*CS*06201*a*01: Network Computing and Efficient Algorithms

Lecture 14 Cloud-Edge-Terminal Cooperation

Xiang-Yang Li and Xiaohua Xu

School of Computer Science and Technology
University of Science and Technology of China (USTC)

August 31, 2021

The Challenges of Cloud Computing

- Massive data computing
- Emerging computing scenarios
- Real-time processing of small data

Cloud computing has some development bottlenecks that require new technologies to break through

The Edge Computing

- Cloud (Internet)
  - central processing of summary data
  - Big data analysis,complex learning model
  - Central control
- Edge (Intra-net)
  - Real-time data processing
  - Real-time control (M2M)
  - Local data filtering and caching
  - At source data visualization
- Sensors and controllers

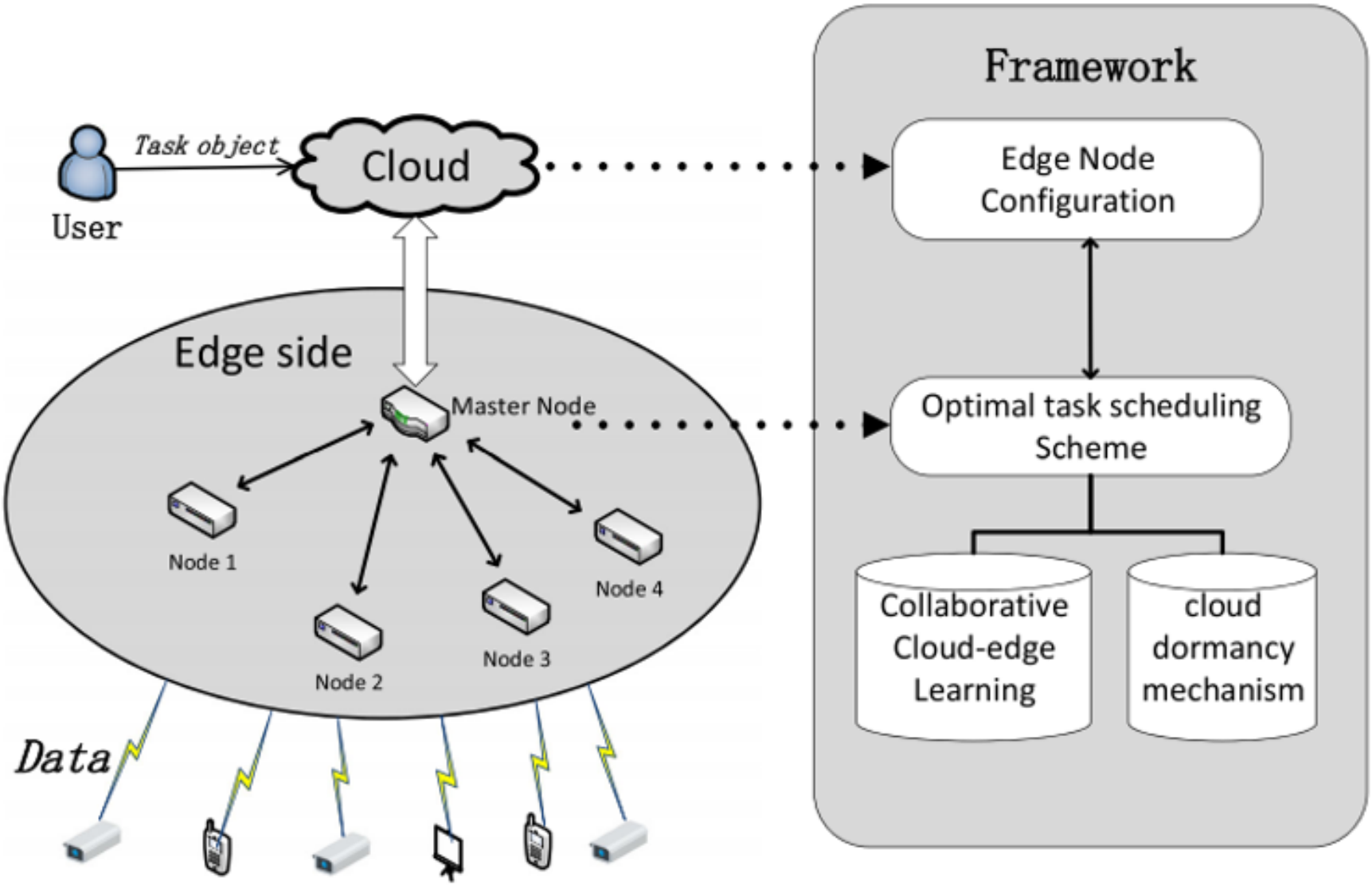
Benefits of Edge Computing

- Faster Response
  - Operating at the source of data
  - Faster response time for triggers
- Secure
  - locally stored
  - No theft during transport
  - Compliance maintained
- Cost Effective
  - No need to transport everything to cloud
  - No recurring cost
- Reliable Operations
  - can work without connectivity

