Optimized Service Function Chain Deployment Design in Mobile Edge Computing Networks (Postprint)

Authors: Chen Jialiang, Wang Feng, Zhang Xiao

Date: 2022-05-18T00:00:00+00:00

Abstract

To address the challenge of excessive system cost expenditure and long response times in user-oriented service function chain deployment algorithms for mobile edge computing networks, this paper proposes a joint optimization design method for service function chain cost and latency. First, during the service function chain construction phase, based on node positions and resource conditions, the current optimal nodes are selected to reduce latency between virtualized network functions (VNFs) and thereby improve the response time of the service function chain. Second, during the service function chain deployment phase, due to limited mapping resources, a node selection algorithm is applied when mapping service nodes to determine the optimal node mapping sequence, and the shortest weighted path is selected as the communication link between VNFs. Experimental simulation results demonstrate that, compared with existing schemes, the proposed method can effectively reduce latency and deployment costs, and significantly improve the success rate of service function chain deployment.

Full Text

Preamble

Vol. 39 No. 10
Application Research of Computers
ChinaXiv Cooperative Journal

Optimization Design of Service Function Chain Deployment in Mobile Edge Computing Networks

Chen Jialiang[1], Wang Feng[1]†, Zhang Xiao[2]
(1. School of Information Engineering, Guangdong University of Technology,

Guangzhou 510006, China;
2. College of Computer Science, South-Central Minzu University, Wuhan 430074, China)

Abstract: To address the challenges of excessive system cost and long response times in user-oriented service function chain (SFC) deployment algorithms for mobile edge computing networks, this paper proposes a joint optimization method for SFC cost and delay. First, during the SFC construction phase, the algorithm selects optimal nodes based on node location and resource status to reduce delays between virtualized network functions (VNFs) and improve SFC response time. Second, during the SFC deployment phase, due to limited mapping resources, a node selection algorithm is applied to determine the optimal node mapping sequence, and the shortest weighted path is selected as the communication link between VNFs. Experimental simulation results demonstrate that compared with existing solutions, the proposed method can effectively reduce delay and deployment cost while significantly improving the success rate of SFC deployment.

Keywords: network function virtualization; virtual network function chain composition; service function chain; operational expenditure; reliability

---

0 Introduction

In the evolution of modern networks, Network Function Virtualization (NFV) and Software-Defined Networking (SDN) are playing increasingly important roles. Currently, network operators utilize numerous specialized network middleware devices to provide various network functions, such as deep packet inspection, firewalls, and load balancers. However, the use of these dedicated network middleware devices faces numerous challenges, including high upgrade costs, lack of framework flexibility, and poor manageability and scalability, making it difficult to meet future network development requirements [?]. Unlike traditional networks that implement functions through dedicated hardware, NFV is a key technology that decouples hardware and software [?], enabling hardware functions in the form of Service Function Chains (SFC) [?]. The goal of NFV is to leverage virtualization technology to implement network functions on general-purpose servers or virtual machines [?, ?, ?], offering excellent flexibility and scalability.

In particular, mobile edge computing networks, where computing and storage resources are located close to user equipment (e.g., at the radio access network side), can effectively reduce network access and computing service latency [?]. However, due to the strong coupling of CPU, storage, and wireless resources, NFV in mobile edge computing networks presents greater challenges. On one hand, edge network nodes are highly heterogeneous, exhibiting significant differences in communication characteristics, computing capabilities, and caching capacities. On the other hand, the number and types of mobile services are growing rapidly, with varying Quality of Service (QoS) requirements (e.g., delay, throughput, and reliability) and distinct network resource demands. The NFV architecture consists of two layers: a physical network layer composed of underlying hardware infrastructure, and a virtual network layer composed of VNFs. A physical network can support multiple virtual networks to enable efficient resource utilization [?, ?]. Serving business requests requires completing the deployment of corresponding SFCs, i.e., designing VNF mapping and routing schemes on network infrastructure under SFC system performance criteria [?]. The SFC deployment process comprises VNF deployment and virtual link mapping.

How to efficiently map and place SFCs onto underlying hardware networks is a widely studied research topic. Current mainstream research focuses on minimizing cost overhead and maximizing network revenue. Reference [?] proposes deploying VNFs using random node sharing in cross-layer network scenarios to average resource allocation, but fails to consider load balancing, resulting in low resource utilization. Reference [?] presents a sub-chain coordinated protection model for reliability-guaranteed SFC deployment, considering the reliability of both VNFs and links, and proposes a heuristic algorithm to achieve lower deployment costs. Reference [?] employs the PageRank algorithm to rank nodes for global optimization of delay and reliability, but does not consider node delay. Reference [?] proposes a high-efficiency VNF placement and chaining algorithm based on centrality points, aiming to minimize provider costs by finding the optimal number and geographic locations of VNFs. Reference [?] uses Integer Linear Programming (ILP) modeling with tools like CPLEX, optimizing for minimal operational expenditure and resource fragmentation, including VNF deployment costs, energy consumption, and traffic overhead. Reference [?] proposes a resource-aware collaborative construction and mapping method for SFCs, utilizing an improved genetic particle swarm algorithm to reduce resource overhead costs during deployment. Reference [?] studies delay-constrained minimum-cost unicast SFC design but does not share nodes. Reference [?] investigates a cost-aware strategy for coordinating SFCs across multiple data centers under deployment center load constraints, establishing an ILP model and proposing a cost-aware SFC orchestration heuristic algorithm to reduce total deployment costs.

