

移动边缘计算网络中联合无线多播的服务功能链部署算法

王侃¹, 赵楠², 李军怀¹, 王怀军¹

(1. 西安理工大学计算机科学与工程学院, 陕西 西安 710048; 2. 大连理工大学信息与通信工程学院, 辽宁 大连 116024)

摘要: 为解决移动边缘计算 (MEC) 网络中面向用户的服务功能链部署 (SFC) 算法系统开销过大、业务拥塞严重等问题, 针对部署 MEC 服务器的多基站多用户边缘网络, 提出了一种面向内容的联合无线多播的 SFC 部署算法。综合考虑数据流、服务器功能维护功耗、服务器功能服务功耗和无线传输功耗 4 种系统开销, 建立波束成形设计和 SFC 映射的联合优化模型。首先, 利用拉格朗日对偶分解技术, 将优化问题解耦为 SFC 部署和波束成形设计 2 个独立子问题; 其次, 利用基于 L_p 范数惩罚项的连续凸近似算法, 将整数形式的 SFC 部署问题松弛为一个等价线性规划问题; 最后, 利用路径跟随技术, 将非凸波束向量优化问题转化为一组凸优化子问题。仿真结果表明, 所提算法具有较好收敛性能, 并在系统开销方面优于传统的最优单播 SFC 部署算法和随机多播 SFC 部署算法。

关键词: 移动边缘计算; 服务功能链; 多播波束成形; 网络功能虚拟化

中图分类号: TN92

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020210

Service function chain embedding algorithm with wireless multicast in mobile edge computing network

WANG Kan¹, ZHAO Nan², LI Junhuai¹, WANG Huaijun¹

1. School of Computer and Science Engineering, Xi'an University of Technology, Xi'an 710048, China

2. School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China

Abstract: To resolve the excessive system overhead and serious traffic congestion in user-oriented service function chain (SFC) embedding in mobile edge computing (MEC) networks, a content-oriented joint wireless multicast and SFC embedding algorithm was proposed for the multi-base station and multi-user edge networks with MEC servers. By involving four kinds of system overhead, including service flow, server function sustaining power, server function service power and wireless transmission power, an optimization model was proposed to jointly design SFC embedding with multicast beamforming. Firstly, with Lagrangian dual decomposition, the problem was decoupled into two independent subproblems, namely, SFC embedding and multicast beamforming. Secondly, with the L_p norm penalty term-based successive convex approximation algorithm, the integer programming-based SFC embedding problem was relaxed to an equivalent linear programming one. Finally, the non-convex beamforming optimization problem was transformed into a series of convex ones via the path following technique. Simulation results revealed that the proposed algorithm has good convergence, and is superior to both the optimal SFC embedding with unicasting and random SFC embedding with multicasting in terms of system overhead.

Key words: mobile edge computing, service function chain, multicast beamforming, network function virtualization

收稿日期: 2020-07-16; 修回日期: 2020-09-24

通信作者: 赵楠, zhaonan@dlut.edu.cn

基金项目: 国家重点研发计划基金资助项目 (No.2018YFB1703000); 国家自然科学基金资助项目 (No.61801379, No.61871065, No.61971347); 西安电子科技大学综合业务网理论及关键技术国家重点实验室开放课题资助项目 (No.ISN21-08)

Foundation Items: The National Key Research and Development Program of China (No.2018YFB1703000), The National Natural Science Foundation of China (No.61801379, No.61871065, No.61971347), The Open Research Fund from State Key Laboratory of Integrated Services Networks, Xidian University (No.ISN21-08)

1 引言

5G 移动通信系统以其灵活性和高效性为用户提供了超高吞吐量和超低时延的服务体验^[1]。作为 5G 的关键技术,网络功能虚拟化(NFV, network function virtualization)和移动边缘计算(MEC, mobile edge computing)已得到学术界和工业界的广泛关注^[2-3]。不同于传统网络通过部署专门硬件来实现网络功能,NFV 将软硬件解耦,通过在商用服务器上部署虚拟网络功能,构造面向各类业务需求的、将虚拟网络功能有序排列的服务功能链(SFC, service function chain),为用户提供差异化和定制化的服务^[2]。MEC 将部分或全部网络功能卸载到靠近用户的边缘网络,将显著提升服务的时延和吞吐量性能,有效降低核心网的业务拥塞^[3]。

NFV 和 MEC 相互补充,通过在边缘网络的基站中部署虚拟网络功能,可使用户就近接受网络服务,从而增强了网络管理的灵活性和可伸缩性^[4]。然而,与云接入网(C-RAN, cloud radio access network)等集中式网络架构相比,边缘网络中 SFC 部署问题将面临如下特有挑战。

1) 与集中式网络架构中的云计算服务器不同,MEC 服务器的计算容量和存储容量通常受限^[5-7];单个 MEC 服务器只能部署 SFC 中的部分网络功能。如何协同利用受限的 MEC 服务器资源,进行虚拟网络功能的有序部署,对 SFC 部署算法带来设计上的挑战。

2) 与核心网中 SFC 部署不同,边缘网络中 SFC 的部署不仅涉及数据流在基站与基站之间基于有线链路的服务路径选择,而且需考虑执行完最后一项网络功能后,数据流经基站到用户的无线链路传输。

因此,基于上述挑战,边缘网络 SFC 部署问题面临的关键技术包括服务路径选择技术和无线链路干扰消除技术。首先,针对实时服务请求,服务路径选择表示引导数据流依次通过能够支持 SFC 中网络功能的服务基站。其次,由于无线链路的广播特性,在执行完 SFC 的最后一项网络功能后,数据流经基站到用户的无线链路存在相互干扰;利用正交频谱资源分配或波束成形技术以消除用户间干扰,是边缘网络 SFC 部署问题的另一项关键技术。如何针对差异化的用户需求和有限的网络资源,设计低开销的、基于干扰消除的边缘网络 SFC

部署算法,已成为学术界和工业界的研究热点。

已有研究工作通常关注端到端的、面向用户的 SFC 部署方案^[8-11],在网络资源受限的约束下,实现虚拟网络功能到 MEC 服务器的有效映射,优化用户的服务体验。文献[8]研究了边缘网络中 SFC 的服务迁移问题,基于用户随机游走模型,考虑了迁移代价和时延约束,提出了一种在线服务迁移算法,以最小化用户的服务中断概率。同样基于用户运动模型,文献[9]研究了面向内容缓存服务的最短路由路径的 SFC 部署问题,将最短路径问题建模为一个整数线性规划(ILP, integer linear programming)问题,提出了一种低复杂度的启发式算法。基于软件定义的边缘网络,文献[10]同时考虑了用户的业务体验和服务的可靠性需求,研究了端到端时延约束下系统的吞吐量优化问题,利用多路径路由策略保障服务的可靠性,并利用主对偶方法找到每条可行路径的分配带宽。基于单路径路由策略,文献[11]同时考虑了节点处理容量和链路带宽约束,研究了最小化数据流开销的 SFC 映射问题。综上所述,已有的 SFC 部署方案通常是面向用户的,为每个用户部署一条专用的、定制化的 SFC。然而,随着内容提供业务(例如,视频、音乐和文件)的持续增长,边缘网络中的流行内容会被多个用户重复下载^[12-13],此时,若对具有相同内容请求的多个用户均部署定制化的 SFC,不仅会增大边缘网络中数据流的开销,造成业务拥塞,而且会导致额外的 MEC 服务器功耗开销。因此,研究基于内容分组的 SFC 部署算法具有重要理论意义和现实价值。

将请求同一内容的用户划分到一组,设计面向内容的 SFC 部署算法,使该组所有用户共享一条 SFC,可有效降低系统数据流开销和功耗开销。然而,与核心网不同,边缘网络无线信道的广播特性使同一内容到多个用户的多播传输存在相互干扰。基于正交频谱资源分配,文献[14]提出了一种两阶段协作多播机制,以降低系统传输功耗,并增强系统覆盖率。首先,利用随机几何理论,采用基于平均接收信号强度的选择合并技术,分析了基站传输功耗和系统传输功耗的相互关系;其次,提出了一种基于扇形环结构的移动中继部署方案,进而推导了基于期望覆盖率的最优基站传输功耗的表达式。基于波束成形技术,文献[15]将请求同一内容的用户进行组划分,形成多播组,对组内所有用户发送相同的数据符号,从而有效节省了无线频谱资源,

