Edge Computing Overview

Wei Gao

Course Information

- Class time: 2:30pm – 5:00pm Tuesdays
- Instructor: Wei Gao, weigao@pitt.edu
  - Office: 1205 Benedum
- Schedule and course materials will be posted at the instructor's website
  - http://www.pitt.edu/~weigao/ece2195/spring2022/schedule.html
- CourseWeb is used for posting announcements, grades and project feedback

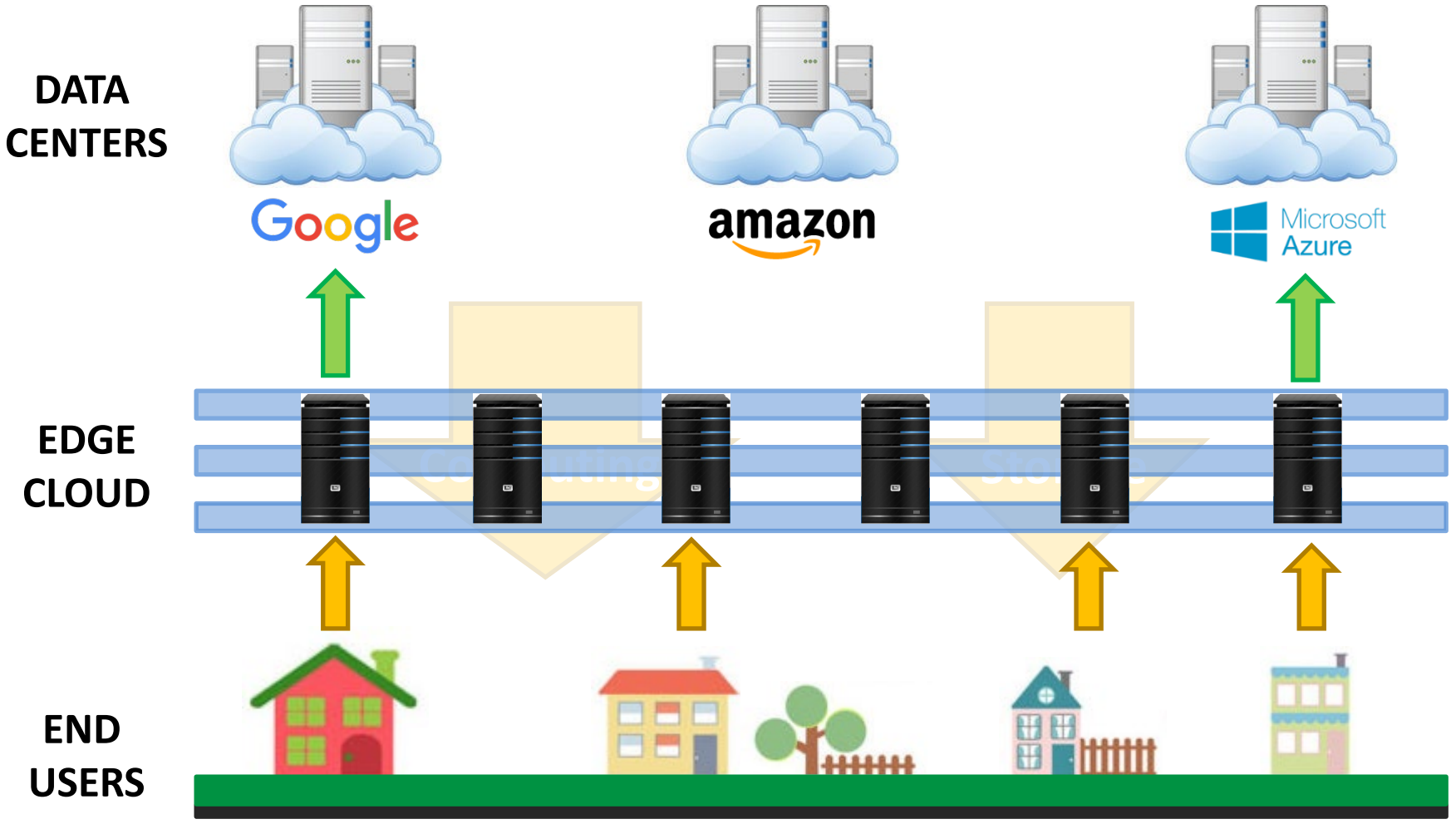
The limits of Cloud Computing

- Network communication latency of MCC
  - Can be up to 400 ms
  - Many mobile apps are delay-sensitive

