Panellist Summary

Overcoming the Edge Constraints with Scheduling Optimization between Edge and Cloud

Zhaobo Zhang, Futurewei Technologies, USA zzhang1@futurewei.com

- Intelligent Edge Computing
- Unified Orchestration between Edge and Cloud
- Scheduling Optimization

→ Fast-growing Internet sensor-based systems and services drive computing to the edge.

→ Edge computing can leverage cloud resources to overcome its resource constraints.

→ Dynamic scheduling among different locations optimizes performance and resource utilization.

Overcoming the Edge Constraints with Scheduling Optimization between Edge and Cloud

Zhaobo Zhang
Futurewei Technologies
zzhang1@futurewei.com
Nov. 2020

IARIA

FUTUREWEI Technologies

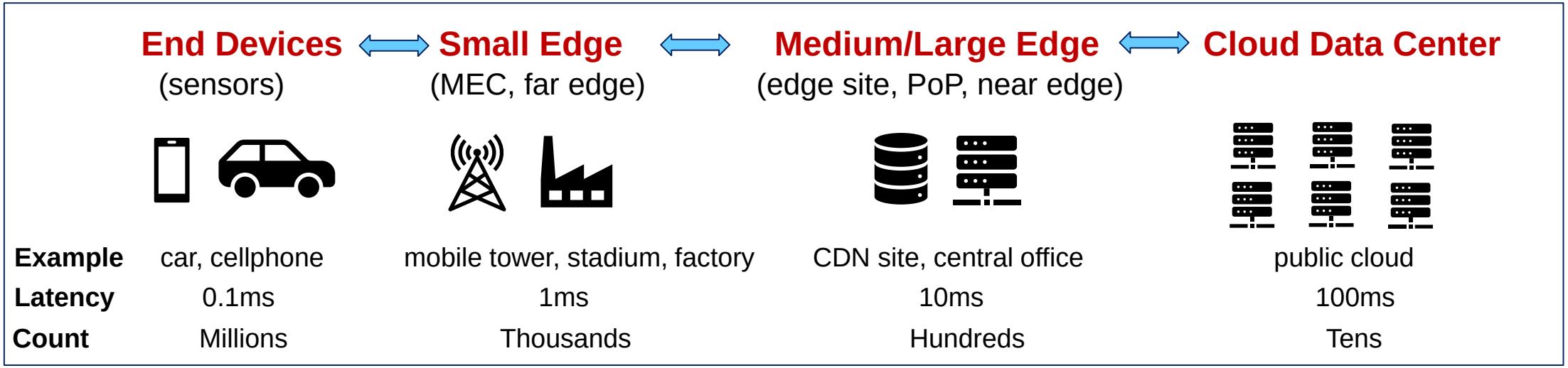
Booming of Internet Sensor-based Systems

- **Fast-growing Internet of Things (IoT) market**
  - Field: Industrial 4.0, Smart City (meter, lights, video), Smart farming (irrigation, animal tracking), Autonomous Vehicle
  - Industrial edge computing market: $11B 2020 → $30B 2025 *
- **5G enables massive machine type communication**
  - High device density: 1 million devices per km^2
  - Low power: up to 10-year battery life for low power IoT devices
- **Intelligent edge enables new set of opportunities**
  - Elastic infrastructure with edge computing and cloud computing
  - Distributed software applications

*https://iot-analytics.com/iot-edge-computing-what-it-is-and-how-it-is-becoming-more-intelligent/

Edge Computing Definition

- **Where is the edge?**



- **Edge Computing:** a distributed computing paradigm, brings computation and data storage closer to users, to improve response times and save bandwidth.

Edge Computing Challenges

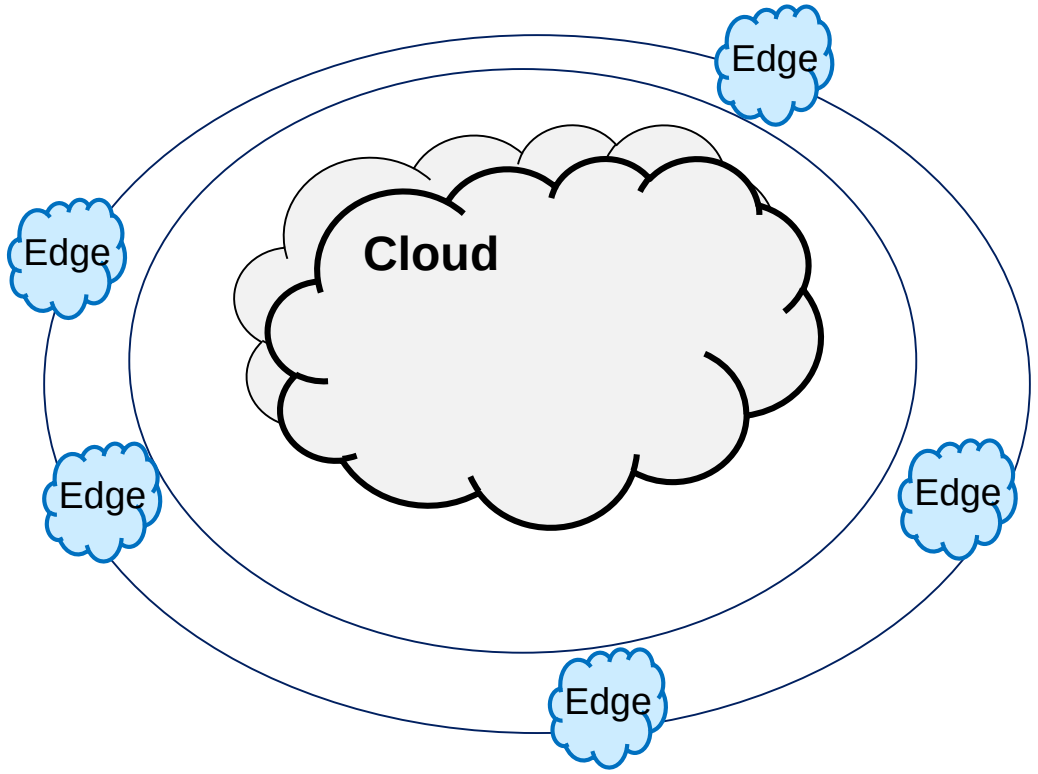
- Open architecture, various communication protocols
- Large volume of data (hard to transport back to cloud)
- Require real-time decision
- Dynamic scenarios, static rule-based solutions not enough
- Resource constraints, limited compute and storage resource
- Limited network connectivity