Due to resource constraints or request location limitations, VNFs comprising an SFC may be distributed across the entire network, particularly in mobile edge networks where edge networks are geographically divided into small edge cloud local area networks. Geographic distances exist between edge clouds and between core clouds and edge clouds. When SFCs are dispersed across core clouds and edge networks, the time cost of data transmission between geographically separated VNF instances is high [?]. Therefore, geographic location factors are critical issues that must be considered when deploying SFCs in edge networks. Moreover, the aforementioned research lacks studies on SFC deployment in mobile edge computing networks based on trade-off criteria between delay and resource costs.

To address these problems and enable more efficient mapping and deployment of SFCs in mobile edge computing networks, it is necessary to fully consider coordinated resource allocation among different VNFs, improve resource utilization, and reduce network deployment costs. To this end, this paper proposes an efficient and reliable SFC deployment method for mobile edge computing networks in hybrid wired and wireless network scenarios. First, through a geographic location-aware node sharing algorithm, the method determines whether VNF positions during deployment satisfy node sharing conditions. Second, using a resource-aware node selection algorithm, nodes are selected based on comprehensive scoring rankings to map VNFs onto physical nodes, reducing delay and improving reliability. Finally, based on a minimum-bandwidth virtual link mapping scheme, the final SFC mapping is completed.

The main innovations of this paper are summarized as follows:

a) For mobile edge network scenarios, this paper proposes an SFC optimization model and solution supporting wireless functionality, optimizing resource allocation for VNFs at different geographic locations to minimize operator costs while guaranteeing user QoS (e.g., delay, reliability). First, under mobile edge network resource constraints, the paper models user service request delays and VNF resource constraints to establish an SFC system overhead cost model. Second, reliability is a crucial performance metric for SFC systems. Due to architectural design and other reasons, edge network devices generally lack disaster recovery backup and other redundancy measures, resulting in much higher failure rates than core clouds. This paper introduces reliability as a constraint to balance reliability with response time and other objective requirements for rational SFC deployment.

b) This paper designs an SFC underlying network resource-aware node selection algorithm. For SFC service requests initiated at the current time, the algorithm ensures that SFC service requests satisfy node sharing conditions. When merging VNFs, it ensures that current physical node resource constraints are met. For VNF geographic layout and routing path planning, the algorithm optimizes current resource allocation, screens shareable physical node sets, designs virtual link mapping algorithms, and reduces SFC link delay.

---

1 Network Model and Problem Description

[Figure 1: see original paper] illustrates the system model of this paper, comprising a virtual logical service demand layer and a physical network layer. The virtual logical service demand layer provides SFC requirements, where SFC requests from source endpoints to destination endpoints consist of different functional VNFs chained together. Based on their resource requirements, VNFs are mapped to the physical network layer using VNF mapping algorithms. The physical network layer provides hardware resources, including computing hardware, storage hardware, and network hardware. The SFC deployment optimization scheme in this paper collects SFC requirements and underlying physical network resource information from both layers to design VNF embedding strategies for mapping nodes and links from SFC requirements to the physical network.

Fig. 1 SFC deployment on a mobile edge network

1.1 Model Establishment

The physical layer network is modeled as an undirected graph $G^p = (N^p, L^p)$, where N^p represents the set of edge physical nodes and L^p is the set of physical links. Physical links are divided into wired and wireless links. In the undirected graph G^p, each physical node can deploy several virtual machines, with each virtual machine instantiating only one type of VNF. Let $S = \{S_k | k = 1, 2, ..., K\}$ denote the set of SFC requests, where an SFC represents a series of network service functions traversed by data flows. Considering K service requests, each SFC consists of multiple types of VNFs, denoted as $S_k = \{v_i | i = 1, 2, ..., |V|\}$, where V represents the VNF set. Physical nodes are numbered sequentially, with physical node $n_j \in N^p$ denoted as n_j, and physical nodes $n_j \in N^p$ with wireless functionality denoted as w_j. Additionally, the physical link connecting nodes n_u and n_v is represented as e_{uv}. This paper considers heterogeneous nodes where, at different times, different nodes have varying mapping capabilities and resource amounts. Node n_j has idle CPU resources $C_j^p(n)$, idle storage resources $M_j^p(n)$, and idle forwarding resources $F_j^p(n)$. Wireless nodes w_j have idle CPU resources $C_j^p(w)$, wireless resources including antenna count $A_j^p(w)$, wireless resource blocks $B_j^p(w)$, and transmission power $P_j^p(w)$. Wireless links are denoted as e_{uv}^w.

