A Contextual Bi-armed Bandit Approach for MPTCP Path Management in Heterogeneous LTE and WiFi Edge Networks

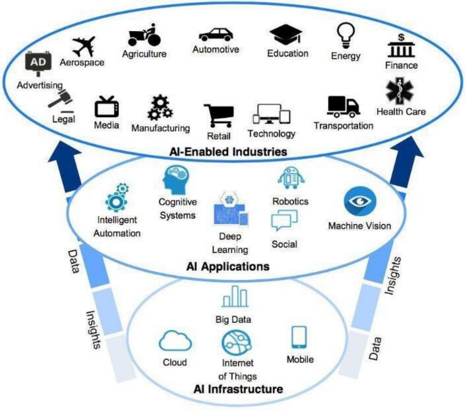
Aziza Al Zadjali[1], Flavio Esposito[2], Jitender Deogun[1]

Department of Computer Science, University of Nebraska-Lincoln[1]
Department of Computer Science, Saint Louis University[2]

October 30, 2020

Machine Learning at the Mobile Edge?

- High demanding requirements 5G networks.
- Enable running analytical and performance tasks closer to Edge devices.
  - Reduce network congestion
  - enhance application performance
- Connect IoT customers from vertical industries:
  - e-health
  - automotive
  - energy
  - agriculture

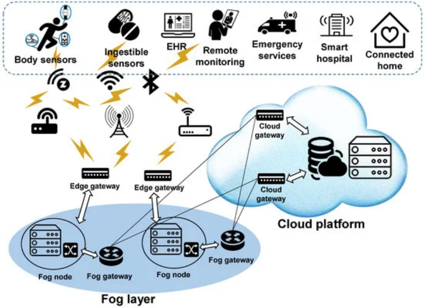


N SEC

Motivation: Dynamic Online Multi Path Transmission

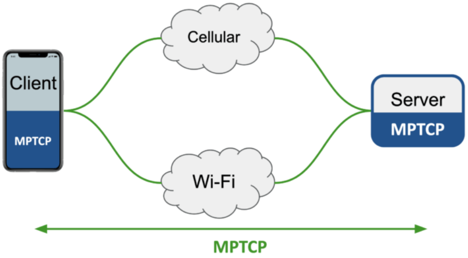
Online machine learning algorithms helps to make precise and effective decisions

1. Explore multiple paths for multiple access technologies (Wifi, LTE, etc).
2. Establish new subflows of multiple paths.
3. Uses online learning theory to take optimal decisions under unpredictable traffic environment.



Gap: Existing Transmission Protocols are Suboptimal

- Do not fit into dynamic and distributed environment.
- Missing adaptability and autonomy for heterogeneous networks.
- Rely on static and predefined rules
- Employ fullmesh to setup subflows between all available pair of interfaces.



Need for Real Time Automation

Automate decision process according to real time system learned rules.

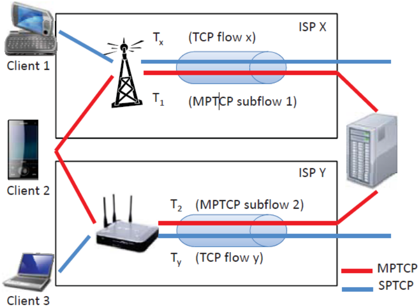
Objective: MPTCP Path Manager via Bi-Armed Bandit

- Design new MPTCP path manager
  - Use machine learning to generate optimal path decision rules under uncertain network conditions.
- Adopt contextual bandit (online active learner) to find MPTCP primary path in heterogeneous networks.

Multi-path TCP (MPTCP)

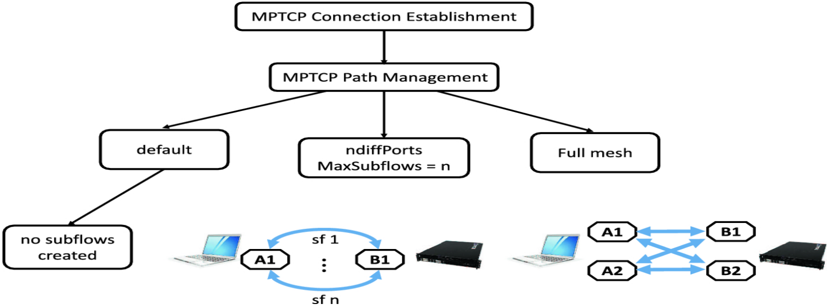
Forms multiple TCP flows over all available network interfaces to simultaneously utilize them.

- Split single data stream transmission across multiple paths.
- concurrent transmissions to increase connectivity resilience and maximizes network resources usage.



MPTCP Path Manager

The throughput of MPTCP relies extensively on its path management mechanism and path characteristics.