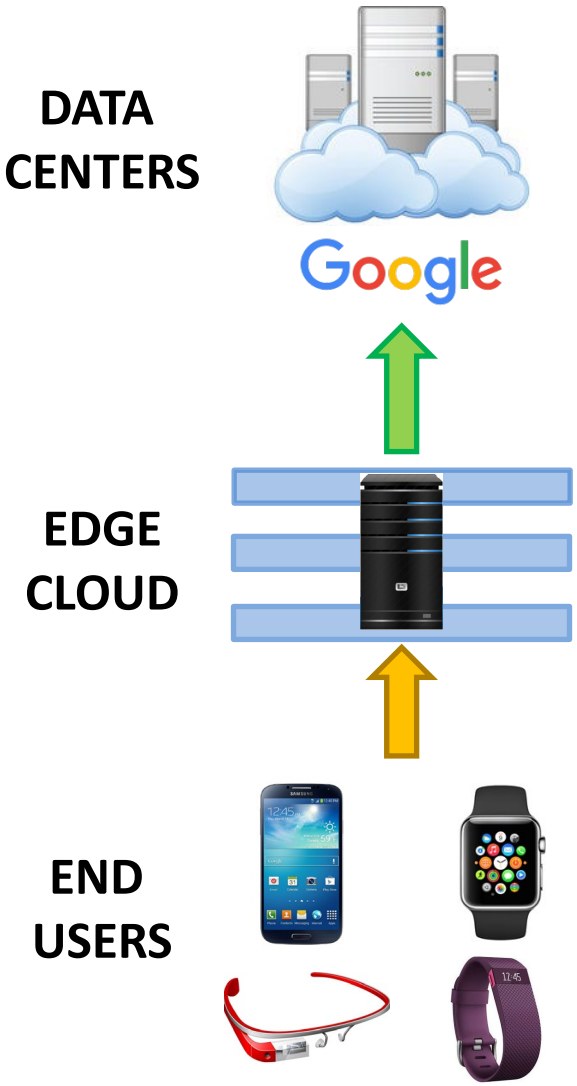
} Performance degrades!

Round trip cities	Max(ms)	Mean(ms)	Min(ms)
Berkeley-Canberra	174.0	174.7	176.0
Berkeley-Troudheim	197.0	197.0	197.0
Pittsburgh-Hong Kong	217.0	223.1	393.0
Pittsburgh-Seatle	83.0	83.9	84.0
Pittsburgh-Dublin	115.0	115.7	116.0

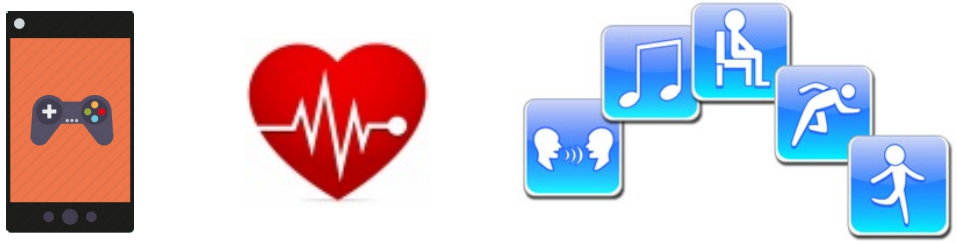
Edge Cloud


DATA CENTERS
Google
amazon
Microsoft Azure
EDGE CLOUD
END USERS

Edge Cloud for Mobile Computing



- Reduced response latency
  - Delay-sensitive mobile applications



- Higher efficiency of resource utilization
  - Distributed processing

Course Organization

- *Goal:* learn the basic concepts, methodology and technical solutions to edge computing
- Classroom lectures
  - Cover different aspects of edge computing research
    - Hardware support
    - Software systems
    - Edge computing and AI
    - Edge computing and 5G

Course Organization

- Research paper presentations (40%)
  - 2 papers presented in each class
  - 45-min presentation + 15-20 min discussions
  - The next week's presenters will be announced in the previous work's class
    - One-week time for preparation
- Individual course project (60%)
  - Your projects are expected to focus on one of the perspectives covered in course lectures
  - Keep your progress on track

Project Expectation

- Perform a system project
  - An innovative idea in emerging edge computing application paradigms
  - Could focus on either hardware or software
    - Hardware prototyping
    - Software programming, debugging and testing
    - Wide coverage of subsystems: computation, communication, sensing and control
  - Demo to the class

Find out A Project Topic

- Among those emerging edge computing application paradigms
  - How can they be better?