Consider a time-slot system where $t \in \{1, 2, ...\}$ represents the time-slot index. In edge network environments, VNFs are combined according to certain functional sequences to form service function chains based on user service requests. Define the VNF traffic change factor as $\eta_i = \frac{r_i^{out}}{r_i^{in}}$, where r_i^{in} and r_i^{out} are the input and output traffic of VNF v_i, respectively. VNF nodes require deployment resources. For example, VNFs related to video rendering require more GPU resources than ordinary VNFs, while image editing requires more CPU resources in edge servers.

Let D_{uv}^w denote wireless transmission delay and D_{uv}^e denote wired link delay. First, wireless transmission delay is defined as the ratio of user downlink traffic to the wireless processing speed that wireless nodes can provide [**?**], i.e.:

$$D_{uv}^w = \frac{C_{in}}{C_{kj}^w}$$

where C_{in} represents the length of the sent data packet, and C_{kj}^w represents the wireless link transmission rate to node n_j [**?**], i.e.:

$$C_{kj}^{w} = B_{kj}^{w} \log_2 \left(1 + \frac{p_{w_k} h_{kj}}{\sigma_j^2}\right), \quad \forall k \neq j \in N$$

where p_{w_k} denotes the wireless transmission power of node w_k, h_{kj} denotes the wireless channel coefficient from node w_k to node n_j [**?**], B_{kj}^{w} denotes the wireless channel bandwidth from node w_k to node n_j, and σ_j^2 denotes the receiver noise variance of node n_j.

Second, the wired link transmission delay between wired nodes n_u and n_v is $D_{uv}^{e} = \frac{r_i^{in} \cdot \eta_i}{T_{uv}^{e}}$, where T_{uv}^{e} represents the wired link transmission rate.

At time t, let the available resources of node n_j be $R_j^p(t) = C_j^p(t) + M_j^p(t) + F_j^p(t)$. In mobile edge computing network environments, wireless nodes $w_j \in N^p$ require optimized design of base station antenna count $A_j^p(w)$ and wireless transmission power $P_j^p(w)$. The wireless resources of node w_j are denoted as $R_j^w(t) = A_j^p(w) + B_j^p(w) + P_j^p(w)$.

Let the deployment resource cost of SFC' s VNF nodes be:

$$C_{ov}^{i} = \alpha C_i^d + \beta M_i^d + \gamma F_i^d$$

where α, β, and γ represent weighting coefficients for deployment costs, ρ_{ij} denotes the number of hops between source and destination nodes, and ξ_{ij} represents the bandwidth resources required for virtual links.

1.2 Evaluation Metrics

For SFC deployment problems in mobile edge computing networks, this paper considers system performance metrics including resource consumption, average request arrival delay, request arrival rate, average reliability, bandwidth overhead, remapped links, and algorithm runtime [**?**, **?**].

1) Deployment Cost

In SFC deployment, mapping cost is modeled as bandwidth loss in physical links. Combining SFC deployment cost (15) and bandwidth resource overhead, the SFC deployment cost is denoted as C_{map}.

2) Deployment Success Rate

The SFC deployment success rate is defined as the ratio of successfully deployed SFC requests to total SFC requests arriving within a certain time period, i.e.:

$$\sigma = \lim_{T \to \infty} \frac{\sum_{t=0}^{T} succ_{SFC}(t)}{\sum_{t=0}^{T} all_{SFC}(t)}$$

where $succ_{SFC}(t)$ and $all_{SFC}(t)$ represent the number of successfully deployed and total SFC requests at time t, respectively.

3) Average Reliability

In NFV environments, component (VNF or physical node) reliability is defined. Let R_k denote the ratio of the average time a component provides service to the sum of the average service time and average downtime for repair, i.e., the probability that the component can provide service without failure within a specified time period [?]:

$$R_k = \frac{MTTF_k}{MTTF_k + MTTR_k}$$

where $MTTF_k$ represents the mean time between component failures, and $MTTR_k$ represents the mean time to repair failed components. Therefore, the reliability of SFC S_k is written as [?]:

$$R^k_{SFC} = \prod_{v_i \in V_k} R_i \cdot \prod_{e_{uv} \in L_k} R_{uv}$$

Let the virtual link set be $L_k = \{e_{uv} | e_{uv} \in L\}$, where e_{uv} represents the virtual link between VNF v_i and v_{i+1}. Let $e_{0,1}$ denote the virtual link between the source and the first VNF, and $e_{|V|,|V|+1}$ denote the virtual link between the last VNF and the destination.

