

DOI:10.20033/j.1003-7241.(2026)04-0146-10

IMS 行政交换 NFV 核心网的网元资源配置方法

柳旭,梅增杨,束一,贾平,祁步仁

(国网江苏省电力有限公司信息通信分公司,江苏南京 210000)

摘要:网络功能虚拟化技术将电信网功能网元以虚拟化的形式部署在通用硬件服务器上,减少对特定硬件的依赖。IMS 行政交换 NFV 核心网在设计轻量级组网方案时,行政交换用户需求与 NFV 核心网算力资源、NFV 核心网元合理配置的适配成为关键问题。针对网络功能虚拟化环境下 IMS 行政交换网的核心网元资源配置问题,为提高网络灵活性、可靠性及资源利用率,提出一种混合组网轻量级优化方法。首先,构建虚拟网元与物理网元的混合架构,结合最小二乘法拟合用户规模与资源需求的动态关系模型;其次,基于贪心算法设计单节点最简配置方案,利用递归算法实现分级冗余部署;最后,通过多活实例同步与动态扩容技术优化双节点冗余配置。实验结果表明,所提方法显著优于传统静态配置,单节点最简配置降低硬件成本 35%,双节点冗余配置在资源利用率损失仅 4%~6%的条件下实现 99.99%可用性,起到了良好的优化效果。

关键词:IMS 行政交换网;网络功能虚拟化;资源配置;多媒体子系统;轻量级 NFV

中图分类号: TP393

文献标志码: A

文章编号: 1003-7241(2026)04-0146-10

Network element resource allocation method of IMS administrative switching NFV core network

LIU Xu, MEI Zengyang, SHU Yi, JIA Ping, QI Buren

(Information and Communication Branch, State Grid Jiangsu Electric Power Company, Nanjing 210000, Jiangsu, China)

Abstract: Network function virtualization (NFV) technology deploys telecommunications network function elements in a virtualized form on general-purpose hardware servers, reducing reliance on specialized hardware. When designing lightweight networking solutions for the IMS administrative switching NFV core network, the compatibility between administrative switching user requirements with the rational allocation of NFV core network computing resources and NFV core network elements becomes a critical issue. To address core network element resource allocation in IMS administrative switching networks under network function virtualization, this paper proposes a lightweight hybrid networking optimization scheme to enhance network flexibility, reliability, and resource utilization. First, a hybrid architecture integrating virtual network elements functions and physical network elements function is established, with a least 2 squares method modeling relationships. Next, a greedy algorithm designs minimal configurations for single nodes, while recursive algorithms achieve hierarchical redundancy deployment. At last, dual-node redundancy is optimized through multi-active instance synchronization and dynamic scaling. Numerical results demonstrate superior performance over static configurations, single-node configurations reduce hardware costs by 35%, while dual-node redundancy maintains 99.99% availability with only 4%~6% resource efficiency loss, achieving effective optimization.

Keywords: IMS administrative exchange networks; network function virtualization; resource allocation; multimedia subsystem; light weight NFV

IP 多媒体子系统 (IP multimedia subsystem, IMS) 是一个由 3GPP 标准化的架构^[1],旨在通过 IP 网络提供融合的多媒体服务,其作为下一代网络的核心技术^[2],推动电信行业向全 IP 化、服务融合的方向发展。IMS 网络主流设备采用 ATCA 架构专用硬件^[3],具有高性价比、模块化、兼容性强、可拓展等特点,使得这种硬件架构广泛应用于大型电信运营企业的现网部署^[4]。尽管有成熟的技术与强大的能力,但这种硬件架构与国家电网公司行政、调度交换网需求有所不同,国家电网公司服务于行业,其应用

人群规模固定且需求较为单一,这使得为电信运营企业设计的硬件设备长期处于空闲状态,硬件使用率偏低,造成了硬件资源及能耗上的浪费^[5]。

网络功能虚拟化 (network function virtualization, NFV) 的核心理念是通过虚拟化技术将网络功能从专用硬件迁移至云化基础设施^[6]。采用虚拟化的技术构建核心网,可将电信网功能网元以虚拟化的形式部署在通用硬件服务器上,从而以一种硬件资源池的方式提供电信网网元功能所需的计算、存储、网络资源,减少对特定硬件的依赖。使

收稿日期:2025-02-24;录用日期:2025-04-11

基金项目:国网江苏省电力有限公司科技项目 (J2024082)

作者简介:柳旭 (1992—),男,工程师,研究方向:电力信息通信技术。

引用本文:柳旭,梅增杨,束一,等. IMS 行政交换 NFV 核心网的网元资源配置方法 [J]. 自动化技术与应用, 2026,45(4):146-154,168. (LIU Xu, MEI Zengyang, SHU Yi, et al. Network element resource allocation method of IMS administrative switching NFV core network [J]. Techniques of Automation and Applications, 2026,45(4):146-154,168.)