N SEC

Contextual Multi Armed Bandits (C-MAB)

Introduced by William R Thompson in 1993:

ON THE LIKELIHOOD THAT ONE UNKNOWN PROBABILITY EXCEEDS ANOTHER IN VIEW OF THE EVIDENCE OF TWO SAMPLES
(Thompson 1933), From the Department of Pathology at Yale University

- Machine learning in a streaming data setting by training a model in consecutive rounds.
    - At each round, the algorithm perform prediction on some input sample.
    - The algorithm verifies prediction correctness and feeds it back to the model.

N SEC

C-MAB Model Settings

Basic C-MAB Model

- At each round T, the algorithm selects an action and collects a reward for that chosen arm.
- For each round $t \in [T]$, the algorithm observes a context x_t, picks an arm a_t from $k = \{1, .., k\}$, and experience a reward $r_t \in [0, 1]$, whose value depends on the context x_t and the chosen arm a_t.

Notations

1. A set of contexts $x_k^t \in X$: $t = rounds, \; k = arms$
2. Policy $\pi : \; (context \; x) \mapsto (action \; a)$
3. Action / Arm a_t
4. Reward r_k^t

N SEC

C-MAB Model Settings (cont'd)

Exploration Vs. Exploitation dilemma.

- Use what is already learnt (exploit), but also learn about actions that look inferior (explore).
- Balance to get good statistical performance.

N SEC

Contextual Bandit Policies

Active Explorer:

With probability p:

Select action $a = argmax\ \hat{f}(x^t)$

Otherwise:

for arm q, Set $w_q = (1 - \hat{f}_q(x^t)||g_q(x^t, 0)|| + \hat{f}_q(x^t)||g_q(x^t, 1)||$

Select action $argmax\ w$

- Predictions are made according to an active learning heuristic:
    - The gradient that the observation would produce on each model predicting a class

N SEC

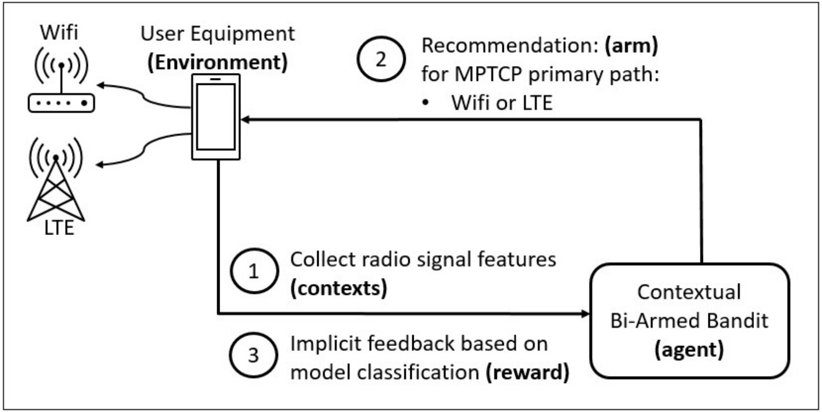
C-MAB Learning objective

Goal: Regret $\mapsto 0$ as fast as possible as $T \mapsto \infty$

Regret (i.e., relative performance) to a policy π

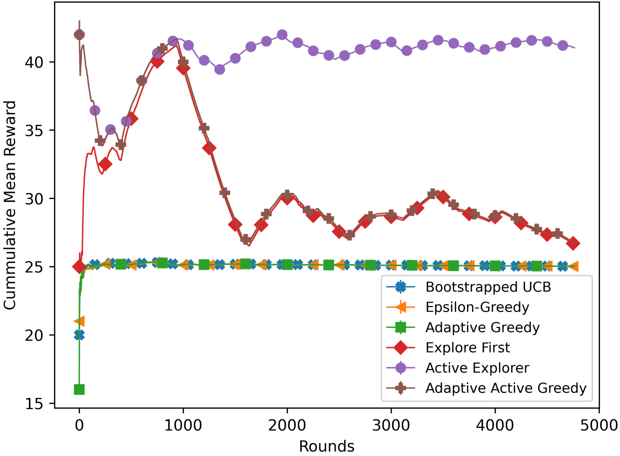
$$\underbrace{\max_{\pi} \frac{1}{T} \sum_{t=1}^{T} r_t(\pi(x_t))}_{\text{average reward of policy } \pi} - \underbrace{\frac{1}{T} \sum_{t=1}^{T} r_t(a_t)}_{\text{average reward of learner}}$$