- Innovations from daily lives
  - Current technologies, tech news, Sci-Fi fictions and movies

Find out A Project Topic

Find out A Project Topic

Find out A Project Topic

- Imagine big, design small
  - Your design could be the key enabler of…

  - And in this course your prototype will be …

- What is an appropriate project scope?

Smart Building

- Occupancy sensing and monitoring
  - Camera, infrared, ultrasound, etc
  - Adjustment based on user needs
- Remote and intelligent control
  - Lighting, HVAC, sound
  - Custom and zonal control
- Information infrastructure
  - Ubiquitous display and feedback
  - Emergency evacuation
- Honeywell's vision: https://www.youtube.com/watch?v=kQ3CJdwP3fY

Smart Cities and Communities

- What is a smart city?
  - https://www.youtube.com/watch?v=vpSLICKnjPc
  - Public safety
    - Gunshot detection: https://www.youtube.com/watch?v=f8jkApBTGd4
  - City surveillance and planning
    - Traffic monitoring and control
    - Air quality and noise monitoring
    - Array of Things in Chicago: https://www.youtube.com/watch?v=pFL5QNwgs6A

Intelligent Transportation System

- Autonomous driving
  - Road sensing
    - Traffic detection, pedestrian detection
  - AI decision and control
    - Following and avoidance

- Communication
  - Vehicle to road side
  - Vehicle to vehicle
  - Toyota's vision: https://www.youtube.com/watch?v=uwLE3csyDAc

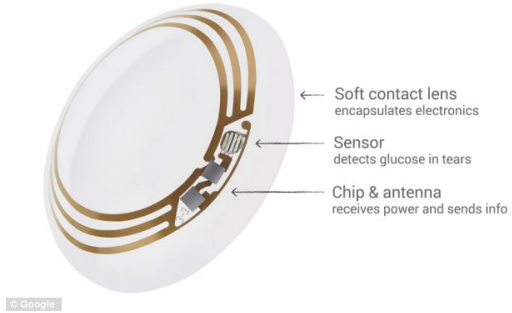
Virtual Reality

- Immersive experience
- Sensing is the key!
  - Headset
    - Gyroscope + accelerometer
    - Eye gaze tracking: https://www.youtube.com/watch?v=ImgfCFk8qy0
    - Emotion sensing: https://www.youtube.com/watch?v=2aXnfxH-anA
  - Hand controllers
    - Motion tracking with accelerometers
    - Google's Soli project: https://www.youtube.com/watch?v=0QNiZfSsPc0
  - More controllers…

Smart Health

- Digital fitness tracking

- Tele medicine and diagnostics



- Surgery assistance
  - Smart brain surgery system: https://www.youtube.com/watch?v=QOafVIkLgyk

Milestones

- Project proposal (10%): **Jan 25**
- 3 Interim milestones (10% each)
  - Once every three weeks
- Final presentation & report (20%): **Apr 26**

Project Proposal

- High-level overview of your project
  - *Project idea:* What are you doing in this project
  - *Project background:* Where do you start with?
  - *Project approach:* What are the major technical components / subsystems?
    - Better if you have some brief ideas about how to do them
  - *Project plan:* An outline of project plan
    - Final project objectives / deliverables
    - A list of project goals for each interim milestone

Project Proposal

- Grading criteria:
  - Design feasibility: 20%
  - Anticipated technical difficulty: 20%
  - Project planning: 40%
  - Presentation clarity: 20%

