# Unified Optical Networks and AI Computing Orchestration (UONACO)

**CCAMP WG, IETF124**

**draft-tan-ccamp-uonaco-problem-statement-00**

**draft-hu-ccamp-uonaco-control-framework-00**

**Author:**

Qiaojun Hu (Beijing University of Posts and Telecommunications)

Yanxia Tan (ChinaUnicom)

Yanlei Zheng (ChinaUnicom)

Wei Wang (Beijing University of Posts and Telecommunications)

Yongli Zhao (Beijing University of Posts and Telecommunications)

Jie Zhang (Beijing University of Posts and Telecommunications)

# Motivation

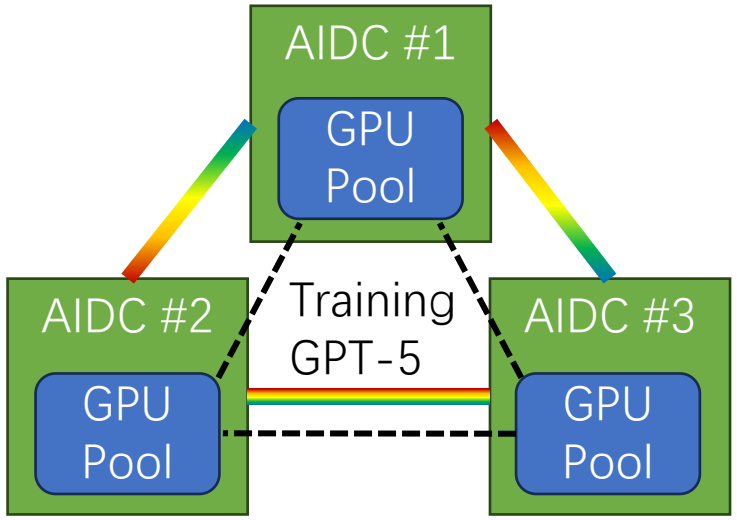
- Coordination between networks and computational resources is a hotpot within the IETF (e.g., CATS and SRv6)
- Key drivers for optical network and AI computing collaboration
  - Latency reduction
  - Bandwidth enhancement
  - Isolation characteristics
- CCAMP WG possesses a solid foundation in optical network control and management, facilitating the process of subsequent works

# Concept

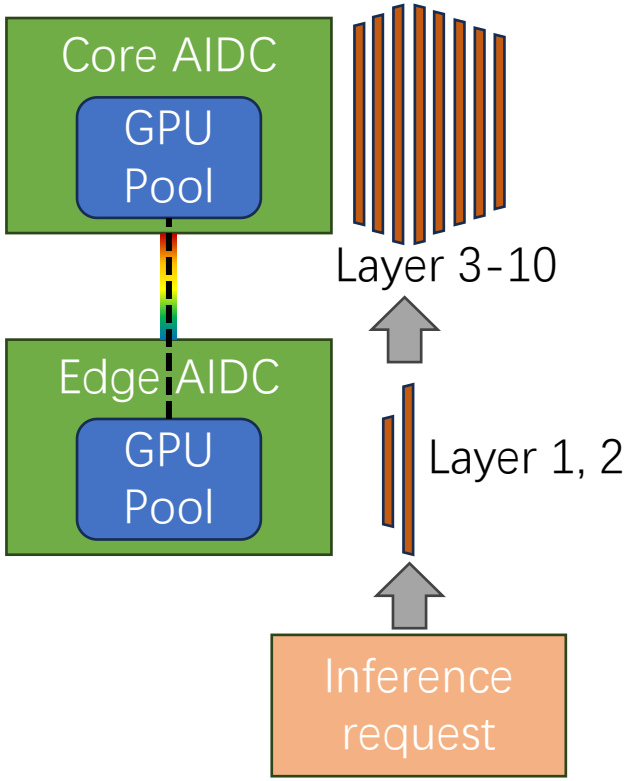
- Unified Optical Networks and AI Computing Orchestration (UONACO) is a framework that enables bidirectional awareness, resource abstraction, and joint orchestration across the compute-optical boundary, resolving the critical isolation between optical transport networks and AI computing infrastructure to optimize compute efficiency and network resource utilization

