# Enabling Flow-level Latency Measurements across Routers in Data Centers

Parmjeet Singh, **Myungjin Lee**
Sagar Kumar, Ramana Rao Kompella

PURDUE UNIVERSITY

# Latency-critical applications in data centers

- Guaranteeing low end-to-end latency is important
  - Web search (e.g., Google's instant search service)
  - Retail advertising
  - Recommendation systems
  - High-frequency trading in financial data centers

- Operators want to troubleshoot latency anomalies
  - End-host latencies can be monitored locally
  - Detection, diagnosis and localization through a network: no native support of latency measurements in a router/switch

# Prior solutions

- Lossy Difference Aggregator (LDA)
  - Kompella et al. [SIGCOMM '09]
  - Aggregate latency statistics

- Reference Latency Interpolation (RLI)
  - Lee et al. [SIGCOMM '10]
  - Per-flow latency measurements

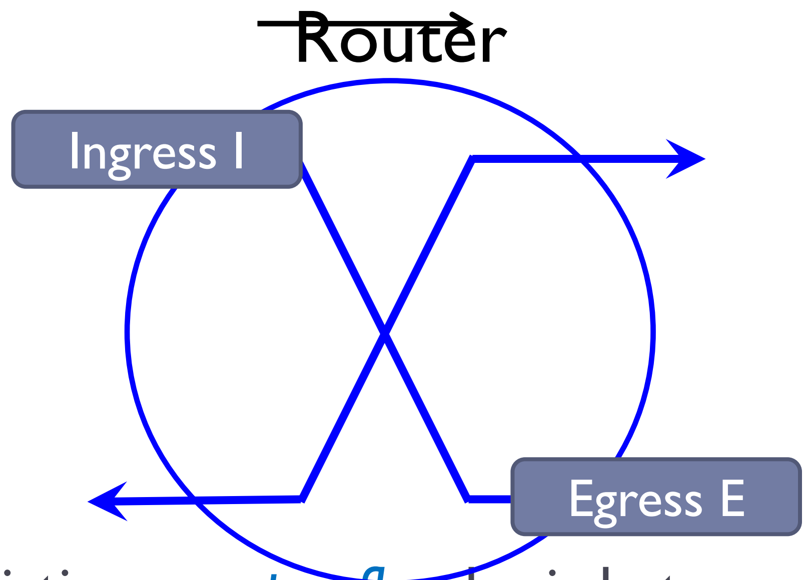
More suitable due to more fine-grained measurements

# Deployment scenario of RLI

- Upgrading all switches/routers in a data center network
- Pros
  - Provide finest granularity of latency anomaly localization
- Cons
  - Significant deployment cost
  - Possible downtime of entire production data centers

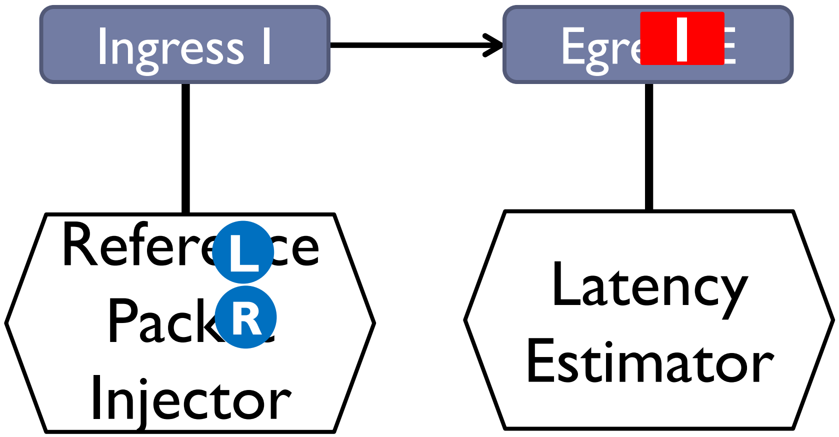
- In this work, we are considering partial deployment of RLI
  - Our approach: *RLI across Routers (RLIR)*

