Neural Emulation: Twinning a 5G Mobile Scheduler via High Resolution Telemetry

Kyle Jamieson with Haoran Wan (Princeton) and
Yaxiong Xie (University at Buffalo)

Princeton Advanced Wireless Systems Lab
paws.princeton.edu

This Talk

Problem

- **End-to-end performance** hinges on **mobile**
- Current network emulation **neglects causality**

Solution

- **Make digital twins** of black-box 5G schedulers
  - Enabled by **high-resolution telemetry** tools

Impact

- **Scaled and broadened** mobile network emulation for **reproducible and generalizable** experiments

Context: 5G is Crucial to Real-Time Application Performance

Rise in real-time applications:


Drones


Teleconferencing

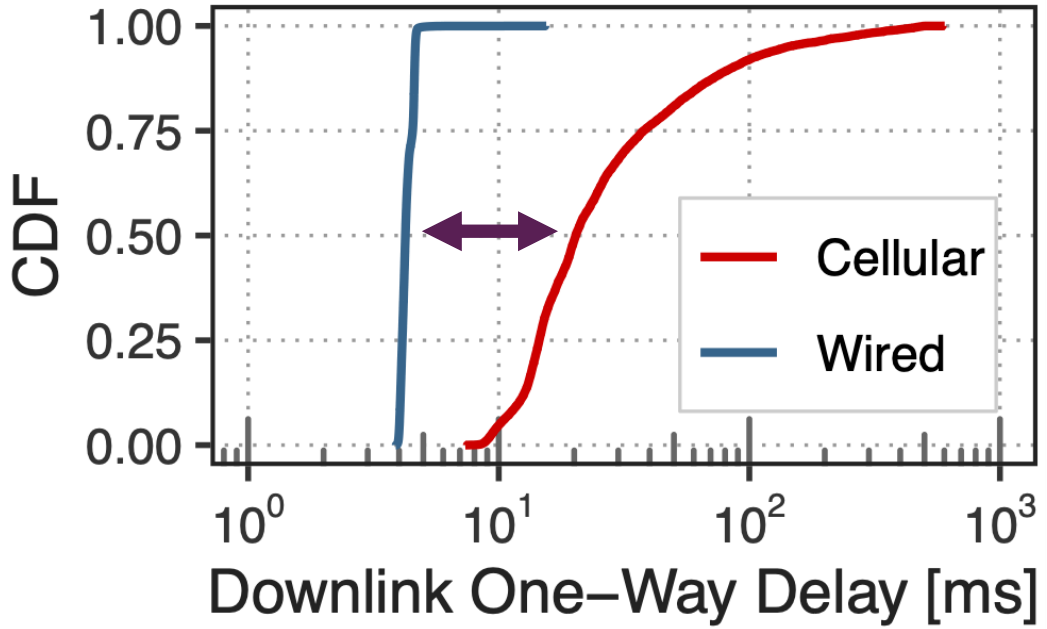

Cloud Gaming


AR/VR

Require ↗ Quality, ↗ Interactivity → ↗ High bandwidth, ↘ Low latency →

Requirements can outpace 5G network capabilities!

Commercial 5G measurements:
T-Mobile 5G New Radio, 15 MHz bandwidth, frequency-duplexing division, Band 71 (622.85 MHz)



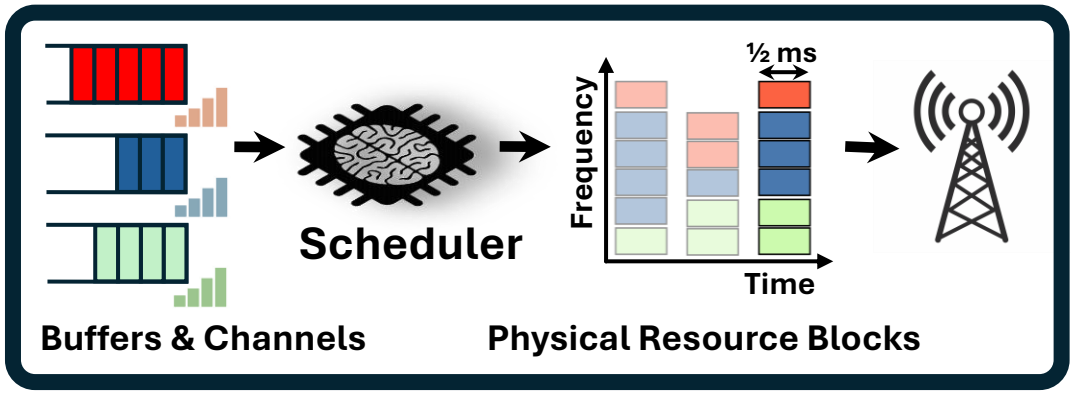
Gold Standard: Test in a Live Mobile Network

Mobile network behavior that application developers need to capture:

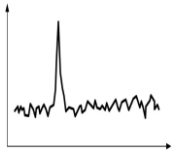
Inputs:

Highly dynamic user contention

Unpredictable, environment-dependent channel conditions



Output:



