POLYCORN: Data-driven Cross-layer Multipath Networking for High-speed Railway through Composable Schedulerlets

Yunzhe Ni, Feng Qian, Taide Liu, Yihua Cheng, Zhiyao Ma, Jing Wang,
Zhongfeng Wang, Gang Huang, Xuanzhe Liu, Chenren Xu

POLYCORN: Data-driven Cross-layer Multipath Networking for High-speed Railway through Composable Schedulerlets

- Extra Networking Challenges From Extreme Mobility
  - Fluctuating, Unpredictable and Heterogeneous
  - Inaccurate Measurements Hurt Performance
- System Design
  - Event-triggered Schedulerlets
  - Composable Scheduling Framework
- Evaluation
  - Deployment on HSR LTE Gateway
  - 3 weeks/52720 km Evaluation on Beijing-Shanghai Route

Fluctuating, Unpredictable and Heterogeneous

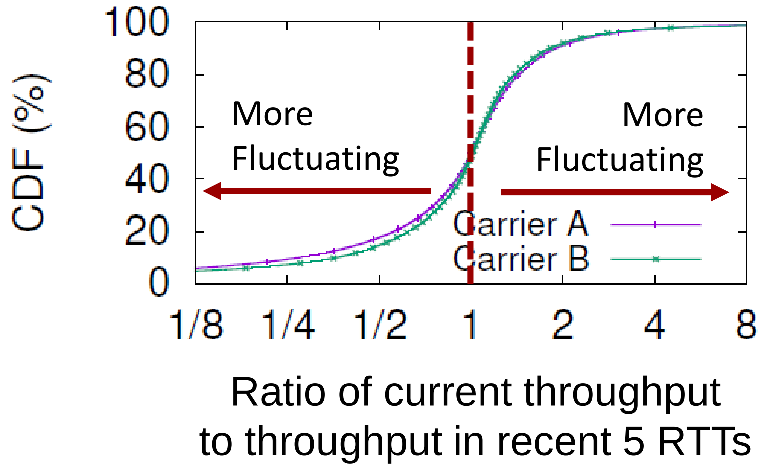
- High-speed railway (HSR) in China
  - Travels at 300-350 km/h
  - 155000 km (until 2022), available in 29 major cities
  - 1.6 billion trips in 2022 and 10+ billion so far
- Internet access on HSR
  - Cellular network
    - Mostly LTE
  - HSR public Wi-Fi
    - Offered by the “Fuxing” HSR train
    - Based on LTE

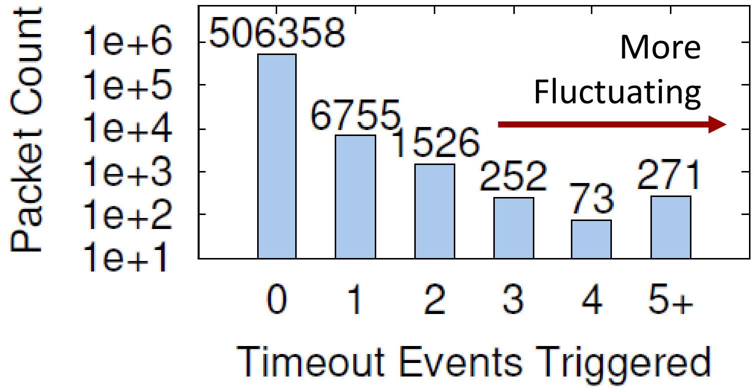
Next: Measurement data (long-lived TCP flows) on HSR LTE network, 2 carriers

Carrier A = China Mobile; Carrier B = China Unicom

Fluctuating, Unpredictable and Heterogeneous

- Single carrier HSR LTE network: Fluctuating





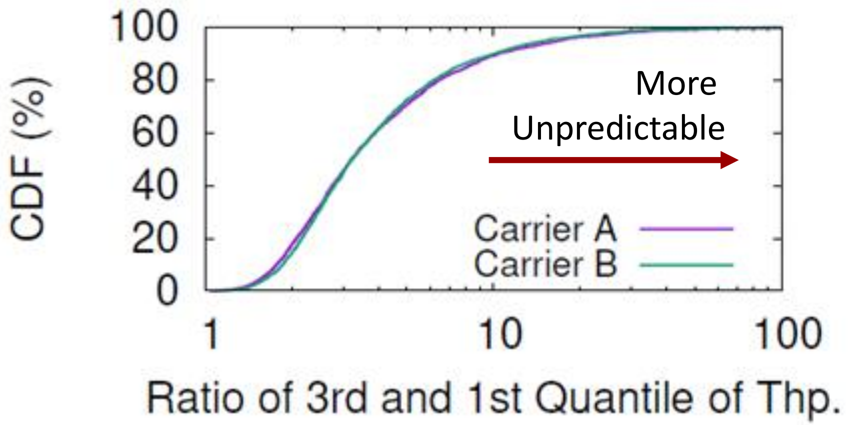
~25% of the cases:
Ratio is lower than 0.5 or higher than 2

~1.8% packets experienced timeout
~24% of them experienced multiple timeouts

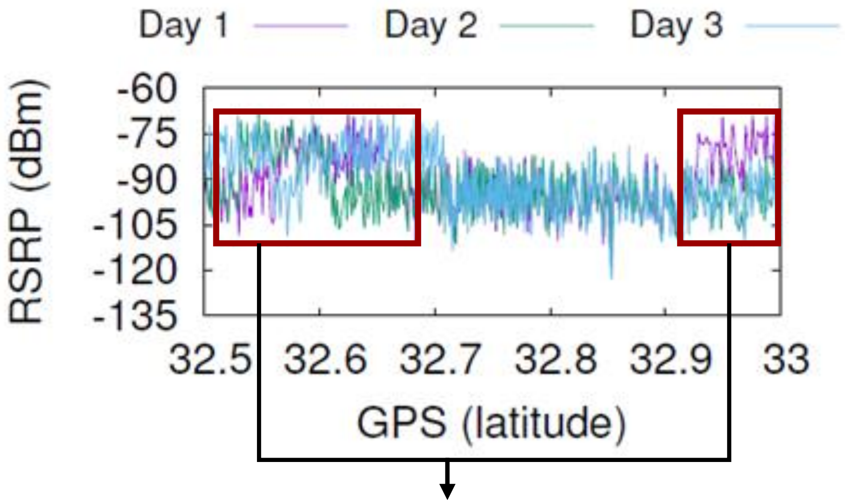
Takeaway: Key performance metrics, such as throughput and RTT, could change significantly In several RTTs

Fluctuating, Unpredictable and Heterogeneous

- Single carrier HSR LTE network: Unpredictable
  - Would fixed rail tracks lead to predictable network performance?