# Use Cases

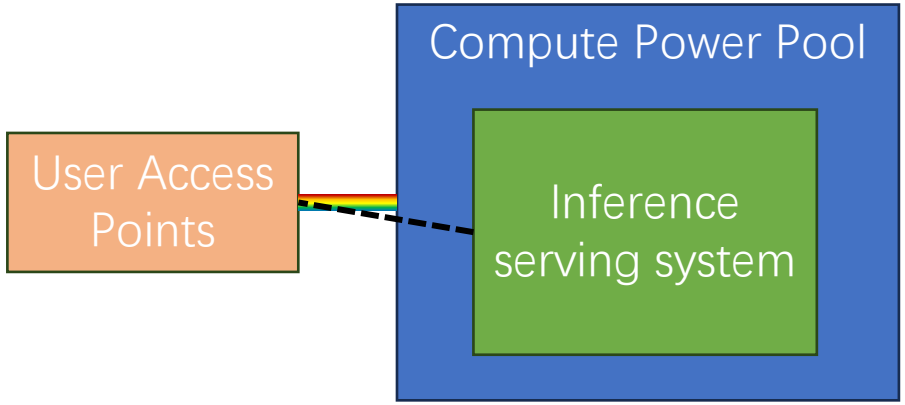
- Distributed AI Training




- Distributed AI Inference



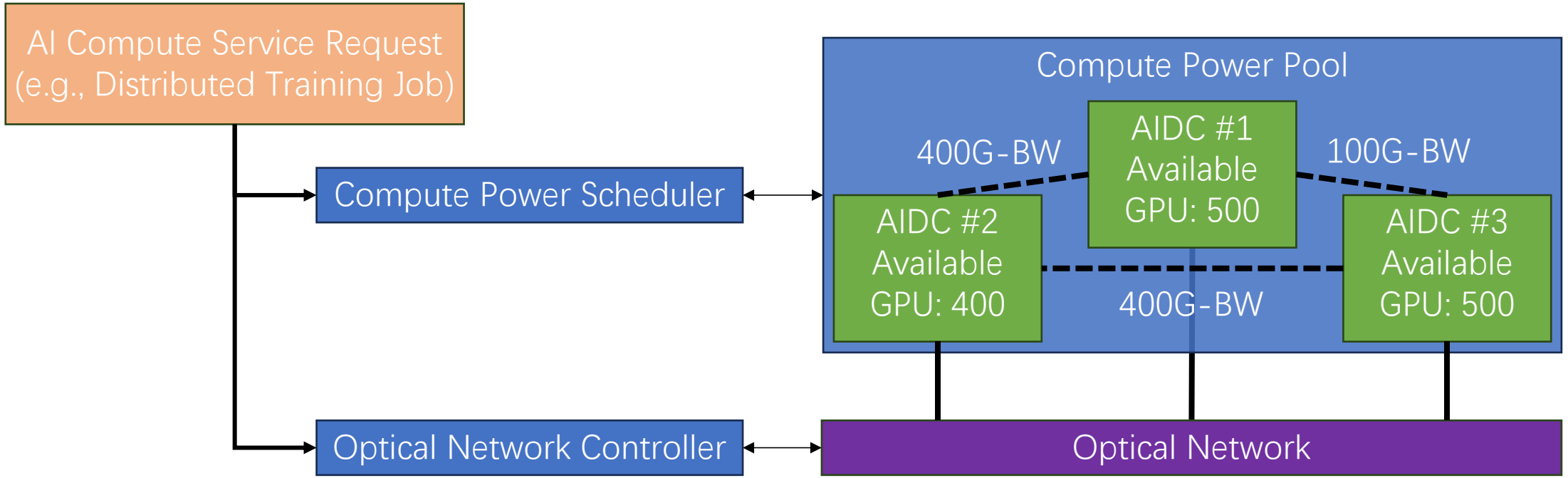

- Accessing Remote AI Service



# Problem Statement

- Isolated Control and Management
- Independent Resource Efficiency Evaluation

**Suboptimal Resource Allocation**

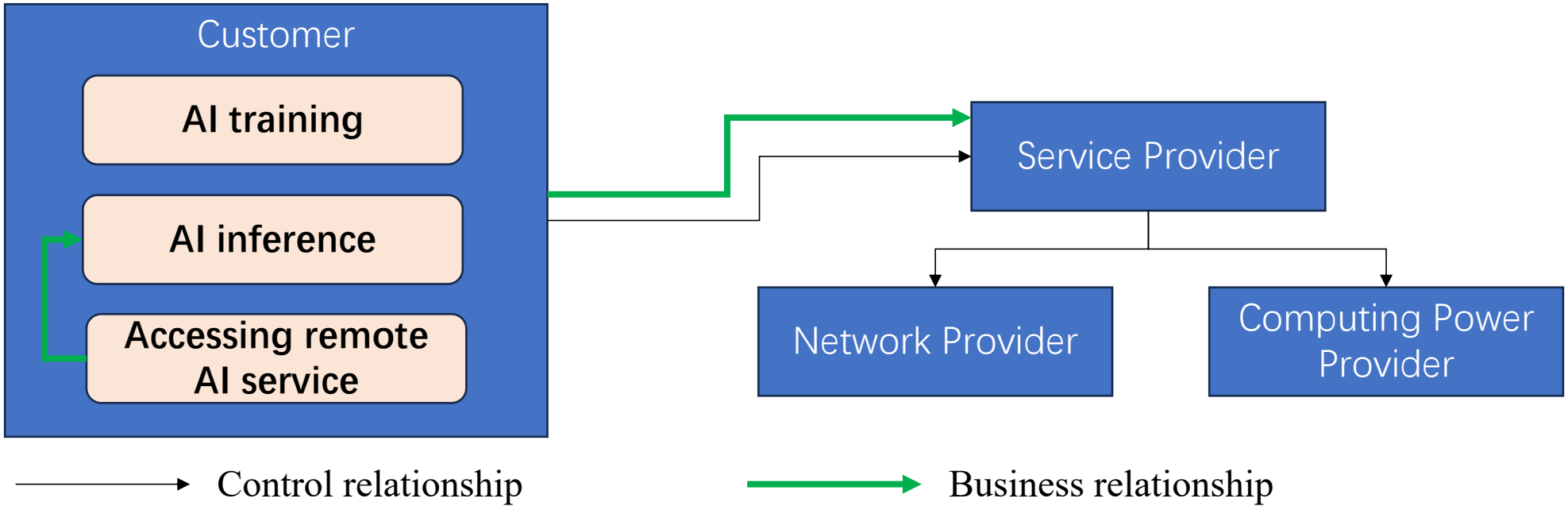


# Requirements

- Integrated Control and Management Architecture
  - An architecture facilitates bidirectional fusion between the compute and network control planes
- Unified Abstraction and Evaluation Framework
  - A common language is needed for both domains to understand each other's capabilities and state
- Scheduling Algorithms for Joint Orchestration
  - Some domain-specific end-to-end resource allocation algorithms