Interim Milestones

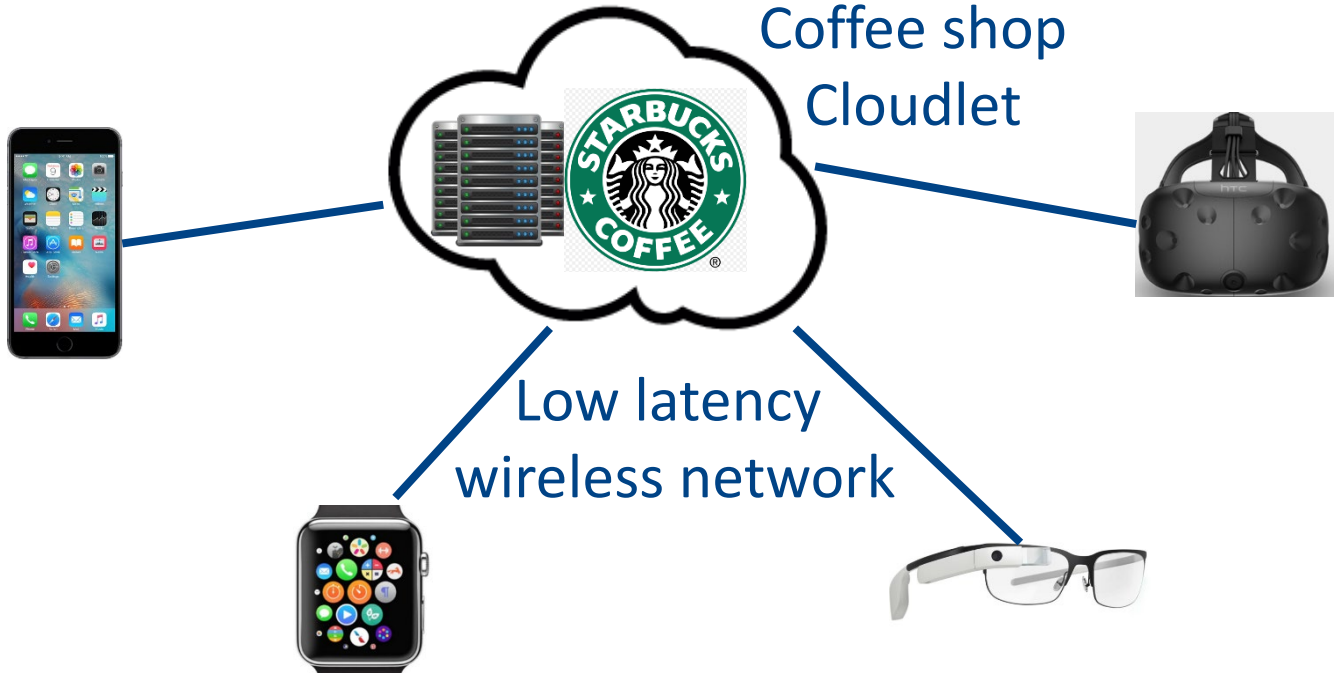
- Maintaining your progress on track!
  - Opportunities to received advices and revise your plan
- Grading criteria
  - Design feasibility: 25%
  - Technical difficulty and efforts: 25%
  - Development completeness: 30%
  - Presentation clarity: 20%

Project Final Report

- Recommended outline
  - Introduction
  - Related Work
  - Overview: motivation, problem formulation, basic idea
  - System design
  - Experimentation: your system setup, evaluation plan, experimental data
  - Discussions & conclusions

Current Solution: Cloudlet

- Small scale cloud servers at the edge
  - Reduce the network latency accessing data center
  - Support user mobility