~3.2 median ratio, up to 100

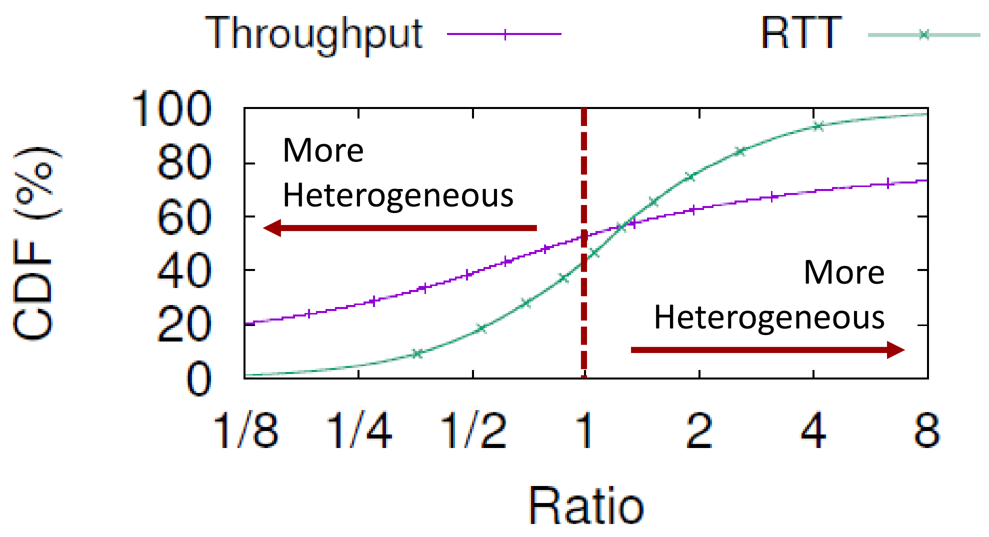


Signal strength varies over days

Takeaway: Fixed rail tracks would not make TCP performance or signal strength predictable

Fluctuating, Unpredictable and Heterogeneous

- Multi-carrier HSR LTE network: Heterogeneous
  - Would different infrastructure deployment lead to different network performance?



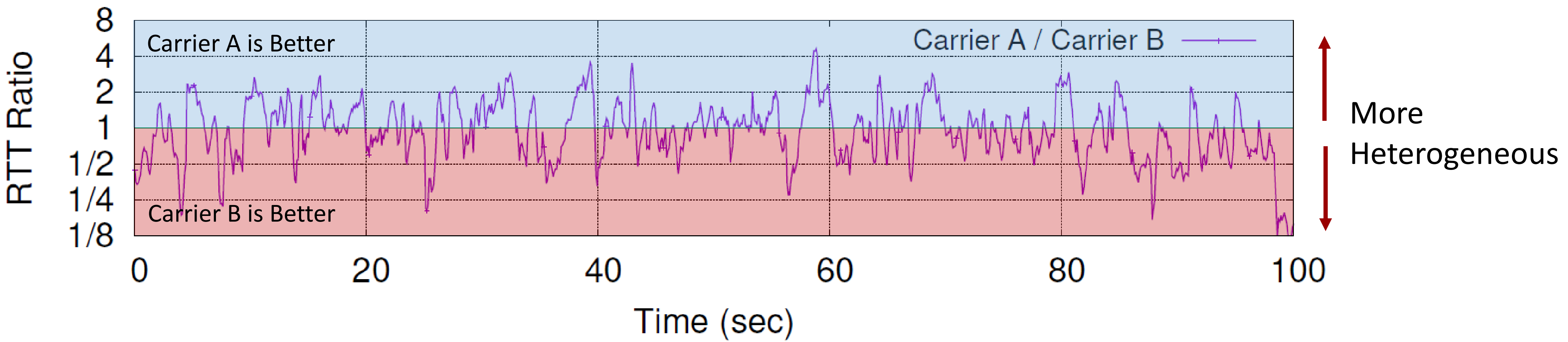
~70% of the cases: Throughput ratio <0.5 or >2

~45% of the cases: RTT ratio <0.5 or >2

Takeaway: When one path performs badly, others may be much better
(Multipath transport that uses multiple carriers at the same time
is a promising approach for optimizing HSR LTE network)

Fluctuating, Unpredictable and Heterogeneous

- Multi-carrier HSR LTE network: Heterogeneous
  - Highly-dynamic interleaved RTT introduces great challenge in multipath scheduling



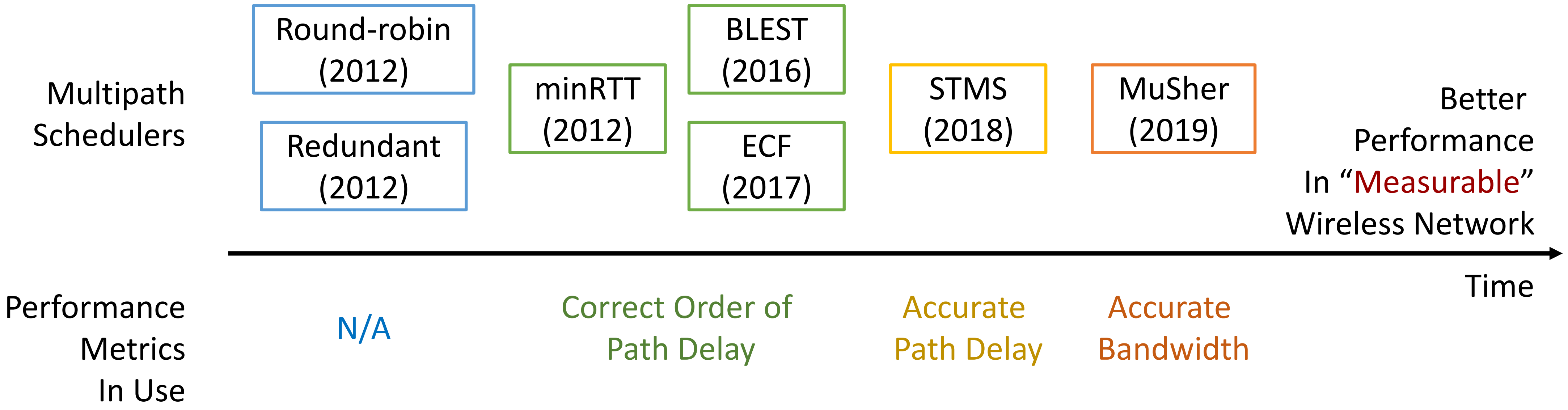
~26% of the cases: ratio <0.5 or >2

“Best” path changes every 2-4 RTTs

Takeaway: Permanently best path is not available;
Choosing a better path is critical because of the disparate performance

Inaccurate Measurements Hurt Performance

- Inaccurately measured path performance challenges multipath schedulers



Inaccurate Measurements Hurt Performance

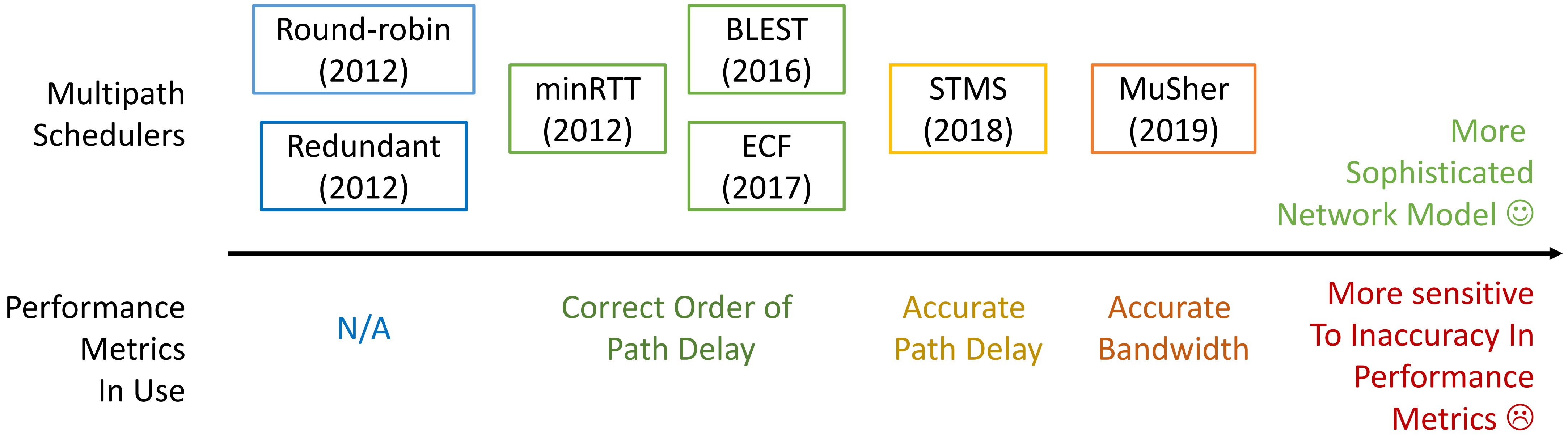
- Inaccurately measured path performance challenges multipath schedulers