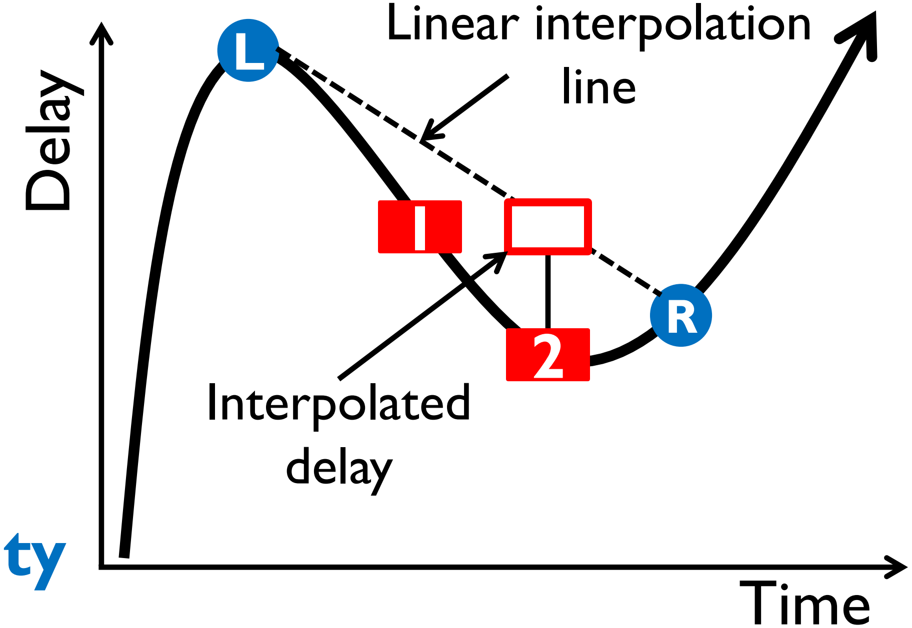
# Overview of RLI architecture




- Goal
  - Latency statistics on a *per-flow* basis between interfaces
- Problem setting
  - No storing timestamp for each packet at ingress and egress due to high storage and communication cost
  - Regular packets do not carry timestamps

# Overview of RLI architecture




- Premise of RLI: **delay locality**
- Approach

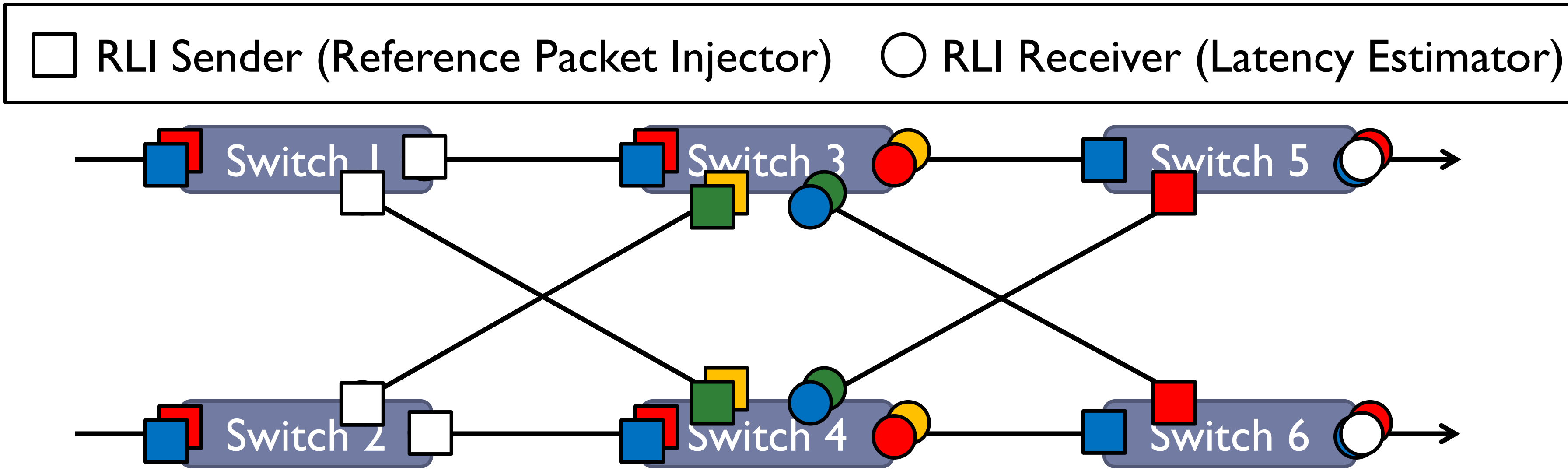
  1) The injector sends reference packets *regularly*

