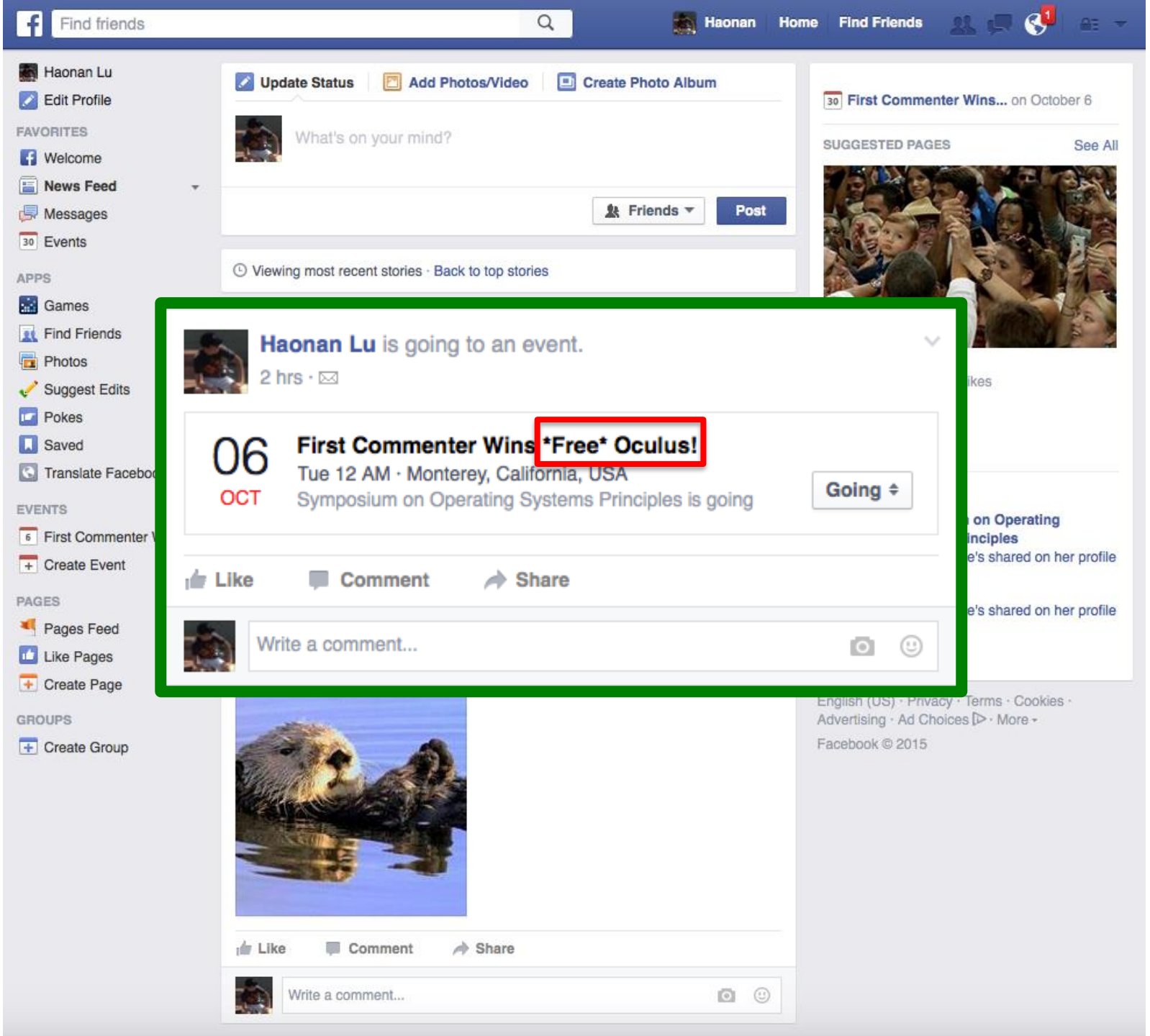
Existential Consistency: Measuring and Understanding Consistency at Facebook

Haonan Lu*†, Kaushik Veeraraghavan†,
Philippe Ajoux†, Jim Hunt†,
Yee Jiun Song†, Wendy Tobagus†,
Sanjeev Kumar†, Wyatt Lloyd*†

*University of Southern California, †Facebook


Find friends
Haonan
Home
Find Friends
Haonan Lu
Edit Profile
FAVORITES
Welcome
News Feed
Messages
Events
APPS
Games
Find Friends
Photos
Suggest Edits
Pokes
Saved
Translate Facebo
EVENTS
First Commenter
Create Event
PAGES
Pages Feed
Like Pages
Create Page
GROUPS
Create Group
Update Status
Add Photos/Video
Create Photo Album
What's on your mind?
Friends
Post
Viewing most recent stories · Back to top stories
Haonan Lu is going to an event.
2 hrs ·
06
OCT
First Commenter Wins *Free* Oculus!
Tue 12 AM · Monterey, California, USA
Symposium on Operating Systems Principles is going
Going
Like
Comment
Share
Write a comment...
Like
Comment
Share
Write a comment...
First Commenter Wins... on October 6
SUGGESTED PAGES
See All
on Operating
inciples
e's shared on her profile
e's shared on her profile
English (US) · Privacy · Terms · Cookies ·
Advertising · Ad Choices · More
Facebook © 2015

Find friends

Haonan | Home | Find Friends

Haonan Lu
Edit Profile

FAVORITES
Welcome
News Feed
Messages
Events

APPS
Games
Find Friends
Photos
Suggest Edits
Pokes
Saved

EVENTS
First Commenter W...
Create Event

PAGES
Pages Feed
Like Pages
Create Page

GROUPS
Create Group

Update Status | Add Photos/Video | Create Photo Album

What's on your mind?