Multipath Schedulers	Round-robin (2012) Redundant (2012)	minRTT (2012)	BLEST (2016) ECF (2017)	STMS (2018)	MuSher (2019)	Better Performance In “Measurable” Wireless Network
Performance Metrics In Use	N/A	Correct Order of Path Delay		Accurate Path Delay	Accurate Bandwidth	Time →

Common design principle:
More sophisticated network model = Better performance

Inaccurate Measurements Hurt Performance

- Inaccurately measured path performance challenges multipath schedulers



The real case:

More sophisticated network model + accurate measurement = Better performance

(not true on HSR)

Inaccurate Measurements Hurt Performance

- Important fact
  - ACK-based feedback provides recent network performance
  - Scheduling requires current network performance!

- The common practice:
  - Use recent performance as an approximation of current performance, ignore the difference

- **Major challenge**: On HSR, recent performance ≠ current performance
  - Causes measurements to be inaccurate
    - Leads to erroneous scheduling decisions that hurts performance

POLYCORN: Data-driven Cross-layer Multipath Networking for High-speed Railway through Composable Schedulerlets

- Extra Networking Challenges From Extreme Mobility
  - Fluctuating, Unpredictable and Heterogeneous
  - Inaccurate Measurements Hurt Performance
- System Design
  - Event-triggered Schedulerlets
  - Composable Scheduling Framework
- Evaluation
  - Deployment on HSR LTE Gateway
  - 3 weeks/52720 km Evaluation on Beijing-Shanghai Route

Event-triggered Schedulerlets

- Reconsider the multipath scheduler design principle for high mobility

Sophisticated network model
+ accurate measurement
= Better performance

⇨

Robust network model
+ accurate performance indicator
= Better performance

Start from a robust "base" scheduler, shape its behavior when specified events are detected

↓ Handles most cases that cannot be correctly understood by the transport protocol in a robust manner

↓ Once happen, very possibly that specified action (schedulerlet) should be taken (activated)

Event-triggered Schedulerlets

- Reconsider the multipath scheduler design principle for high mobility

Sophisticated network model
+ accurate measurement
= Better performance

⇒

Robust network model
+ accurate performance indicator
= Better performance

Start from a robust "base" scheduler, shape its behavior when specified events are detected

↓ Handles most cases that cannot be correctly understood by the transport protocol in a robust manner

↓ Once happen, very possibly that specified schedulerlet should be activated

↳ Shapes the behavior of multipath scheduler by manipulating its input and output

Event-triggered Schedulerlets

- Reconsider the multipath scheduler design principle for high mobility

Sophisticated network model
+ accurate measurement
= Better performance

⇨

Robust network model
+ accurate performance indicator
= Better performance

Start from a robust "base" scheduler, shape its behavior when specified events are detected

↓ Handles most cases that cannot be correctly understood by the transport protocol in a robust manner

↓ Once happen, very possibly that specified action (schedulerlet) should be taken (activated)

"Try to understand the network only when it is understandable."

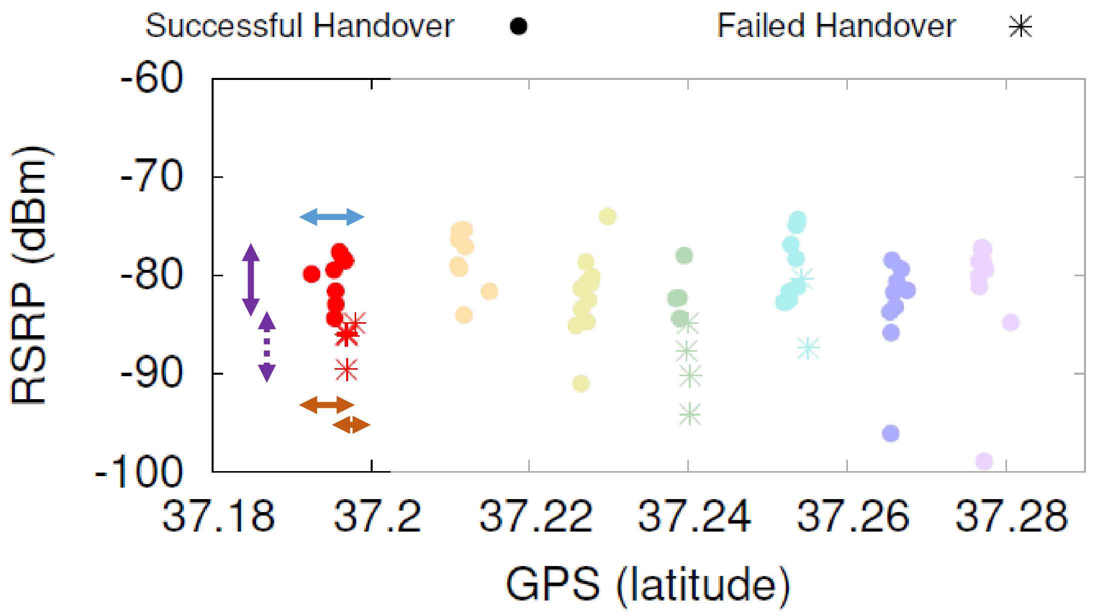
Event-triggered Schedulerlets

- **Event #1**: Handover failure
  - Successful/Failed handovers could be classified with a simple SVM with signal strength and location as feature
  - For each cell, predict a location, and the time $\widehat{t_{HOF}}$ when the train would pass the location. If handover did not happen before $\widehat{t_{HOF}}$, predict the handover to be fail.
  - Refer to the paper for details!

Handovers typically happen at similar location

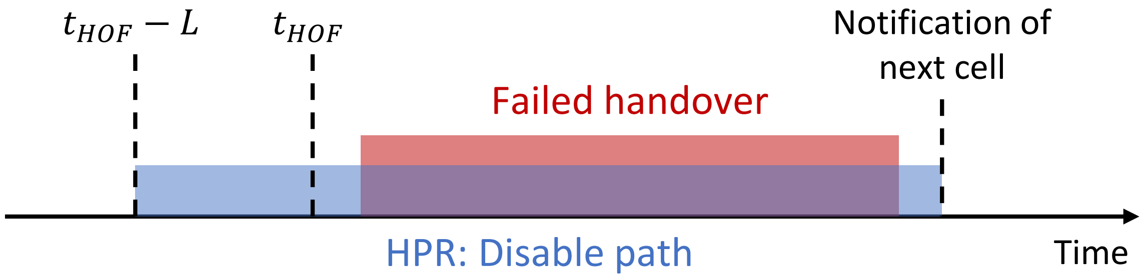
Lower signal strength, lower handover success rate