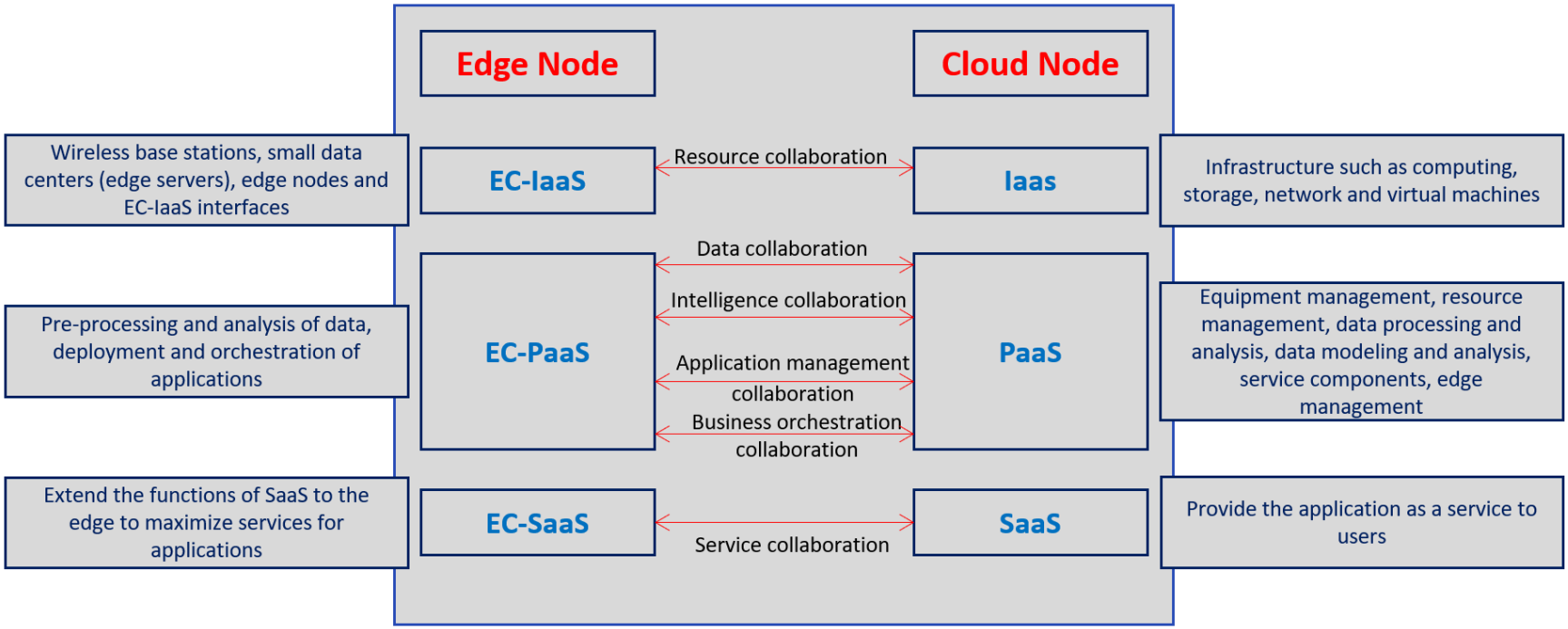
Four Edge Computing Challenges

- Data
  - Integration
  - Governance
  - Analytics
- Diversity
  - Use cases
  - Topology
  - Technologies
  - Standards
- Protection
  - Security
  - Privacy
  - Compliance
- Reliable Operations
  - Scale
  - Environmental
  - Remote Management
  - Autonomy

The Collaborative Cloud-Edge Computing Framework.



The Collaborative Cloud-Edge Computing Architecture



Cloud-edge Offloading Schematic

- The neural network architecture that is divided into two parts for deployment at the edge and the cloud as an example.
- Task scheduling schematic. How the adaptive task scheduling algorithm optimally schedules tasks onedge nodes

Service Collaboration

- Edge Service
  - ECS (Edge Computing Service)
  - QoS (Quality of Service)
- Service distribution strategy
  - Use the heuristic algorithm to solve the problem of edge service allocation.
  - MEC system that provides computing power for user services in a limited area.
  - Solve the best service allocation strategy through Markov, use the least resources to save energy consumption.
  - Implement different service distribution processing for applications with different resource requirements and delay sensitivity according to the load results of the entire task.