Significant jitter due to complex uplink and downlink scheduler patterns

- The scheduler is a “black box:” not standardized, **mobile vendor’s** intellectual property
- **Mobile operator** configures scheduler, possibly resulting in intra-vendor diversity

Live Mobile Network Testing: Unmet Needs

Scaling & Broadening Evaluation ("Reproducibility")

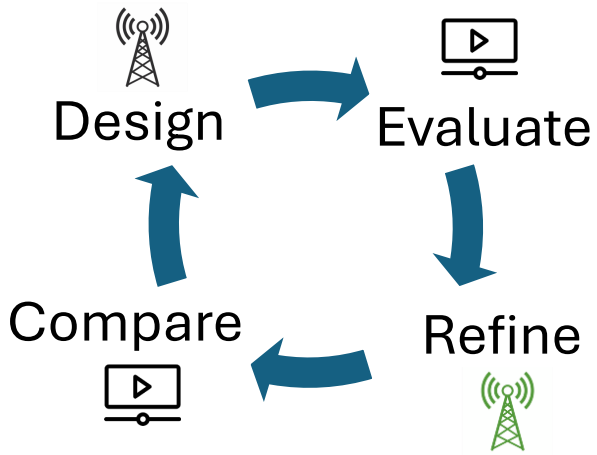
Design

✅ Evaluation
{Urban A, T-Mobile, 20 MHz, x 5}

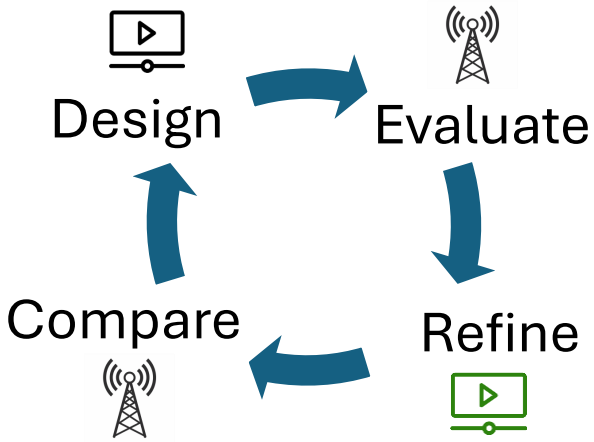
? Evaluate
{Rural, AT&T, 40 MHz, x 50}

? Evaluate
{Urban B, O2, 100 MHz, x 20}

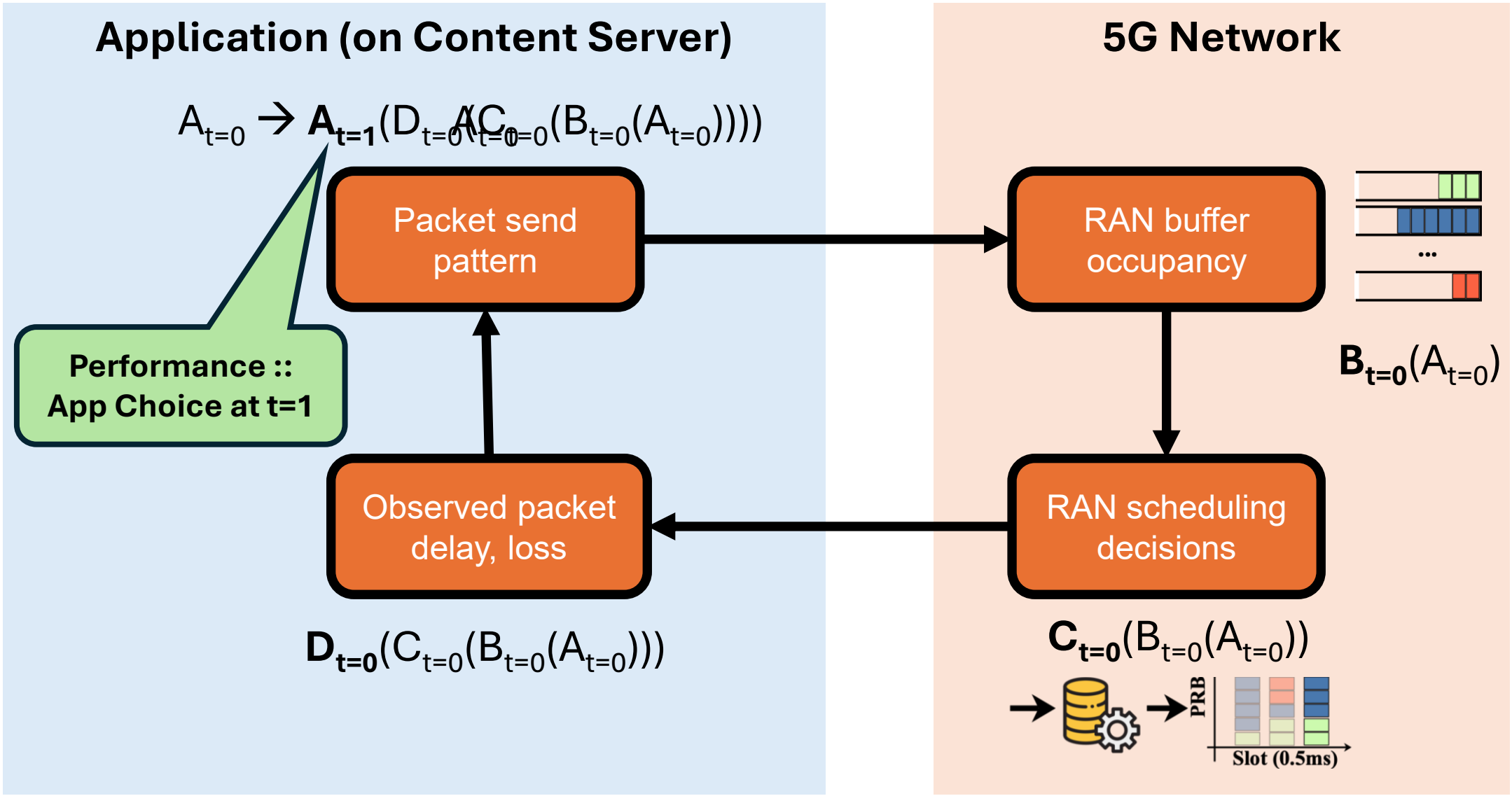
Cell Scheduler Refinement



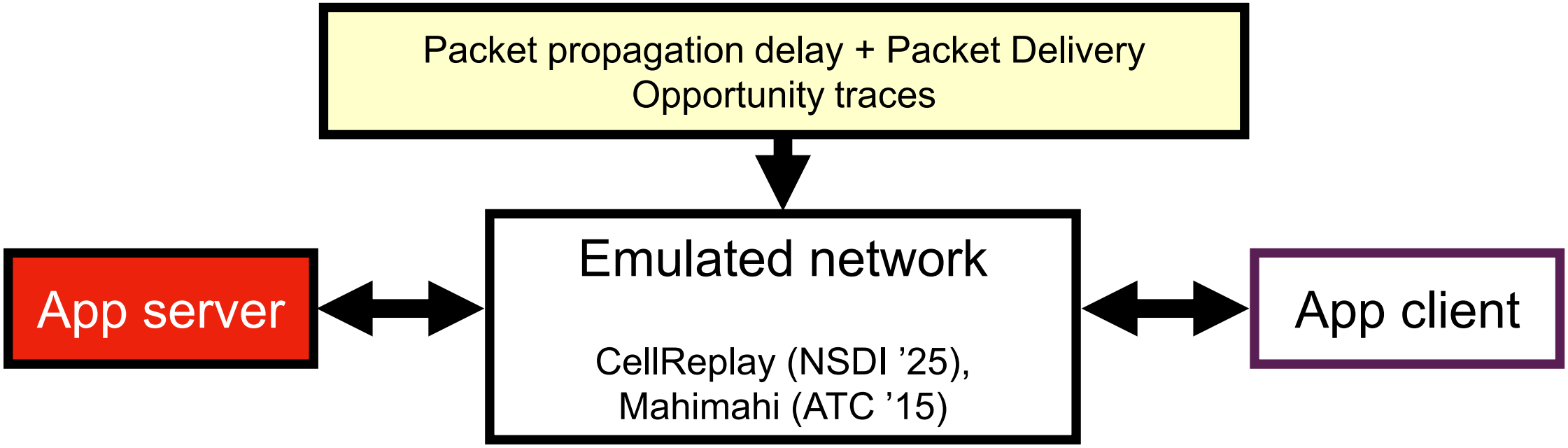
Application Refinement