Later handover, lower handover success rate



Event-triggered Schedulerlets

- **Schedulerlet #1**: Handover-failure-aware Path Rejection (HPR)
  - Triggered when the train is approaching $\widehat{t_{HOF}}$.
  - Disable path to avoid packet losses
  - Drain the queue before handover to avoid spurious losses caused by loss of ACKs
  - Re-enable the path when the train enters next cell
  - Expected impact
    - > Faster delivery during link disconnection
    - > TCP timeout (and slow start) avoidance



- Before t_{HOF}: Drain queue, avoid spurious RTO
- After t_{HOF}: Avoid sending on disconnected path

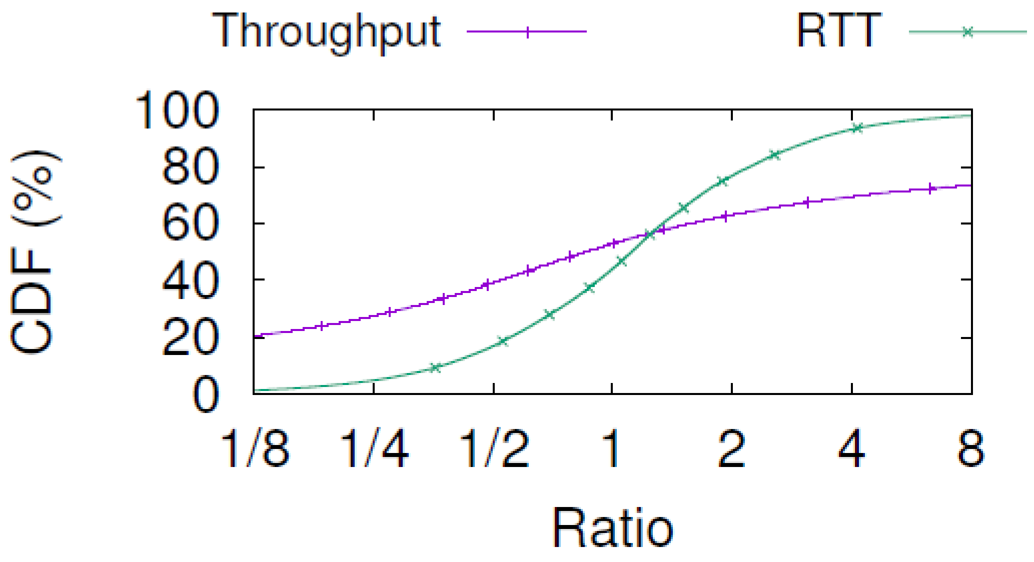
$$\mathrm{L} \coloneqq SRTT + \frac{E_{GPS}}{V_{HSR}}$$

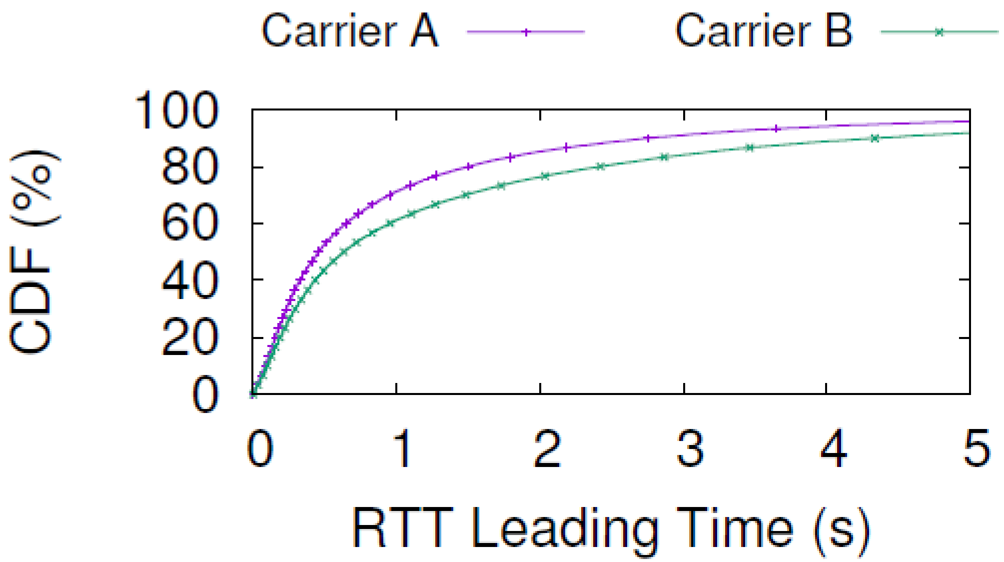
(Loosed RTT estimation)

t_{HOF}: Predicted with Linear SVM, location as feature

Event-triggered Schedulerlets

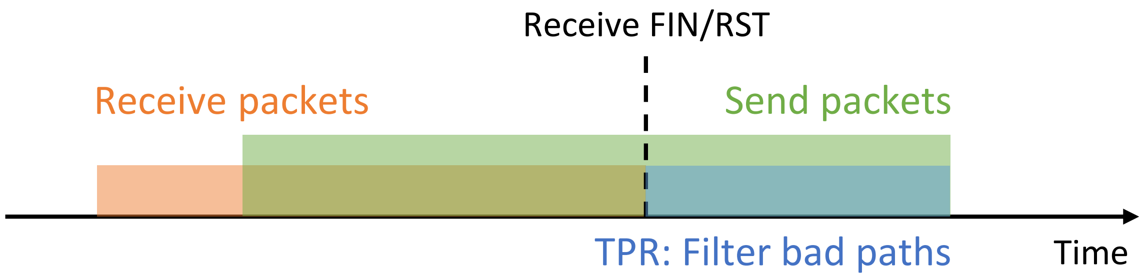
- **Event #2**: Incoming session tail
  - Path diversity may cause large tail delay, which hurts short session user experience
  - Large RTT difference in different paths; fast-changing "best" path





Event-triggered Schedulerlets

- **Schedulerlet #2**: Tail-aware Path Rejection (TPR)
  - Triggered for a specified user session when the end of the session (FIN/RST) is detected
  - Favor good full link over bad available link during the session tail
    - > Sacrifice full interface(s) utilization for reduced out-of-order delay
    - > Discard if link delay > total time needed by the best path to complete the session
  - Expected impact
    - > Faster (short) flow completion



- Before FIN/RST: Send on all paths
- After FIN/RST: Avoid sending on clogged paths