Let $x^v_{i,j} \in \{0,1\}$ indicate whether VNF v_i is deployed on edge node n_j, and $y^l_{i,j} \in \{0,1\}$ indicate whether virtual link e_{uv} is deployed through physical link e_{uv} in the underlying physical network, where:

$$x^v_{i,j} = \begin{cases} 1, & \text{if deployed on edge node } n_j \\ 0, & \text{otherwise} \end{cases}$$

$$y^l_{i,j} = \begin{cases} 1, & \text{if deployed on wireless link } e_{uv} \\ 0, & \text{otherwise} \end{cases}$$

This paper considers the average reliability probability of SFCs as:

$$\bar{R}_{SFC} = \frac{1}{|K_s|} \sum_{k \in K_s} R^k_{SFC}$$

4) Average Request Arrival Delay

Define the average SFC request arrival delay as the ratio of total delay to the number of successfully deployed SFCs. SFC delay equals the sum of transmission delay, node delay, and delay of links where virtual links are deployed:

$$D_{SFC}^k = \sum_{e_{uv} \in L_k} D_{uv}^e + \sum_{v_i \in V_k} D_i^d + \sum_{e_{uv} \in L_k} D_{uv}^d$$

The average SFC request arrival delay is:

$$\bar{D}_{SFC} = \frac{1}{|K_s|} \sum_{k \in K_s} D_{SFC}^k$$

Mapping virtual links to physical links requires certain bandwidth resources. Different node combinations and mapping methods produce different link costs. Let the SFC mapping cost be:

$$C_{map}^{k,s} = \sum_{v_i \in V_k} C_{ov}^i \cdot x_{i,j}^v + \sum_{e_{uv} \in L_k} \big(\phi(e_{uv}) + \tau_{hop} \cdot b(e_{uv})\big) \cdot y_{i,j}^l$$

where the first term represents the mapping bandwidth resources of physical link e_{uv} and the link cost of service request S_k. The second term represents the inherent link consumption between v_i and v_{i+1}, and the third term represents the remaining bandwidth of link e_{uv}, where $\phi(e_{uv})$ denotes bandwidth overhead when $e_{uv} \in L^p$, $b(e_{uv})$ is the bandwidth of link e_{uv}, and τ_{hop} represents the shortest hop count from the current node to the next node.

To avoid data transmission congestion and hardware failures, load balancing of physical node resources is considered. The resource cost of wireless physical node w_j is defined as [**?**]:

$$C_{node}^j = \omega_1 \frac{\sum_{v_i \in V} H(v_i) \cdot x_{i,j}^v}{C_j^p(w)} + \omega_2 \frac{A_j^{used}(w)}{A_j^p(w)} + \omega_3 \frac{B_j^{used}(w)}{B_j^p(w)} + \omega_4 \frac{P_j^{used}(w)}{P_j^p(w)}$$

where $H(v_i)$ represents the wireless resources requested by virtual network function v_i, $A_j^{used}(w)$, $B_j^{used}(w)$, and $P_j^{used}(w)$ represent the occupied antenna count, wireless resource blocks, and wireless power of physical node w_j, respectively, while $A_j^p(w)$, $B_j^p(w)$, and $P_j^p(w)$ represent the maximum available antennas, wireless resource blocks, and power of physical node w_j.

In summary, the network service deployment cost is expressed as:

$$\Psi = \omega_1 C_{map} + \omega_2 \bar{D}_{SFC} + \omega_3 C_{node} + \omega_4 \sum_{e_{uv} \in L^p} hop_\tau(e_{uv})$$

where ω_1, ω_2, ω_3, and ω_4 are weighting coefficients.

5) Bandwidth Overhead

The bandwidth required during the process from SFC request to successful deployment is defined as:

$$B_{SFC}^{k} = \sum_{e_{uv} \in L_k} b(e_{uv}) \cdot y_{i,j}^{l}$$

where $b(e_{uv})$ denotes the bandwidth consumption required for successful SFC deployment.

---

2 Problem Formulation

2.1 Objective Function

When deploying network service requirements, the optimal network function deployment location is calculated based on available resources to minimize total node deployment cost. Simultaneously, to meet SFC reliability requirements, the selected link delay must satisfy user delay constraints. The optimization objective function for SFC deployment is established as follows:

$$\min \quad \mu_1 \Psi + \mu_2 \bar{D}_{SFC}$$

where μ_1 represents the weighting coefficient for SFC service request delay, and μ_2 represents the weighting coefficient for SFC deployment cost.

2.2 Constraints

Any VNF on an SFC can be deployed on either wireless or wired networks in the physical layer, but can only be instantiated on a single physical node:

$$\sum_{n_j \in N^p} x_{i,j}^{v} = 1, \quad \forall v_i \in V_k$$

The resources consumed by mapping all VNFs on each node cannot exceed the node' s maximum resource capacity. Let the mapping cost of each VNF be C_i^v. The total capacity C_j^p of resources provided by physical node n_j for VNF mapping must satisfy:

$$\sum_{v_i \in V_k} C_i^v \cdot x_{i,j}^{v} \le C_j^p, \quad \forall n_j \in N^p$$

To satisfy communication resource constraints, the bandwidth required for virtual link deployment cannot exceed the idle bandwidth resources in the link at the current time:

$$\sum_{e_{uv} \in L_k} b(e_{uv}) \cdot y_{i,j}^{l} \leq B_{uv}^{p}(t), \quad \forall e_{uv} \in L^p$$