并降低了数据流开销。文献[16]研究了 C-RAN 中运营商的收益优化问题，同时考虑了高带宽和低时延 2 种典型 5G 业务，针对高带宽的内容提供业务，利用波束成形技术实现组内用户干扰抑制；针对低时延业务，利用正交频谱资源消除用户间干扰。同样基于 C-RAN，文献[17]研究了基于混合时间尺度的内容缓存服务，利用协作多播技术以抑制组内用户干扰，综合考虑系统功耗和用户吞吐量需求等约束，以最大化系统的长期收益。综上所述，针对面向内容的无线多播技术的研究通常只考虑内容从接入基站到用户的无线传输这一环节。然而，与云计算服务器相比，MEC 服务器的存储和计算资源通常受限，单个接入基站只能部署 SFC 中的部分功能。因此，基于内容分组，引导数据流依序通过多个服务基站，服务路径规划需联合考虑基站之间的有线链路和接入基站到用户的无线链路，构造“第一个服务基站→第二个服务基站→…→最后一个服务基站（接入基站）→用户”的端到端服务路径。

本文主要的研究工作如下。

1) 面向内容提供服务，建立边缘网络中联合无线多播的 SFC 部署模型。最大化系统中数据流开销和功耗开销，并满足链路带宽约束、处理容量约束、吞吐量需求约束、SFC 部署序列约束、波束向量与信号处理功能的耦合约束，以及最大发射功率约束。综合考虑数据流、服务器功能维护功耗、服务器功能服务功耗和无线传输功耗这 4 种系统开销，建立波束成形设计和 SFC 部署的联合优化问题。该问题是一个 NP (non-deterministic polynomial) 难问题，很难找到多项式时间求解算法。

2) 利用拉格朗日对偶分解技术，将原优化问题转化为 2 个独立子问题。利用基于 L_p 范数惩罚项的连续凸近似算法，将 ILP 项式的 SFC 部署问题松弛为一个线性规划 (LP, linear programming) 问题，并给出了原问题和松弛问题的最优解等价性证明；利用路径跟随技术，将非凸波束向量优化问题转化为一组凸子问题，并给出了算法的单调性分析。

3) 仿真结果表明，本文算法具有良好收敛性能。将本文算法与最优单播 SFC 部署算法和随机多播 SFC 部署算法进行对比，验证了本文算法的有效性。

2 系统模型

在边缘网络中，通过在基站的 MEC 服务器中部署 NFV 技术，基站不仅能提供基带处理单元

(BBU, base-band unit) 功能，而且可支持缓存、计算、防火墙和网络地址转换等多种虚拟网络功能。考虑一个多基站部署的边缘网络，基站集合表示为 $\mathcal{N} = \{1, 2, \dots, N\}$ ，如图 1 所示。基站之间通过 X2+ 链路实现互联，同时每个基站配置 I 根天线和一个商用服务器。边缘网络中分布 K 个用户，用集合 $\mathcal{K} = \{1, 2, \dots, K\}$ 表示。假设所有用户均请求内容提供服务，并将请求同一内容的用户分配到同一多播组。多播组用集合 $\mathcal{M} = \{1, 2, \dots, M\}$ 表示，而多播组 m 的所属用户集合用 $\mathcal{G}(m)$ 表示。

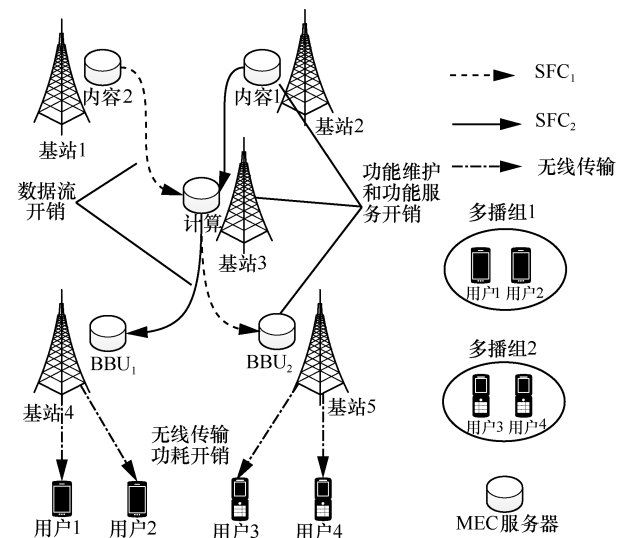


图 1 多基站部署的边缘网络

2.1 SFC 模型和容量约束

在基于 NFV 技术的边缘网络中，每项内容提供服务可映射为一条端到端的 SFC 部署，即“第一个服务基站→第二个服务基站→…→最后一个服务基站（接入基站）→用户”的端到端服务路径。在多播组 m 中，任一数据流在被基站天线发送到用户之前，需遍历 SFC 的每一项网络功能。分别用 f_1^m 和 f_L^m 表示该条 SFC 中第一和最后一项功能，则可用 $\mathcal{F}(m) = (f_1^m \rightarrow \dots \rightarrow f_i^m \rightarrow \dots \rightarrow f_L^m)$ 来描述该条 SFC。多播组 m 的任一数据流需起源于 f_1^m （例如，图 1 中的内容 1 和内容 2 的 MEC 服务器），依序遍历 $\mathcal{F}(m)$ 中其他功能，最终以 f_L^m （例如，图 1 中部署 BBU1 和 BBU2 的 MEC 服务器）终止服务。如图 1 所示，边缘网络中部署了 2 条基于内容提供服务的 SFC，每个数据流需依次遍历缓存功能、计算功能和信号处理功能，最终由基站无线传输到终端用户。

首先,为实现 SFC 在边缘网络的有序部署,对 $\mathcal{F}(m)$ 中的任一项功能 f_l^m , 定义一个二进制变量 $y_{l,n}^m$: 若 f_l^m 在基站 n 获取服务, 则 $y_{l,n}^m=1$; 否则, $y_{l,n}^m=0$ 。定义 \mathcal{N}_l^m 表示可提供功能 f_l^m 的所有基站集合。为确保 $\mathcal{F}(m)$ 中的任一项功能 f_l^m 只在一处基站的 MEC 服务器中获取服务, 则 $y_{l,n}^m$ 需满足式(1)所示约束。

$$\sum_{n \in \mathcal{N}_l^m} y_{l,n}^m = 1, \forall f_l^m \in \mathcal{F}(m), \forall m \quad (1)$$

其次, 定义二进制变量 $x_{f,n}$, 表示基站 n 是否提供并维护功能 f 。若基站 n 在当前时刻维护功能 f , 则 $x_{f,n}=1$; 否则, $x_{f,n}=0$ 。显然, $y_{f,n}^m$ 和 $x_{f,n}$ 之间满足式(2)所示约束。

$$y_{f,n}^m \leq x_{f,n}, \forall f, \forall m, \forall n \quad (2)$$

再次, 将基站 n 的 MEC 服务器中所有功能集合表示为 \mathcal{F}_n 。针对每项功能 $f \in \mathcal{F}_n$, 为其设置一个处理容量 $\mu_{f,n}$, 即功能 f 在当前时刻所服务的数据流总吞吐量不得超过 $\mu_{f,n}$ 。将多播组 m 的吞吐量需求表示为 R_m , 则所有多播组在功能 $f \in \mathcal{F}_n$ 处获取服务的总吞吐量需满足式(3)。

$$\sum_m y_{f,n}^m R_m \leq \mu_{f,n}, \forall f \in \mathcal{F}_n, \forall n \quad (3)$$

最后, 将任意基站 n 和 s 之间的链路带宽定义为 $\mu_{n,s}$, 并定义二进制变量 $z_{l,n,s}^m$ 。 $z_{l,n,s}^m=1$ 表示多播组 m 的第 l 项功能 f_l^m 和第 $l+1$ 项功能分别由基站 n 和 s 依序服务。因此, 所有多播组在链路 $n \rightarrow s$ 上传输的总吞吐量需满足式(4)。

$$\sum_m \sum_{f_l^m \in \mathcal{F}(m)} z_{l,n,s}^m R_m \leq \mu_{n,s}, \forall n, \forall s \quad (4)$$