If $T^-_{i,f} > T^+_{i,f}$, then path i is clogged for packet f

$$T^-_{i,f} \coloneqq owd_i + \frac{buf_i}{bw_i}$$

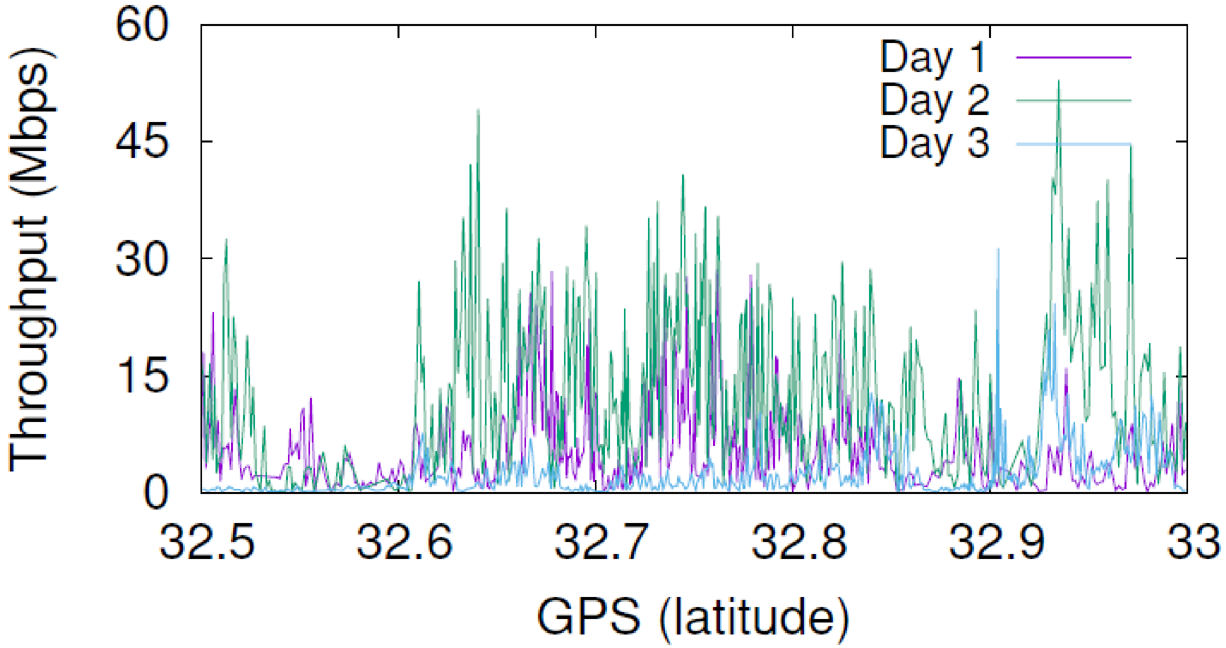
(Lower bound of packet delivery time)

$$T^+_{i,f} \coloneqq \min_{j \neq i} \left\{ \left(owd_j + \frac{buf_j + remain_f}{bw_j} \right) (1 + \eta_j) \right\}$$

(Upper bound of single-path flow completion time)

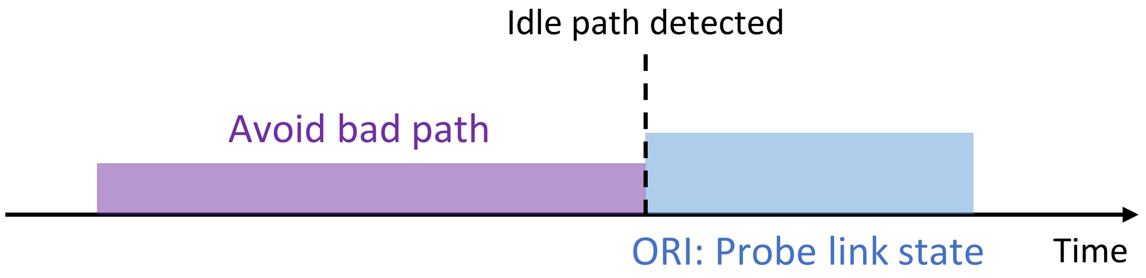
Event-triggered Schedulerlets

- **Event #3**: Path idle
  - Performance metrics on paths varies quickly over time
  - Performance metrics on idle paths could be very inaccurate



Event-triggered Schedulerlets

- **Schedulerlet #3**: Opportunistic Redundant Traffic Injection (ORI)
  - Triggered when a path was kept idle for some period
  - Send redundant copies on idle (bad) paths to detect their "recovery" state earlier
    - Keep network metrics updated
    - Opportunistically reduce session-level packet loss and out-of-order delay
  - Expected Impact
    - Higher interface utilization rate
    - Better scheduling decision



- Before detection: Avoid sending on bad paths
- After detection: Send probe packets (redundant copies)

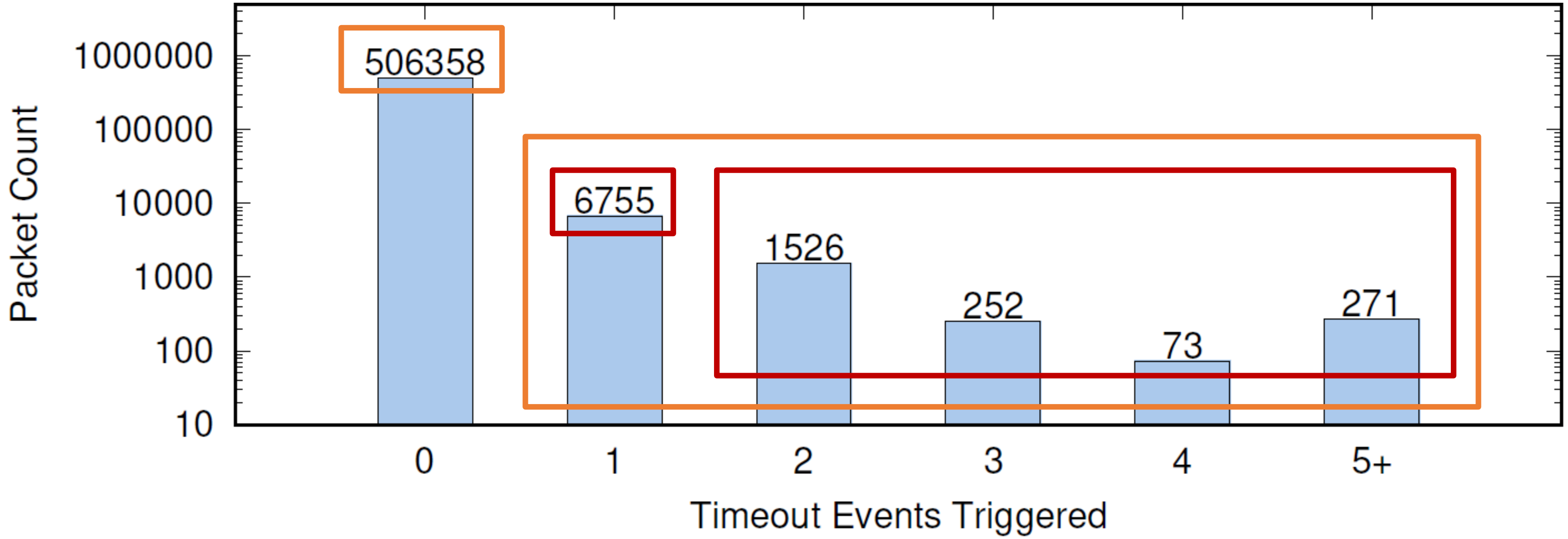
If no traffic is scheduled over a path for α seconds, or β bytes worth of data, the path is idle

Send at most τ probe packets onto idle path to refresh the network performance metrics

$\alpha = 1, \beta = 8\text{KB}, \tau = 16$

Event-triggered Schedulerlets

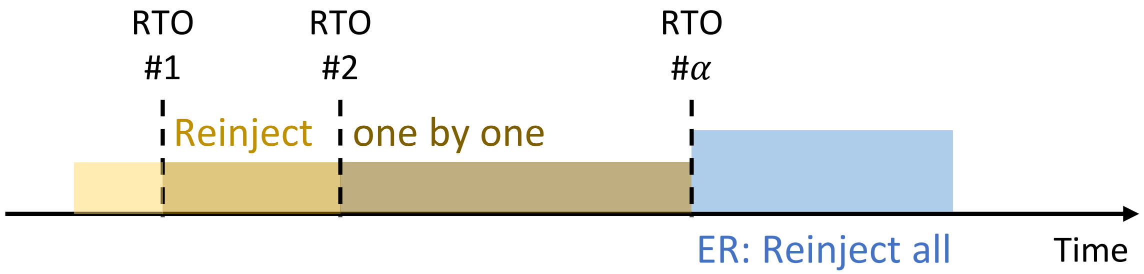
- **Event #4**: Repeated retransmission timeout
  - Frequent timeout; Multiple timeouts on a single packet



~1.8% packets experienced timeout; ~24% of them experienced multiple timeouts.