Challenge: Feedback Loop between Applications, Network



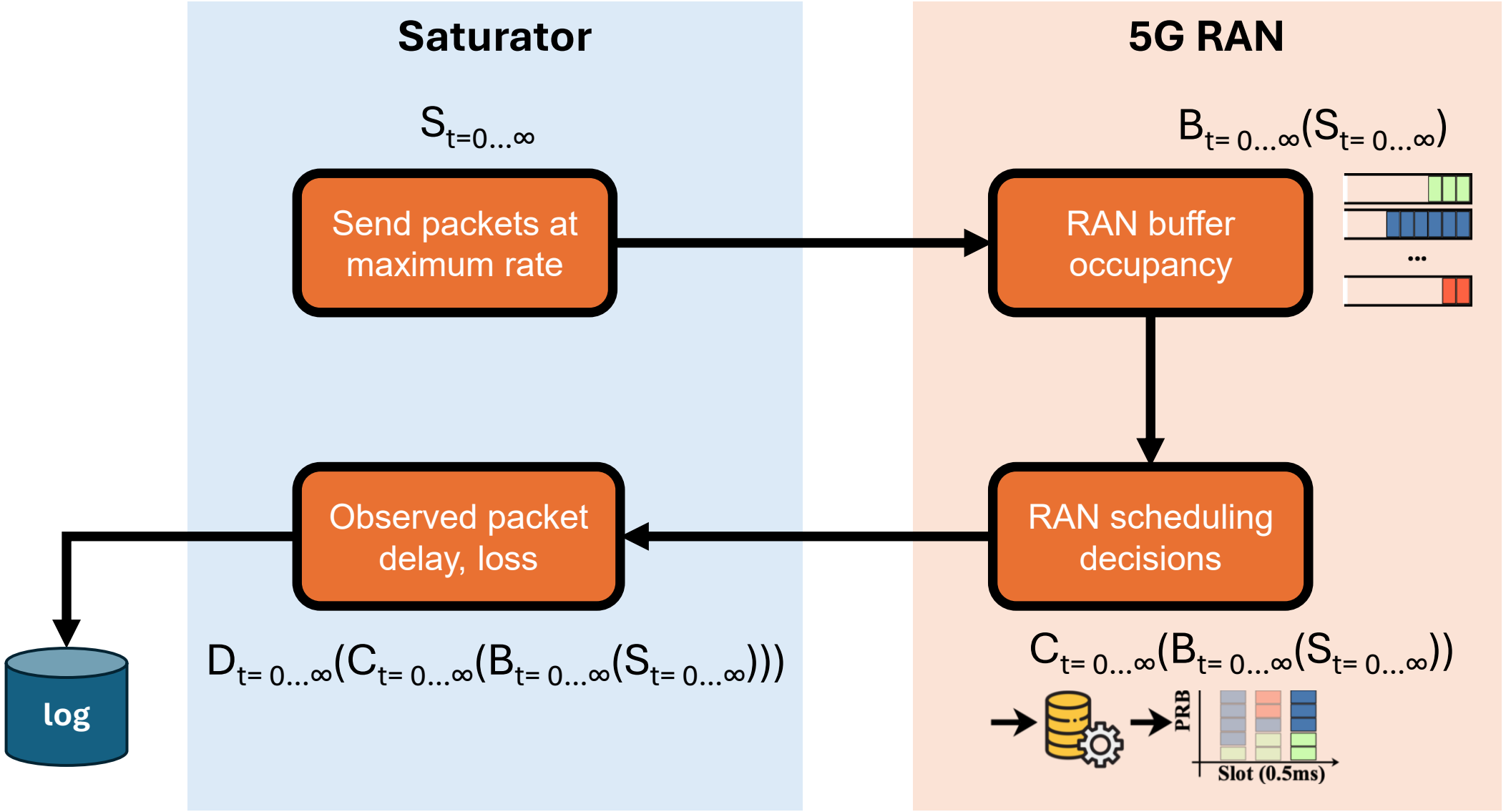
Evaluating Performance with “Record-and-Replay” Emulators



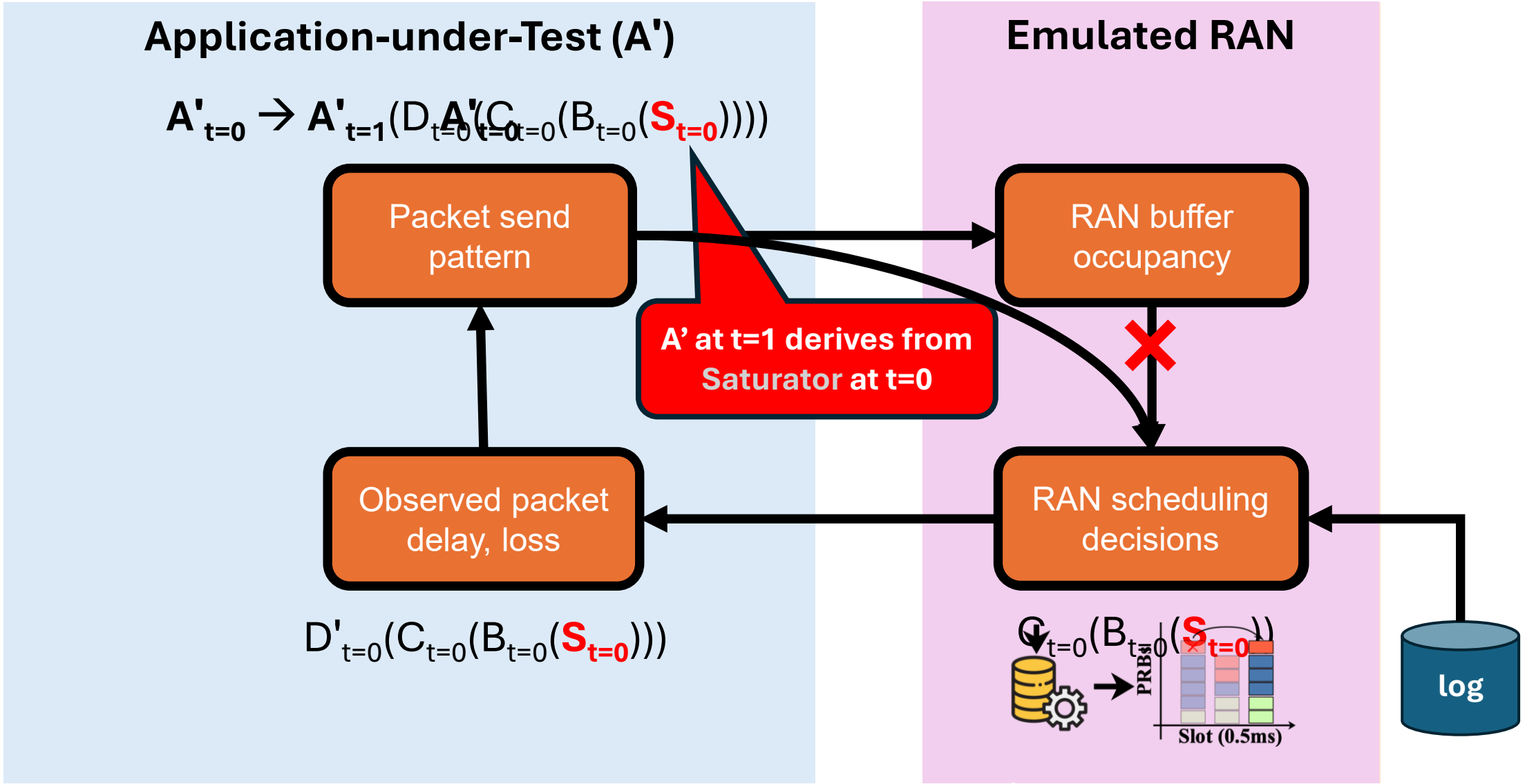
- **Record** fixed network traces under **conditions of interest** (e.g., contention, mobility) with a **predefined workload** (e.g., iperf3, packet trains).