只有当 $y_{l,n}^m=1$ 和 $y_{l+1,s}^m=1$ 同时成立时, $z_{l,n,s}^m=1$ 才成立。因此, $y_{l,n}^m$ 、 $y_{l+1,s}^m$ 和 $z_{l,n,s}^m$ 需满足式(5)所示约束。

$$z_{l,n,s}^m \geq y_{l,n}^m + y_{l+1,s}^m - 1, \forall f_l^m \in \mathcal{F}(m), \forall m, \forall n, \forall s \quad (5)$$

2.2 多播模型

针对 $\mathcal{F}(m)$ 中最后一项功能 f_L^m , 考虑二进制变量 $y_{L,n}^m$ 。若多播组 m 的请求内容在基站 n 处实现信号处理功能, 则 $y_{L,n}^m=1$; 否则, $y_{L,n}^m=0$ 。多播组 m 中的任一用户 k 从基站 n 处获得的接收信号为

$$u_{k,n} = \mathbf{h}_{k,n}^H \mathbf{w}_{m,n} u_m + \sum_{j \neq m} \mathbf{h}_{k,n}^H \mathbf{w}_{j,n} u_j + n_k, \quad \forall k \in \mathcal{G}(m), \forall m, \forall n \quad (6)$$

其中, $\mathbf{h}_{k,n} \in \mathbb{C}^{1 \times 1}$ 为基站 n 到用户 k 的信道向量, $\mathbf{w}_{m,n} \in \mathbb{C}^{1 \times 1}$ 为基站 n 到多播组 m 的传输波束向量, u_m 为当前时刻多播组 m 的接收数据符号, n_k 为服从复高斯分布的信道噪声。只有当 $y_{L,n}^m=1$ 时, 波束向量 $\mathbf{w}_{m,n}$ 才能取非零值, 因此, $\mathbf{w}_{m,n}$ 的传输功率需满足式(7)。

$$\|\mathbf{w}_{m,n}\|_2^2 \leq P_n^{\text{th}} y_{L,n}^m \quad (7)$$

其中, P_n^{th} 为基站 n 的传输功率门限值。

与协同多点传输 (CoMP, coordinated multiple-point transmission) [15-20] 技术不同, 本文假设每个基站在自己的天线组内独立进行多播波束成形, 对每个基站分配正交的频谱资源, 从而有效消除了小区间干扰。基于单天线 CoMP 技术, 文献[18-19]提出了宏分集协同多点传输 (MD-CoMP, macro diversity coordinated multi-point transmission) 概念, 将小区频率复用因子设为 1, 通过对同小区用户分配正交频谱资源, 以消除小区内干扰。因此, 用户所受干扰为来自其他基站簇的小区间干扰。本文没有考虑基站之间的 CoMP, 这是因为, 若考虑 CoMP, 则每个用户将选择多个接入基站, 本文考虑的单服务路径规划问题将成为一个多服务路径规划问题, 进一步增大优化问题的求解难度。定义每个基站的无线频谱带宽为 B , 则用户 k 从基站 n 处得到的接收信号信干噪比 (SINR, signal to interference plus noise ratio) $\Gamma_{k,n}$ 为

$$\Gamma_{k,n} = \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}|^2}{\sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma_k^2}, \forall k \in \mathcal{G}_m, \forall m, \forall n \quad (8)$$

其中, $|\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}|^2$ 表示用户 k 接收到的有用信号功率, $\sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2$ 表示用户 k 受到的来自其他多播组的干扰。本文 3.3 节设计了一种波束成形算法, 以有效抑制小区内组间用户干扰。在同一多播组内的多个用户之间, 接收的是同一内容和同一数据符号; 即使同一组内多个用户之间共同使用带宽 B , 也不存在组内用户干扰。

最后, 用户 k 在当前时刻从基站 n 接收的吞吐量可表示为 $R_{k,n} = B \ln(1 + \Gamma_{k,n})$ 。为保障正确的数据传输和解码, $R_k = \sum_n R_{k,n}$ 不得小于所在多播组的吞吐量需求 R_m , 即

$$R_k \geq R_m, \forall k \in \mathcal{G}(m), \forall m \quad (9)$$

2.3 问题描述

一方面，为最小化 SFC 路径中的数据流开销，每条 SFC 的功能应集中部署在尽可能少的 MEC 服务器中，然而受限于 MEC 服务器的处理容量，同一 SFC 的不同功能往往部署到不同的基站，这将不可避免地增大数据流开销。另一方面，不仅基站的发送天线带来无线传输功耗开销，MEC 服务器的功能维护和功能服务也将导致服务器功耗开销。系统开销应综合考虑数据流开销和功耗开销。因此，可将系统总开销 C 定义为

$$C = \sum_{n,s} \sum_m \sum_{f_i^m \in \mathcal{F}(m)} z_{l,n,s}^m R_m + \eta \left(\sum_n \sum_{f \in \mathcal{F}_n} x_{f,n} P_{f,n} + \sum_n \sum_m \sum_{f_i^m \in \mathcal{F}(m)} y_{l,n}^m P_{l,n}^m + \sum_n \sum_m \|\mathbf{w}_{m,n}\|_2^2 \right) \quad (10)$$

其中， η 为平衡数据流开销和功耗开销的折中系数， $P_{f,n}$ 为基站 n 维护功能 f 的功耗， $P_{l,n}^m$ 为基站为多播组 m 提供功能 f_i^m 的服务功耗。式(10)中，系统开销分别为数据流开销、服务器功能维护功耗开销、服务器功能服务功耗开销和无线传输功耗开销。首先，相邻 2 个网络功能之间的数据传输将导致数据流开销；为实现 SFC 的有序部署，每个基站需为多播组的服务请求提供网络功能，将导致功能维护开销和功能服务开销。上述 3 种功耗发生在边缘网络的 MEC 服务器中或基站之间的有线链路中。其次，为实现端到端（从边缘网络基站到多播用户）的数据传输，多播组的任一数据流需遍历 SFC 中的每一项网络功能，最终到达用户的所属服务基站，使用波束成形等信号处理，通过无线多播的方式到达用户。因此，无线传输功耗发生在无线多播环境中服务基站到多播用户的最后一跳中。4 种系统开销的关系如图 1 所示。

为最小化 SFC 部署的系统开销，节省边缘网络资源，将优化问题 \mathcal{P}_0 描述如下。

$$\begin{aligned} \mathcal{P}_0 : \min C \\ \text{s.t. 式(1)~式(4), 式(5), 式(7), 式(9)} \\ \text{var } \mathbf{w}_{m,n}, \forall m, \forall n \\ x_{f,n} \in \{0,1\}, \forall f, \forall n \\ y_{l,n}^m \in \{0,1\}, \forall f_i^m \in \mathcal{F}(m), \forall m, \forall n \\ z_{l,n,s}^m \in \{0,1\}, \forall f_i^m \in \mathcal{F}(m), \forall m, \forall n, \forall s \end{aligned} \quad (11)$$

观察 \mathcal{P}_0 可得，波束成形和基站间的 SFC 部署之间存在如下折中关系。

1) 基站之间的 SFC 部署将决定用户的接入基

站选取。边缘网络的 SFC 部署将规划边缘网络到用户的端到端服务路径，从而为每个用户选取一个接入基站，以无线传输方式完成数据流到用户的最后一跳。若只优化基站之间的 SFC 部署开销，不考虑用户的接入基站选取，将导致无线传输功耗开销过高。

2) 用户的接入基站选取将影响基站之间的 SFC 部署。作为 SFC 部署的最后一个服务节点，用户的接入基站将数据流经无线空口传输到用户。同时，接入基站的位置将影响上一个服务基站的选取，从而依次逆序影响其他所有服务基站的选取。因此，若只考虑波束成形算法的无线传输功耗，不考虑用户的接入基站选取，将导致基站之间 SFC 部署开销过高。