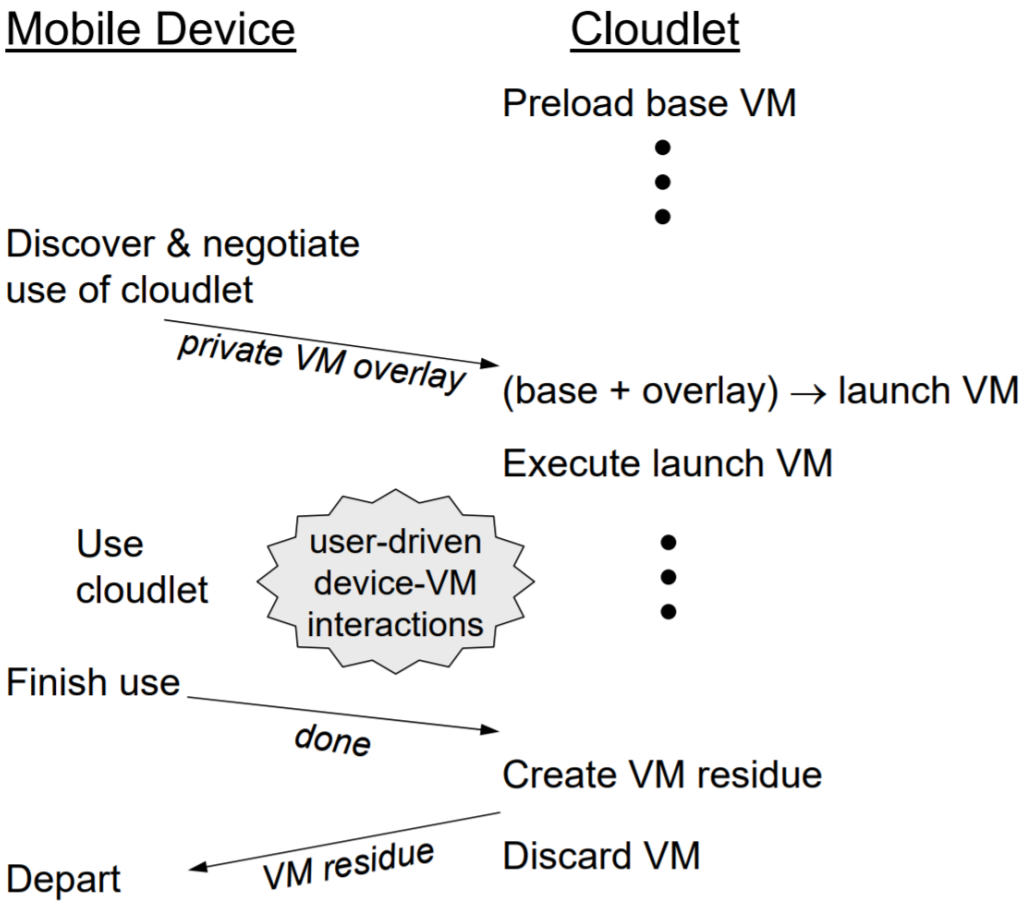


Advice

- Start early and work on it regularly!
- Discuss with me often for feedbacks and directions

System Implementation

- Each user application being submitted to the cloudlet is encapsulated into a Virtual Machine (VM)



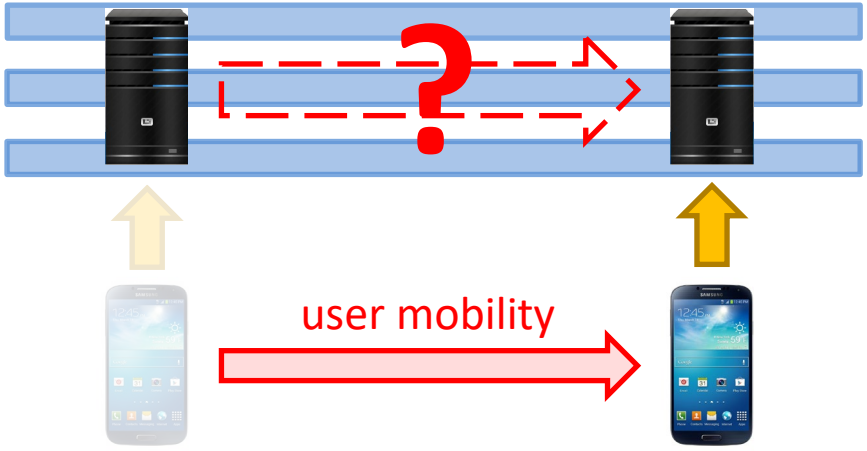
Challenges

- Adaptability
  - Optimized performance?
  - Minimized cost?

