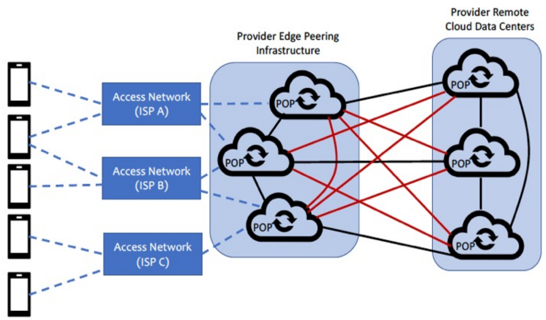
EdgeBoost: Data-Driven Traffic Engineering and Configuration Verification for Cloud Service Assurance

PI: Prof. Mahesh K. Marina
Project team members: Leyang Xue, Waleed Ahsan

BACKGROUND

- Most networked applications on end devices partially or wholly served by remote processing and storage in the cloud
- To better support such cloud-dependent applications, service/content providers extending their infrastructure towards the edge
- Providers extend their footprint to the edge through their own private wide area network (WAN) while also peering with public Internet at the edge points of presence (POPs)
- Edge infrastructure not just for networking but also augmented with compute and storage, making it a cloud closer to end devices

TARGET END-TO-END SCENARIO



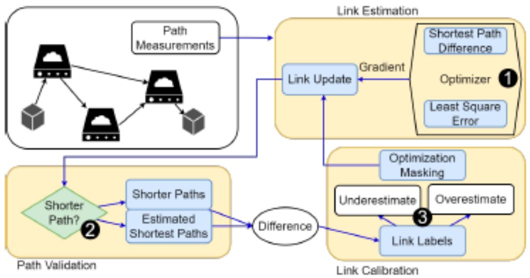
- Provider's remote and edge POPs form a fully connected "overlay" mesh network
- Each overlay link can span multiple physical network links
- Black colored overlay links in the figure are solely through provider's private WAN
- Red colored ones include links from public Internet

PROJECT OBJECTIVES

- Broad aim of this project is *end-to-end service assurance* in the target scenario outlined above, where "service" refers to a set/class/group of application flows that run on end-devices with similar performance requirements (e.g., bounded delay)
- Meeting this objective involves addressing several challenges:
  1. Accurate prediction of service demands and link performance characteristics over time
  2. Adaptive traffic engineering (TE) at service level: path selection and bandwidth allocation for each service's traffic to meet its SLA
  3. Identifying faulty and anomalous path segments
  4. Enforcing and verifying control plane routing decisions over the data plane

PAINT: PATH AWARE ITERATIVE NETWORK TOMOGRAPHY [ICNP'22]

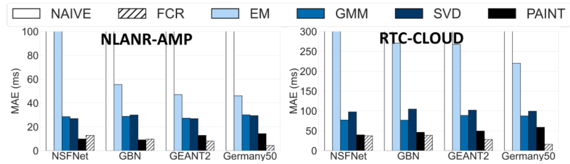
- Knowledge of link level performance key to assuring quality of cloud based and other services via TE, robust network operations and beyond
- Direct measurement of each link incurs too much overhead or is infeasible due to various reasons
- Existing approaches to network tomography make unrealistic assumptions on link stability, path controllability and visibility
- Motivated by this, we propose a new online iterative network tomography algorithm called PAINT



- With PAINT, the link metrics are iteratively estimated by minimizing their least square error (LSE) and calibrated based on the comparison of weight between the estimated shortest paths (SPs) and best-known paths from end-to-end path measurements
- *Key insight: when there is inconsistency between the estimated and measured paths, then weights of links on the estimated SP are likely mis-estimated, triggering a further round of estimation to refine the estimated link metrics*

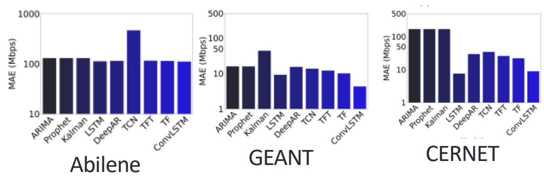
PAINT ACCURACY EVALUATION

- We evaluated PAINT for link delay estimation using four different real network topologies and two real-world delay measurement datasets



- *Takeaway: PAINT yields up to 3x gain in absolute link delay estimation accuracy compared to existing approaches*

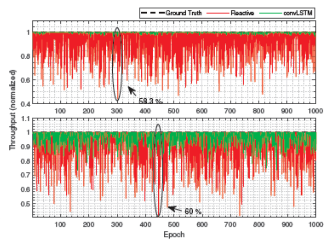
COMPARISON OF DEMAND MATRIX FORECASTING METHODS



- *Takeaways:*
  - *Forecasting method performance dependent on demand matrix data characteristics*
  - *Simpler deep learning based forecasting methods (LSTM and ConvLSTM) more effective than statistical or complex deep learning models*

BENEFIT OF DEMAND MATRIX FORECASTING FOR TE

- Examined the impact of demand matrix forecasting to enable proactive TE
- Results with throughput maximization and latency constrained throughput maximization with GEANT dataset shown on right



- *Takeaway: Augmenting common TE algorithms with DM forecasting ability leads to significant TE efficiency gain by up to 60%*