Friends | Post

Viewing most recent stories · Back to top stories

Haonan Lu is going to an event.
2 hrs ·

06
OCT

First Commenter Wins *Free* Oculus!
Tue 12 AM · Monterey, California, USA
Symposium on Operating Systems Principles is going

Going

Like | Comment | Share

Haonan Lu Mine! yeah~ lucky!

1 min · Like

Write a comment...

First Commenter Wins... on October 6

GAMES YOU MAY LIKE See All

Clash of Kings
1 million players
Play Now

War Commander
100,000 players
Now

y · Terms · Cookies ·
ces · More


Find friends
Haonan
Home
Find Friends
Haonan Lu
Edit Profile
FAVORITES
Welcome
Update Status
Add Photos/Video
Create Photo Album
What's on your mind?
Consistency
Performance
APPS
Games
Find Friends
Photos
Suggest Edits
Pokes
Saved
EVENTS
First Commenter W...
Create Event
PAGES
Pages Feed
Like Pages
Create Page
GROUPS
Create Group
Haonan Lu is going to an event.
1 hr
06
OCT
First Commenter Wins *Free* Oculus!
Tue 12 AM · Monterey, California, USA
Symposium on Operating Systems Principles is going
Going
Like
Comment
Share
Wyatt Lloyd I wouldn't mind taking one.
IS THIS REAL LIFE?
1 min · Like
Haonan Lu Mine! yeah~ lucky!
GOOD LUCK!
Just now · Like
Write a comment...
Symposium on Operating Systems Principles updated her profile picture.
2 hrs
First Commenter Wins... on October 6
GAMES YOU MAY LIKE
See All
10 million players
Play Now
English (US) · Privacy · Terms · Cookies · Advertising · Ad Choices · More
Facebook © 2015

Fundamental Tension

Consistency VS Performance

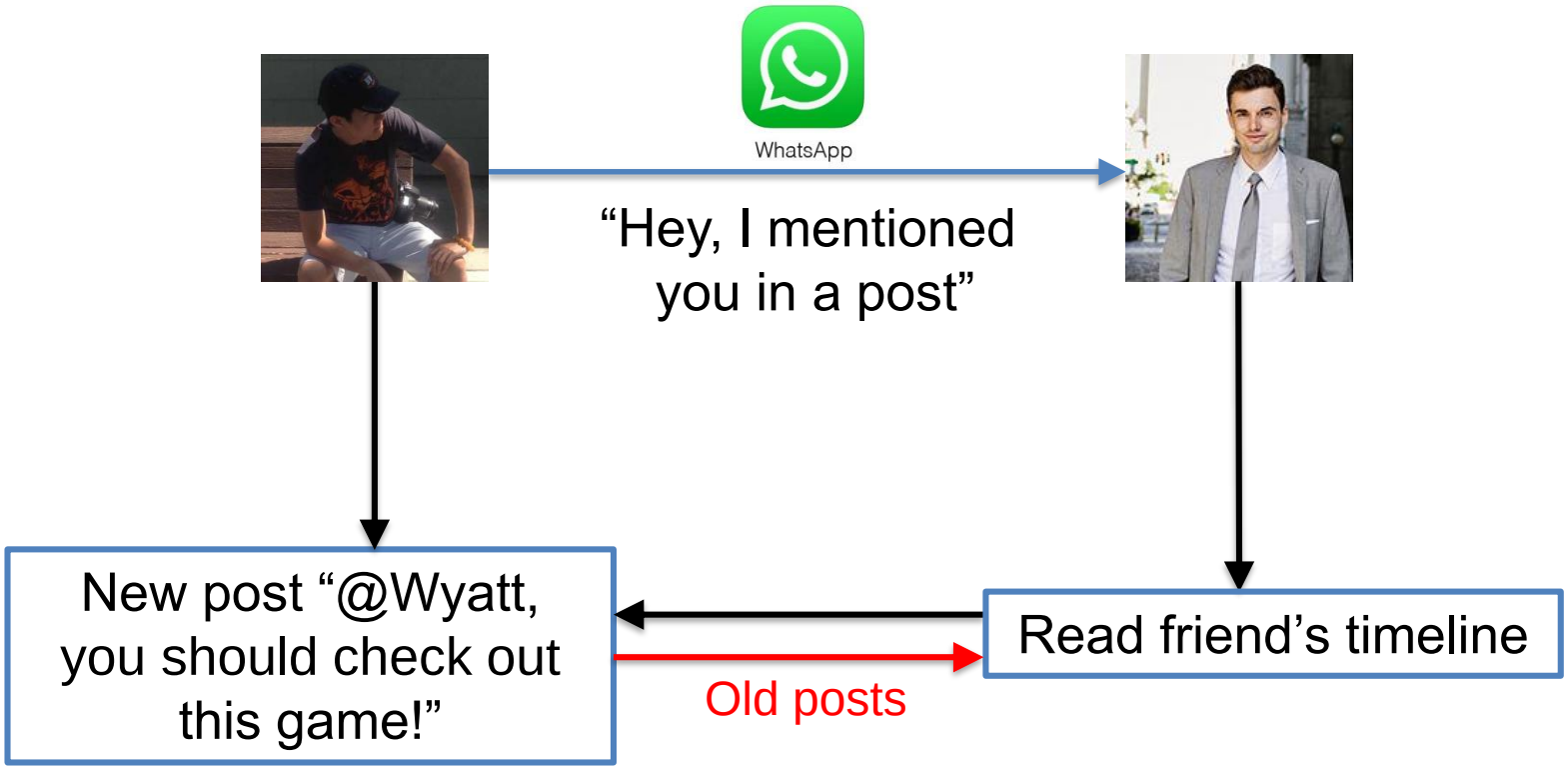
- Eliminates anomalies (Oculus example)
- Makes systems easier to program
- Difficult to quantify

- Lower latency
- Higher throughput
- Simple to quantify

First study of consistency in a large-scale, production system – Facebook TAO

Anomaly: Unexpected Behavior

Post Example



Anomaly: Unexpected Behavior

Oculus Example

1. “Mine! yeah~ lucky!”