因此，无线传输的波束成形和基站之间的 SFC 部署存在折中关系，需联合优化。

3 算法设计

问题 \mathcal{P}_0 可规约为一个无容量约束设施选址 (UFL, uncapacitated utility location) 问题，而 UFL 问题已被证明是一个 NP 难问题^[9]，因此， \mathcal{P}_0 也是一个 NP 难问题，很难找到多项式时间求解算法。通过观察所有约束，发现波束变量和功能部署变量只在式(7)相耦合。因此，本文采用拉格朗日对偶分解技术，通过将 $\mathbf{w}_{m,n}$ 和 $y_{l,n}^m$ 解耦合，从而将 \mathcal{P}_0 分解为 2 个独立的子问题。

3.1 拉格朗日对偶分解

首先，引入拉格朗日对偶乘子 $\boldsymbol{\lambda} = \{\lambda_{m,n}\}$ ，将式(7)叠加到 \mathcal{P}_0 的目标函数，即

$$\begin{aligned} \sum_{n,s} \sum_m \sum_{f_i^m \in \mathcal{F}(m)} z_{l,n,s}^m R_m + \eta \left(\sum_n \sum_{f \in \mathcal{F}_n} x_{f,n} P_{f,n} + \sum_n \sum_m \sum_{f_i^m \in \mathcal{F}(m)} y_{l,n}^m P_{l,n}^m \right) + \sum_n \sum_m (\eta P_{l,n} - P_n^{\text{th}} \lambda_{m,n}) y_{l,n}^m + \sum_n \sum_m (\eta + \lambda_{m,n}) \|\mathbf{w}_{m,n}\|_2^2 \end{aligned} \quad (12)$$

其中， $\bar{\mathcal{F}}(m) = \mathcal{F}(m) \setminus \{y_{l,n}^m\}$ 表示 $\mathcal{F}(m)$ 中除信号处理功能之外的所有其他功能集合。借助拉格朗日对偶方法，可将求解 \mathcal{P}_0 转化为求解对偶问题 \mathcal{P}_1 。

$$\begin{aligned} \mathcal{P}_1 : \max L(\boldsymbol{\lambda}) = L_1(\boldsymbol{\lambda}) + L_2(\boldsymbol{\lambda}) \\ \text{s.t. 式(1)~式(5), 式(9)} \\ \text{var } \lambda_{m,n} \geq 0, \forall m, \forall n \end{aligned} \quad (13)$$

其中，子问题 \mathcal{P}_2 和 \mathcal{P}_3 可分别描述为

$$\begin{aligned} \mathcal{P}_2: L_1(\lambda) = & \min \sum_{n,s} \sum_m \sum_{f_l^m \in \mathcal{F}(m)} z_{l,n,s}^m R_m + \eta \left(\sum_n \sum_{f \in \mathcal{F}_n} x_{f,n} P_{f,n} + \right. \\ & \left. \sum_n \sum_m \sum_{f_l^m \in \mathcal{F}(m)} y_{l,n}^m P_{l,n} \right) + \sum_n \sum_m (\eta P_{l,n} - P_n^{\text{th}} \lambda_{m,n}) y_{l,n}^m \\ \text{s.t. 式(1)~式(5)} \\ \text{var } x_{f,n} \in & \{0,1\}, \forall f \in \mathcal{F}_n, \forall n \\ y_{f,n}^m \in & \{0,1\}, \forall f \in \mathcal{F}(m), \forall m, \forall n, n \\ z_{l,n,s}^m \in & \{0,1\}, \forall f_l^m \in \mathcal{F}(m), \forall m, \forall n, s \end{aligned} \quad (14)$$

$$\begin{aligned} \mathcal{P}_3: L_2(\lambda) = & \min \sum_n \sum_m (\eta + \lambda_{m,n}) \| \mathbf{w}_{m,n} \|^2 \\ \text{s.t. 式(9)} \\ \text{var } \mathbf{w}_{m,n}, & \forall m, \forall n \end{aligned} \quad (15)$$

由于 $L(\lambda)$ 是关于变量 λ 的非光滑不可导的分段线性函数, 本文采用次梯度法^[11,16]求解 \mathcal{P}_1 。对于给定的对偶变量 λ , 次梯度法将独立求解 \mathcal{P}_2 和 \mathcal{P}_3 , 并迭代更新对偶变量, 迭代式如下。

$$\lambda_{m,n}^{[t+1]} = \left[\lambda_{m,n}^{[t]} + \alpha^{[t]} (\| \mathbf{w}_{m,n}^{[t]} \|^2 - P_n^{\text{th}} y_{L,n}^{m[t]}) \right]^+ \quad (16)$$

其中, $\| \mathbf{w}_{m,n}^{[t]} \|^2 - P_n^{\text{th}} y_{L,n}^{m[t]}$ 为第 t 次迭代的次梯度方向, $\alpha^{[t]}$ 为第 t 次迭代的步长。若 \mathcal{P}_2 和 \mathcal{P}_3 可被求解, 则迭代更新 λ , 以得到最优解。次梯度方法具体步骤如算法 1 所示。

算法 1 基于对偶分解的联合无线多播的 SFC 部署算法

初始化 定义 $t=0$, 给定最大迭代次数 T 和精度参数 ε_{tol} ; 令 $\lambda_{m,n}=0, \forall m, \forall n$

步骤 1 迭代次数更新, $t=t+1$

步骤 2 求解 \mathcal{P}_2 , 得到 SFC 部署变量的最优解 $y_{l,n}^{m*}$, 记作 $y_{l,n}^{m[t]}$

步骤 3 求解 \mathcal{P}_3 , 得到波束向量的最优解 $\mathbf{w}_{m,n}^*$, 记作 $\mathbf{w}_{m,n}^{[t]}$

步骤 4 根据式(16), 更新次梯度方向和对偶变量 循环执行步骤 1~步骤 4, 直到算法满足收敛

条件或达到最大迭代次数, 即 $\left| \frac{\lambda_{m,n}^{[t+1]} - \lambda_{m,n}^{[t]}}{\lambda_{m,n}^{[t]}} \right| \leq \varepsilon_{\text{tol}}$ 或 $t \geq T$ 成立。

然而, \mathcal{P}_2 和 \mathcal{P}_3 均难以求解。首先, \mathcal{P}_2 仍可规约为 NP 难的 UFL 问题; 其次, 尽管 \mathcal{P}_3 的优化目标是一个凸函数, 但式(9)是一个非凸约束。因此, 本文提出基于惩罚项的 SFC 部署算法和基于路径跟随播波束成形算法, 分别求解 \mathcal{P}_2 和 \mathcal{P}_3 。

3.2 基于惩罚项的 SFC 部署算法

因为 \mathcal{P}_2 为一个 ILP 问题, 将 \mathcal{P}_2 中的变量 \mathbf{x} 、 \mathbf{y} 和 \mathbf{z} 松弛, 则 \mathcal{P}_2 将成为一个 LP 问题。然而, \mathcal{P}_2 通常和其 LP 形式不等价, 求解 \mathcal{P}_2 对应的 LP 问题无法确保其最优解为二进制变量。而求解 ILP 问题通常采用分支定界法或割平面法^[21-22], 计算复杂度较高, 难以求解较大网络规模的 SFC 部署问题。因此, 本文采用基于 L_p ($0 < p < 1$) 范数的连续凸近似算法, 构造一个 L_p 范数惩罚项, 并将该惩罚项叠加到 \mathcal{P}_2 的目标函数, 以保证松弛问题的最优解为二进制变量。

首先, 定义向量 $\mathbf{y}_l^m = \{y_{l,n}^m\}_{n \in \mathcal{N}_l^m}$, 则式(1)可写为

$$\| \mathbf{y}_l^m \|_1 = 1, \forall f_l^m \in \mathcal{F}(m), \forall m \quad (17)$$

考虑如下 LP 优化问题

$$\begin{aligned} \max \quad & \| \mathbf{y}_l^m + \varepsilon \mathbf{I} \|_p^p \\ \text{s.t. } \quad & \| \mathbf{y}_l^m \|_1 = 1 \\ \text{var } \quad & 0 \leq y_{l,n}^m \leq 1 \end{aligned} \quad (18)$$