  2) Reference packet carries *ingress timestamp*

  3) *Linear interpolation*: compute per-packet latency estimates at the latency estimator

  4) Per-flow estimates by aggregating per-packet estimates

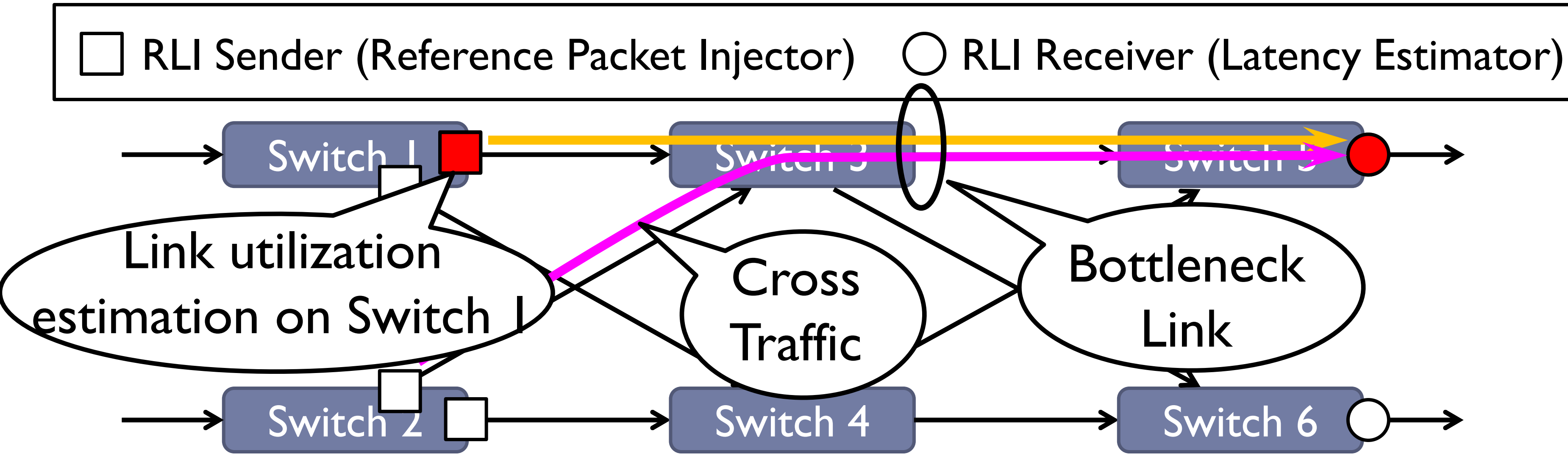
# Full vs. Partial deployment




- Full deployment: 16 RLI sender-receiver pairs
- Partial deployment: 4 RLI senders + 2 RLI receivers