Provisioning for the peak load

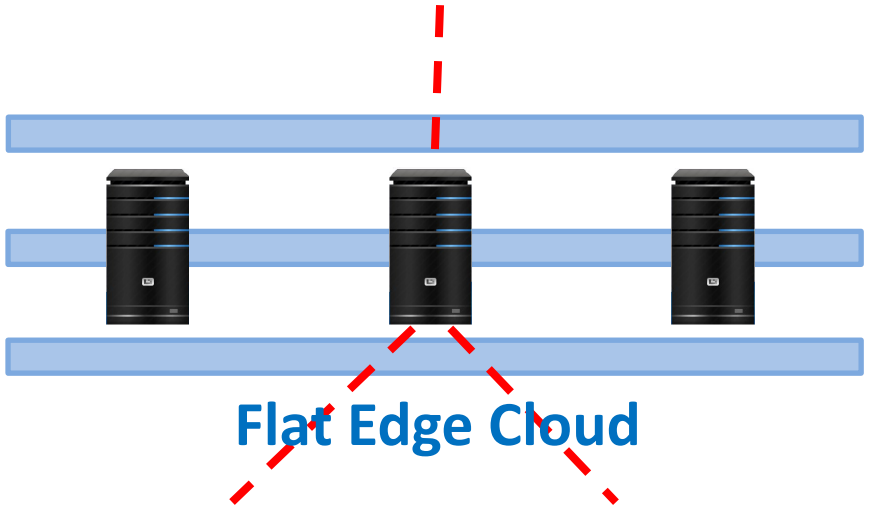
- User mobility
  - Minimized cost?

Complete move of data and program



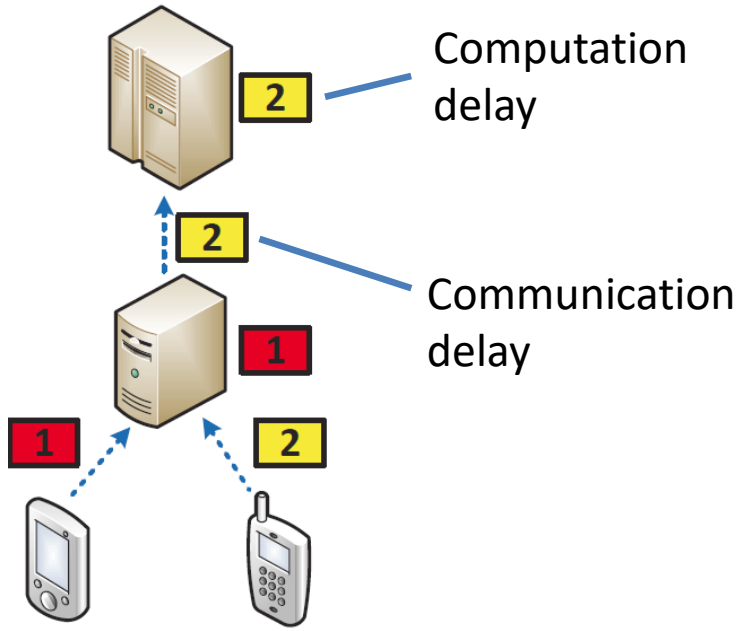
A Better Solution: Hierarchical Edge Cloud

- Adaptability
  - Aggregation of peak load
- User mobility
  - Partial migration of data and program



Optimal Workload Placement

- Our focus: minimized response latency
  - Where to place a workload
  - How much capacity for a workload
- Challenge
  - Delay tradeoff