其中, $p \in (0,1)$ 且 ε 为任意正数。由文献[11]可知, 尽管式(18)的可行域已被松弛, 但其最优解仍为二进制变量。因此, 式(18)的最优值计算如下。

$$c_{\varepsilon, f_l^m} = (1+\varepsilon)^p + (|\mathcal{N}_l^m| - 1)\varepsilon^p \quad (19)$$

然后, 根据 L_p 范数的上述性质, 将一个基于 L_p 范数的惩罚项 $P_\varepsilon(\mathbf{y})$ 叠加到 \mathcal{P}_2 的目标函数, 并将变量松弛。则 \mathcal{P}_2 可转化为

$$\begin{aligned} \mathcal{P}_{2-L_p}: \min C_\sigma = & \sum_{n,s} \sum_m \sum_{f_l^m \in \mathcal{F}(m)} z_{l,n,s}^m R_m + \eta \left(\sum_n \sum_{f \in \mathcal{F}_n} x_{f,n} P_{f,n} + \right. \\ & \left. \sum_n \sum_m \sum_{f_l^m \in \mathcal{F}(m)} y_{l,n}^m P_{l,n}^m \right) + \sum_n \sum_m (\eta P_{l,n}^m - P_n^{\text{th}} \lambda_{m,n}) y_{l,n}^m + \\ & \underbrace{\sigma \sum_m \sum_{f_l^m \in \mathcal{F}(m)} (\| \mathbf{y}_l^m + \varepsilon \mathbf{I} \|_p^p - c_{\varepsilon, f_l^m})}_{P_\varepsilon(\mathbf{y})} \\ \text{s.t. 式(1)~式(5)} \\ \text{var } 0 \leq & x_{f,n} \leq 1, \forall f \in \mathcal{F}_n, \forall n \\ 0 \leq & y_{f,n}^m \leq 1, \forall f \in \mathcal{F}(m), \forall m, \forall n \\ 0 \leq & z_{l,n,s}^m \leq 1, \forall f_l^m \in \mathcal{F}(m), \forall m, \forall n, s \end{aligned} \quad (20)$$

其中, σ 为惩罚参数。由式(20)可知, 若最优解 $y_{f,n}^{m*} \in \{0,1\}$, 则由于约束式(2)的存在和 \mathcal{P}_{2-L_p} 的最小化性质, 有 $x_{f,n}^* \in \{0,1\}$ 成立。同理可得, 由于约束式(5)的存在, 有 $z_{l,n,s}^{m*} \in \{0,1\}$ 成立。因此, 针对 \mathcal{P}_{2-L_p}

的最优解 $U^* = \{x^*, y^*, z^*\}$ ，本文给出等价性定理如定理 1 所示，证明当 $\sigma \rightarrow \infty$ 时， U^* 也构成 \mathcal{P}_2 的最优解。

定理 1 对于任意正数 ε ，令 $\{\sigma_v\}$ 和 U^v 分别表示单调递增的惩罚参数序列和 \mathcal{P}_{2-L_p} 的最优解。若 $\sigma_v \rightarrow \infty$ ，则 U^v 也构成 \mathcal{P}_2 的最优解。

证明 见附录。

由于惩罚项 $P_\varepsilon(\mathbf{y})$ 的存在， \mathcal{P}_{2-L_p} 是一个非凸问题。尽管其可行域是一个凸集，但是优化目标是一个凹函数。因此，本文采用连续凸近似算法，求出原目标函数的一阶泰勒展开式，将 \mathcal{P}_{2-L_p} 转化为一个凸优化问题。针对惩罚项 $P_\varepsilon(\mathbf{y})$ ，其一阶泰勒近似函数可表示为

$$P_\varepsilon(\mathbf{y}) \leq P_\varepsilon(\mathbf{y}^v) + \nabla P_\varepsilon(\mathbf{y}^v)^\top (\mathbf{y} - \mathbf{y}^v) \quad (21)$$

其中， \mathbf{y}^v 表示第 $v+1$ 次迭代的初始可行解， $\nabla P_\varepsilon(\mathbf{y}^v)$ 表示 $P_\varepsilon(\mathbf{y})$ 在 \mathbf{y}^v 处的梯度方向。因此， \mathcal{P}_{2-L_p} 可转化为

$$\begin{aligned} \mathcal{P}_{2-S}: \quad & \min C + \sigma_{v+1} \nabla P_\varepsilon(\mathbf{y}^v)^\top \mathbf{y} \\ & \text{s.t. 式(1)~式(5)} \\ & \text{var } 0 \leq x_{f,n} \leq 1, \forall f \in \mathcal{F}_n, \forall n \\ & 0 \leq y_{f,n}^m \leq 1, \forall f \in \mathcal{F}(m), \forall m, \forall n, n \\ & 0 \leq z_{l,n,s}^m \leq 1, \forall f_l^m \in \mathcal{F}(m), \forall m, \forall n, s \end{aligned} \quad (22)$$

至此，得到一个 LP 问题 \mathcal{P}_{2-S} ，该问题可用单纯形法或内点法^[22]高效求解。基于惩罚项的 SFC 部署算法的总体流程如算法 2 所示。

算法 2 基于惩罚项的 SFC 部署算法

初始化 定义 $v=0$ ，给定最大迭代次数 V ；定义 σ_1 和 ε_1 ；定义乘数因子 $0 < o < 1 < \varrho$

步骤 1 $\varepsilon_{v+1} = o\varepsilon_v$ 且 $\sigma_{v+1} = \varrho\sigma_v$

步骤 2 求解 \mathcal{P}_{2-S} ，得到最优解 \mathbf{y}^{v+1}

步骤 3 迭代次数更新 $v = v+1$

循环执行步骤 1~步骤 3，直到算法满足收敛条件或达到最大迭代次数，即 $\mathbf{y}^{v+1} \in \{0, 1\}$ 或 $v \geq V$ 成立。

3.3 基于路径跟随的波束成形算法

由式(8)可知，吞吐量 $R_{k,n}$ 是关于变量 \mathbf{w} 的非凸函数。若能将 $R_{k,n}(\mathbf{w})$ 转化为一个凹函数，则式(9)将构成一个凸约束。文献[23]提出了一种路径跟随技术，求出 $R_{k,n}(\mathbf{w})$ 的一个下界凹函数，记为 $R_{k,n}^{[v]}(\mathbf{w})$ (与算法 2 类似，同样用 v 表示迭代次数)。则式(9)可转化为

$$\sum_{n \in \mathcal{N}} R_{k,n}^{[v]}(\mathbf{w}) \geq R_m, \forall k \in \mathcal{G}_m, \forall m \quad (23)$$

首先，列出路径跟随算法的下界凹函数不等式，即

$$\begin{aligned} \text{lb} \left(1 + \frac{1}{ab} \right) & \geq \text{lb} \left(1 + \frac{1}{\bar{a}\bar{b}} \right) + \\ & \frac{1}{\bar{a}\bar{b}} \left(2 - \frac{a}{\bar{a}} - \frac{b}{\bar{b}} \right) \\ & \frac{1}{1 + \frac{1}{\bar{a}\bar{b}}} \end{aligned} \quad (24)$$

$$\forall a > 0, b > 0, \bar{a} > 0, \bar{b} > 0$$

然后，进行变量替换，即

$$a = \frac{1}{|\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}|^2}, \bar{a} = \frac{1}{|\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}^{[v]}|^2} \quad (25)$$

$$b = \sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma_k^2, \bar{b} = \sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}^{[v]}|^2 + \sigma_k^2 \quad (26)$$

则原吞吐量 $R_{k,n}(\mathbf{w})$ 可转化为

$$\begin{aligned} R_{k,n}^{[v]}(\mathbf{w}) & = B \text{lb} \left(1 + \alpha_{k,n}^{[v]} \right) + \frac{\alpha_{k,n}^{[v]}}{1 + \alpha_{k,n}^{[v]}} \left(2 - \right. \\ & \left. \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}^{[v]}|^2}{2\Re\{(\mathbf{w}_{m,n}^{[v]})^H \mathbf{h}_{k,n} \mathbf{h}_{k,n}^H \mathbf{w}_{m,n}\}} - |\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}^{[v]}|^2 \right. \\ & \left. \frac{\sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma_k^2}{\sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}^{[v]}|^2 + \sigma_k^2} \right) \end{aligned} \quad (27)$$