Resource Collaboration

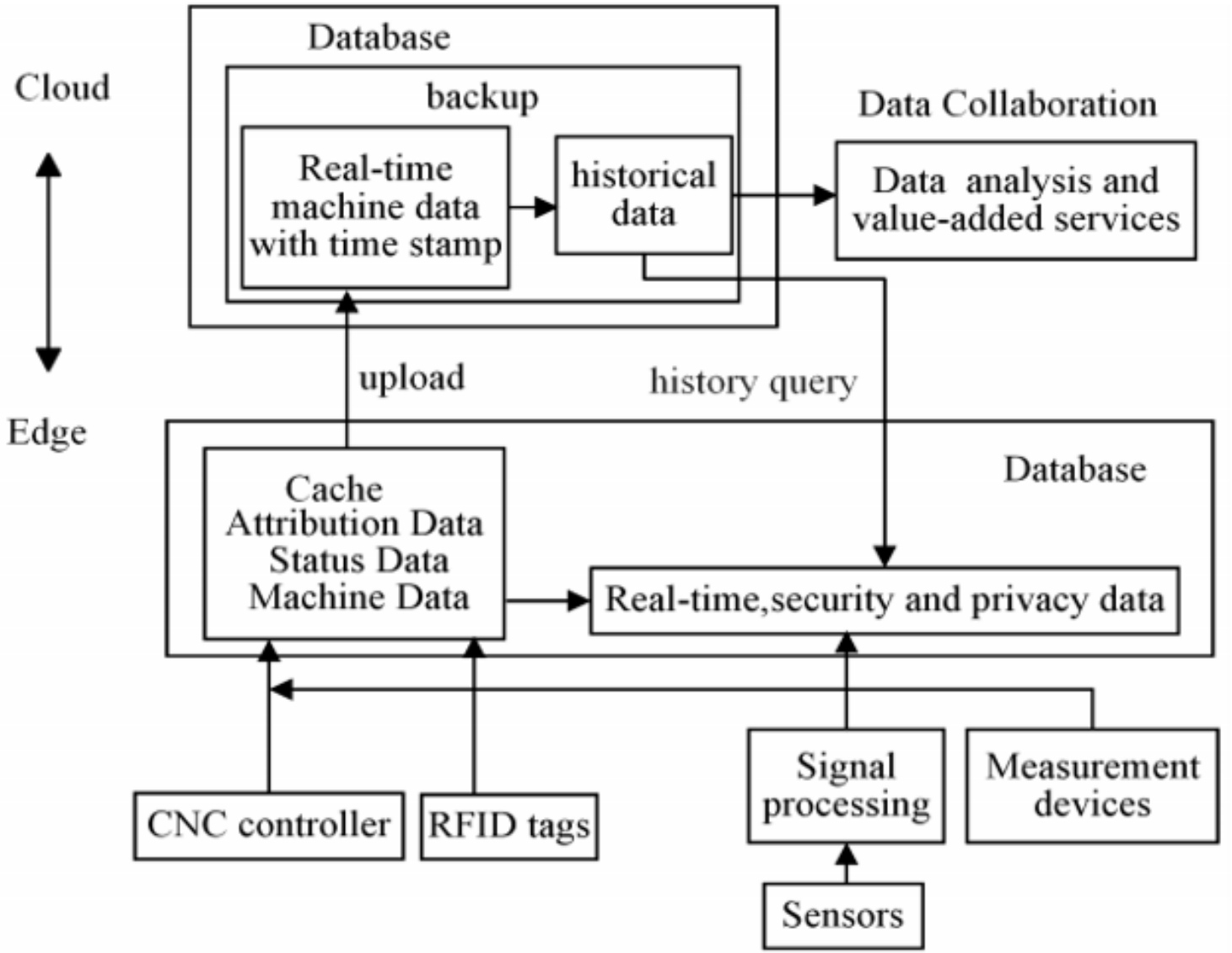
Computing Offload

- Four kinds of delay
  - Transmission delay of mobile devices
  - Computing latency of edge nodes
  - Transmission delay at edge nodes
  - Cloud server computing latency
- Resources are allocated to the i-th device serving the j-th edge node, how to compute the total delay of the i-th device served by the j-th edge node?