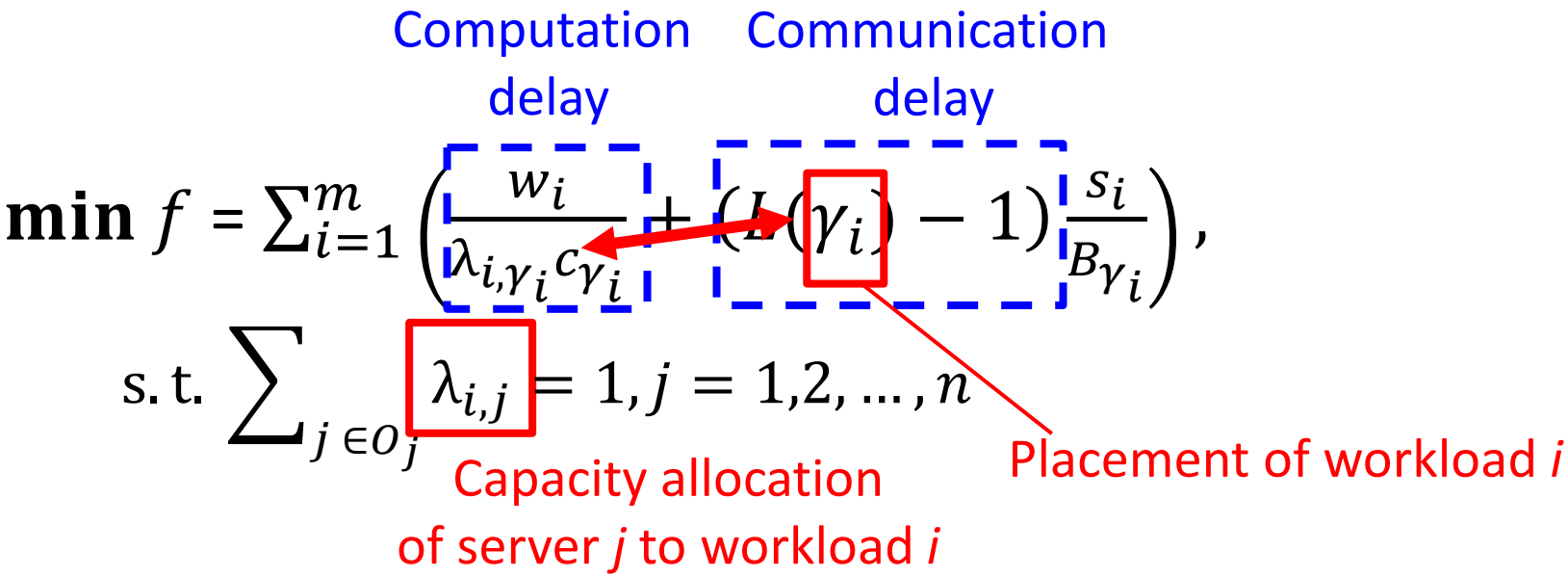


Response latency = Computation + Communication

High tiers

Optimal Workload Placement

- Distributed optimization

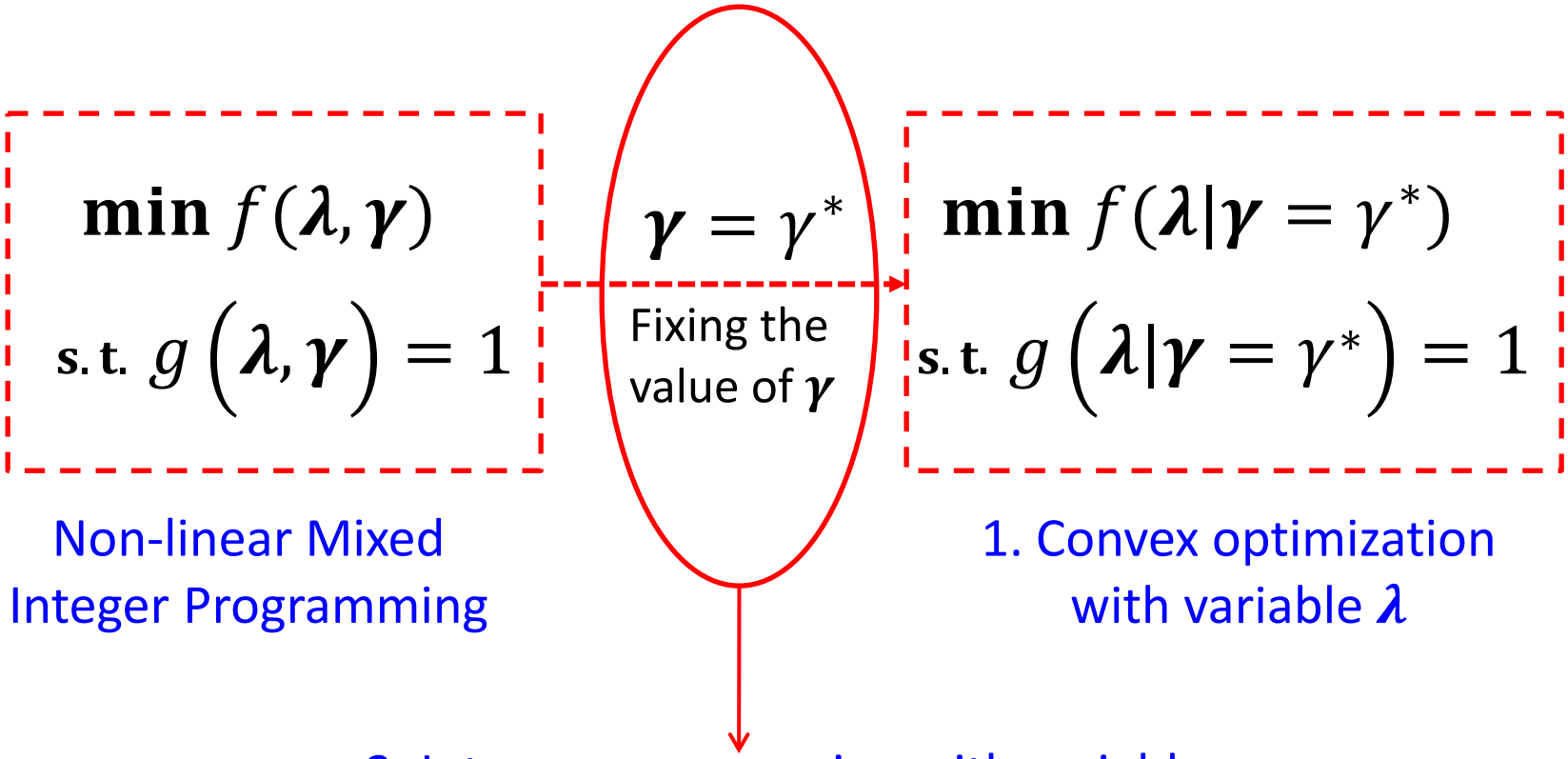


Non-linear mixed integer programming