其中， $\alpha_{k,n}^{[v]} = \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}^{[v]}|^2}{\sum_{j \neq m} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}^{[v]}|^2 + \sigma_k^2}$ 。同时，为保证式(27)

中第三项的分母大于 0，设置一个信任域，使 $\mathbf{w}_{m,n}$ 满足如下约束。

$$\begin{aligned} 2\Re\{(\mathbf{w}_{m,n}^{[v]})^H \mathbf{h}_{k,n} \mathbf{h}_{k,n}^H \mathbf{w}_{m,n}\} - |\mathbf{h}_{k,n}^H \mathbf{w}_{m,n}^{[v]}|^2 & > 0, \\ \forall k \in \mathcal{G}_m, \forall m, \forall n \end{aligned} \quad (28)$$

至此，优化问题 \mathcal{P}_3 可转化为

$$\begin{aligned} \mathcal{P}_{3-S}: \quad L_2(\lambda) & = \min \sum_n \sum_m (\eta + \lambda_{m,n}) \|\mathbf{w}_{m,n}\|_2^2 \\ & \text{s.t. 式(23), 式(27)} \\ & \text{var } \mathbf{w}_{m,n}, \forall m, n, \end{aligned} \quad (29)$$

由 $R_{k,n}^{[v]}(\mathbf{w})$ 的定义可知， $\mathbf{w}^{[v]}$ 满足式(23)和式(28)，自然也是优化问题 \mathcal{P}_{3-S} 的一个可行解。因此，第 $v-1$ 次迭代的最优解 $\mathbf{w}^{[v]}$ 可作为第 v 次迭代的初始解。在第 v 次迭代中，构造一个 $\mathbf{w}^{[v]}$ 的可行域，由初始解 $\mathbf{w}^{[v]}$ 开始搜索，得到一个最优解 $\mathbf{w}^{[v+1]}$ 。因此，随着迭代次数的增加， \mathcal{P}_{3-S} 的目标函数将单调递减。

同时, 因为 $R_{k,n}^{[v]}(\mathbf{w})$ 是原吞吐量 $R_{k,n}(\mathbf{w})$ 的一个下界凹函数, 所以 $R_k(\mathbf{w}^{[v+1]}) \geq R_m$ 必然成立, 即每次迭代得到的最优解 $\mathbf{w}^{[v+1]}$ 是式(9)中最小吞吐量约束的一个可行解。

3.4 计算复杂度分析

在基于惩罚项的 SFC 部署算法(即算法 2)中, 每次迭代需要用内点法求解子问题 $\mathcal{P}_{2,S}$ 。经观察, 变量个数主要由 $z_{l,n,s}^m$ 决定。文献[24]使用内点法求解 $\mathcal{P}_{2,S}$ 的计算复杂度为 $\Phi_1 = \mathcal{O}((MN^2F)^3)$, 其中, M 是多播组的组数, N 是边缘网络中所有基站的数量, F 为边缘网络中所有网络功能的数目, MN^2F 为 $z_{l,n,s}^m$ 的变量数目。

同理, 基于路径跟随的波束成形算法的计算复杂度^[23]可表示为

$$\Phi_2 = \mathcal{O}((MNI)^2((N+1)K)^{2.5} + ((N+1)K)^{3.5}) \quad (30)$$

其中, MNI 为波束向量中的变量数目, $(N+1)K$ 为式(23)和式(27)中的约束数目。

假设算法 2 和波束成形算法的最大迭代次数分别为 $V_{\max 1}$ 和 $V_{\max 2}$, 次梯度法最大迭代次数为 T_{\max} 。则本文算法的计算复杂度为

$$\mathcal{O}(T_{\max}(V_{\max 1}\Phi_1 + V_{\max 2}\Phi_2)) \quad (31)$$

4 仿真分析

为验证本文算法的有效性, 首先分别验证基于惩罚项的 SFC 部署算法、基于凹函数近似的波束成形算法和次梯度方法的收敛性能。其次, 将本文算法与文献[9]中的最优单播 SFC 部署算法和文献[15]中的随机多播 SFC 部署算法进行对比。文献[15]中并未考虑 SFC 的最优部署, 所以将文献[15]中的多播波束成形机制和 SFC 的随机部署相结合。

本文采用的仿真工具是 MATLAB R2014b, 并使用了凸优化工具包 CVX。网络模型为一个经典的六边形七小区蜂窝网络^[15,25], 基站间隔设为 200 m, 每个基站配置 10 根天线和 5 MHz 频谱带宽, 基站最大发射功率为 46 dBm。针对信道模型, 用户和基站之间的路径损耗模型为 $PL = 32.45 + 10\lg d$, PL 单位为 dB, d 单位为 m, 小尺度衰落模型服从瑞利分布, 对数阴影衰落设置为 8 dB, 信道热噪声功率谱密度为 -174 dBm/Hz。针对内容缓存, 使用 U-vMOS 视频体验标准^[12]的前 5 种等级, 假设系统存在 5 种不同内容, 其吞吐量需求分别

为 0.5 Mbit/s、1 Mbit/s、4 Mbit/s、5 Mbit/s 和 10 Mbit/s; 每个基站的 MEC 服务器随机选取 2 项内容进行缓存。针对内容缓存之外的其他功能, 假设系统中存在 6 项功能, 每个 MEC 服务器随机选取 3 项功能, 每项功能的处理容量服从 5~15 Mbit/s 的随机均匀分布, 每项功能的维护功耗和服务功耗服从 1~4 W 的均匀分布。针对有限链路, 任意 2 个基站之间的链路带宽服从 10~50 Mbit/s 的均匀分布。

算法参数设置中, 算法 1 中次梯度方法的初始步长设为 100, 并按照 0.4 的衰减系数逐次衰减; 算法 2 中, ε 和 σ 的初始值分别设为 0.01 和 10, 乘数因子 $\rho = 0.8$, $\varrho = 3$ 。

图 2 给出了不同折中系数 η 下的基于惩罚项的 SFC 部署算法的收敛性能。设用户数为 20, 每个用户等概率请求一项内容服务, 每项服务的 SFC 中均包含 4 项有序功能。为衡量算法性能, 使用相邻两次 \mathcal{P}_{2,L_p} 问题的最优解的范数差, 即 $\|U^{v+1} - U^v\|_2$, 作为性能指标。如图 2 所示, 在不同 η 设置下, 本文算法均具有良好收敛性能, 在迭代 10 次之内达到稳定值。

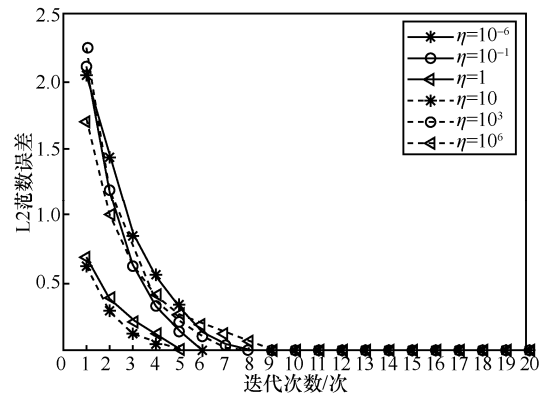


图 2 基于惩罚项的 SFC 部署算法的收敛性能

图 3 给出了不同折中系数 η 下的基于路径跟随的波束成形算法的收敛性能。设用户数为 20, 且等概率请求一项内容服务, 每项服务的 SFC 均包含 4 项有序功能。因为该部分算法仅涉及了基站的波束成形设计, 所以使用无线传输功耗作为性能指标。由图 3 可看出, 在不同 η 设置下, 本文算法均在迭代 15 次之内收敛, 且功耗值单调递减。这也验证了 3.3 节有关算法单调性分析的正确性。

图 4 给出了不同折中系数 η 下的次梯度方法的收敛性能。用户数和 SFC 的设置均与图 2 和图 3 一致。为公平衡量算法性能, 克服 η 过大或过小对问

题 \mathcal{P}_0 的优化目标的影响, 仅使用系统功耗开销作为性能指标。首先, 由图 4 可知, 尽管次梯度方法相较于基于惩罚项的 SFC 部署算法和基于路径跟随的波束成形算法的收敛性能较差, 但均在迭代 30 次左右趋于稳定值。其次, 与图 3 相比, 系统总功耗开销和无线传输功耗开销之间的差值表现为功能维护功耗开销和功能服务功耗开销。最后, 随着 η 的增大, 系统总功耗开销逐渐降低。这是因为原始优化问题 \mathcal{P}_0 的优化目标为最小化系统开销, η 的增大将使功耗开销的权重增大, 数据流开销的权重减小, 从而使最优值中的功耗部分降低。