- **Replay** the recorded trace in the **emulated network** for an **Application-under-Test**

Record and Replay: Saturate and Log

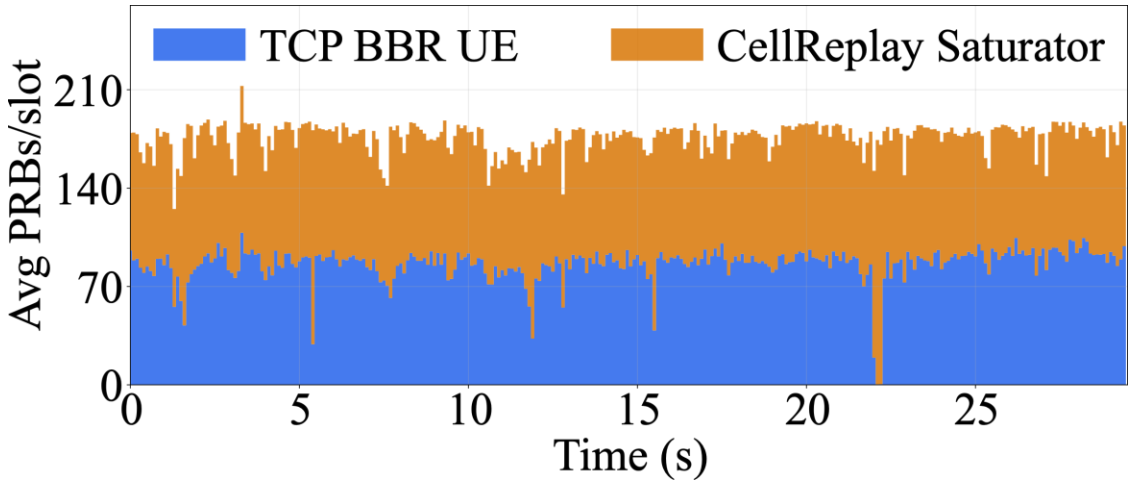


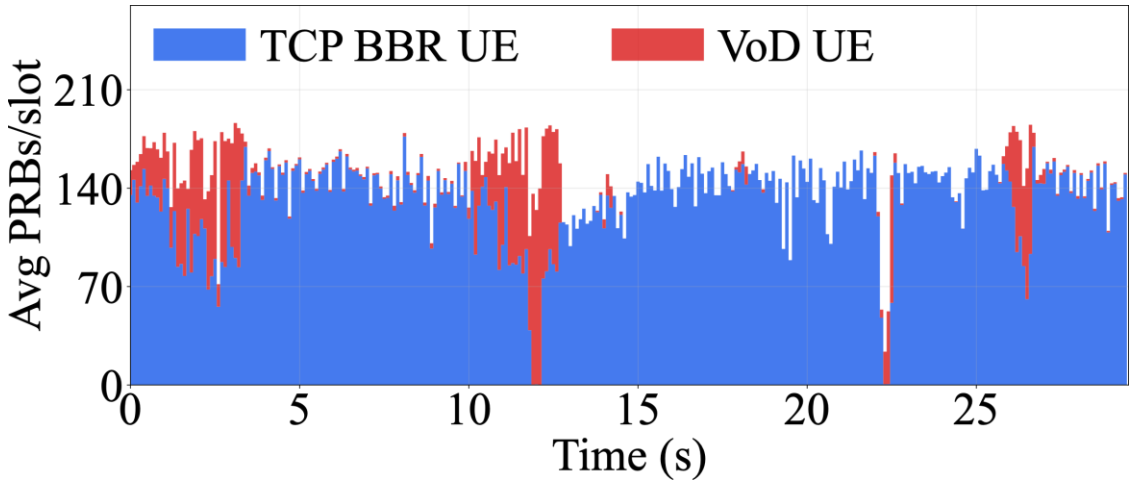
Record and Replay: Breaks the Causality Cycle



