Accelerating Applications in the Fast-moving Devices with Proactive Provisioning

HyunJong (Joseph) Lee*, HeeWon Lee§, Moo-Ryong Ra‡, Yu Xiang§, Jason Flinn*^

* University of Michigan   §AT&T Labs - Research   ‡Amazon   ^ Facebook

Motivation

- Emerging applications in fast-moving devices
  - latency-sensitive & computation-heavy
  - **stateful**
- Edge-nodes at celll towers are densely populated
- Migrating previous to current *nearby* edge node is essential to retain low latency benefit of edge

Problem

- Migrating the app state just-in-time, on-demand
  - incurs **downtime**
  - degrades overall user-experience
- Fast-moving devices frequently switch edge-node
- Difficult to precisely predict next edge-node due to random mobility pattern and various use-cases

Proposed Solution: Proactive Provisioning

- Forecast next edge nodes based on only current context
  → construct projectory from location and velocity
  - the next node prediction algorithm is device-agnostic
  - no need for ISPs to expose strategy-sensitive info.
- Employ redundant proactive provisioning on multiple nodes
  → selectively choose nodes by computing estimated utility
  - **latency benefit** by migrating to future forecasted node(s)
  - **resource cost** (extra BW) to provision ahead-of-time
- Throughout the proactive provisioning stage, the state of applications on current edge nodes continues to change
  - provisioning too early wastes resource to sync the state
  - provisioning too late results in high downtime
  → must provision as late as possible but before switching
- Decouple control plane from mechanism plane allows
  - fine-grained control on where and when to provision
  - easy to support use-case specific provisioning algs.
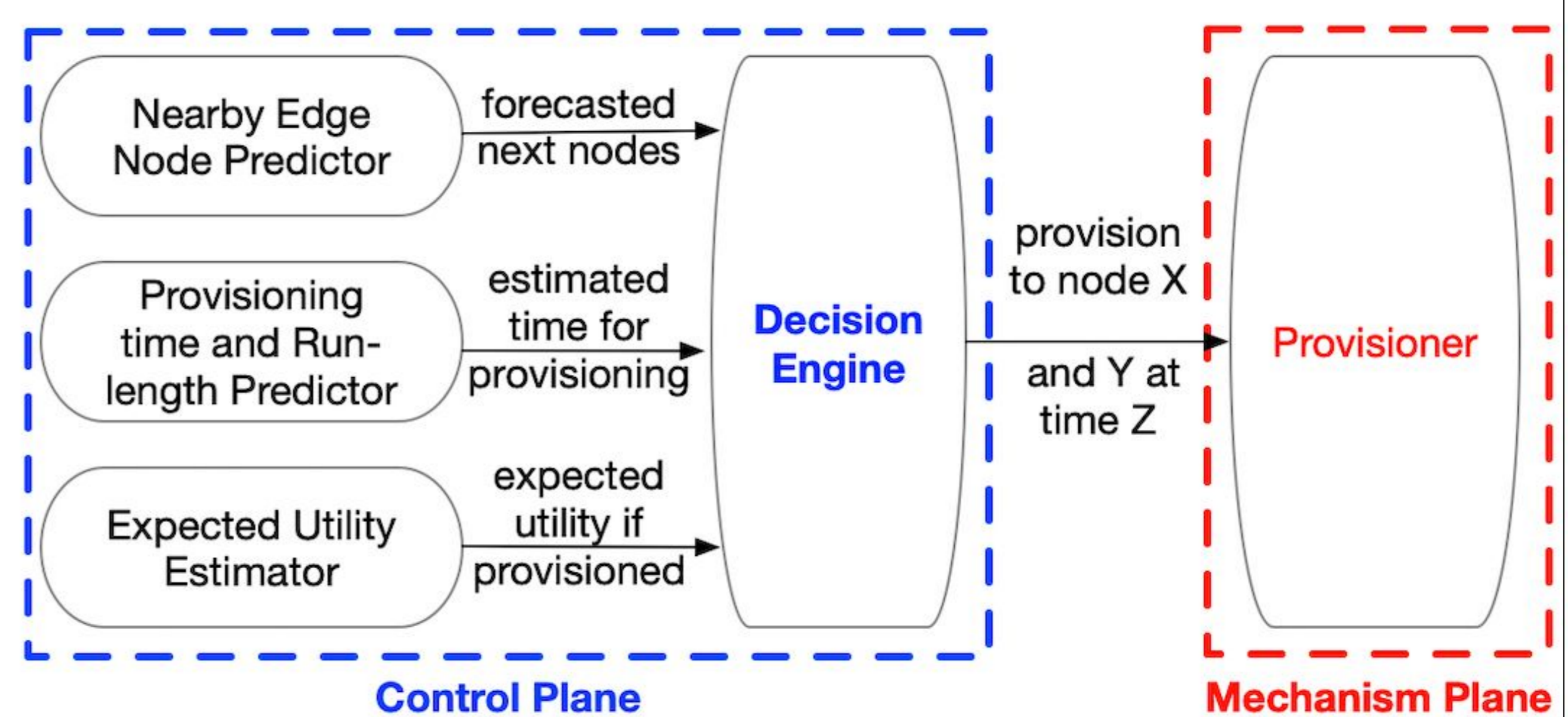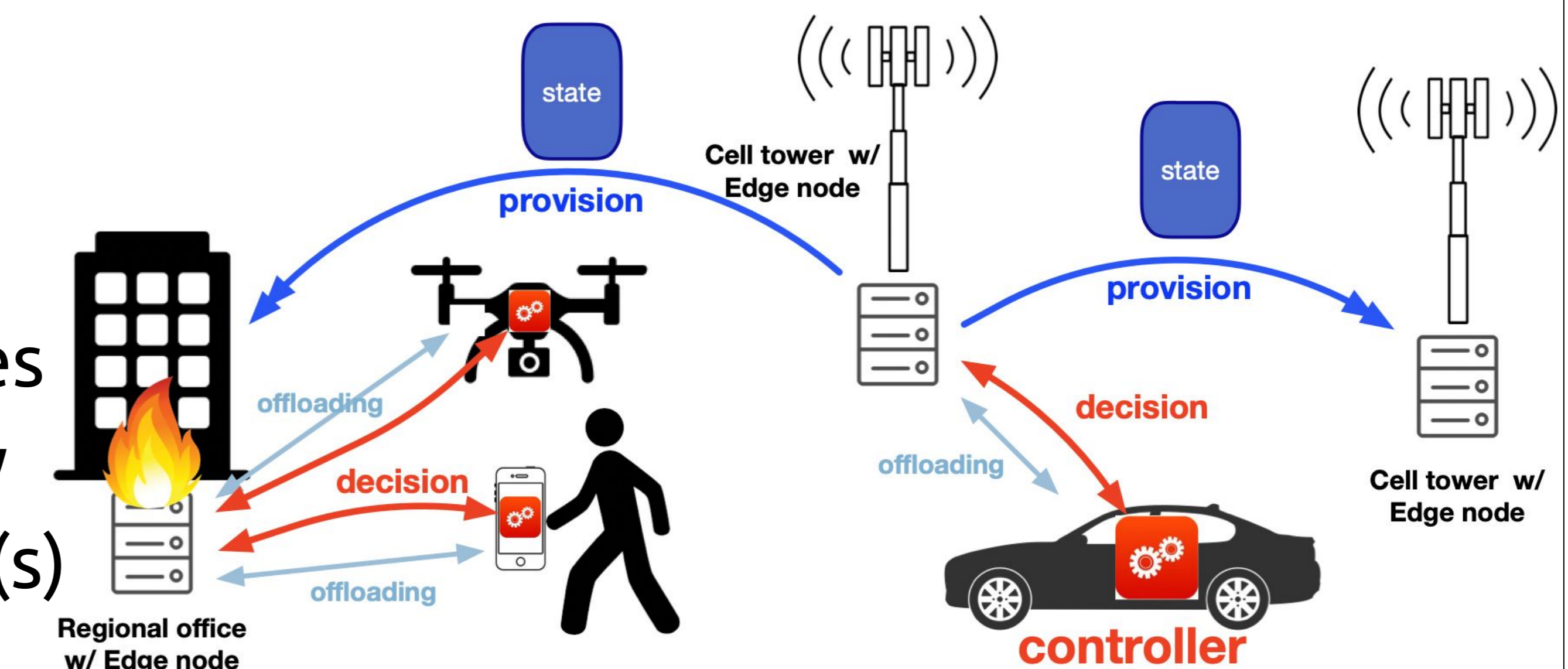
Preliminary Evaluation

- Augmented Reality (AR) navigation application hosted on simulated moving vehicle

- Simple proactive provisioning scheme initiated 10 mins prior, in comparison to just-in-time provisioning, reduces
  - median latency by 50ms
  - tail latency by 20%