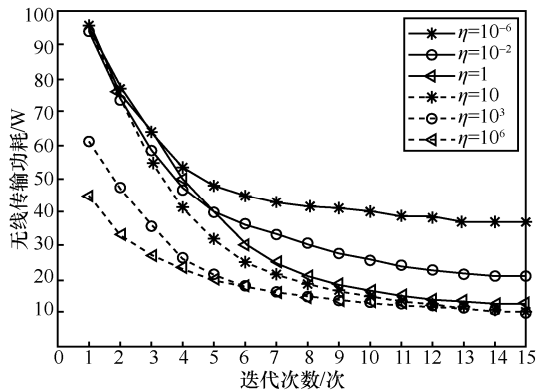


图 3 基于路径跟随的波束成形算法的收敛性能

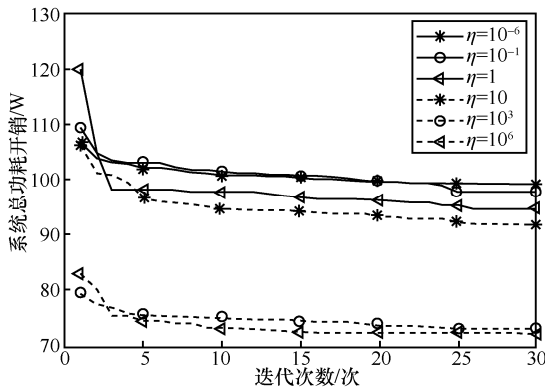


图 4 次梯度方法的收敛性能

图 5 给出了不同折中系数 η 下算法 1 功耗开销-数据流开销折中曲线。与图 4 结论类似, 当 $\eta=10^6$ 时, 可认为 η 趋近于无穷大, 系统开销只考虑功耗开销, \mathcal{P}_0 简化为功耗开销最小化问题, 而算法 1 成为功耗开销最小化算法。因此, 优化目标中不考虑数据流开销, $\eta=10^6$ 时数据流开销超过了 60 Mbit/s。在 $\eta=10^{-6}$ 时, 可认为 η 趋近于无穷小, 系统开销只考虑数据流开销, \mathcal{P}_0 简化为数据流

最小化问题, 优化目标中不涉及功耗开销, 故 $\eta=10^{-6}$ 时功耗开销达到 105 W。图 4 和图 5 的结论为合理、灵活选择折中系数 η 提供了理论依据和经验参考。

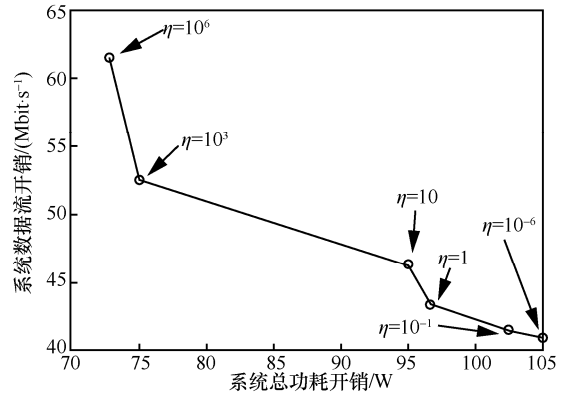


图 5 算法 1 的功耗开销-数据流开销折中曲线

图 6 给出了本文算法和比较算法在不同用户数下的系统总功耗性能。SFC 的设置均与图 2~图 4 一致, $\eta=10^3$ 。由图 6 可看出, 本文算法性能优于最优单播 SFC 部署算法。这是因为, 若为每个用户独立分配一个波束向量, 将显著增大基站的无线传输功耗, 使系统总功耗提升。同时, 在只考虑系统功耗开销时, 随机多播 SFC 部署算法的性能优于本文算法。这是因为, $\eta=10^3$ 时本文算法需折中考虑功耗开销和数据流开销, 最优值中的数据流开销部分将导致功耗开销部分增大; 而随机多播 SFC 部署算法不需要考虑数据流开销。

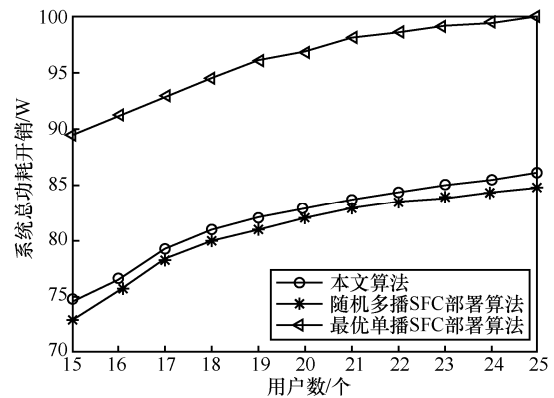


图 6 系统总功耗与用户数之间关系

图 7 给出了本文算法和比较算法在不同用户数下的数据流开销性能。SFC 的设置均与图 2~图 4 一致, $\eta=10^3$ 。如图 7 所示, 曲线呈分段线性的折线形式, 这是因为, 数据流开销需取 U-vMOS 吞吐

量需求的整数倍,无法像功率分配一样取任意连续值。在只考虑系统数据流开销时,最优单播 SFC 部署算法的性能优于本文算法。这是因为, $\eta = 10^3$ 时本文算法需折中考虑功耗开销和数据流开销,最优值中的功耗开销部分同样会导致数据流开销部分的增大;而最优单播 SFC 部署算法不需要考虑功耗开销。

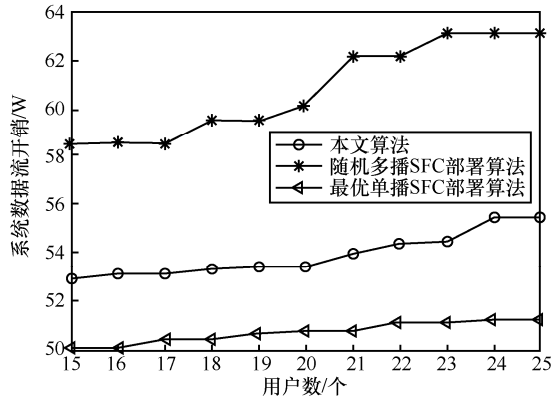


图 7 系统数据流开销与用户数之间关系

5 结束语

本文提出了一种面向内容的联合无线多播的 SFC 部署算法,建立了波束成形设计和 SFC 映射的联合优化模型。利用拉格朗日对偶分解技术,将优化问题转化为 SFC 部署和波束成形设计 2 个独立子问题;利用基于 L_p 范数惩罚项的连续凸近似算法,将整数形式的 SFC 部署问题松弛为一个等价线性规划问题,并给出了最优解的等价性证明;利用路径跟随技术,将非凸波束向量优化问题转化为一组凸优化问题,并给出了算法单调性分析。仿真结果表明,所提算法在系统开销方面优于最优单播 SFC 部署算法和随机多播 SFC 部署算法。未来工作中,将进一步研究联合协作多播波束成形的多服务路径 SFC 部署问题。

附录 定理 1 证明

为符号描述简便,将式(10)中 \mathcal{P}_{2-L_p} 的优化目标写为 $C(\mathbf{U}) = C_\sigma(\mathbf{U}) - P_\epsilon(\mathbf{y})$ 。令 v 表示迭代次数。当惩罚参数为 σ_v 时, \mathbf{U}^v 是 \mathcal{P}_{2-L_p} 的最优解。因此,有式(32)和式(33)成立。

$$C_{\sigma_v}(\mathbf{U}^v) \leq C_{\sigma_v}(\mathbf{U}^{v+1}), \forall v \quad (32)$$

$$C_{\sigma_{v+1}}(\mathbf{U}^{v+1}) \leq C_{\sigma_{v+1}}(\mathbf{U}^v), \forall v \quad (33)$$

