UNIVERSITEIT VAN AMSTERDAM

Sundial: Fault-tolerant Clock Synchronization for Datacenters

Large Systems

Sundial

Zsombor Benedek & Diogo Marques

12 december 2024

Originally presented at OSDI 2020 by researchers from Google, Harvard, and Lilac Cloud

Clock Synchronization: The Backbone of Datacenters

- Distributed databases
- Consistent snapshots
- Network telemetry
- Congestion control
- Distributed logging

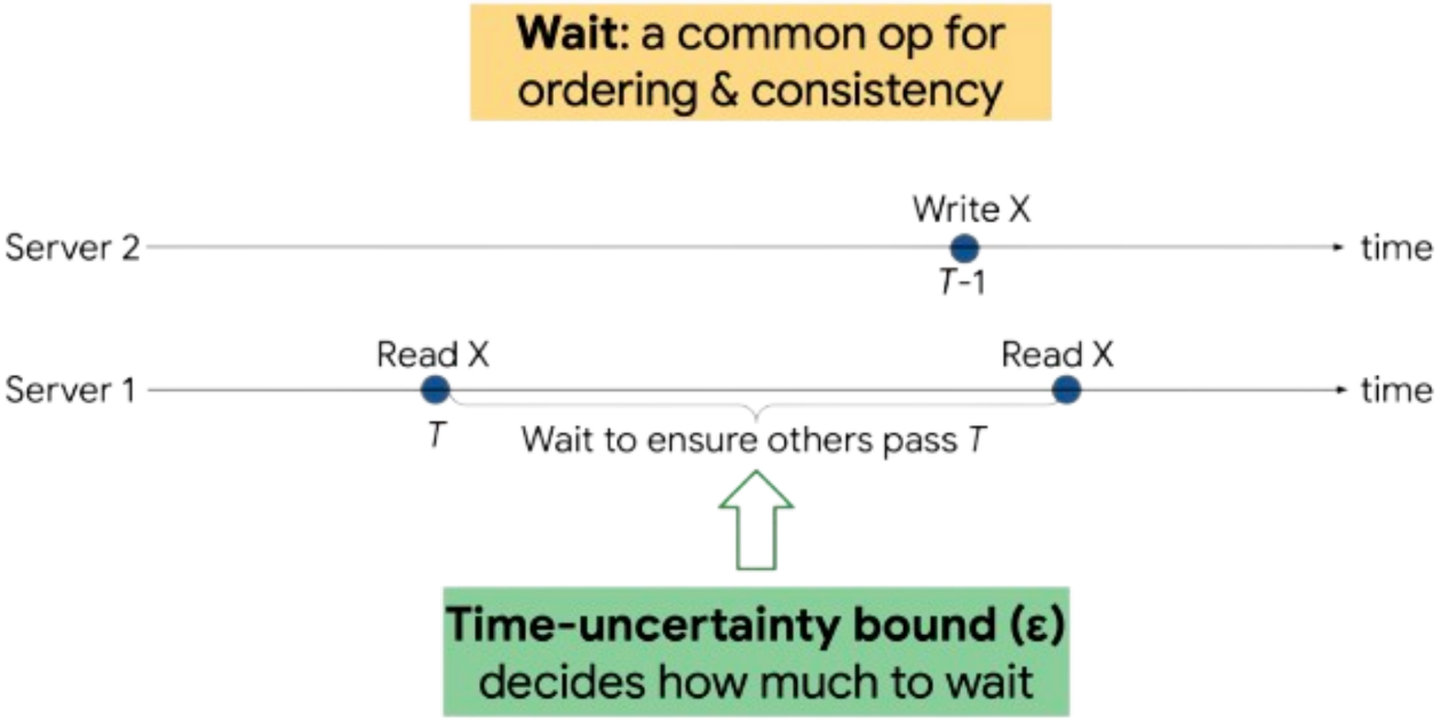
Spanner

FaRMv2

Amazon RDS

https://www.usenix.org/sites/default/files/conference/protected-files/osdi20_slides_li-yuliang.pdf

Time-uncertainty bound (ε)