What operational models and technologies will be able to effectively unlock the edge potential?

Industry Movements

- **Cloud Providers**
  - Extend cloud to edge, with the same cloud services on at the edge, uniformly manage edge computing as part of a broader cloud footprint
  - AWS Outposts; Azure Arc, Stack Edge; Google Anthos
- **Challengers**
  - CDN (content delivery networks), colocation and telecom providers challenge from the edge with their global network locations and experience
  - Akami, Equinix, AT&T place edge sites near customers and offer cloud-like services

Intelligent Edge with Cloud Computing



- **Elastic Infrastructure**, workloads move seamlessly between Cloud and Edge, and Edge to Edge
- **Edge Intelligence**, data preprocessing and compression, time/location/context aware
- **Global Optimization**, scheduler, scaler, continuous resource allocation optimization
- **Resource-aware Application Design**, split into microservices based on the requirements of latency, computation complexity, state, etc.

Continuous Optimization, Federation and Collaboration

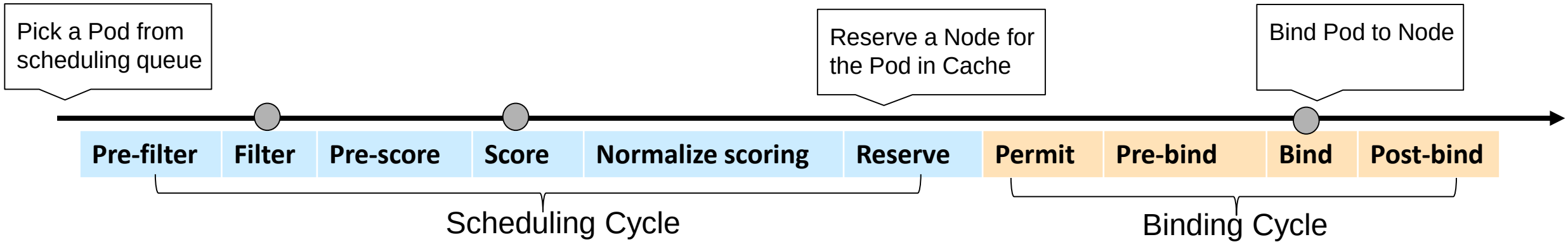
Scheduling Optimization between Edge and Cloud

Application Type	Schedule Requirements	Optimization Goal	Algorithms
• Scientific workflow • Large-scale data • Real-time • Cloud storage • Throughput intensive • Delay sensitive • Network services	• Workload dependency • Resource affinity • Resource quota • Latency, geo-proximity • Throughput	• Deployment cost • Service reliability • Resource utilization • Response time • Energy cost • Operation cost	• Decision tree • ILP (Integer Linear Programming) • Graph neural network • Greedy algorithm • Dynamic programming • Genetic algorithm • Particle swarm optimization

- Find a tradeoff between cost and performance
- Scheduled jobs could be dynamically rescheduled for optimization purpose

Kubernetes-based Scheduling Example

- Kubernetes is a popular container orchestration platform, considered as Cloud OS
- Scheduling framework: filter (by predicate) and score (by priority) with rich extension



- Scheduling Framework Optimization:
  - **Co-scheduling**, schedule a group of pods instead of one considering workload dependency
  - **Multi-profile scheduling**, combine different filter and score function for different workload
  - **Priority function extension**, add customized functions to be location/latency/energy aware
  - **De-scheduling**, periodically audit scheduling requirements, evict unsatisfied pods

Takeaways

- Edge is the next computing frontier after Cloud
- Edge computing strategies are critical for unleashing the full potential of Internet Sensor-based Systems and Services
- Global scheduling optimization can bring the best out of Edge and Cloud
- Resource-aware applications are easy to deploy and achieve better performance