When SFC virtual functions are mapped to physical nodes, the sum of delays from wired and wireless links must be less than the specified user delay D_0:

$$\sum_{e_{uv} \in L_k} D_{uv}^{e} \cdot y_{i,j}^{l} + \sum_{v_i \in V_k} D_i^d \cdot x_{i,j}^{v} \leq D_0, \quad \forall S_k \in S$$

When mapping service chains to the underlying edge network, SFCs must satisfy reliability constraints:

$$R_{SFC}^{k} \geq R_0, \quad \forall S_k \in S$$

For wireless nodes w_j, the wireless resource blocks, antenna count, and wireless power required to deploy all VNFs cannot exceed their available ranges:

$$\sum_{v_i \in V_k} H(v_i) \cdot x_{i,j}^{v} \leq A_j^p(w), \quad \forall w_j \in N^p$$

$$\sum_{v_i \in V_k} R_i^{RB} \cdot x_{i,j}^{v} \leq B_j^p(w), \quad \forall w_j \in N^p$$

$$\sum_{v_i \in V_k} P_i^{req} \cdot x_{i,j}^{v} \leq P_j^p(w), \quad \forall w_j \in N^p$$

where $A_j^p(w)$, $B_j^p(w)$, and $P_j^p(w)$ represent the maximum number of wireless resource blocks, maximum antenna count, and maximum base station power that wireless node w_j can provide, respectively.

---

3 Algorithm Design

The SFC deployment problem (23) belongs to the class of virtual network embedding problems [**?**], which is NP-hard with very high computational complexity. To meet the low-latency business requirements of mobile edge computing networks, low-complexity SFC deployment solutions are needed. First, when random SFC requests arrive at time t, the SFC types are classified to distinguish between wired SFCs and SFCs in hybrid wired-wireless network scenarios. This paper proposes a geographic location-aware service function chain placement optimization algorithm to determine whether VNF positions and types in the SFC satisfy sharing initial conditions, thereby reducing inter-VNF delay and improving SFC response time. If conditions are not met, deployment proceeds under normal circumstances. Second, under current mapping resource constraints, a resource-aware node selection algorithm is used to screen and evaluate node sets that satisfy resource constraints, determining the optimal node mapping sequence. Finally, using a bandwidth-demand-aware virtual link mapping algorithm, the path with minimum bandwidth and overhead is selected from candidate path sets, and the shortest weighted path is chosen as the communication link between VNFs.

3.1 Geographic-Aware Node Sharing Algorithm

In edge networks, node sharing not only improves SFC reliability but also converts data transmission between adjacent VNFs into intra-node data transfer, reducing transmission delay and bandwidth pressure.

Therefore, it is necessary to sense whether VNF positions in SFCs belong to the same type of VNF to reduce communication delay caused by geographic location. When adjacent VNFs select shared nodes, it must be considered whether they belong to the same VNF type (i.e., wired or wireless VNF). When an SFC request arrives, this paper proposes a geographic-aware node sharing algorithm. First, based on the SFC request, the algorithm searches whether VNFs at each position in the current SFC satisfy the sharing initial condition, i.e., whether adjacent VNFs are functionally mutually exclusive. Then, according to each VNF's traffic change factor, the algorithm determines whether merging conditions exist. Among VNF sequence sets with compatible functions, VNFs whose traffic change factor product satisfies the condition are merged, while corresponding VNF resource requirements are combined before node deployment.

Algorithm 1: Geographic-Aware Node Sharing Algorithm
Input: Edge information network G^p, VNF location information in SFC request S_k
Output: VNF set for node sharing

```
FOR all VNFs to be deployed
    IF current VNF has wireless function
        FOR each VNF in wireless set
            Screen continuous VNF sets satisfying wireless functional constraints
            Enter into merge queue set
            FOR each VNF set in the queue
                Merge VNFs in the current set
            ELSE
                Dissolve current set
            END IF
        END FOR
    END IF
END FOR

FOR each VNF in wired set
```

```
    Screen continuous VNF sets satisfying wired functional constraints
    Enter into merge queue set
    FOR each VNF set in the queue
        Merge VNFs in the current set
    ELSE
        Dissolve current set
    END IF
END FOR
END IF
```

3.2 Resource-Aware Node Selection Algorithm

After perceiving underlying network resources and VNF geographic locations, the algorithm traverses the VNF set of the current SFC. First, it screens node sets that satisfy resource constraints, such as CPU, storage, and forwarding resources. If a VNF node has wireless functionality, it must ensure wireless resource constraints are satisfied. Second, using linear weighted analysis, the algorithm scores and ranks the mapping from VNFs to be deployed in the current SFC to corresponding physical nodes, deploying the current VNF on the physical node with the highest comprehensive score. This process continues until all VNFs in the SFC are deployed. Physical node scoring metrics include node reliability, queuing delay, processing delay, and transmission delay. Finally, the algorithm updates the remaining resources of physical nodes occupied by SFC deployment at the current time.