Event-triggered Schedulerlets

- **Schedulerlet #4**: Extended Reinjection (ER)
  - Triggered when packets experienced α repeated timeouts
  - Reinject all unACKed packets upon α timeouts on a single packet
    - > Balancing performance gain and overhead
      - ~ E.g., α = 3 incurs 0.4% overhead, while α = 2 incurs 15% (which is unacceptable)
  - Expected Impact
    - > Less packets with extremely prolonged delay and shorter end-to-end delay

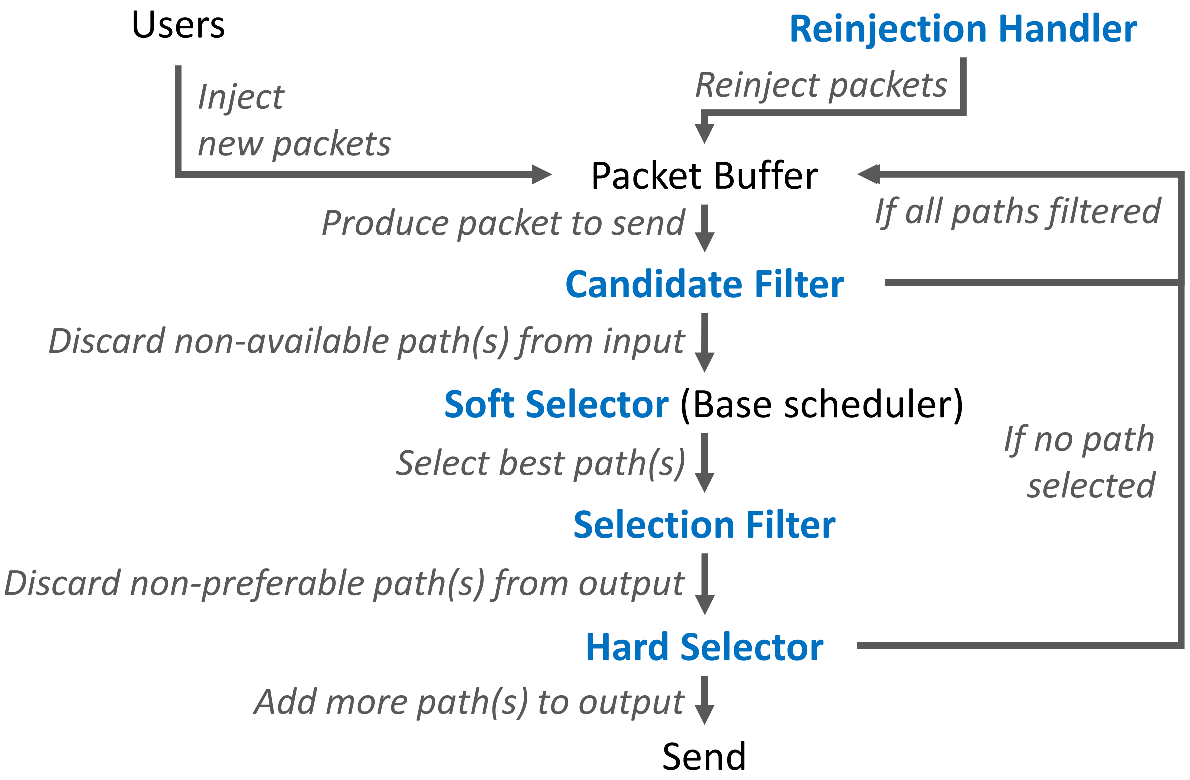


- Before RTO #α: MPTCP-flavored reinjection
- Upon RTO #α: Reinject all unACKed packets sent on corresponding path

$\alpha = 3$

Composable Scheduling Framework

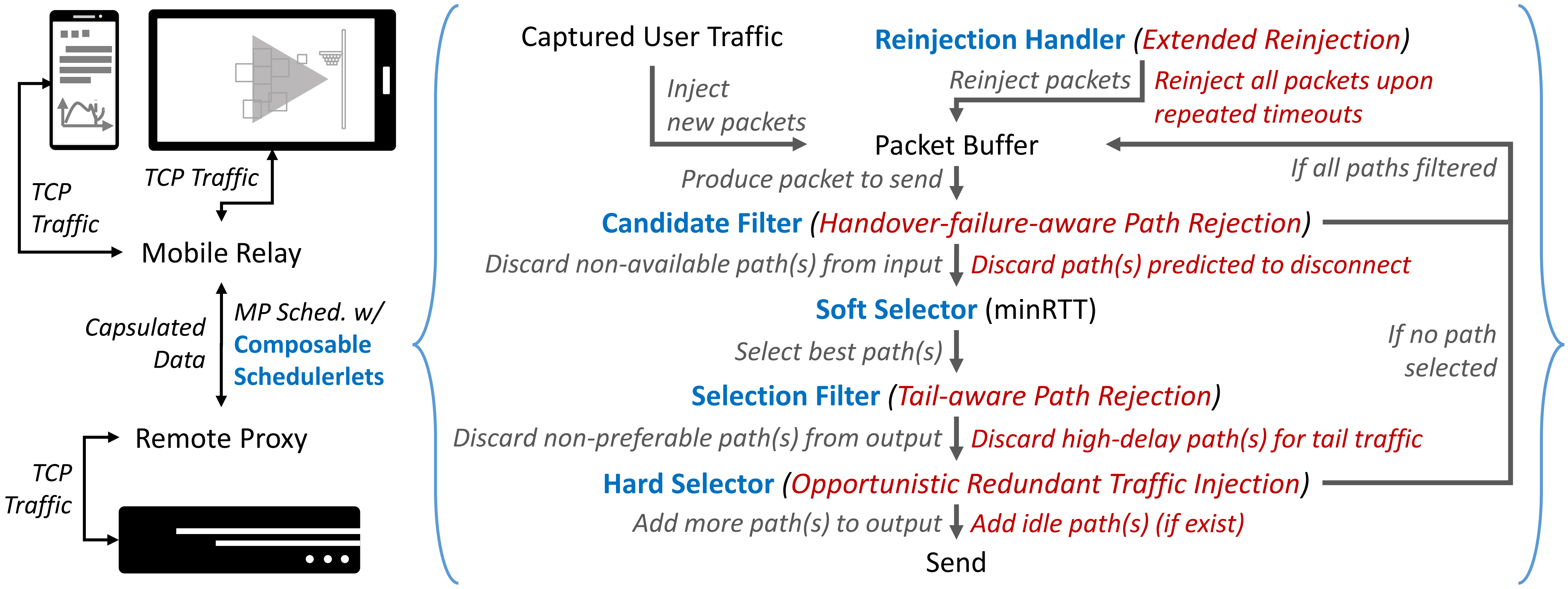
- System Overview (Refer to the paper for detailed design!)



- Base scheduler + Event-triggered **Schedulerlets**
- Completeness: By properly applying Schedulerlets, one can convert any specified multipath scheduler A into any other multipath scheduler B

Composable Scheduling Framework

- System Overview (Refer to the paper for detailed design!)