Record-and-Replay Distorts Contention

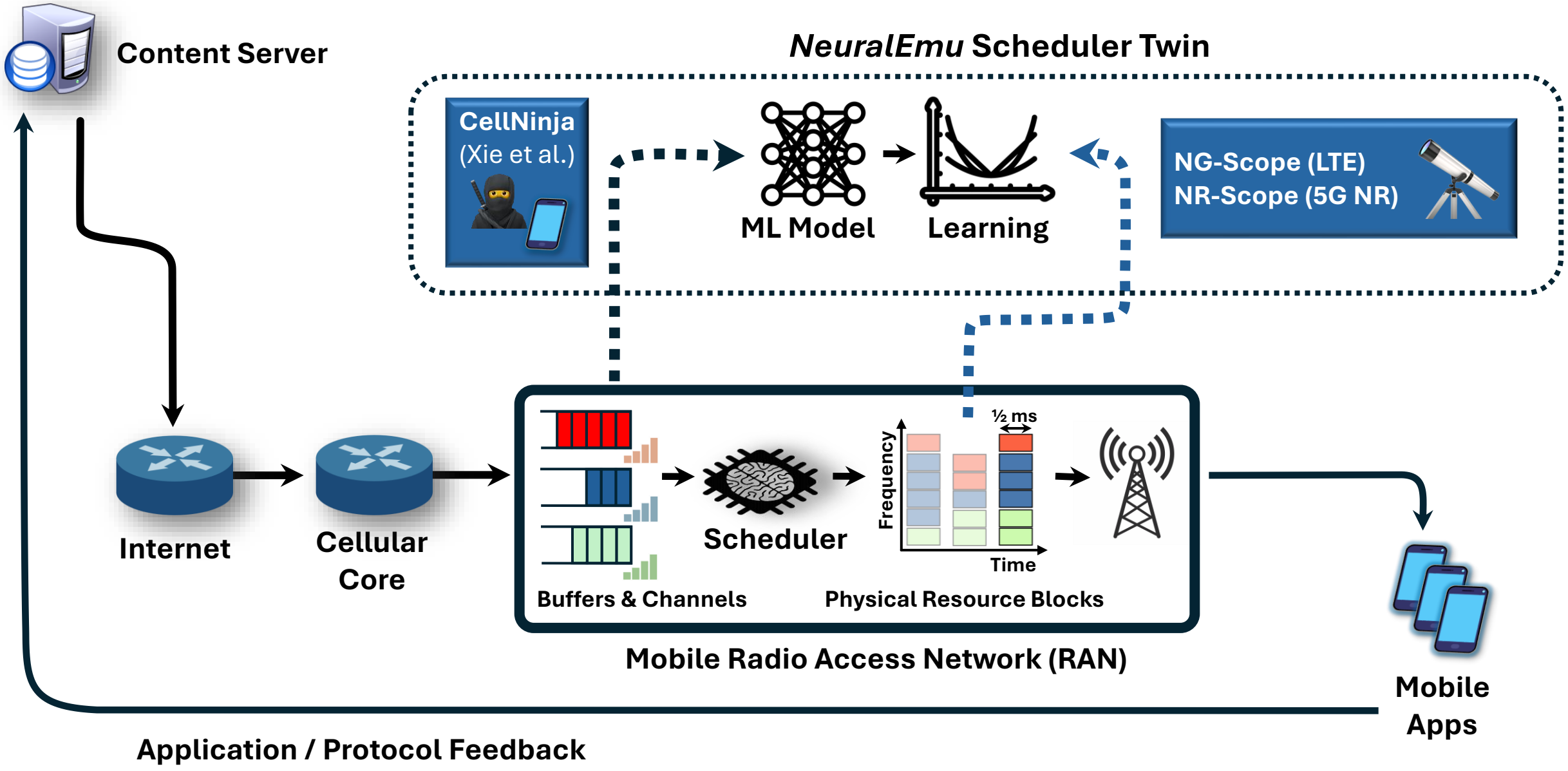
Example: emulate performance of **Video on Demand (VoD)** in the presence of **BBR background traffic**