Our Solution: MPTCP Path Manager via Bi-Armed Bandit



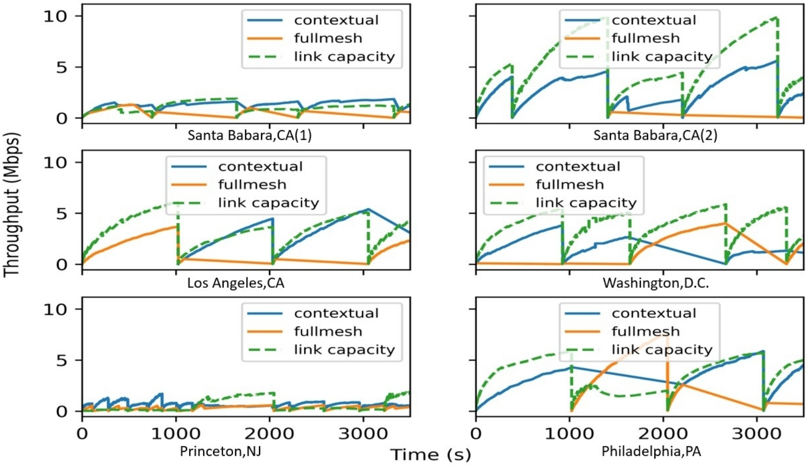
N SEC

C-MAB MPTCP Results and Evaluation



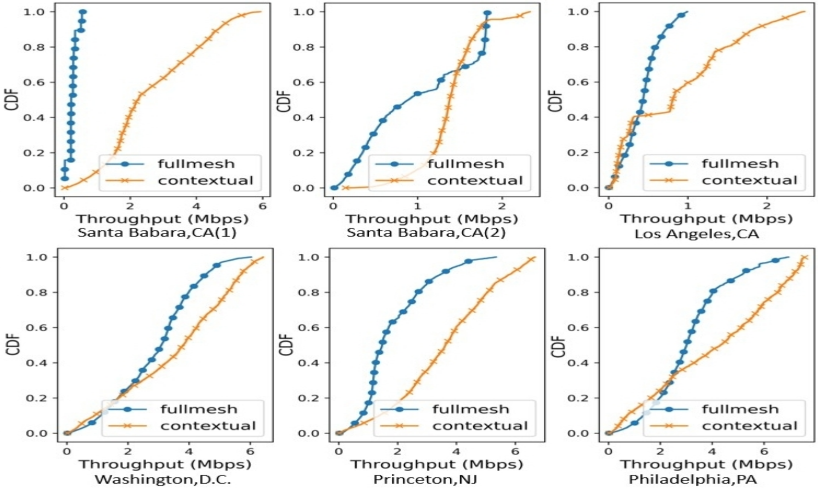
The mean cumulative reward (and its error upto 95% confidence level) is calculated for each policy over its 50 batch online simulations.

C-MAB MPTCP Results and Evaluation (cont'd)



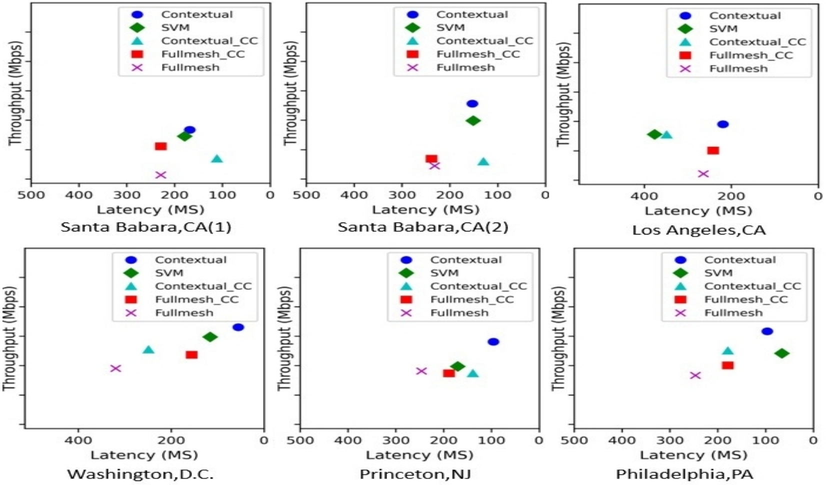
Contextual bandit path manager maximize utilization of available resource within given capacity limit.

C-MAB MPTCP Results and Evaluation (cont'd)



The throughput of contextual bandit approach is higher at a rate of around 50% of the times in average for all locations.

C-MAB MPTCP Results and Evaluation (cont'd)



The Top-right part of the graph indicate better performance.

Conclusion

- Designed MPTCP path manager selection strategy to decide primary path under rapid wireless signal fluctuations in heterogeneous edge networks.
    1. Online contextual bandit algorithm using Stochastic Gradient Descend classification as an oracle to decide the optimal primary MPTCP path for each new connection.
    2. A patch to the MPTCP protocol that allows overwrites to the path manager module.

N SEC

Thank You

Aziza Alzadjali: aalzadjali@cse.unl.edu