1. “I wouldn’t mind…”

😢

😀

1. “I wouldn’t mind…”
2. “Mine! yeah~ lucky!”

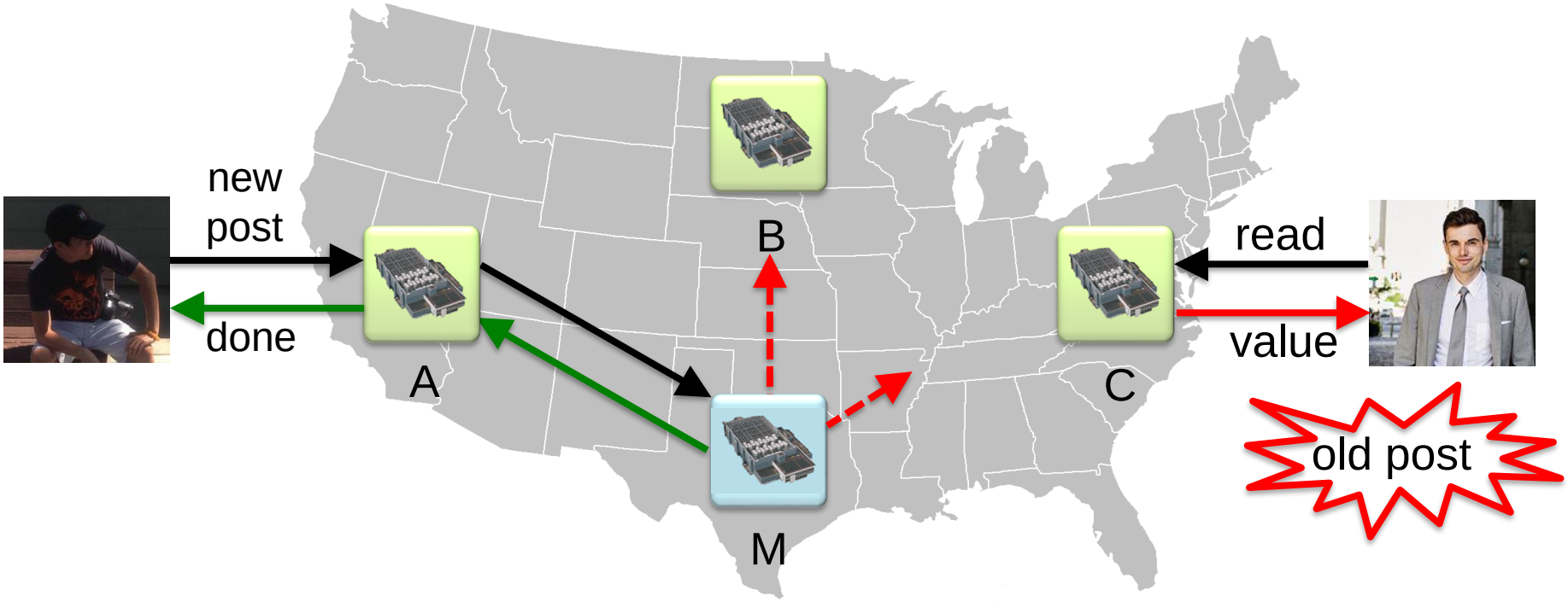
Does Facebook have consistency anomalies?

How many?

What type?

TAO: Eventually Consistent Cache

Vulnerability window: time during asynchronous replication when anomalies can happen



Quantifying Anomalies

- How often do anomalies occur?
  - Collect trace of requests to TAO
- What consistency would prevent them?
  - Run anomaly checkers on the trace

Trace Collection

- Collect trace on web servers

- Challenges in tracing production system
  - Volume of requests
  - Time skew between web servers
  - Missing requests

Challenge: Volume of Requests

- Billions of requests per second [ATC ’13]
  - Too many to log
- Sample on objects
  - Object: vertex in social graph
  - Log all requests to objects in sample
  - Sufficient for local consistency models

Local Property Enables Sampling

- "… the system as a whole satisfies P whenever each individual object satisfies P."[1]

Local consistency models can be checked on a per object basis

- Local
  - Linearizability
  - Per-Object Sequential
  - Read-After-Write
- **Non-local**
  - **Strict Serializability**
  - **Causal**

[1] M. P. Herlihy and J. M. Wing "Linearizability: A Correctness Condition for Concurrent Objects." ACM TOPLAS, 1990

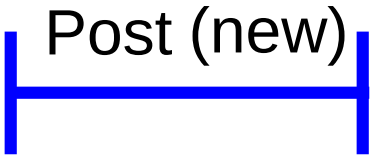
Challenge: Time Skew

- Time skew across web servers
  - 99.9 percentile for 1 week: 35ms
- Add time skew to request's duration
  - More overlapped requests
  - Eliminates false positives

Logging Details

- Logged information:
  - Start time
  - Finish time
  - Read or write
  - Value: match read with write

} Determine real time ordering of requests



- Sampling rate: 1 out of 1 million objects
  - ~100% of requests to sampled objects