Resource management strategy

- Calculate the cost of renting the node based on the predicted workload and the price of the edge node, and then use the penalty function to constrain the resource to achieve the effect of reducing the cost of remote cloud computing and storage resources.

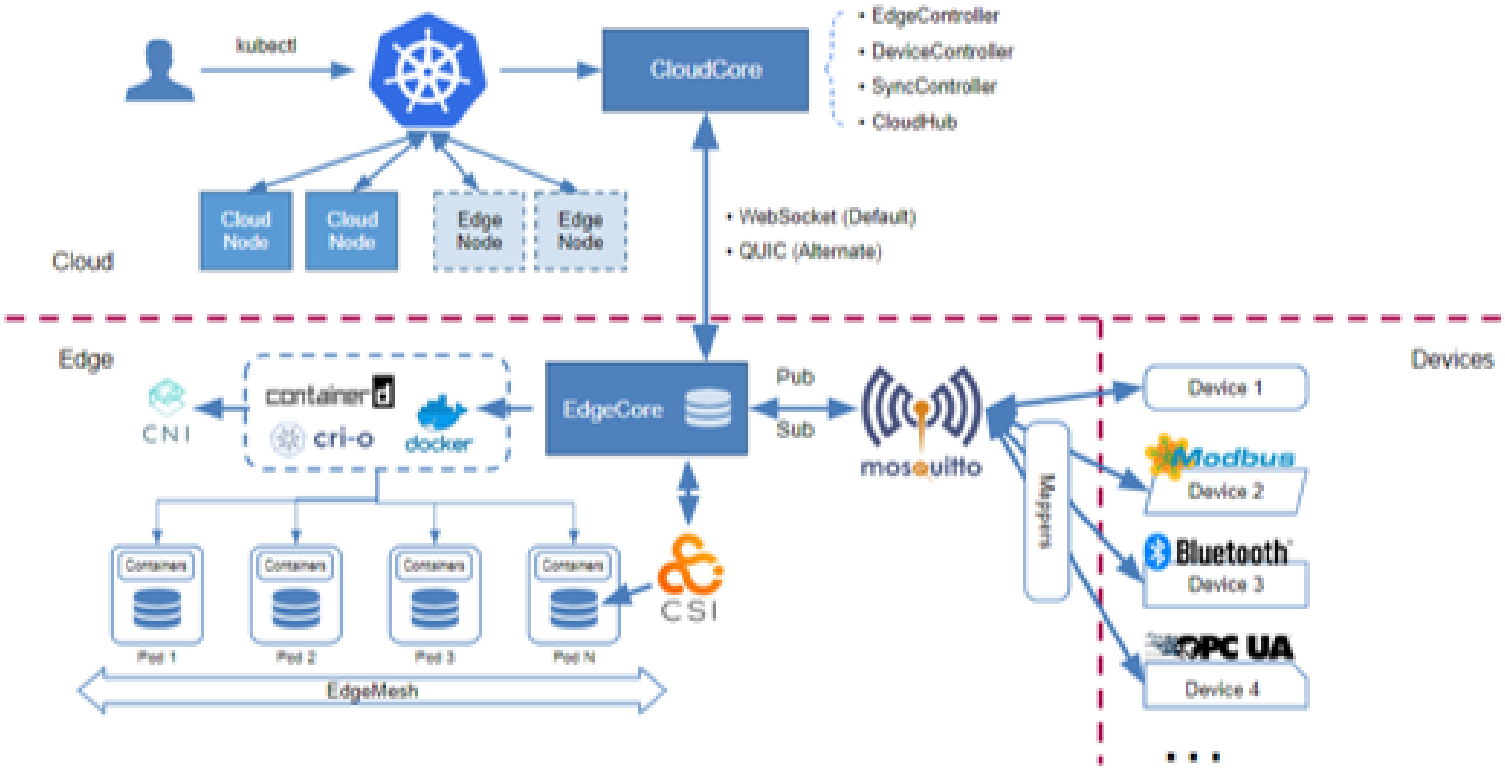
Data Collaboration



AI Collaboration

- Inference optimization technology
  - Model compression
  - Model segmentation technology and early exit mechanism
- Model training optimization
  - Improve model training efficiency
  - minimize errors
- Federated learning
  - Distributed training
  - Data privacy protection

Manage Collaboration



Choreography Collaboration

- Instance of application
  - Modular decomposition
  - Deploy a large number of microservice instances
- Business Orchestration
  - MiCADO, a dynamic Choreography framework based on microservice applications