Algorithm 2: Resource-Aware Node Selection Algorithm
Input: Merged SFC request S'_k
Output: Service function chain construction scheme

```
FOR each VNF in merged set
    IF current VNF does not have wireless function
        FOR each node in set N^p
            Screen node set satisfying CPU, storage, forwarding resource constraints
            Remove source nodes, target nodes, wireless nodes from set
            Remove already mapped nodes from set
            Screen node with highest evaluation metrics in set to map current VNF
            Update remaining resources of each physical node
        END FOR
    ELSE
        FOR each node in set N^p
            Screen node set satisfying CPU, storage, forwarding resources and wireless resou
            Remove source nodes, target nodes from set
            Remove already mapped nodes from set
            Calculate delay generated by current wireless node set for this VNF
            Calculate weighted decision matrix using current node reliability and delay
            Screen node with highest metrics to map this wireless VNF
            Update remaining resources of each physical node
```

```
        END IF
    END FOR
END FOR
```

3.3 Bandwidth-Demand-Aware Virtual Link Mapping Algorithm

After successful VNF deployment, routing selection is performed to map corresponding links. The specific algorithm steps are shown in Algorithm 3. First, when all VNFs in the SFC have been assigned optimal mapping physical nodes, based on the current bandwidth matrix, it is converted to sparse matrix format. Links that do not satisfy current SFC bandwidth requirements and those that do not interconnect between optimal nodes to be mapped are removed to save memory and improve search efficiency. The shortest path algorithm is used to find the shortest path between physical nodes to be mapped, and virtual links are deployed accordingly. Finally, after virtual link mapping is completed, the total delay of the SFC under the current mapping path is calculated. If the service minimum delay requirement is not satisfied, the SFC request is rejected.

Algorithm 3: Bandwidth-Demand-Aware Virtual Link Mapping Algorithm
Input: Service function chain construction scheme
Output: Service function chain mapping scheme

```
FOR links between all mapped nodes in S_k'
    Calculate weighted path bandwidth matrix using shortest path algorithm
    Obtain shortest path hop count d and node set p along the path
    Calculate consumed bandwidth and remapped links
END FOR

IF this S_k' does not have wireless function
    FOR nodes and VNFs in construction scheme
        Calculate wired delay
    END FOR
ELSE
    FOR nodes and VNFs in construction scheme
        IF current VNF does not have wireless function
            Calculate wired delay
        ELSE
            Calculate wireless delay
        END IF
    END FOR
END IF

Calculate total delay of current mapping scheme
IF current SFC delay does not meet its delay requirement
    Mapping fails, SFC request not arrived
ELSE
```

```
    Mapping succeeds, SFC request arrived, update remaining link bandwidth resources
END IF
```

---

4 Simulation Experiments

The SFC deployment simulation platform was built on a desktop computer configured with an Intel Core i5-10500 CPU and 16GB RAM. To comprehensively evaluate algorithm performance, metrics including SFC deployment resource consumption, request arrival rate, average reliability, average bandwidth overhead, average request delay, number of remapped links, and algorithm runtime were considered. In experiments, the random algorithm, the traversal-selection algorithm proposed in reference [?], the dynamic SFC deployment algorithm proposed in reference [?], and the greedy algorithm proposed in reference [?] were used as baseline comparison schemes. The random algorithm shares nodes but does not perform shortest weighted link mapping.

In the simulation platform, the edge network consists of 100 generated nodes, with 40 nodes enabling wireless functionality. SFC requests are assumed to arrive dynamically, following a Poisson distribution with an arrival rate of 0.05. To reduce random factor effects, simulation results adopt the average of 100 experiments, with each experiment requiring 50,000 time units, as shown in Table 1 .

[Figure 2: see original paper] shows the impact of changes in the number of successfully deployed SFC requests on resource utilization. Figure 2(a) demonstrates that as the number of successfully deployed SFC requests increases, the deployment resource consumption of the proposed method is lower than the other four methods. The algorithms proposed in references [?] and [?] do not share nodes during deployment and do not consider node load balancing, resulting in higher resource consumption. Reference [?] randomly selects portions of SFCs for node sharing but ignores load balancing, leading to relatively high resource consumption. The proposed algorithm, considering the limited resources in edge networks, shares physical nodes among more VNFs, reducing unnecessary hops between nodes. Additionally, the algorithm optimizes link resource allocation, resulting in minimal deployment resource overhead. Figure 2(b) shows the comparison of average delay after successful SFC deployment. The proposed algorithm and baseline random algorithm are significantly superior to the other two algorithms because node sharing can transform data transmission between VNFs into intra-node data transfer, reducing certain transmission delays. The proposed algorithm considers node and link capacity constraints in edge networks, incorporating comprehensive deployment delay into the node screening weight algorithm to achieve lower node delay. In node ranking evaluation weights, the algorithms proposed in references [?] and [?] consider delay weights during deployment, thus outperforming the algorithm in reference [?] which does not consider deployment link delay. Figure 2(c) shows that the proposed algorithm has less bandwidth overhead because it shares nodes according to constraints and uses mapping algorithms that consider load balancing. The baseline random algorithm also performs node sharing, outperforming the other three baseline algorithms, but the improvement is not significant because it does not perform shortest weighted link mapping during deployment. Finally, Figure 2(d) shows that the proposed mapping algorithm uses shortest weighted link mapping, employing bandwidth matrix constraints during shortest path searching to help avoid loops between nodes, resulting in fewer remapped links than the other four methods.