Trace Statistics

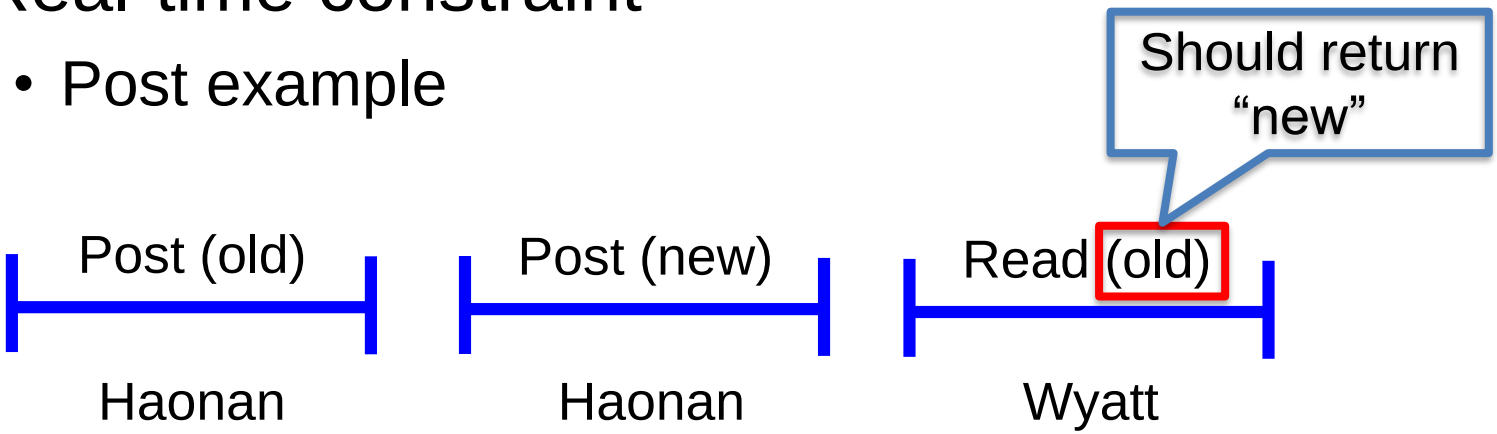
- 12 days (8/20 – 8/31)
- 17 million objects
- 3 billion requests

Check Trace for Anomalies

- Linearizability checker
  - Paxos provides
- Per-Object Sequential checker
  - PNUTS provides
- Read-After-Write checker
  - TAO provides within a cluster

Linearizability

- Strongest non-transactional consistency
  - Real-time constraint
    - Post example



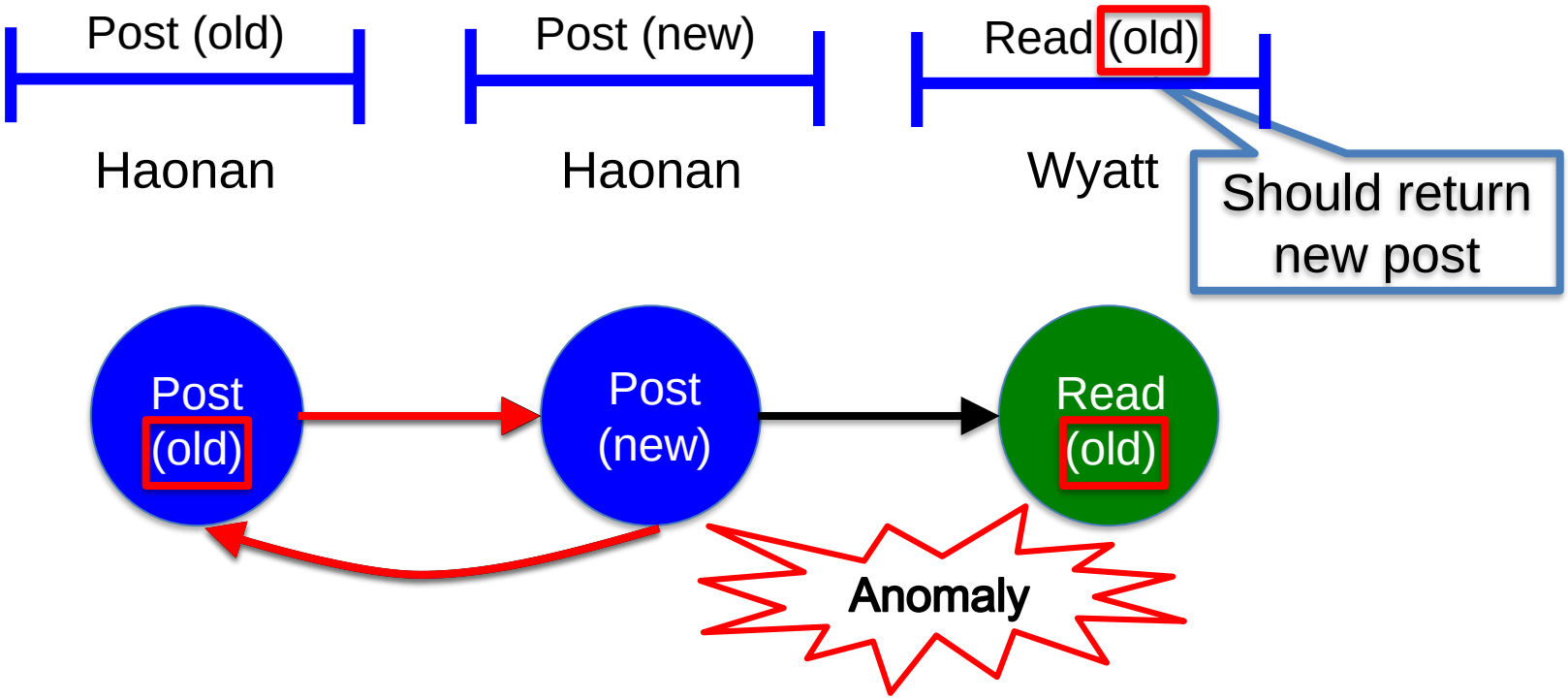
  - Total order constraint
    - Oculus example!