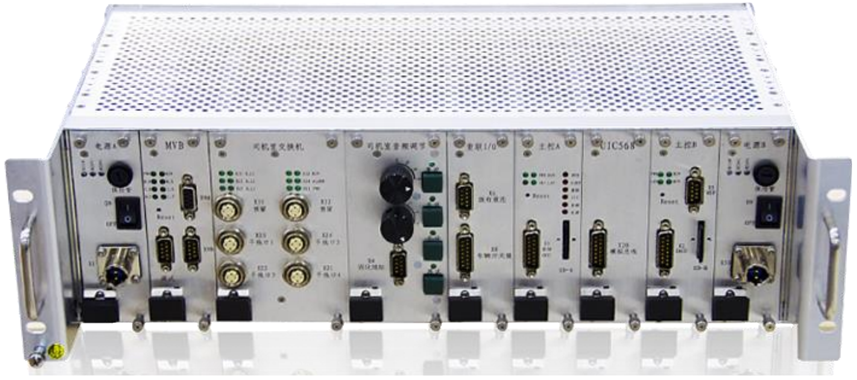


POLYCORN: Data-driven Cross-layer Multipath Networking for High-speed Railway through Composable Schedulerlets

- Extra Networking Challenges From Extreme Mobility
  - Fluctuating, Unpredictable and Heterogeneous
  - Inaccurate Measurements Hurt Performance
- System Design
  - Event-triggered Schedulerlets
  - Composable Scheduling Framework
- Evaluation
  - Deployment on HSR LTE Gateway
  - 3 weeks/52720 km Evaluation on Beijing-Shanghai Route

Deployment on HSR LTE Gateway

- The HSR LTE gateway
  - Per-cabin Wi-Fi AP
  - Per-train LTE Gateway (shown in the figure)
    - Fixed antenna on top of the cabin
    - Multiple prioritized LTE interfaces from major cellular carriers
  - We acquired exclusive access to 4 LTE interfaces from 2 carriers for our evaluation
    - 2 for Polycorn, 2 for the baseline solution



- Where did we evaluate Polycorn?
  - Beijing-Shanghai HSR route
    - 1318km total length, busiest HSR route in China
    - 3 weeks, 40 trips (52720km)

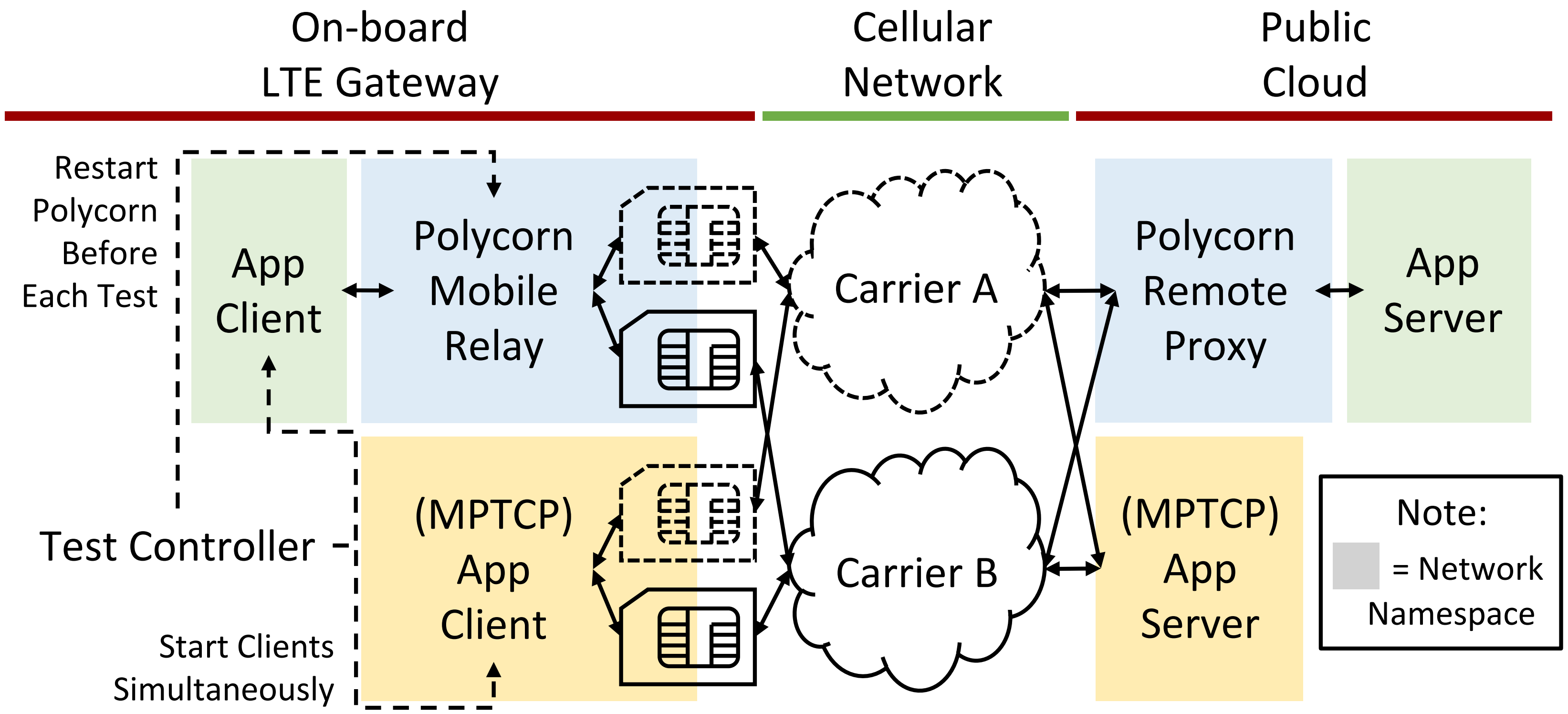
Deployment on HSR LTE Gateway

- The HSR LTE gateway
  - Per-cabin Wi-Fi AP
  - Per-train LTE Gateway (shown in the figure)
    - Fixed antenna on top of the cabin
    - Multiple prioritized LTE interfaces from major cellular carriers
  - We acquired exclusive access to 4 LTE interfaces from 2 carriers for our evaluation
    - 2 for Polycorn, 2 for the baseline solution
- Where did we evaluate Polycorn?
  - Beijing-Shanghai HSR route
    - 1318km total length, busiest HSR route in China
    - 3 weeks, 40 trips (52720km)

Beijing-Shanghai HSR route (1318km)

Deployment on HSR LTE Gateway

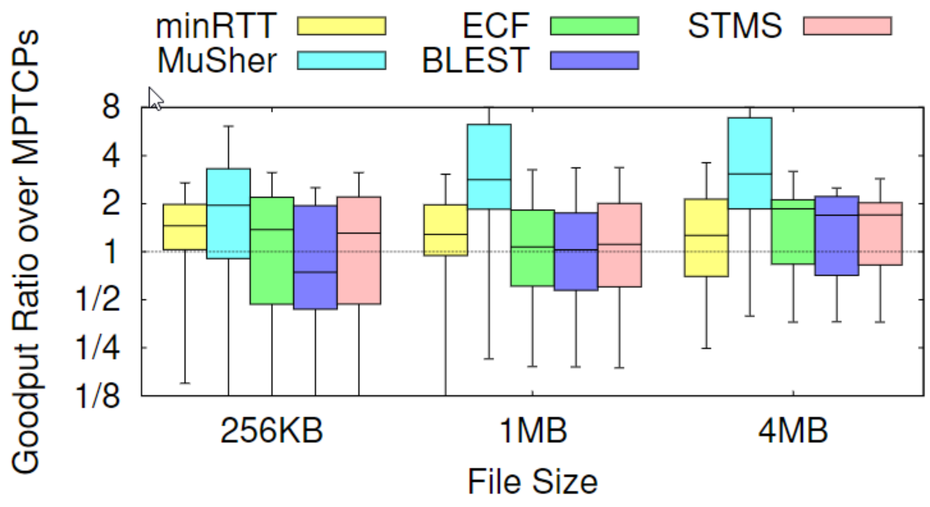
- Experimental setup: fair pairwise tests