Table 1 Simulation setup

Parameter	Value (Unit)
Node transmission delay	[2,5] ms
Node CPU, storage, forwarding resources	[50,100] Mb
Number of wired/wireless VNFs per service chain	[3,6], [0,3]
SFC lifecycle	Exponential distribution {1000}
Node reliability	[0.85,0.98]
Available antennas per node	[1,4]
Available wireless resource blocks per node	[20,30]
Available power per node	[1000,2000] mW
VNF storage, CPU, forwarding resources	[3000,5000] Mbps
Initial bandwidth demand	[10,20] Mb
Wireless VNF antenna requirement	400 Mbps
Wireless VNF wireless resource block requirement	[4,10]
Wireless VNF power requirement	[3,6]
	[60,200] mW

[Figure 3: see original paper] shows the impact of algorithm performance variations. Figure 3(a) displays the curves of SFC request arrival rate over time for five different deployment algorithms. The algorithm proposed in reference [?] does not comprehensively consider delays in various parts of the deployment process. The baseline random algorithm and the algorithm proposed in reference [?] cannot simultaneously satisfy reliability and delay constraints during deployment, resulting in lower request arrival rates. The algorithm proposed in reference [?], although sharing node resources, randomly selects portions of SFCs for sharing without considering load balancing, increasing resource consumption and thus having the lowest request arrival rate. The proposed algorithm considers node load balancing, uses traffic change factor characteristics to merge some VNFs on the same physical node, shares resources, reduces deployment overhead, and improves request arrival rate. Figure 3(b) shows the curve of average SFC reliability versus the number of successfully deployed SFCs. The average reliability coefficients of the proposed algorithm and baseline random algorithm are higher than 0.8 because the node sharing algorithm considers node reliability weights and prioritizes nodes with higher reliability. Deploying multiple VNFs on a single physical node increases certain weights during node screening, helping improve SFC reliability. The algorithm in reference [?] selectively shares nodes during deployment, increasing SFC reliability and thus outperforming the remaining two baseline algorithms. The method in reference [?] considers network topology and reliability but does not share nodes. The algorithm in reference [?] does not consider node reliability and cannot guarantee reliability when screening node sets. Figure 3(c) shows that the proposed algorithm has the lowest time complexity. During screening of optimal candidate nodes for VNF deployment, both the proposed algorithm and baseline algorithms have time complexity O(N), where N represents the number of physical nodes. However, during shortest path mapping, unlike baseline algorithms, the proposed algorithm converts the current path matrix to a sparse matrix, significantly reducing the time complexity of shortest path generation.

---

5 Conclusion

This paper studied the SFC deployment problem in mobile edge computing networks and developed a joint cost-delay optimization design method. During the SFC construction phase, optimal nodes are selected based on node location and resource status to reduce inter-VNF delay and improve SFC response time. During the SFC deployment phase, due to limited mapping resources, a node selection algorithm is applied to determine the optimal node mapping sequence, and the shortest weighted path is selected as the communication link between VNFs. Experimental results demonstrate that compared with existing schemes, the proposed algorithm can effectively reduce SFC service request delay and improve the success rate of reliable SFC deployment.

Future research directions will focus on device-to-device (D2D) wireless communication networks, considering wireless channel fading and user interference to design low-power SFC deployment schemes and resource management algorithms.

References

[1] Zu Jiachen, Hu Guyu, Yan Jiajie, et al. Resource management research of service function chain under network function virtualization [J]. *Computer Research and Development*, 2021, 58 (01): 137-152.

[2] Saraiva de Sousa N F, Lachos Perez D A, Rosa R V, et al. Network service orchestration: a survey [J/OL]. *Computer Communications*, 2019, 142: 69-94. https://doi.org/10.1016/j.comcom.2019.04.008.

[3] Hantouti H, Benamar N, Taleb T, et al. Traffic steering for service function chaining [J]. *IEEE Communications Surveys & Tutorials*, 2018, 21 (1): 487-507.

[4] Han Bo, Gopalakrishnan V, Ji Lusheng, et al. Network function virtual-

ization: challenges and opportunities for innovations [J/OL]. *IEEE Communications Magazine*, 2015, 53 (2): 90-97. https://doi.org/10.1109/MCOM.2015.7045396.

[5] ETSI G. Network functions virtualisation (nfv); use cases [J]. V1, 2013, 1: 2013-10.

[6] Yeganeh S H, Tootoonchian A, Ganjali Y. On scalability of software-defined networking [J]. *IEEE Communications Magazine*, 2013, 51 (2): 136-141.