Linearizability Checker

- Graph captures state transitions
  - Vertex: write operations
  - Edge: real-time order
- Merge read with its write
  - Captures state transitions seen by users
- Anomaly if merge causes a cycle
  - Cycle indicates user's view ≠ system view

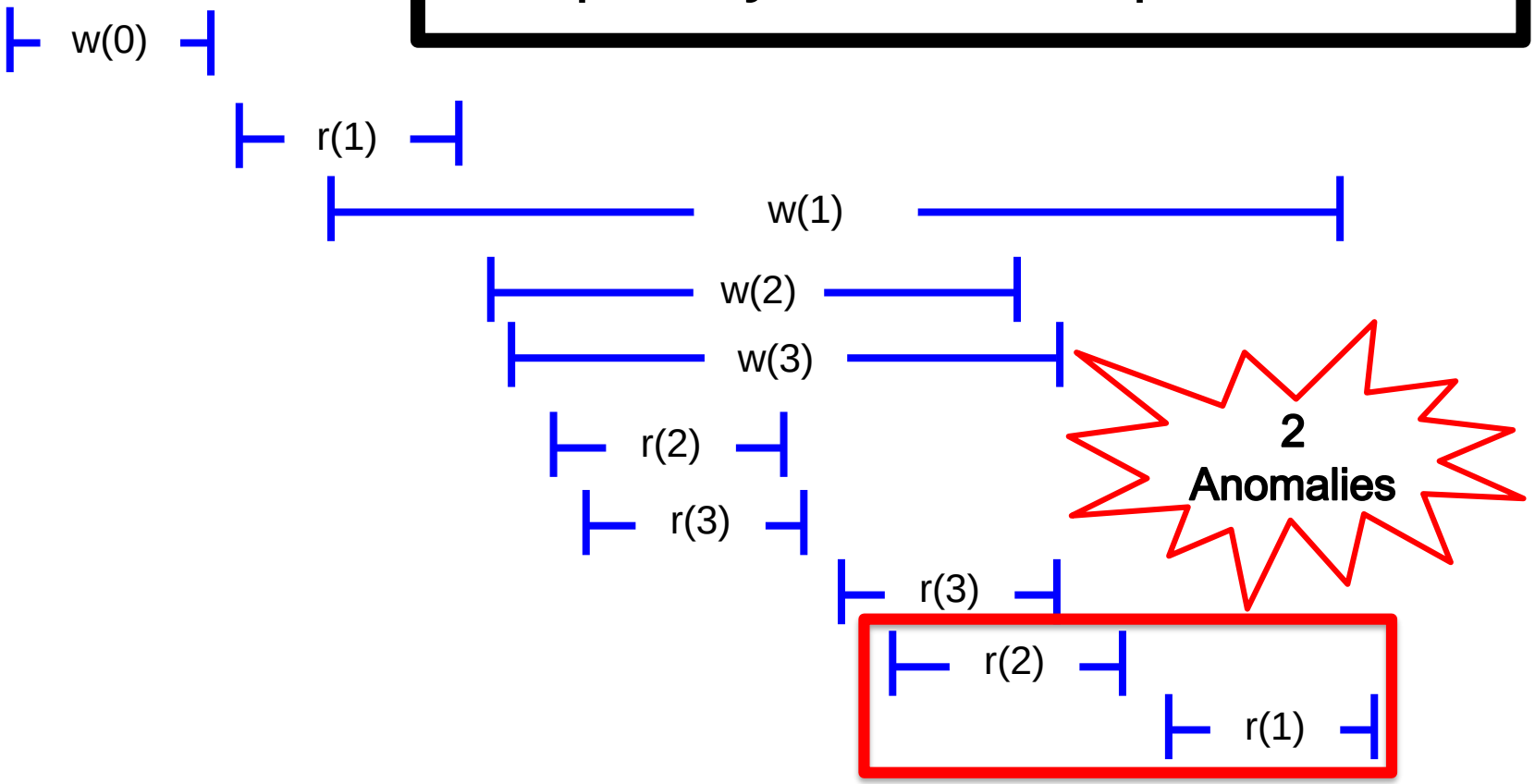
Linearizability Checker

- Captures real-time constraint
  - Read should return new post instead



More Complex Cases

http://tinyurl.com/sosp15-demo



Result Overview

- Linearizability
- Per-Object Sequential
- Read-After-Write
- Bounds on non-local consistency models