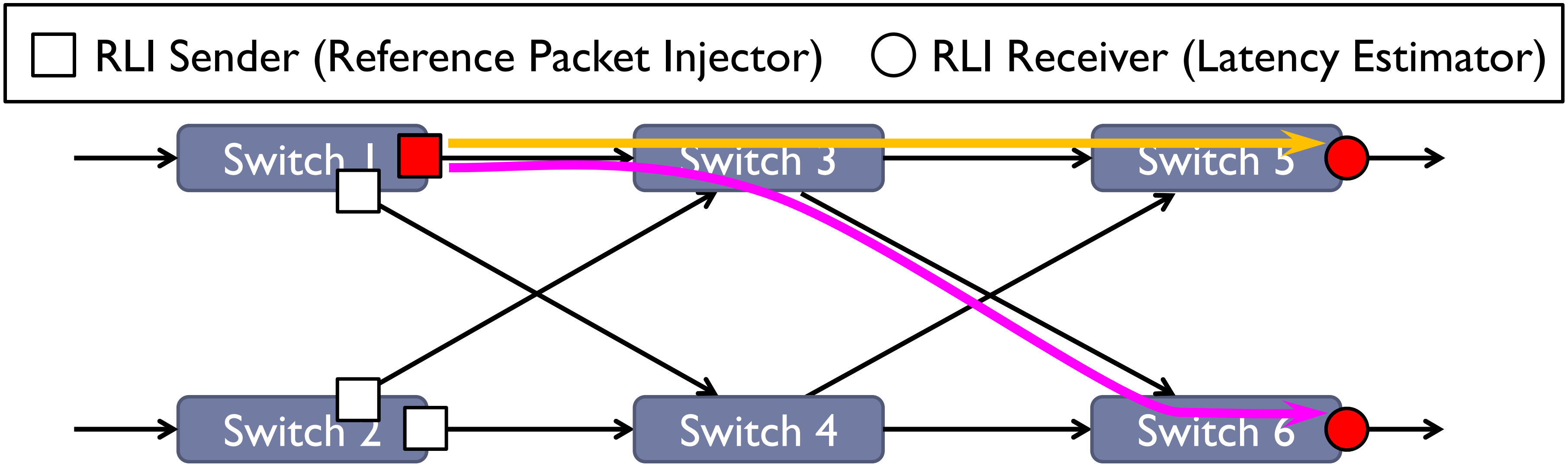
- 81.25 % deployment cost reduction

# Case 1: Presence of cross traffic




- **Issue**: Inaccurate link utilization estimation at the sender leads to high reference packet injection rate
- Approach
  - Not actively addressing the issue
  - Evaluation shows no much impact on packet loss rate increase
  - Details in the paper

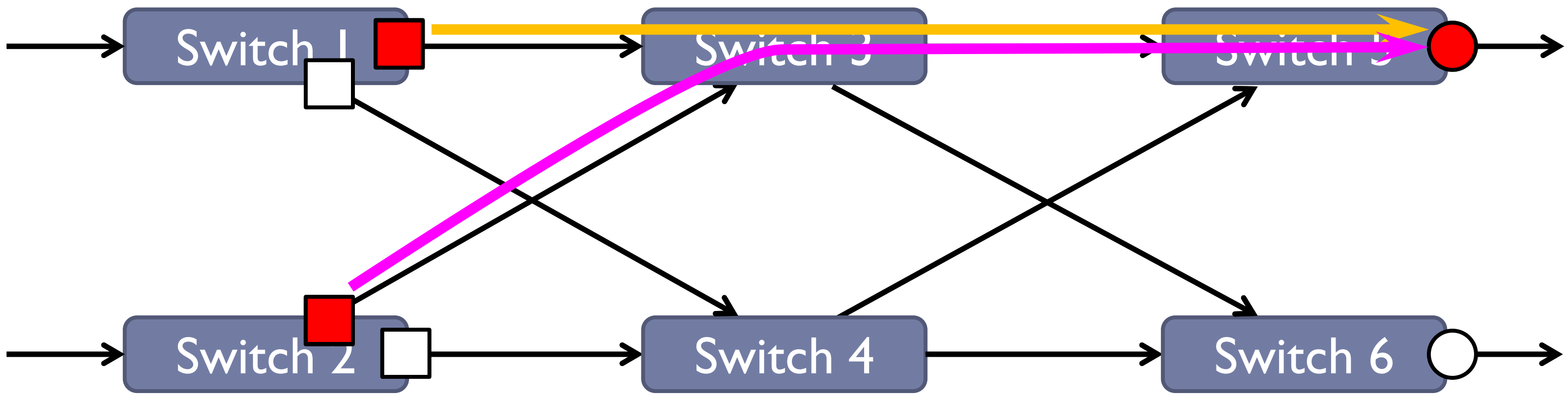
# Case 2: RLI Sender side




- Issue: Traffic may take different routes at an intermediate switch
- Approach: Sender sends reference packets to all receivers