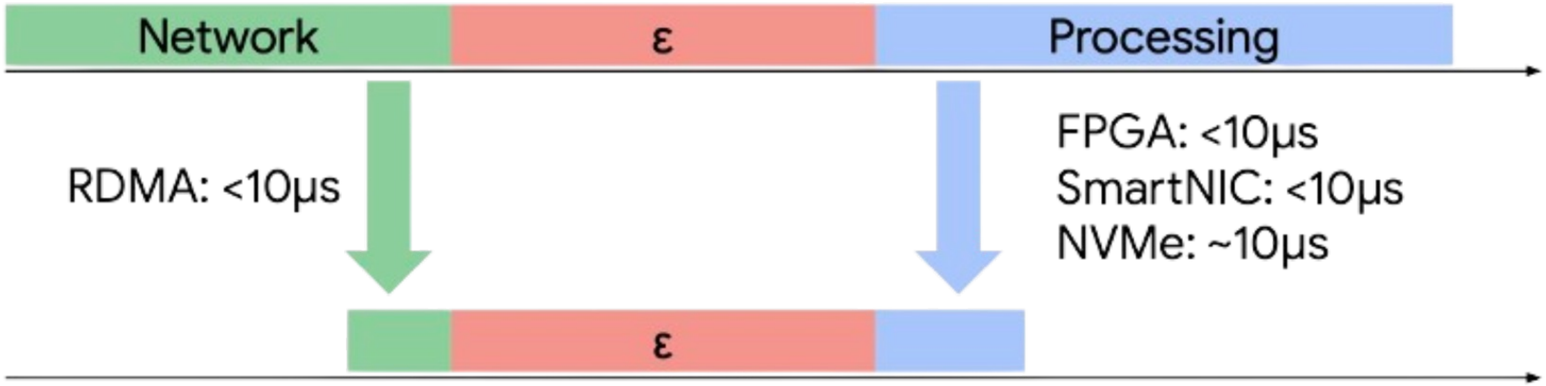


Wait: a common op for ordering & consistency

Time-uncertainty bound (ε) decides how much to wait

https://www.usenix.org/sites/default/files/conference/protected-files/osdi20_slides_li-yuliang.pdf

Time-uncertainty bound (ε)



Example: FaRMv2 latency increases 25% with 10-20 μs uncertainty.

https://www.usenix.org/system/files/osdi20-li_yuliang.pdf

The Roadblocks to Precision

- Clock drift due to environmental changes
- Failures causing connectivity disruptions
- Balancing precision and fault tolerance

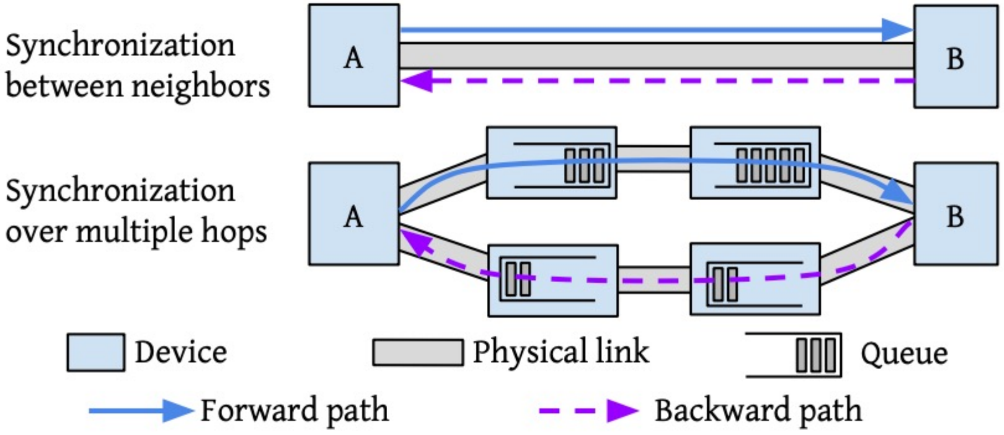

Synchronization between neighbors
A
B
Synchronization over multiple hops
A
B
Device
Physical link
Queue
Forward path
Backward path

Figure 2: Benefit of synchronization between neighbors: symmetric forward and backward paths, and no noises from queuing delay.

https://www.usenix.org/system/files/osdi20-li_yuliang.pdf

What sets Sundial apart?

- Existing systems deliver millisecond-level bounds
- Reduces uncertainty to 100 ns (2-3 orders of magnitude lower than state-of-the-art solutions)
- Enables advanced applications like high-resolution telemetry and debugging

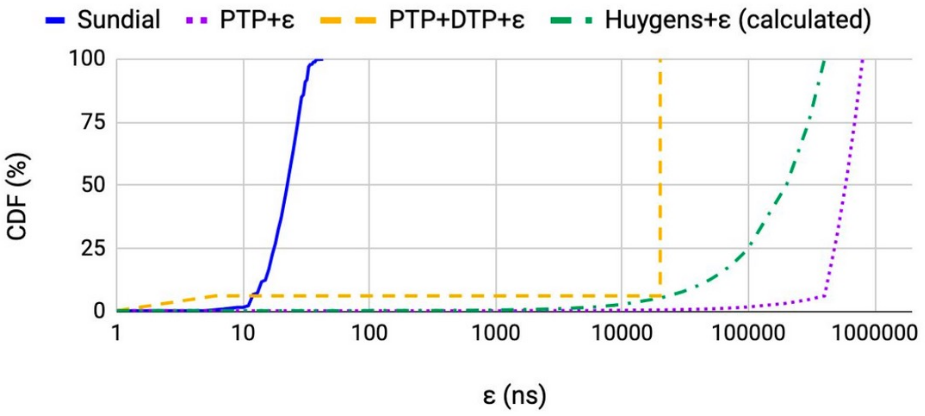
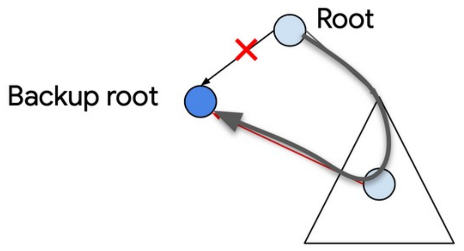