Anomalies found for all consistency models
– adopting them would have benefits

Linearizability Results

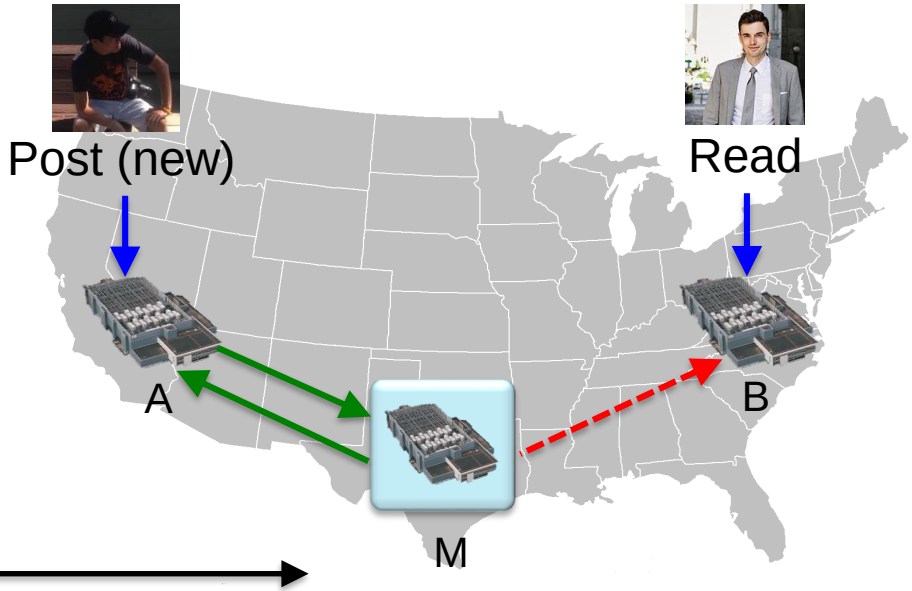
- 5 anomalies per million reads
  - Prevented by Paxos-based implementation

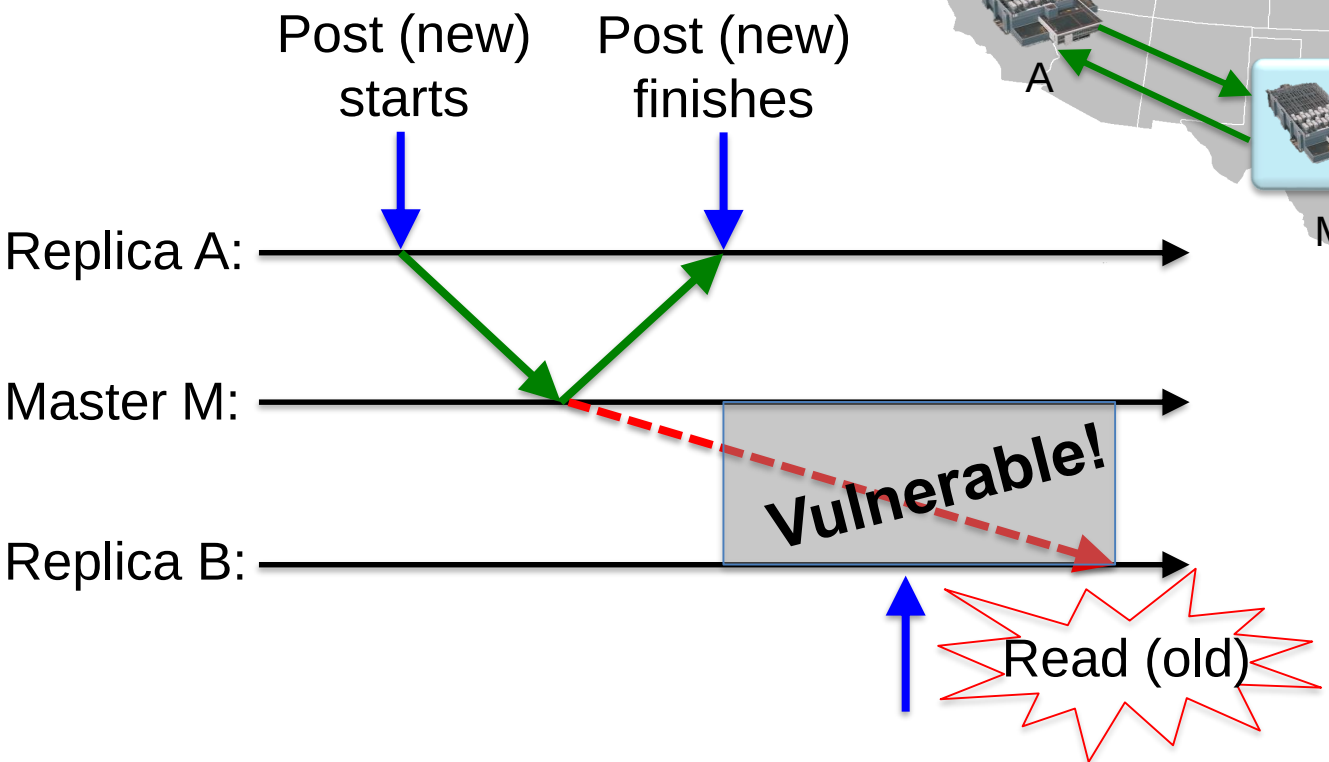
- Upper bound on TAO anomalies
  - Strongest consistency we checked