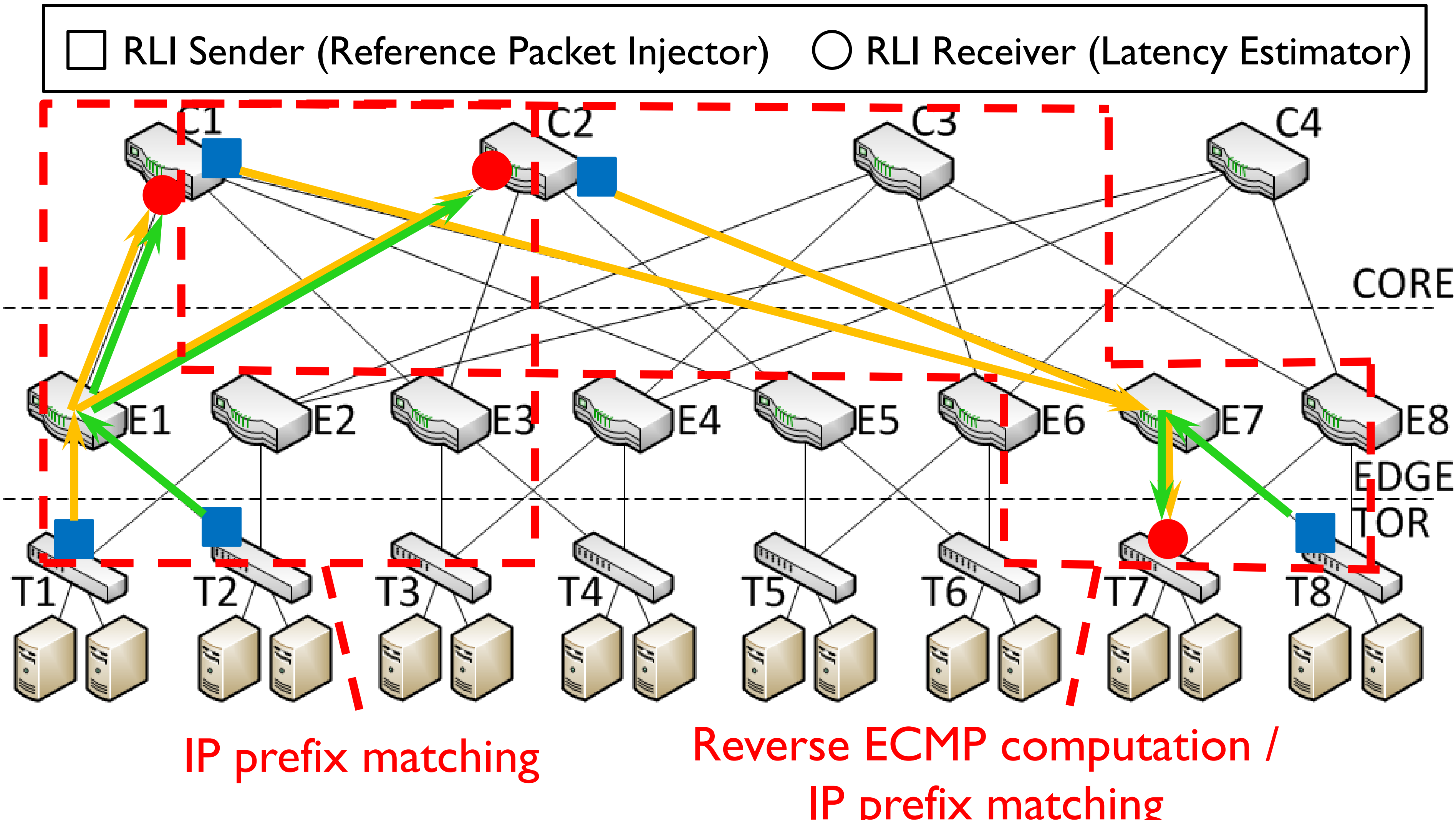
# Case 3: RLI Receiver side

□ RLI Sender (Reference Packet Injector) ○ RLI Receiver (Latency Estimator)