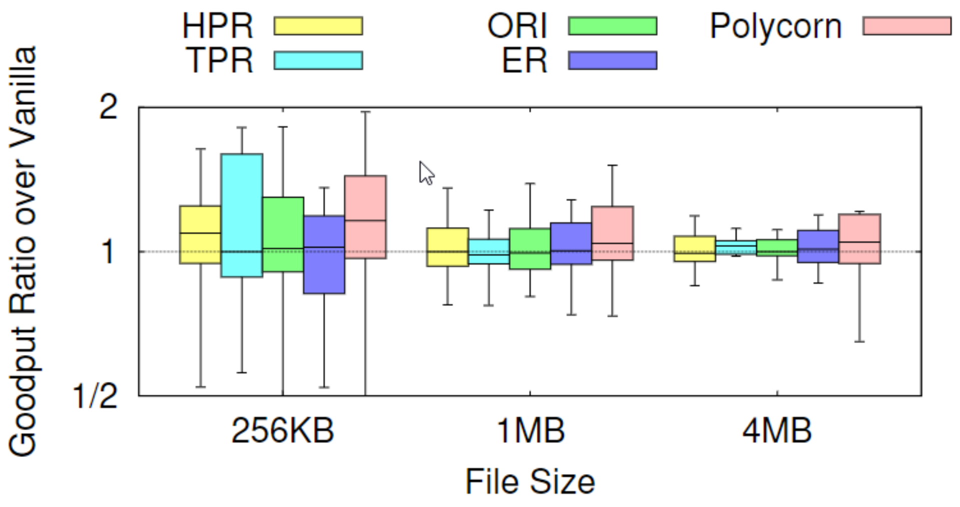


- Polycorn vs. MPTCP as an example
- Also Polycorn vs. Polycorn (microbenchmark), Polycorn vs. SPTCP (instant messaging)
- Run same test on Polycorn and baseline solution at the same time
- Each result is collected from 50 individual tests

On-board Evaluation

- Single user bulk data download performance





Goodput gain over MPTCP minRTT:

+41% (256 KB), +47% (1 MB), +78% (4 MB)

57% (average)

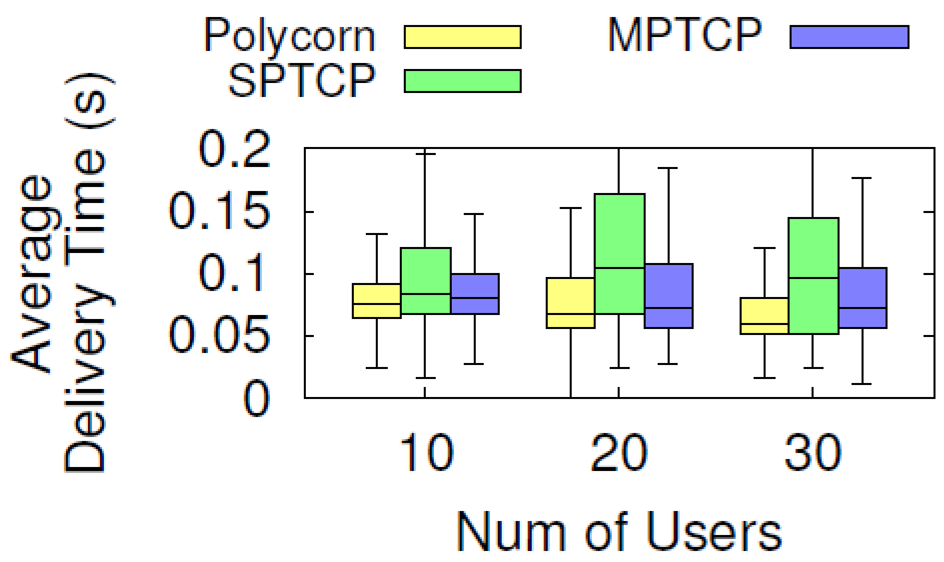
Microbenchmark goodput gain:

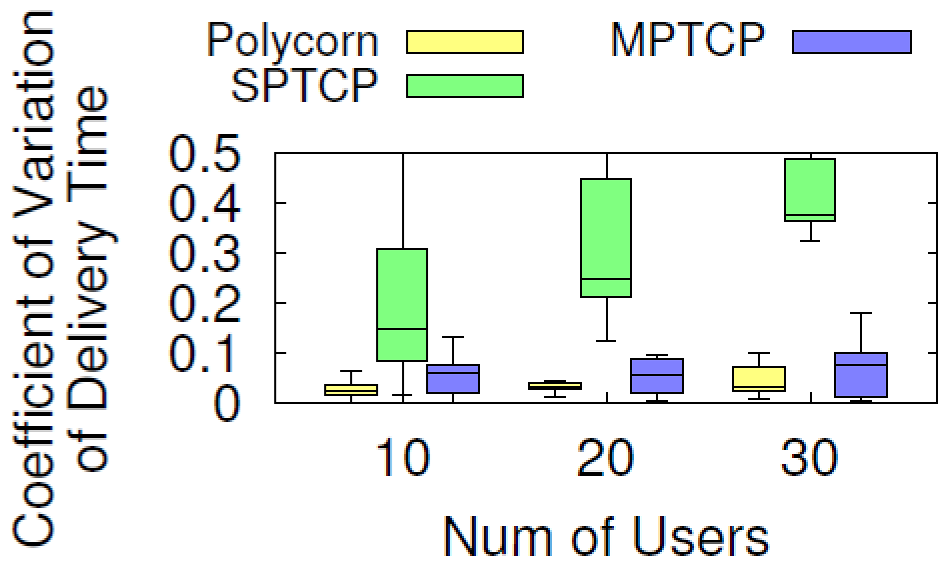
+7% (HPR), +18.8% (ORI),

+4.2% (TPR), 2.9% (ER)

On-board Evaluation

- Multi-user instant messaging performance





Delivery time reduction:

45% over SPTCP and 16% over MPTCP

Coefficient of variance reduction:

86% over SPTCP and 49% over MPTCP

POLYCORN: Data-driven Cross-layer Multipath Networking for High-speed Railway through Composable Schedulerlets

Yunzhe Ni, Feng Qian, Taide Liu, Yihua Cheng, Zhiyao Ma, Jing Wang, Zhongfeng Wang, Gang Huang, Xuanzhe Liu, Chenren Xu

- Conclusion + Take-away messages
  - HSR LTE networks are fluctuating, unpredictable and heterogeneous.
    - In this work, we study to what extent these features are present, and how to derive a multipath scheduler design based on them.
    - Also, we show that handover failures could be classified and predicted using historical data for the first time. (Refer to the paper!)
  - Event-driven approach that tries to understand the network only when it is understandable works.
    - Start from a robust "base" scheduler, shape its behavior when specified events are detected.
  - Polycorn, with its composable scheduler framework and event-triggered schedulerlets, achieved 57% (average) better goodput compared to its base scheduler, outperformed SOTA multipath schedulers, and preserved user-level fairness in multi-user scenario.
    - Evaluated on HSR LTE gateway, Beijing-Shanghai HSR route.
- Thanks for listening!