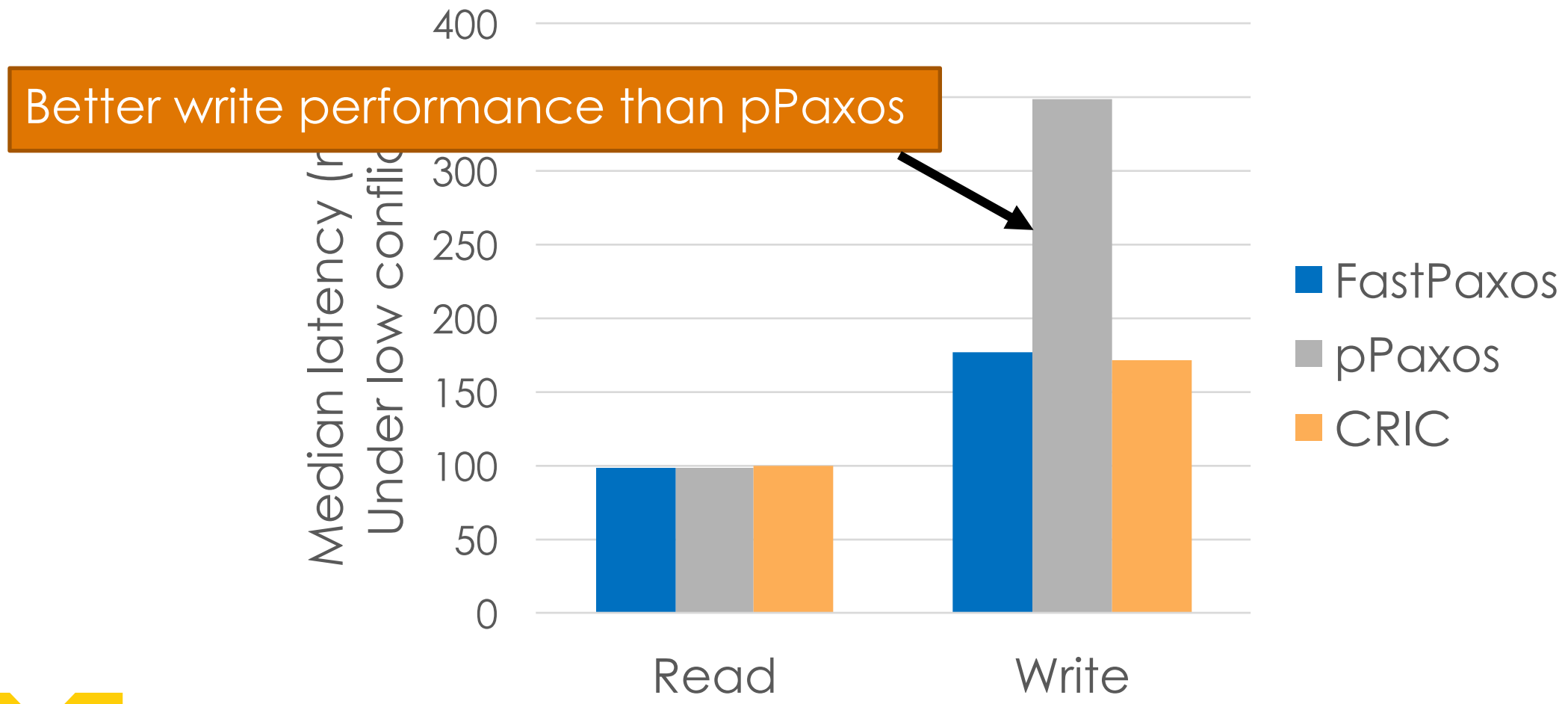
… without Sacrificing Performance



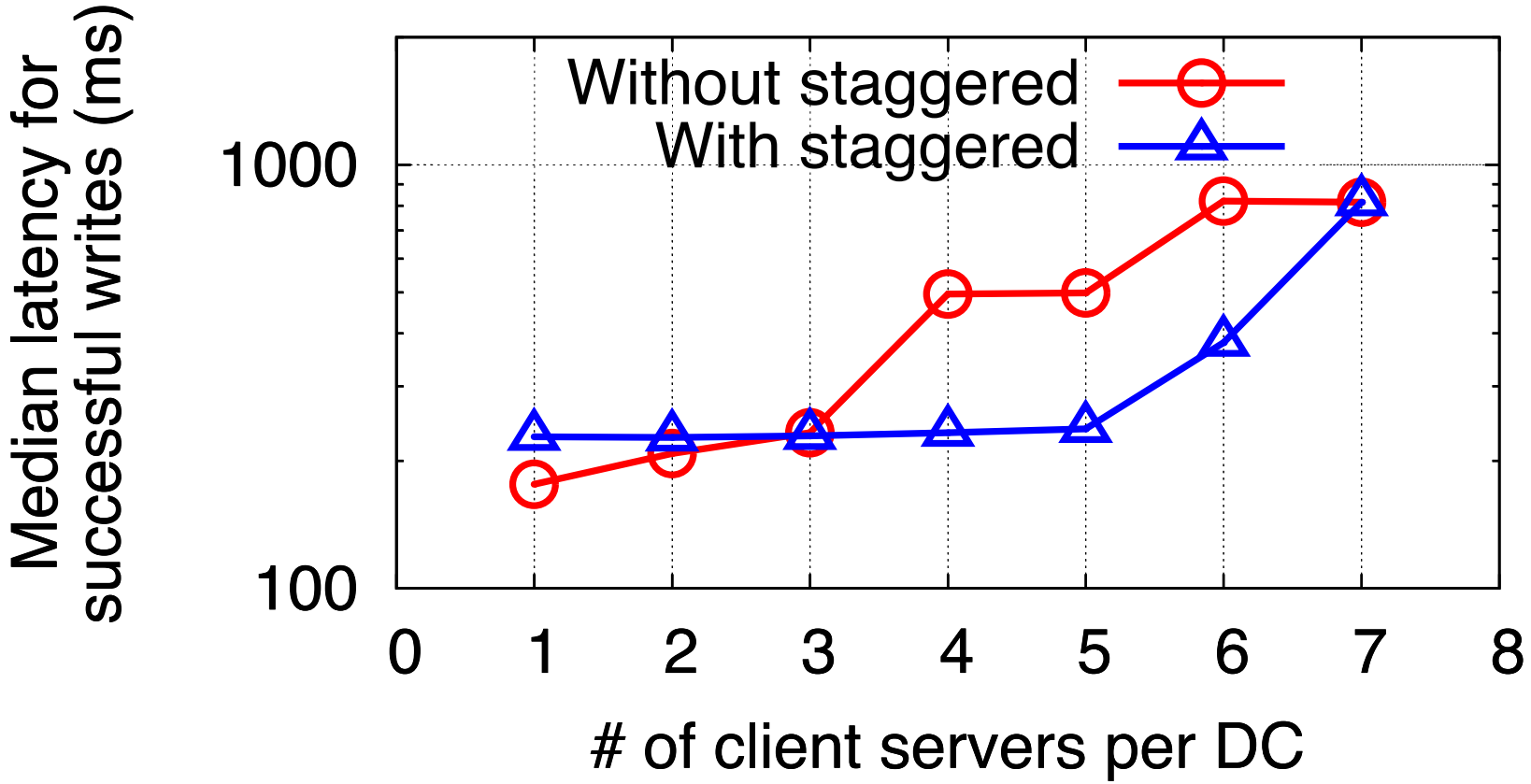
… without Sacrificing Performance



… without Sacrificing Performance

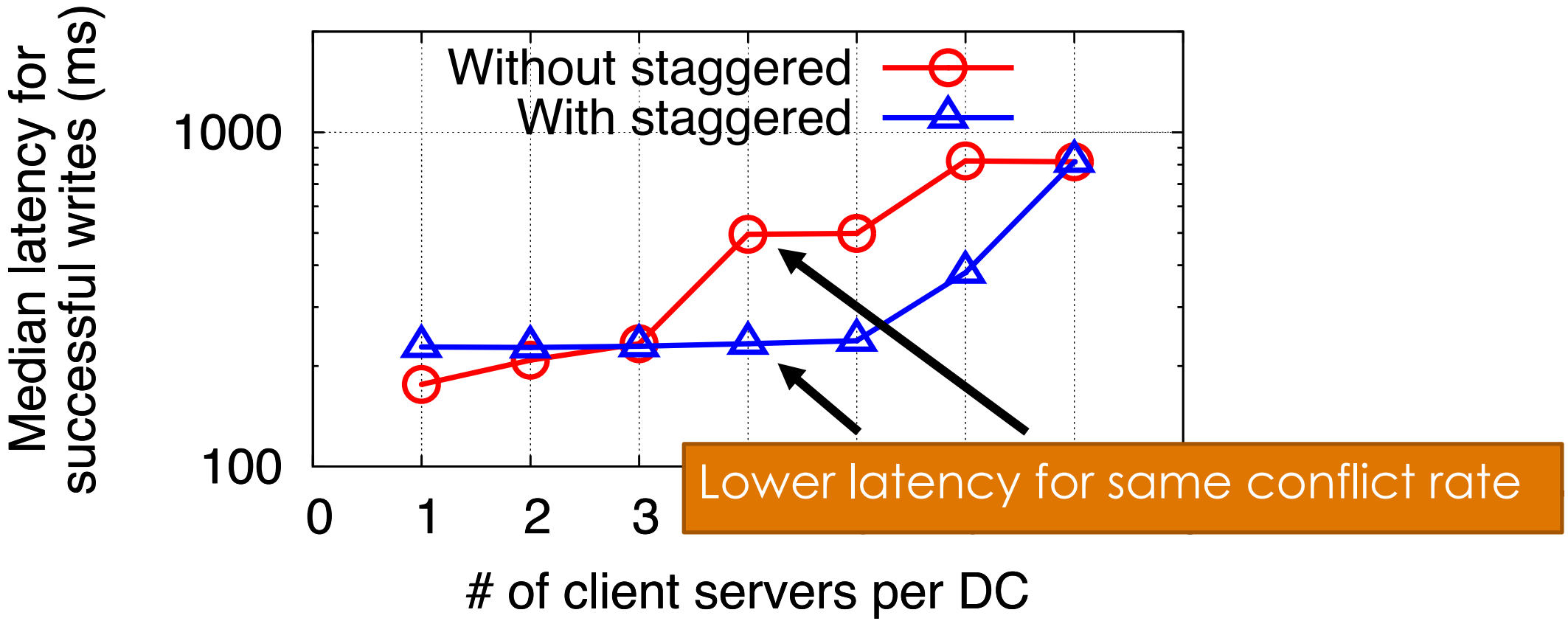


Staggered Requests Lower Latency Under Conflict



Increasing conflict rate

Staggered Requests Lower Latency Under Conflict



Conclusions

- **C***onsistent* ***R****eplication* ***I****n the* **C***loud*
  - **Compatible with cloud storage interface**
  - **One round read/write in common case**
  - **Low cost**

Thank you

towuzhe@gmail.com