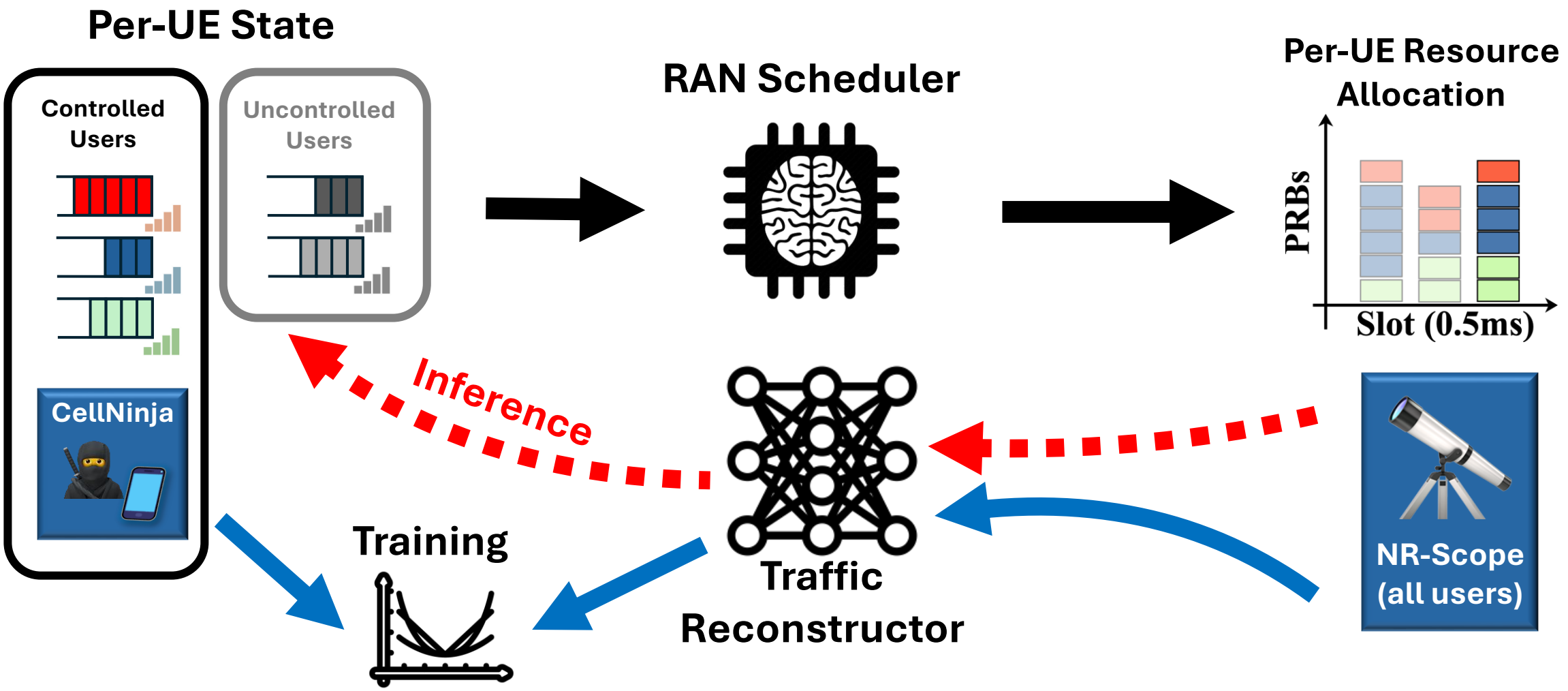


The **Saturator competes** with background traffic **differently** than Video on Demand (application under test)

NeuralEmu "Twins" any RAN Scheduler



NeuralEmu: Uncontrolled Traffic Reconstructor

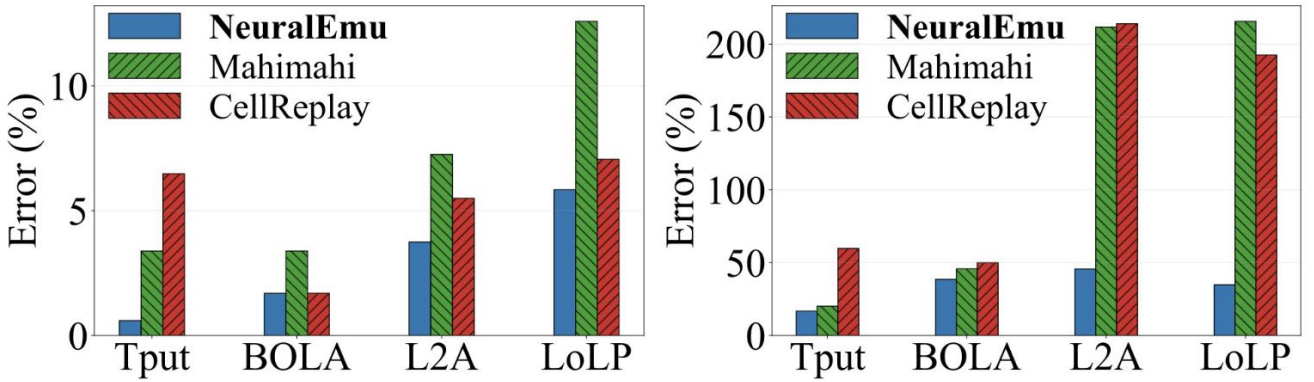


Now we have traffic data for every user!

