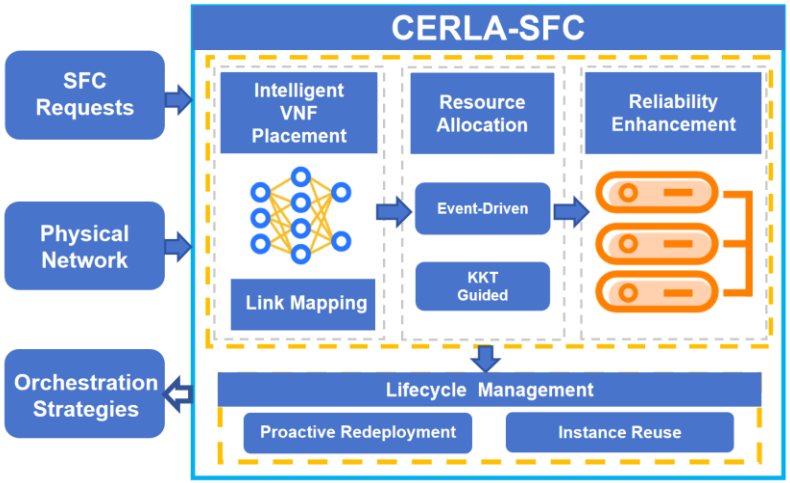
CERLA-SFC: Hierarchical Orchestration of Cost-Efficient, Reliable and Low-Latency SFCs in Multi-Access Edge Computing

Yuanfei Xiao, Zhenli He, Xiaolong Zhai, Junjie Wu, Libo Feng, Cheng Xie, Keqin Li

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50964-2

Problems & Ideas

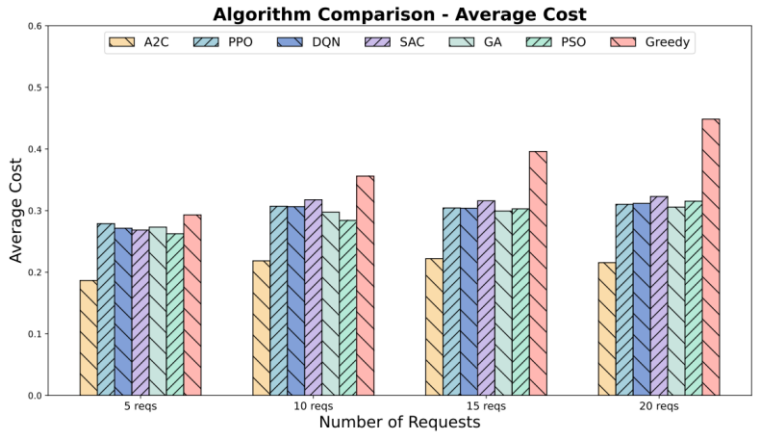
- Problems of existing SFC orchestration approaches:
  - Lifecycle control is fragmented.
  - The non-linear interplay among delay, reliability, and cost is often simplified.
- A unified, hierarchical orchestration framework that integrates learning-based placement, topology-aware routing, event-driven resource slicing, and selective redundancy within a single, closed control loop.

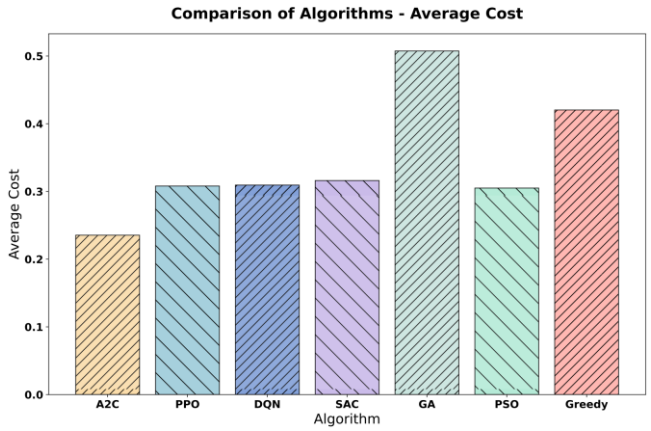


The CERLA-SFC control pipeline. It unifies an A2C-based placer, a KKT-based resource slicer, and a redundancy manager in a single feedback loop. Dashed boxes mark the algorithm stages. Solid arrows indicate algorithm execution steps; hollow arrows indicate feedback returned for online adjustment.

Main Contributions

- Contributions:
  - We propose a tightly-coupled control loop that jointly optimizes VNF placement, routing, and resource slicing, explicitly modeling latency-reliability-cost trade-offs;
  - Our hybrid approach combines an A2C agent for intelligent VNF placement with KKT-based real-time resource tuning, cutting OPEX while ensuring latency targets;
  - A selective redundancy mechanism duplicates only critical VNFs and enables instance sharing across chains, boosting reliability with marginal cost increase.





CERLA-SFC reduces OPEX by 10%–53% while maintaining millisecond latency and near-zero violations, as proven in a 21-node edge-cloud testbed. The left and right figures below show the cost comparison with baseline methods under static and dynamic loads, respectively, under urgent latency constraints.