用通用服务器完成原有专用设备的功能,降低了硬件的成本,在资源调度方面可以充分利用虚拟化的优势,实现资源的按需分配,提高设备运行效率,降低运行成本^[7]。结合云计算及大数据分析,可进一步提高云平台的信息融合度和资源配置能力^[8],提升资源调度的均衡性,并实现合理的信息化管理^[9]。

IMS演进的方向是在现有IMS网络的基础上,基于NFV技术,进一步融入IT领域的创新成果,构建一张云化的IMS核心网。与传统的IMS网络节点相比,基于NFV的虚拟多媒体子系统网络功能建立在开放的虚拟化平台上,从而避免了对硬件的依赖性,实现了对物理资源的按需分配,可以提供更高的网络部署灵活性。然而,电力行业IMS行政交换网采用轻量化核心网和省级集中部署方式^[10],在用户容量、组网模式等方面与运营商IMS核心网存在较大差异,适配电力IMS的云化部署方案存在难点和挑战。

尤其是由于不同业务场景往往具有差异化的服务质量(quality of service, QoS)需求(例如延迟、带宽和安全性等QoS需求),当多个虚拟网络功能(virtualized network function, VNF)合并部署在同一个服务器上时,为了处理具有不同性能的网络需求,VNF之间往往会因为资源竞争而产生性能干扰,从而导致相应网元节点吞吐量下降、延迟增加^[11]。此外,网络运营商需要根据线路内并发用户数初步预判网元资源需求,进而根据业务需求动态调整VNF的服务实例数量。具体而言,当已部署的VNF服务实例无法提供充足的处理能力时,需要基于可用网元节点创建新的VNF服务实例来满足用户的流量请求,同时,当某节点上的VNF出现故障时,需要实时进行VNF服务迁移与容错恢复,以保证服务的连续性,即为网元节点配置一定冗余度,以提高系统鲁棒性^[12]。

目前,国内外围绕电力IMS行政交换网NFV平台部署、物资供应^[13]、灾备模式^[14]、资源调配^[15]和运维方案等方面的研究已取得一定进展,而具体有关行政交换用户需求与NFV核心网算力资源适配分析及建模和适配行政交换网的NFV核心网元合理配置问题仍少有学者展开进一步探讨。本文以IMS行政交换核心网为研究对象,针对NFV环境下的网元资源配置问题,提出一种轻量化混合组网优化方案。核心贡献包括:

- 1) 设计NFV混合组网架构,用于实现流量管理、资源配置、用户割接等重要功能。
- 2) 建立网元资源需求预测模型,用于分析数据分布特征并实时计算核心网资源需求。
- 3) 设计分级冗余策略,结合贪心算法和递归算法,为核心网元节点配置合理冗余。
- 4) 设计多活实例同步与动态扩容融合机制,优化服务持续能力以及系统鲁棒性。

1 系统模型

1.1 网络架构

考虑一个采用单网元-单虚机的核心网,其架构为基

于NFV的IMS行政交换混合组网,拓扑结构如图1所示。设核心网使用二层交换机进行东西流量汇聚,VNF间通信通过二层交换机的VLAN隔离,同时采用三层路由器做南北流量转发,对接到CE路由器,用户接入流量通过三层路由器的策略路由(policy-based routing, PBR)分至不同网元池;物理网络功能(physical network function, PNF)与VNF分别组pool以负责不同区域的业务,二者的OMC均独立设置;用户割接^[16]时新用户优先分配至VNF pool,旧用户分批次迁移。