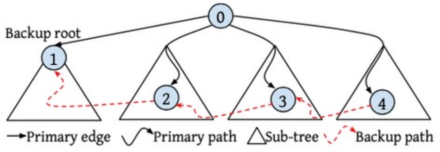
Figure 18: CDF of ε measured across devices without failures.

https://www.usenix.org/system/files/osdi20-li_yuliang.pdf

Innovative Techniques Driving Sundial

- Synchronous messaging for consistent updates.
- Precomputed generic backup plans.
- Autonomous root election during failures.





https://www.usenix.org/system/files/osdi20-li_yuliang.pdf

https://www.usenix.org/sites/default/files/conference/protected-files/osdi20_slides_li-yuliang.pdf

Performance Evaluation

- Testbed: 552 servers
- Simulation: 80,064 nodes
- Results: Sundial’s sub-100 ns uncertainty outperforms alternatives like PTP and Huygens.

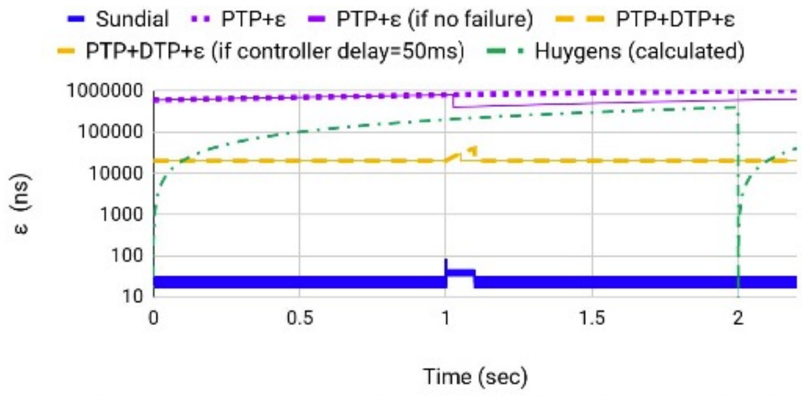
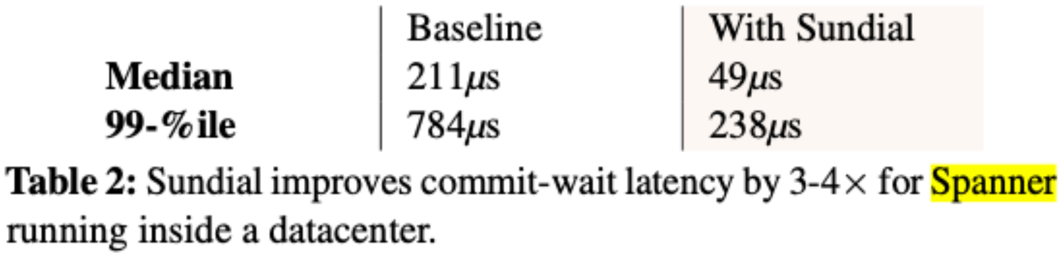


Figure 19: Time series of ε of a device affected by a link failure. The failure happens at 1s and the controller reacts to it near 1.1s.

https://www.usenix.org/system/files/osdi20-li_yuliang.pdf

Transformative Results

- Spanner: 3-4x reduction in commit-wait latency.
- Swift: 60% throughput improvement during congestion.



	Baseline	With Sundial
Median	211μs	49μs
99-%ile	784μs	238μs

Table 2: Sundial improves commit-wait latency by 3-4× for Spanner running inside a datacenter.

https://www.usenix.org/system/files/osdi20-li_yuliang.pdf

Sundial: A New Benchmark

- Precision and resilience redefine datacenter synchronization.
- Potential for ultra-low-latency innovations.
- Drawback: requires additional hardware.

"Sundial is the first submicrosecond-level clock synchronization system that is resilient to failures."

https://www.google.com/about/datacenters/data-security/

Question time

https://www.usenix.org/conference/osdi20/presentation/li-yuliang

Thank you!