[7] Wang Feng, Xu Jie, Ding Zhiguo. Multi-antenna NOMA for computation offloading in multiuser mobile edge computing systems [J]. *IEEE Trans on Communications*, 2018, 67 (3): 2450-2463.

[8] Yu Minlan, Yi Yung, Rexford J, et al. Rethinking virtual network embedding: substrate support for path splitting and migration [J]. *ACM SIGCOMM Computer Communication Review*, 2008, 38 (2): 12.

[9] Hu Qian, Wang Yang, CAO Xiaojun. Resolve the virtual network embedding problem: a column generation approach [C]// *Proc of the Proceedings IEEE INFOCOM*. Piscataway, NJ: IEEE Press, 2013: 410-414.

[10] Siasi N, Jasim M, Crichigno J, et al. Container-based service function chain mapping [C]// *Proc of the SoutheastCon*. Piscataway, NJ: IEEE Press, 2019: 1-6.

[11] Chen Jia, Liu Hongqiao, Jia Haiyu. Cross-layer resource allocation in wireless-enabled NFV [J/OL]. *IEEE Wireless Communications Letters*, 2020, 9 (6): 879-883. https://doi.org/10.1109/LWC.2020.2974198.

[12] Zhang Yuncan, He Fujun, Oki E. Availability-aware service chain provisioning with sub-chain-enabled coordinated protection [C]// *Proc of the IFIP/IEEE International Symposium on Integrated Network Management (IM)*. Piscataway, NJ: IEEE Press, 2021: 638-642.

[13] Tang Lun, Zhao Peipei, Zhao Guofan, et al. Dynamic deployment algorithm of service function chain based on QoS guarantee [J]. *Journal of Beijing University of Posts and Telecommunications*, 2018, 41 (06): 90-96.

[14] Ahvar S, Phyu H P, Buddhacharya S M, et al. CCVP: Cost-efficient centrality-based VNF placement and chaining algorithm for network service provisioning [C/OL]/ *Proc of the IEEE Conference on Network Softwarization (NetSoft)*. Piscataway, NJ: IEEE Press 2017: 1-9. https://doi.org/10.1109/NETSOFT.2017.8004104.

[15] Bari F, Chowdhury S R, Ahmed R, et al. Orchestrating virtualized network functions [J/OL]. *IEEE Trans on Network and Service Management*, 2016, 13 (4): 725-739. https://doi.org/10.1109/TNSM.2016.2585462.

[16] Sun Shiqing, Peng Jianhua, You Wei, et al. Collaborative construction and mapping algorithm of resource-aware service function chain under 5G network [J]. *Journal of Xi' an Jiaotong University*, 2020, 54 (08): 140-148.

[17] Yuan Bo, Ren Bangbang. Embedding the minimum cost SFC with end-to-end delay constraint [C/OL]/ *Proc of the 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE).* Piscataway, NJ: IEEE Press. 2020: 2299-2303. https://doi.org/10.1109/ICMCCE51767.2020.00497.

[18] Zhong Xuxia, Wang Ying, Qiu Xuesong, et al. Cost-aware service function chain orchestration across multiple data centers [C/OL]/ *Proc of the NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium.* Piscataway, NJ: IEEE Press. 2018: 1-7. https://doi.org/10.1109/NOMS.2018.8406174.

[19] Niu Meng. Research on service function chain scheduling mechanism for reliability and response time in edge networks [D/OL]. Beijing University of Posts and Telecommunications, 2021.

[20] Shannon C E. A mathematical theory of communication [J]. *The Bell system technical journal*, 1948, 27 (3): 379-423.

[21] Qu Long, Khabbaz M, Assi C. Reliability-aware service chaining in carrier-grade softwarized networks [J/OL]. *IEEE Journal on Selected Areas in Communications*, 2018, 36 (3): 558-573. https://doi.org/10.1109/JSAC.2018.2815338.

[22] Ayoubi S, Zhang Yanhong, Assi C. RAS: Reliable auto-scaling of virtual machines in multi-tenant cloud networks [C]// *Proc of the IEEE 4th International Conference on Cloud Networking (CloudNet).* Piscataway, NJ: IEEE Press. 2015: 1-6.

[23] Jia Haiyu. Research on the dynamic deployment mechanism of service demand-oriented functions in smart collaborative wireless networks [D/OL]. Beijing Jiaotong University, 2019.

[24] Zhai Dong, Meng Xiangru, Kang Qiaoyan, et al. Service function chain deployment method for delay and reliability optimization [J]. *Journal of Electronics and Information*, 2020, 42 (10): 2386-2393.

[25] Yang Yong, Meng Xiangru, Kang Qiaoyan, et al. Deployment Method of Reliable Service Function Chain Based on Traffic Optimization [J]. *System Engineering and Electronic Technology*, 2021, 43 (10): 3017-3025.

[26] Amaldi E, Coniglio S, Koster A M, et al. On the computational complexity of the virtual network embedding problem [J]. *Electronic Notes in Discrete Mathematics*, 2016, 52: 213-220.

Note: Figure translations are in progress. See original paper for figures.