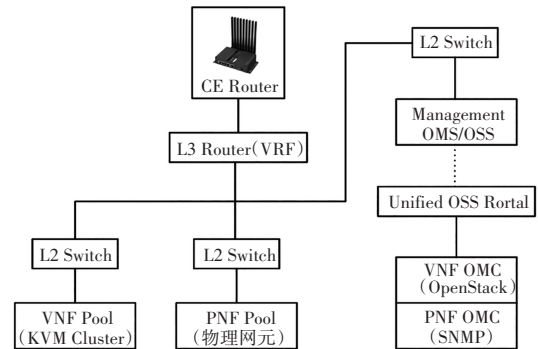


图1 核心网拓扑结构

Fig. 1 Core network topology structure

1.2 核心网元资源需求计算方法

在IMS行政交换轻量化NFV核心网中,呼叫会话控制功能(call session control function, CSCF)、出口网关控制功能(breakout gateway control function, BGCF)、媒体网关控制功能(media gateway control function, MGCF)和归属签约用户服务器(home subscriber server, HSS)等主要功能实体的节点包含多个网元^[17],定义 $X \triangleq \{x_1, x_2, \dots, x_N\}$ 表示系统中网元的集合,其中, x_n 表示系统中第 n 个网元, N 表示网元总个数。

NFV技术允许网络资源根据需求动态调整,当用户数增加时,核心网可以自动扩展以支持更多呼叫^[18]、避免容错瓶颈,因此将核心网用户数作为决策变量,定义 $u \triangleq \{u_1, u_2, \dots, u_M\}$ 表示核心网用户数的样本集合,其中, u_m 表示第 m 个样本对应的用户数, M 表示样本总数。

定义 $R \triangleq \{r_1^n(u_m), r_2^n(u_m), \dots, r_L^n(u_m)\}$ 表示当核心网用户数为 u_m 时,网元 x_n 对系统软硬件资源(CPU、内存、网络带宽、虚机参数等)的需求集合,其中, $r_l^n(u_m)$ 表示核心网元 x_n 对第 l 种资源的需求, L 表示资源种数。

假设系统中网元、用户数及网元资源需求根据测试和数据收集先验已知,且测试周期足够长使得数据具有明显分布特征,为了建立NFV网元需求与用户数关系并进一步预判在特定用户数下的网元资源需求,可通过最小二乘法对网元资源需求函数进行近似。首先,构造如下的 K 次多项式。

$$f_l^n(u_m) = \theta_0^{n,l} + \theta_1^{n,l} u_m + \dots + \theta_K^{n,l} (u_m)^K \quad (1)$$

式中, $f_l^n(u_m)$ 表示当用户数为 u_m 时网元 x_n 对资源 $r_l^n(u_m)$ 的需求拟合值,是关于变量 u_m 的 K 次函数,定义 $\theta_l^n = (\theta_0^{n,l}, \theta_1^{n,l}, \dots, \theta_K^{n,l})^T$ 表示 $(K+1) \times 1$ 维的待定系数向量,

其中, $\theta_k^{n,l}$ 表示向量内第 $k+1$ 个待定系数, $(\cdot)^T$ 表示转置运算。为了寻找待定系数向量 θ_l^n 内各系数的最优估计值, 使得资源需求拟合值 $f_l^n(u_m)$ 最接近实际测试得到的 $r_l^n(u_m)$, 可建立如下无约束线性最小二乘问题优化目标。

$$\min_{\theta_l^n} \sum_{k=0}^K |f_l^n(u_m) - r_l^n(u_m)|^2 \quad (2)$$

式中, n, l, K, u_m 和 $r_l^n(u_m)$ 为已知量, θ_l^n 和 $f_l^n(u_m)$ 为待求解目标。

为得到网元资源需求函数 $f_l^n(u_m)$ 的最佳拟合形式, 先通过线性回归推导出 θ_l^n 的解析解, 将已知的 M 个具有 K 维特征的网元资源需求样本值代入式(1), 并建立如下结构。

$$f_l^n(U) = U \theta_l^n = \begin{pmatrix} 1 & u_1 & u_1^2 & \cdots & u_1^K \\ 1 & u_2 & u_2^2 & \cdots & u_2^K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_M & u_M^2 & \cdots & u_M^K \end{pmatrix} \cdot \begin{pmatrix} \theta_0^{n,l} \\ \theta_1^{n,l} \\ \vdots \\ \theta_K^{n,l} \end{pmatrix}^T \quad (3)$$

式中, $f_l^n(U)$ 为 $K \times 1$ 维的向量, 其中的元素表示网元资源需求拟合值, U 为 $M \times (K+1)$ 维的矩阵, 则目标损失函数的矩阵定义为

$$J(\theta_l^n) = \|f_l^n(U) - V_l^n\|^2 = \|U \theta_l^n - V_l^n\|^2 = (U \theta_l^n - V_l^n)^T (U \theta_l^n - V_l^n) \quad (4)$$