TAO is highly consistent

Linearizability Results
Real-Time Constraint Violations

- 4 per million reads

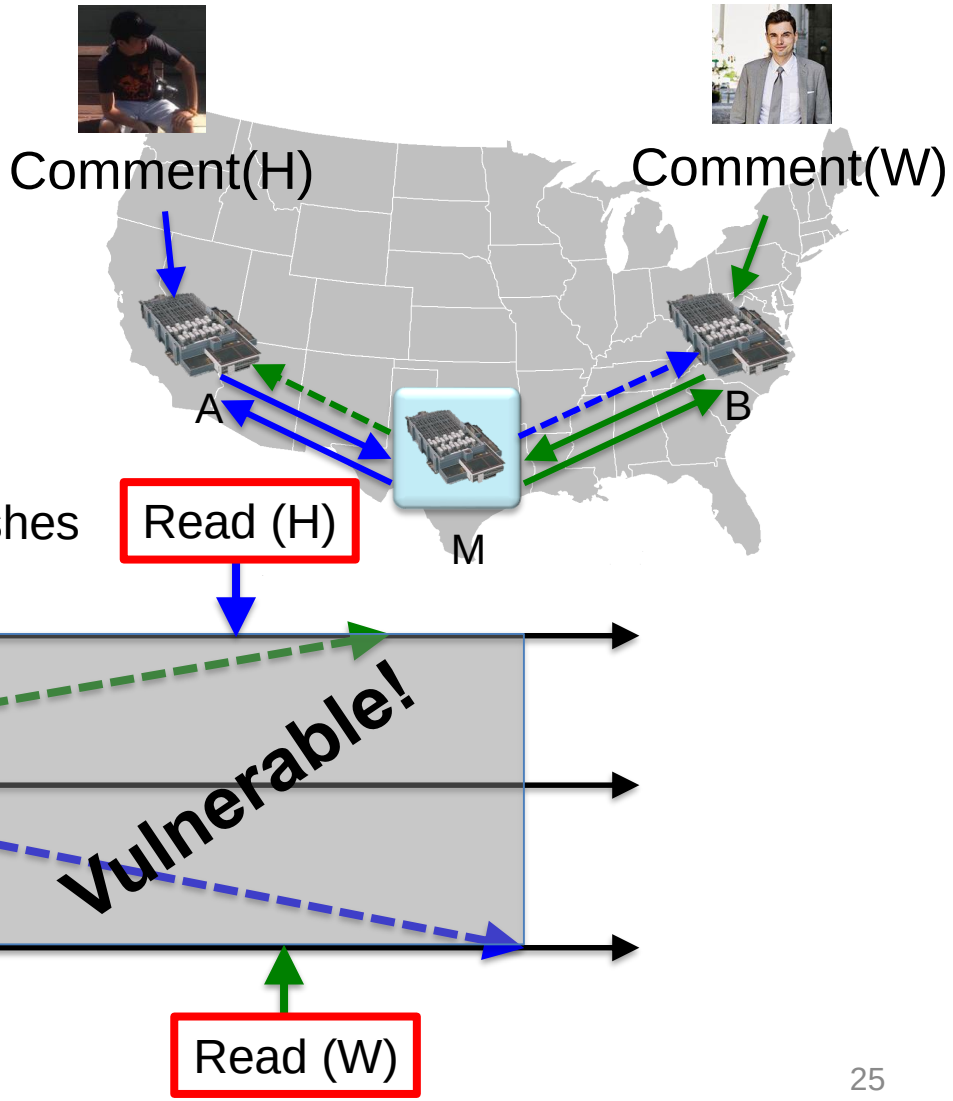




Linearizability Results

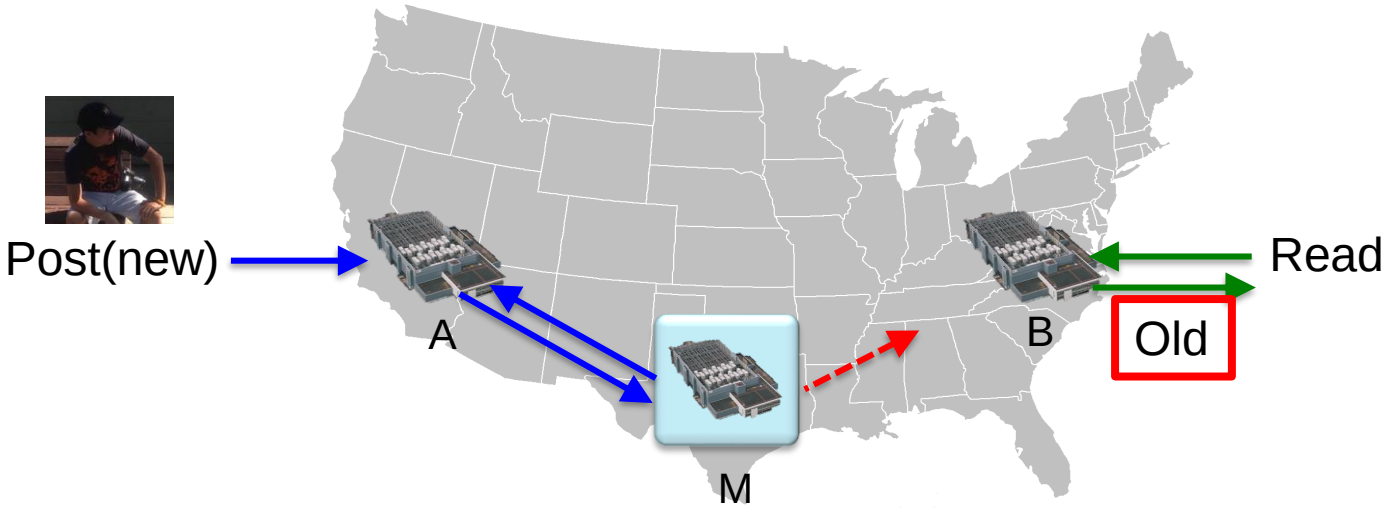
Total Order Constraint Violations

- 1 per million reads

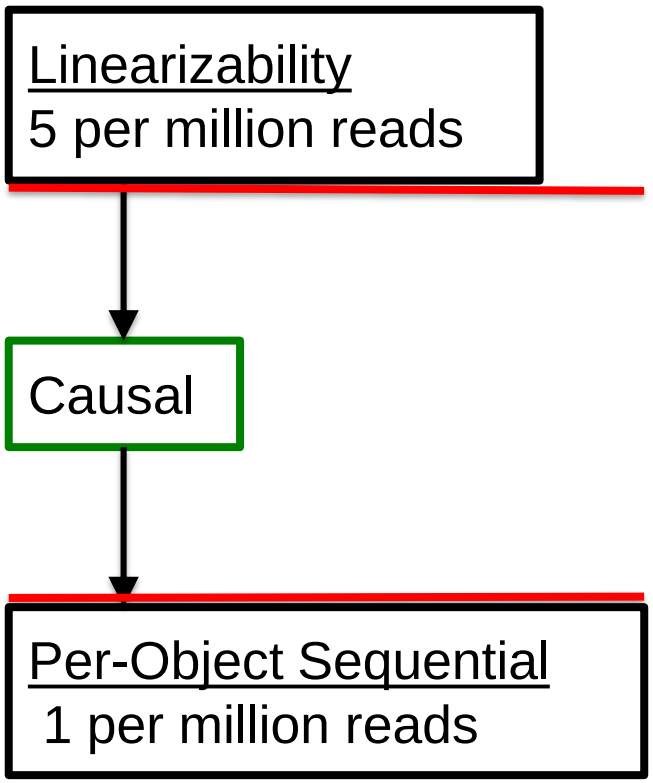


Per-Object Sequential Results

- 1 anomaly per million reads
  - Total order constraint
  - User session constraint (1 per 10 million)
    - Users should see their writes



Infer Bounds on Causal



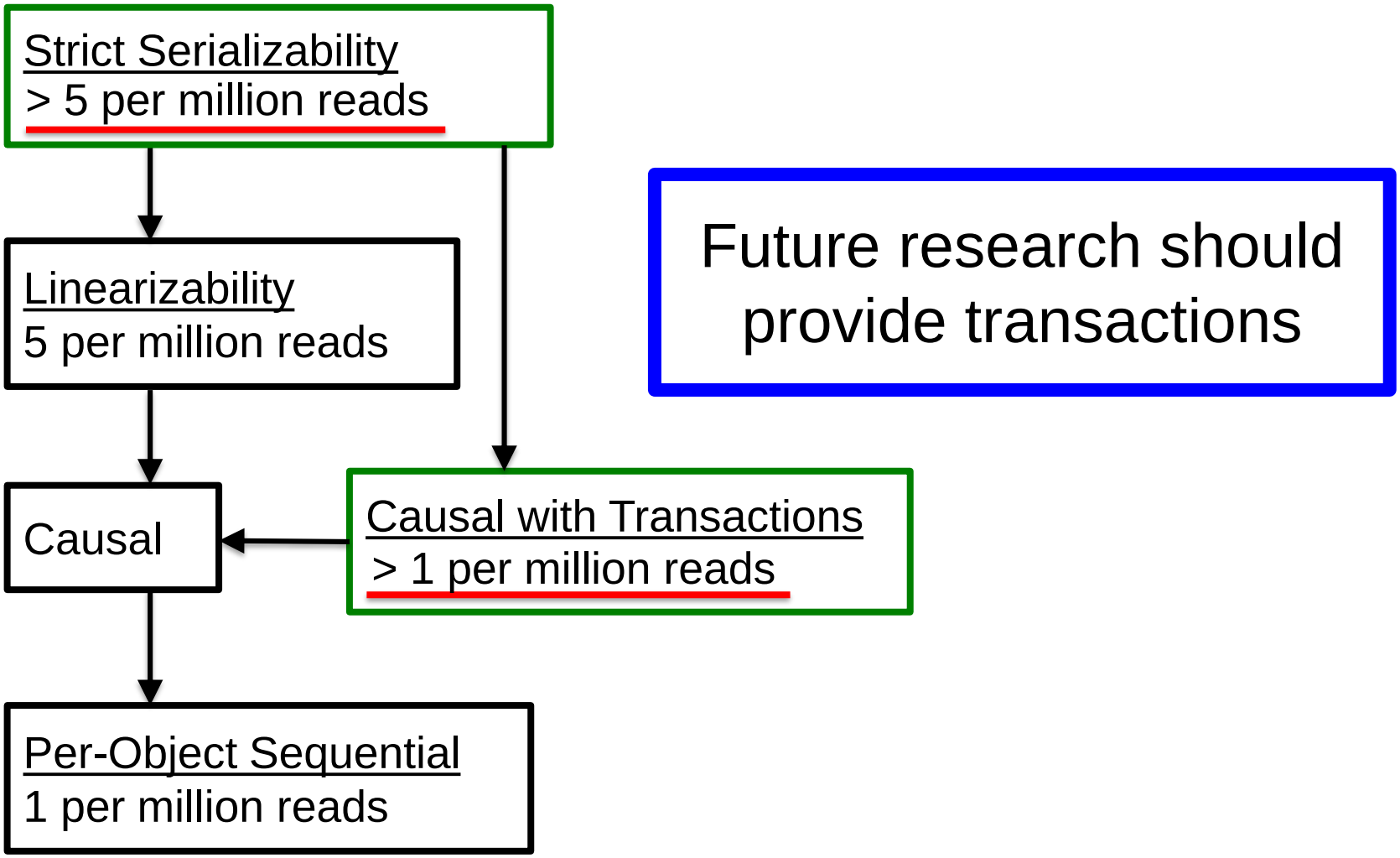
Superset of causal anomalies

≤ 5 per million reads

≥ 1 per million reads

Subset of causal anomalies

Lower Bounds on Transactions



Real-Time Consistency Monitor

- Checkers cannot run in real-time
- Φ-consistency
  - Measure convergence of replicas
- A real-time health monitor
  - Alarms when a replica falls behind

Conclusion

- Benefits of consistency are hard to quantify
  - First study of a large-scale production system
- Measure Facebook's TAO system
  - Collect trace and run anomaly checkers
  - Real-world challenges
- Results
  - TAO is highly consistent
  - Benefits of adopting stronger consistency exist
  - Research should provide transactions