NeuralEmu Reduces Emulation Error up to 5x

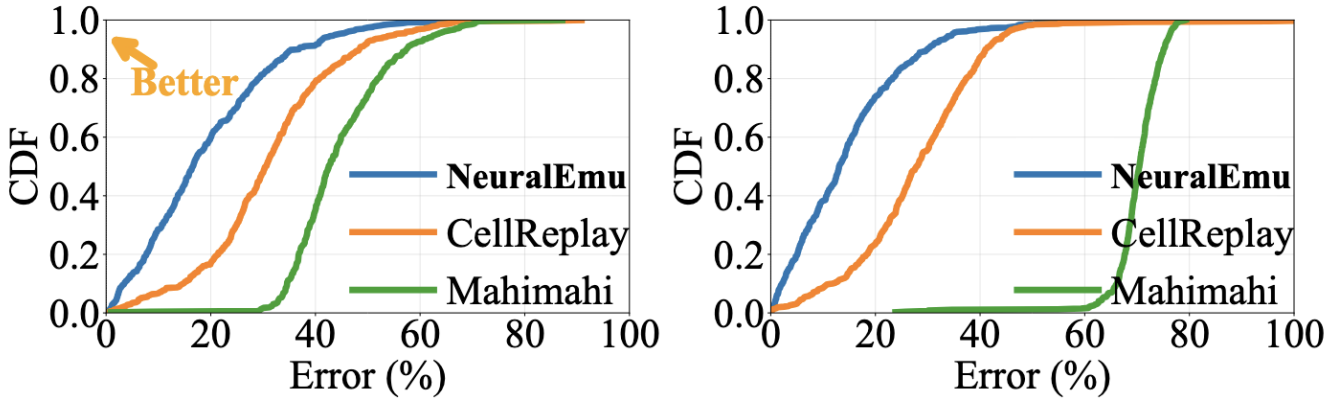
- Compare application performance distributions
  - Emulation v. live T-Mobile 5G New Radio network (100 MHz bandwidth, band 41)
- ***Emulation error:*** average distance between distributions
  - Follows CellReplay methodology

Video-on-Demand



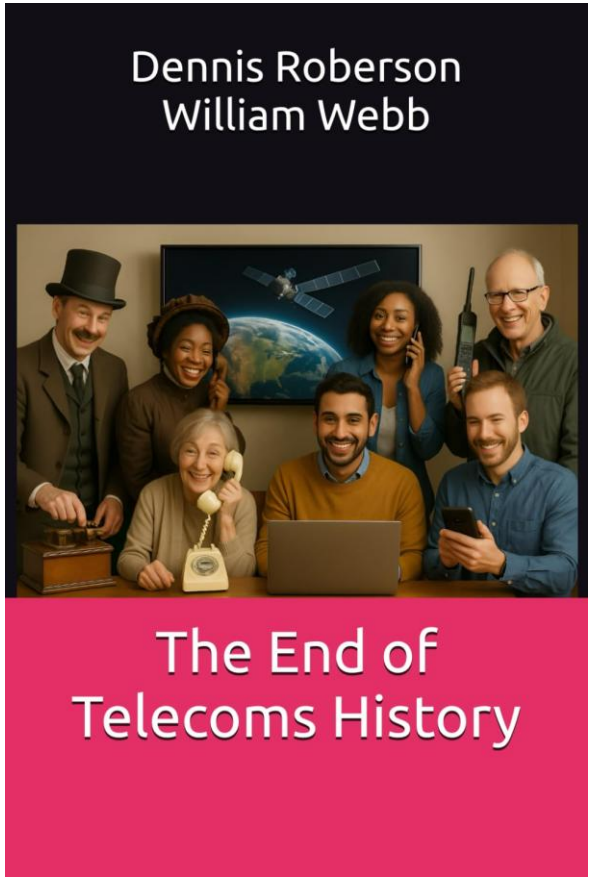
(a) VoD bit rate selection (kbps). **(b)** VoD buffer level (seconds).

Webpage Load



(a) HTTP/1.1 **(b)** HTTP/2

Over a Pint: Roberson & Webb's Position on Mobile Networks



- Exponential growth in **mobile data usage** **will not continue** from now to [2035-2045]
- You don't need more than ~20-30 Mbit/sec
  - can't watch more video on demand!
- **They also rebut:** AR/VR/IoT, the rise of AI, Fixed Wireless Access, developing world economy, etc.

What's next: Resilience & Reliability

Conclusion

- Bolun Zhang will talk about **causality in RAN configurations** in 30 minutes!

- *CausalSim* [NSDI '23] makes coupling observation but **in context of video streaming bit rate adaptation** (coarse, four-second chunk time-scales)
  - Millisecond-resolution telemetry brings concept to packet-level timescales

- Our Next Steps: MILCOM demo, scale to multiple RAN vendors & compare

CellNinja [GitHub]

NG-Scope [GitHub]

NR-Scope [GitHub]

NeuralEmu (this work) [arxiv]

This material is based upon work supported by the National Science Foundation under grants CNS-2223556 and OAC-2429485. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.