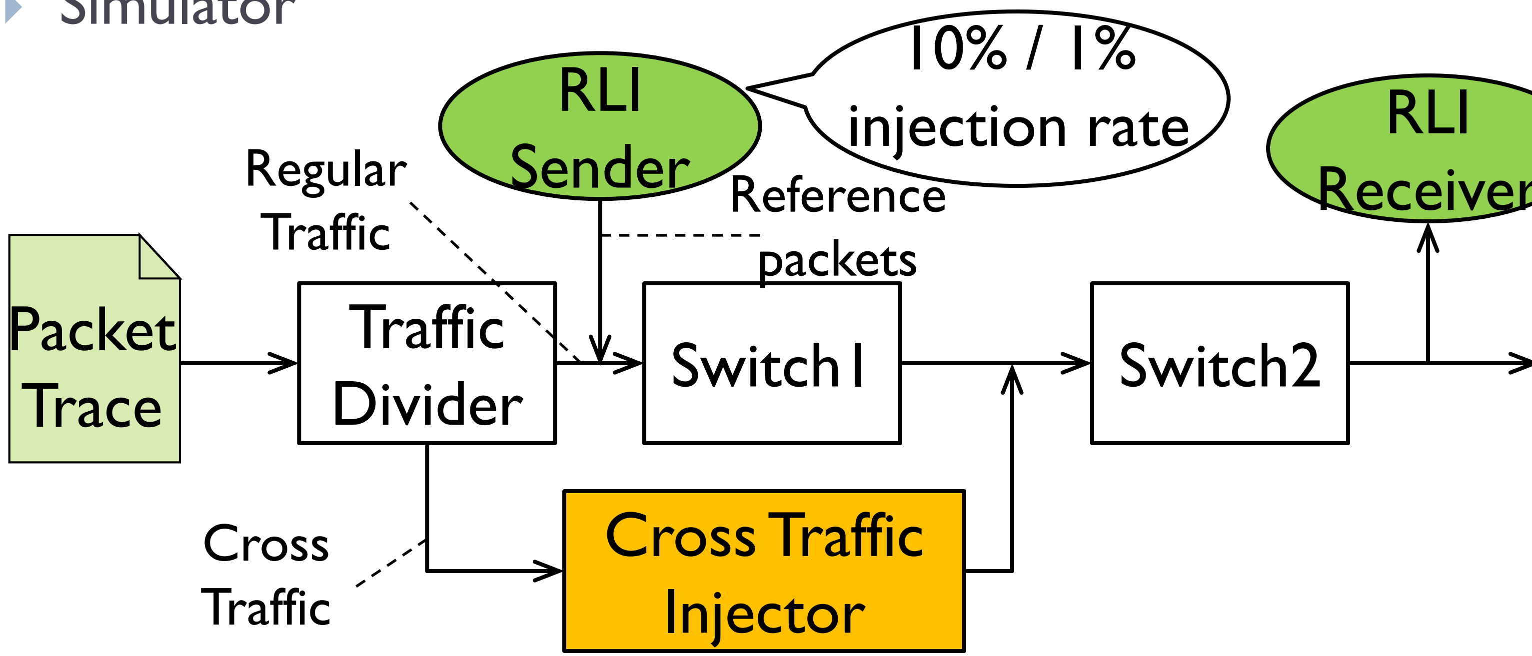

- Issue: Hard to associate reference packets and regular packets that traversed the same path
- Approaches
  - Packet marking: requires native support from routers
  - Reverse ECMP computation: 'reverse' engineer intermediate routes using ECMP hash function
  - IP prefix matching at limited situation