?

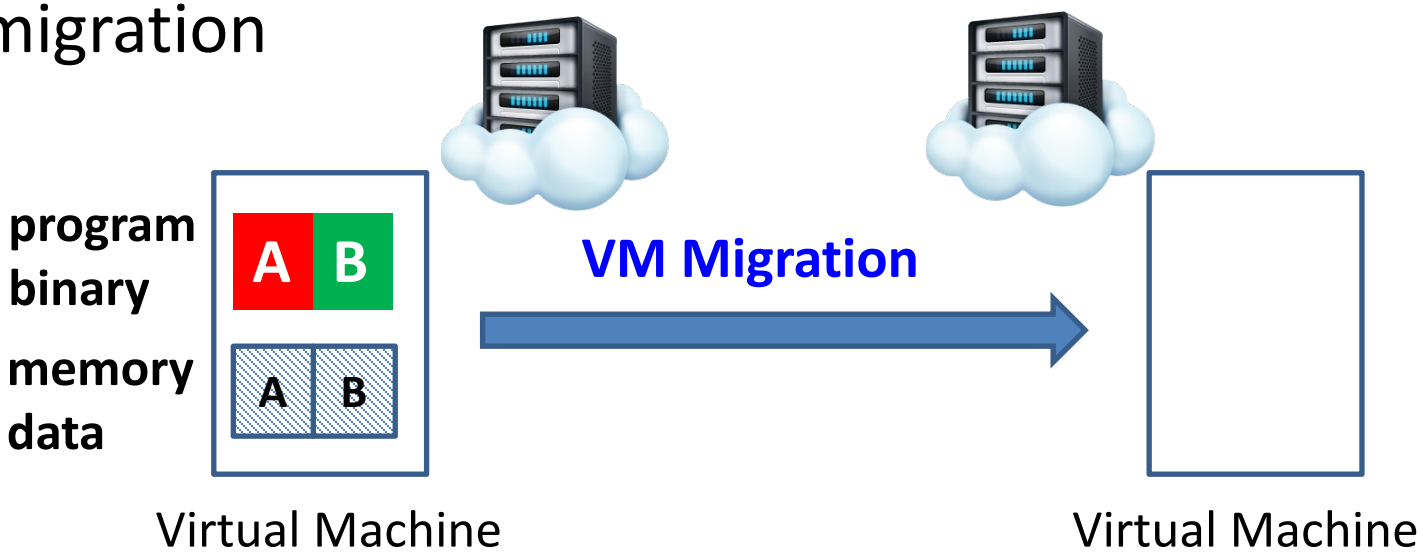
Optimal Workload Placement

- Problem transformation



Supporting User Mobility

- Remote program execution with least context migration



Supporting User Mobility