同时,已知 $\sigma_v \leq \sigma_{v+1}$ 成立,则由式(32)和式(33)可得

$$\sigma_v(P_\epsilon(\mathbf{y}^v) - P_\epsilon(\mathbf{y}^{v+1})) \leq$$

$$C(\mathbf{U}^{v+1}) - C(\mathbf{U}^v) \leq \sigma_{v+1}(P_\epsilon(\mathbf{y}^v) - P_\epsilon(\mathbf{y}^{v+1})), \forall v \quad (34)$$

由式(34)可知,随着 v 增大, $C(\mathbf{U}^v)$ 逐步增大,而惩罚项 $P_\epsilon(\mathbf{y}^v)$ 逐步减小。

令 \mathbf{U}^* 表示原问题 \mathcal{P}_2 的最优解,将 \mathbf{U}^* 代入 \mathcal{P}_{2-L_p} ,则 \mathcal{P}_{2-L_p} 中的惩罚项 $P_\epsilon(\mathbf{y}^*) = 0$ 。同时,已知 \mathbf{U}^v 是惩罚项为 σ_v 时 \mathcal{P}_{2-L_p} 的最优解,则有式(35)成立。

$$C(\mathbf{U}^v) + \sigma_v P_\epsilon(\mathbf{y}^v) \leq C_{\sigma_v}(\mathbf{U}^*) = C(\mathbf{U}^*), \forall v \quad (35)$$

由式(35)可知,随着 v 增大,有 $\sigma_v P_\epsilon(\mathbf{y}^v) \rightarrow 0$ 和 $P_\epsilon(\mathbf{y}^v) \rightarrow 0$ 成立。因此,不妨设序列 $\{\mathbf{U}^v\}$ 的一个极限点为 \mathbf{U}^V ,则有 $V \rightarrow \infty$ 时, $P_\epsilon(\mathbf{y}^V) = 0$ 一定成立,且 \mathbf{y}^V 是二进制变量。同时,由于式(2)和式(7)的存在以及 \mathcal{P}_{2-L_p} 目标函数的最小化性质, \mathbf{x}^V 和 \mathbf{z}^V 也必然是二进制变量。因此, $\mathbf{U}^V = \{\mathbf{x}^V, \mathbf{y}^V, \mathbf{z}^V\}$ 是二进制变量且满足 $C(\mathbf{U}^V) = C(\mathbf{U}^*)$,说明 \mathbf{U}^V 构成原问题 \mathcal{P}_2 的一个最优解。

证毕。

参考文献:

- [1] 周一青, 李国杰. 未来移动通信系统中的通信与计算融合[J]. 电信科学, 2018, 34(3): 1-7.
ZHOU Y Q, LI G J. Convergence of communication and computing in future mobile communication systems[J]. Telecommunications Science, 2018, 34(3): 1-7.
- [2] 孙晨, 毕军, 郑智隆, 等. MicroNF: 基于微服务的异构网络功能虚拟化框架[J]. 通信学报, 2019, 40(8): 54-59.
SUN C, BI J, ZHENG Z L, et al. MicroNF: a microservice-based hybrid framework for NFV[J]. Journal on Communications, 2019, 40(8): 54-59.
- [3] SONG S, LEE C, CHO H, et al. Clustered virtualized network functions resource allocation based on context-aware grouping in 5G edge networks[J]. IEEE Transactions on Mobile Computing, 2020, 19(5): 1072-1083.
- [4] WANG Y, SHENG M, WANG X, et al. Mobile-edge computing: partial computation offloading using dynamic voltage scaling[J]. IEEE Transactions on Communications, 2016, 64(10): 4268-4282.
- [5] QIU X, LI H, WU C, et al. Cost-minimizing dynamic migration of content distribution services into hybrid clouds[J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(12): 3330-3345.
- [6] TAO O, ZHI Z, XU C. Follow me at the edge: mobility-aware dynamic service placement for mobile edge computing[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(10): 2333-2345.
- [7] TRAN T X, HAJISAMI A, PANDEY P, et al. Collaborative mobile edge computing in 5G networks: new paradigms, scenarios, and challenges[J]. IEEE Communications Magazine, 2017, 55(4): 54-61.
- [8] CHEN Y T, LIAO W. Mobility-aware service function chaining in 5G wireless networks with mobile edge computing[C]//2019 IEEE International Conference on Communications. Piscataway: IEEE Press, 2019: 1-6.
- [9] ZHENG G, TSIPOPOULOS A, FRIDERIKOS V. Optimal VNF chains

- management for proactive caching[J]. IEEE Transactions on Wireless Communications, 2018, 17(10): 6735-6748.
- [10] YU R, XUE G, ZHANG X. QoS-aware and reliable traffic steering for service function chaining in mobile networks[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(11): 2522-2531.
- [11] ZHANG N, LIU Y F, FARMANBAR H, et al. Network slicing for service-oriented networks under resource constraints[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(11): 2512-2521.
- [12] LIANG C, HE Y, YU F R, et al. Enhancing video rate adaptation with mobile edge computing and caching in software-defined mobile networks[J]. IEEE Transactions on Wireless Communications, 2018, 17(10): 7013-7026.
- [13] WANG K, YU F R, LI H. Information-centric virtualized cellular networks with device to device communications[J]. IEEE Transactions on Vehicular Technology, 2016, 65(11): 9319-9329.
- [14] ZHOU Y, LIU H, PAN Z, et al. Two-stage cooperative multicast transmission with optimized power consumption and guaranteed coverage[J]. IEEE Journal on Selected Areas in Communications, 2014, 32(2): 274-284.
- [15] TAO M, CHEN E, ZHOU H, et al. Content-centric sparse multicast beamforming for cache-enabled cloud RAN[J]. IEEE Transactions on Wireless Communications, 2016, 15(9): 6118-6131.
- [16] TANG J, SHIM B, QUEK T Q S. Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(4): 881-895.
- [17] TANG J, QUEK T Q S, Chang T, et al. Systematic resource allocation in cloud RAN with caching as a service under two timescales[J]. IEEE Transactions on Communications, 2019, 67(11): 7755-7770.
- [18] LIU L, ZHOU Y, GARCIA V, et al. Load aware joint CoMP clustering and inter-cell resource scheduling in heterogeneous ultra dense cellular networks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(3): 2741-2755.
- [19] GARCIA V, ZHOU Y, SHI J, et al. Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering[J]. IEEE Transactions on Wireless Communications, 2014, 13(8): 4297-4308.
- [20] PU L, JIAO L, CHEN X, et al. Online resource allocation, content placement and request routing for cost-efficient edge caching in cloud radio access networks[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(8): 1751-1767.
- [21] MIRCHANDANI P B, FRANCIS R L. Discrete location theory[M]. New Jersey: John Wiley & Sons, 1990.
- [22] BOYD S, VANDENBERGHE, L. Convex optimization[M]. Cambridge: Cambridge University Press, 2004.
- [23] SHENG Z, TUAN H D, DUONG T Q, et al. Beamforming optimiza-

tion for physical layer security in MISO wireless networks[J]. IEEE Transactions on Signal Processing, 2018, 66(14): 3710-3723.

- [24] El-ABSI M, SHAAT M, BADER F, et al. Interference alignment with frequency-clustering for efficient resource allocation in cognitive radio networks[J]. IEEE Transactions on Wireless Communications, 2015, 14(12): 7070-7082.
- [25] LIU Y, ZHOU Y, YUAN J, et al. Economically optimal MS association for multimedia content delivery in cache-enabled heterogeneous cloud radio access networks[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(7): 1584-1593.

[作者简介]



王侃（1985-），男，山东济宁人，博士，西安理工大学讲师，主要研究方向为无线网络虚拟化、无线资源分配等。



赵楠（1982-），男，辽宁大连人，博士，大连理工大学教授、博士生导师，主要研究方向为无人机通信、边缘计算等。



李军怀（1969-），男，陕西宝鸡人，博士，西安理工大学教授、博士生导师，主要研究方向为物联网技术、云计算等。



王怀军（1981-），男，山东滕州人，博士，西安理工大学讲师，主要研究方向为智能感知、行为识别、智慧水利等。