# Deployment example in fat-tree topology



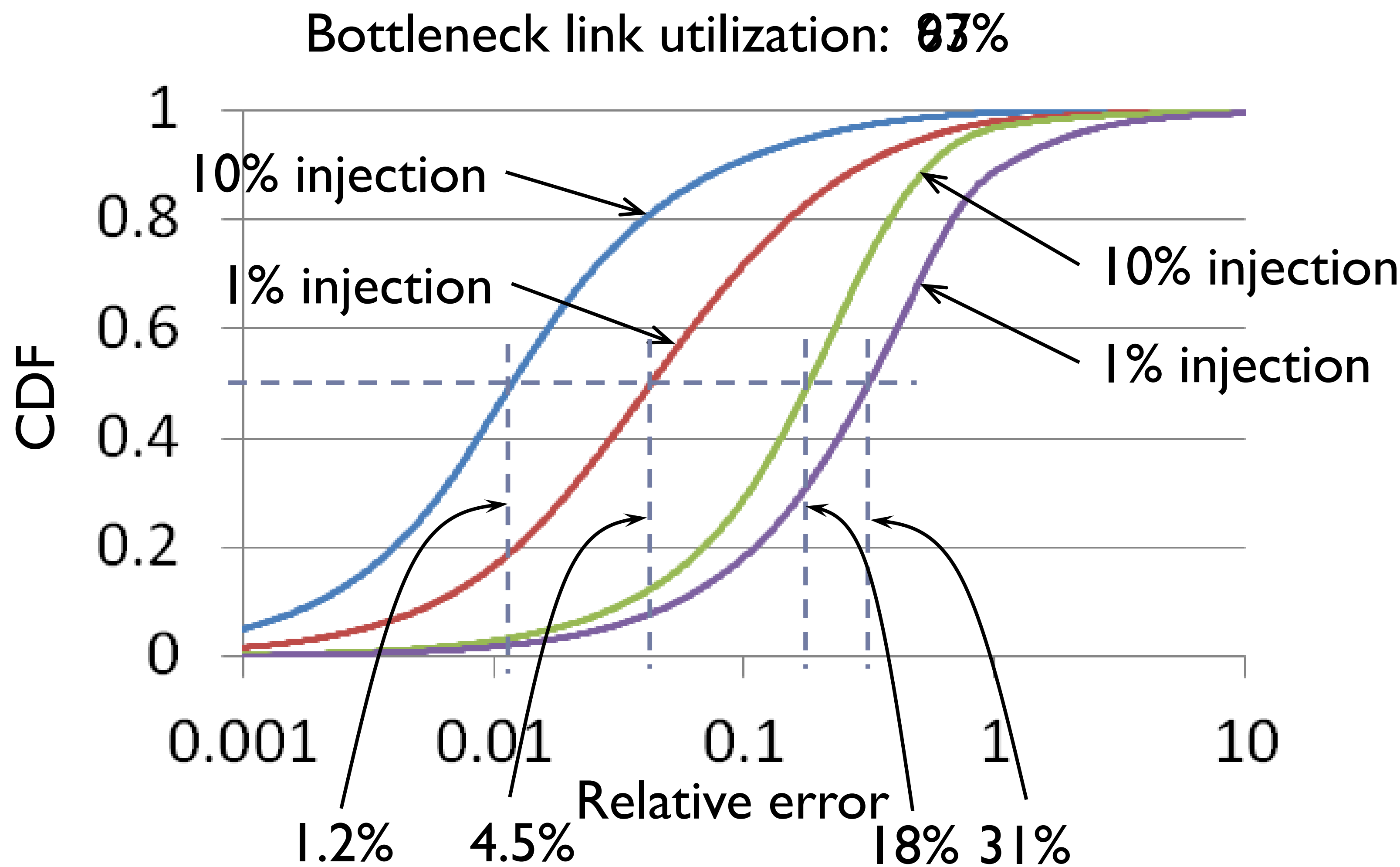
# Evaluation

- Simulation setup
  - Trace: regular traffic (22.4M pkts) + cross traffic (70M pkts)
  - Simulator




- Results
  - Accuracy of per-flow latency estimates

# Accuracy of per-flow latency estimates



# Summary

- Low latency applications in data centers
  - Localization of latency anomaly is important
- RLI provides flow-level latency statistics, but full deployment (i.e., all routers/switches) cost is expensive
- Proposed a solution enabling partial deployment of RLI
  - No too much loss in localization granularity (i.e., every other router)

# Thank you! Questions?