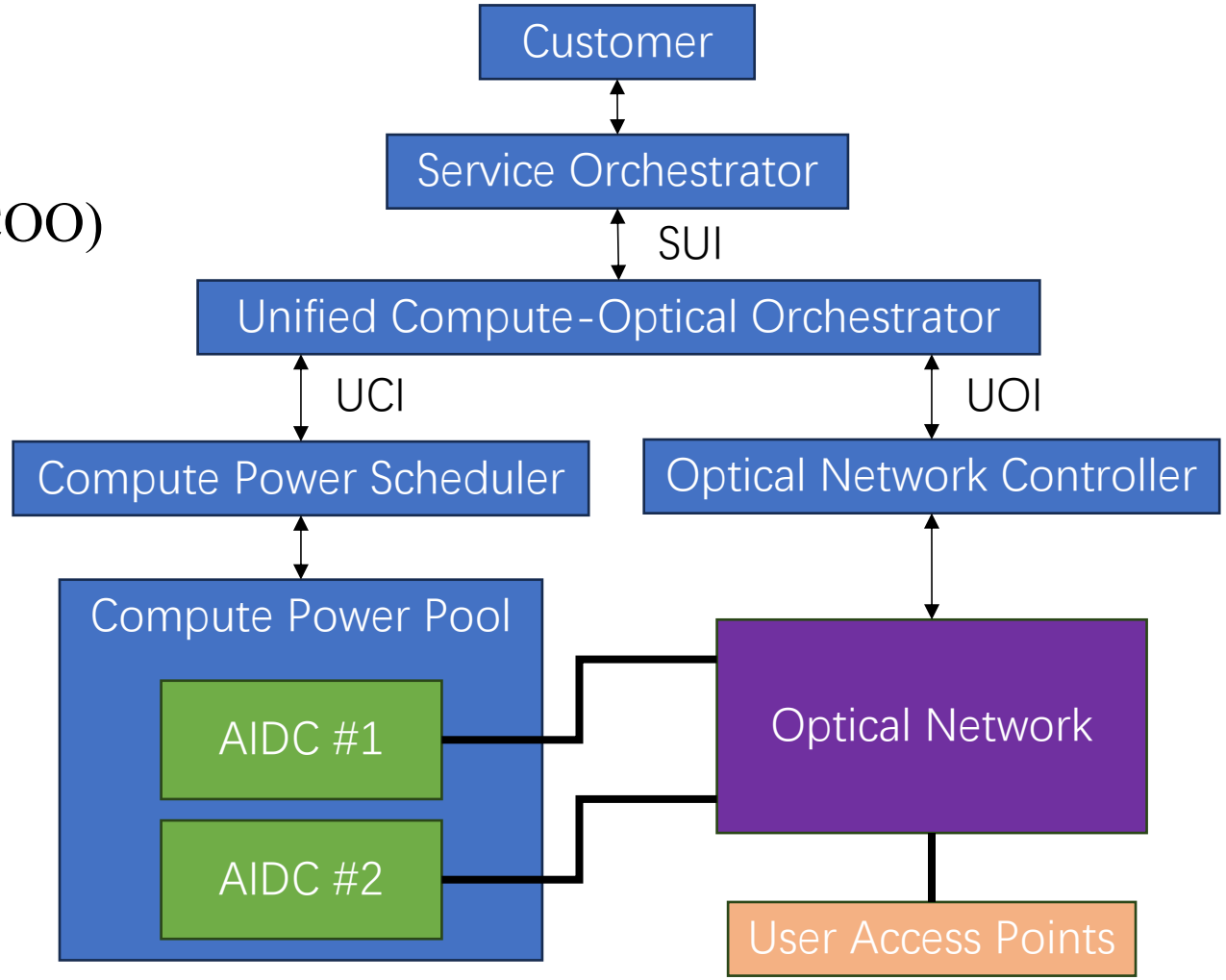
# Service Model

- Customer
  - AI training: specify model, deadline, performance, scale, and data privacy requirements
  - AI inference: lease computing resources to deploy and operate inference models
  - Accessing remote AI service: invoke pre-deployed inference APIs offered by third parties



# UONACO Control and Management Architecture

- Service Orchestrator (SO)
- Unified Compute-Optical Orchestrator (UCOO)
- Optical Network Controller (ONC)
- Computing Power Scheduler (CPS)
- UONACO Interfaces
  - SUI (SO-UCOO Interface)
  - UOI (UCOO-ONC Interface)
  - UCI (UCOO-CPS Interface)



# Next Step

- First draft (version 00)--Open for feedback

- Next steps:
  - Improve the content of the draft upon comments and suggestions
  - UONACO Interfaces and Yang Data Model Design
  - Unified Resource Abstraction and Evaluation Framework

# Thank You