Special Topics: Machine Learning (ML) for Networking

COL867

Holi, 2025

Course Project Ideas

Tarun Mangla

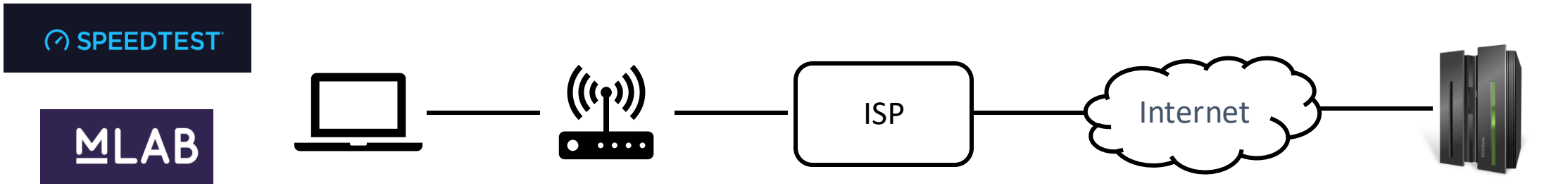
Project Logistics

- Work in a pair
- Project weight: 35%
- Three milestones
  - #1: Project proposal (3%) – Feb 8
  - #2: Progress review (12%) – March 18
  - #3: Final Submission (20%) – Apr 26
- Proposal: Title, steps, and timeline
- Progress review: Short report discussing progress and bottlenecks
- Final submission: 10-minute presentation, working code, report

#1: Early Termination Strategy for Speed Test
#2: Contextualization of Speed Test Data
#3: ML and UDP-based speed tests

Background

- Speed tests are used to diagnose network quality
- Anatomy of a speed test



Flooding-based tools, differ in
- Number of threads
- Test duration
- Sampling etc.

Access Medium
- Wireless
- Wired

#1: Early Termination Strategy for Speed Test

- Speed tests have significant data overhead
- Data used proportional to test duration and link speed
- Trade off:
  - Longer duration → More information about network quality
  - Short duration → Less data usage
- **Goal:** Can we get maximum information about the network quality with minimal data overhead
- **Methodology:** Formulate early termination as a classification/regression problem
- **Evaluation**: Quantify the data saving vs information loss

#2: Contextualization of Speed Data

- Existing speed test data lacks metadata information about the test context:
  - Wired vs WiFi vs Cellular
  - Cellular: 4G vs 5G (access technology)
  - Speed Tier

- Understanding the speed test context useful for downstream task

- **Goal:** Can you infer the test context post hoc using packet-level speed test data?

#3: ML and UDP-based speed testing

- UDP-based methods exist that are used to measure the available bandwidth
- **Challenge**: These methods don't work well under some conditions, including interrupt coalescing, presence of multiple bottlenecks.
  - Mostly because they make strong assumptions about the underlying network
- **Solution:** Augment these methods with machine learning
  - Related paper: Machine learning for measurement-based bandwidth estimation
  - Uses a simple Neural network model
- **Goal**: Improve the paper

#4: Foundation Model for Network Data

- Foundational model have been shown promising results for a range of use cases

- Recent attempts in the networking community to develop foundational models for networking (e.g., netFound)

- **Goal:** Implement and evaluate a foundational model for network data on campus network data

#4: Foundation Model for Network Data

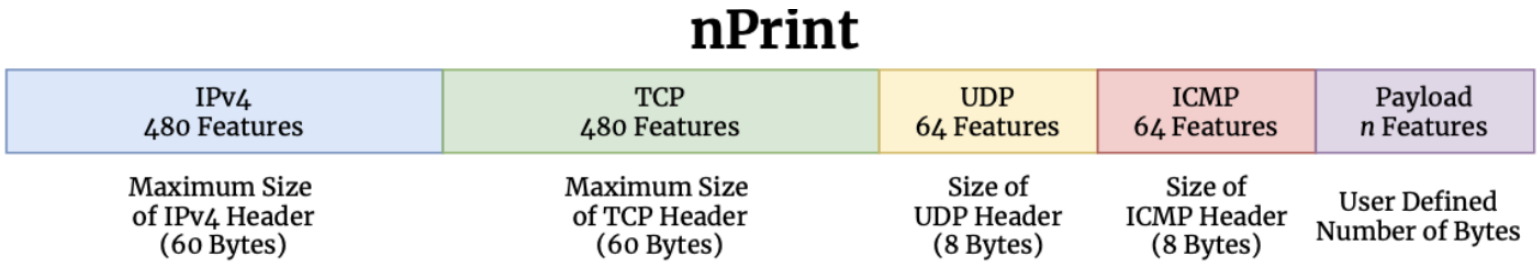
- Multiple papers in recent times: netFound, netLLM, ..
- Train the model using campus network data (will be provided)
- Evaluate the model for a learning task (preferably application classificaiton) and report the accuracy
  - How much data is needed for training the model?
  - What is the accuracy compared to a model with feature engineering

#5: Synthetic Network Data Generation

- Organizations are often comfortable sharing synthetic data derived from their network instead of actual data
- Traditional approaches to generate synthetic data are not low fidelity and fail to capture heavy-tailed behavior
- **Solution:** Two popular models
  - NetShare: Use Generative Adverserial Networks (GANs)
  - NetDiffusion: Use diffusion models
- **Scope:**
  - Reproduce the results from one of these papers using campus network data

#6: Compare nPrintML with a Feature Engineering method

- nPrint is a standard data representation for network traffic
- The code is available online: Link
- **Scope:**
  - For video conferencing performance prediction, compare performance of nPrintML with a feature engineering-based solution
  - Compare the system-level cost
  - Use Trustee to explain the results of nPrint



Other Ideas

- Pick an ML-based paper from Sigcomm 2020-2023 and implement it
- Or if you have your own idea, please discuss it with me

Project Logistics

- Work in a pair
- Project weight: 35%
- Three milestones
  - #1: Project proposal (3%) – Feb 8
  - #2: Progress review (12%) – March 18
  - #3: Final Submission (20%) – Apr 26
- Proposal: Title, steps, and timeline
- Progress review: Short report discussing progress and bottlenecks
- Final submission: 10-minute presentation, working code, report