式中, $V_l^n = [r_l^n(u_1), r_l^n(u_2), \dots, r_l^n(u_M)]^T$ 是样本的输出向量, 维度为 $M \times 1$ 。根据最小二乘法的原理, 对损失函数 $J(\theta_l^n)$ 关于 θ_l^n 求导并取 0, 得到

$$\frac{\partial}{\partial \theta_l^n} J(\theta_l^n) = \frac{\partial [(\theta_l^n)^T U^T U \theta_l^n - (\theta_l^n)^T U^T V_l^n - (V_l^n)^T U \theta_l^n + (V_l^n)^T V_l^n]}{\partial \theta_l^n} = 2U^T U \theta_l^n - 2U^T V_l^n = 0 \quad (5)$$

当矩阵 $U^T U$ 可逆时, θ_l^n 的解析解即为最优估计下的向量, 可表示为

$$\tilde{\theta}_l^n = (U^T U)^{-1} U^T V_l^n \quad (6)$$

进而, 网元资源需求函数 $f_l^n(u_m)$ 最优拟合的表达式可写为

$$\tilde{f}_l^n(u_m) = (1, u_m, u_m^2, \dots, u_m^K) \tilde{\theta}_l^n \quad (7)$$

因此, 网元 x_n 对系统第 l 种资源的需求函数值在并发用户数为 $u > 0$ 时的计算公式为

$$\tilde{f}_l^n(u) = (1, u, u^2, \dots, u^K) \tilde{\theta}_l^n \quad (8)$$

式中, K 为所采用的拟合函数的次数。该式可用于预测合理范围内任意用户数条件下的核心网元资源需求值。

1.3 混合组网模型

IMS 行政交换核心网的 VNF 总资源需求总量可通过各网元的资源需求叠加得到, 当系统并发用户数为 u 时, 设核心网正常运行的节点配置共需要 $N^* \leq N$ 个单网元节点, 对应的网元集合定义为 $X^* \subseteq X$, 对其中每个资源

维度独立计算, 相应的公式为

$$F_l(u) = \sum_{x_n \in X^*} \tilde{f}_l^n(u), l = 1, 2, \dots, L \quad (9)$$

式中, $F_l(u)$ 表示系统并发用户数为 u 时 VNF 对第 l 种资源的总需求量。为得到核心网单节点最简配置, 本文以最小化网元节点数为目标构建混合组网模型, 模型中主要符号及其含义如表 1 所示。

表 1 混合组网模型符号说明

Tab. 1 Description of mixed networking model symbols			
符号	含义	类型	详细描述
G	网络拓扑图	图结构	描述核心网的整体架构, 包含节点集合和连接关系 包括: 物理网元 (PNF)、虚拟网元 (VNF)、二层交换机 (L2_Switch)、三层路由器 (L3_Router)、CE 设备、网管系统 (OMC/OSS)
V	节点集合	集合	包含虚拟网卡连接 (VNF 间) 和物理链路 (PNF 间)
E	边集合	集合	VNF 集合中的单个实例, 需映射到虚拟机资源
v	虚拟网元实例	对象	

则模型的优化目标可建立为最小化核心网的网络拓扑图 $G = (V, E)$ 中节点集合 V 所包含网元节点的个数。

$$\min_{\theta_l^n} \sum_{x_n \in X^* \subset V} N^* \quad (10)$$

需满足的约束条件如下:

1) 虚拟机资源约束。对于网元节点 $x_n \in X^*$ 中的虚拟网元实例 v , 其所对应的 VNF 对于 CPU 资源和内存资源的需求量不超过服务器分配给虚拟机实例的 CPU 资源和内存资源容量。

2) 流量守恒约束。通过网元节点 $x_n \in X^*$ 的总入流量等于网元节点 $x_n \in X^*$ 的总出流量与服务器的存储变化量之和。

求解该问题所得到的最小网元节点数对应的网元节点组合即为单节点的最简配置方案。

1.4 冗余配置

基于系统容错能力和负载均衡考量, 核心网应在局部故障或负载激增时系统仍能维持服务可用性和性能稳定^[19-20], 实现此目标需在满足 VNF 基础需求的条件下继续部署额外资源, 即在单节点最简配置的基础上添加冗余配置。为保证在配置冗余后, 核心网单节点故障不影响业务连续性的同时避免资源过载导致的性能下降, 且支持业务增长时的动态资源扩展, 考虑采取一种分级冗余配置方法^[21]。首先, 对系统的冗余层级进行划分, 如表 2 所示。

设求解式(10)所表示的优化目标得到的单节点最简配置, 即核心网基础资源配置集合为 $F = \{F_1(u), F_2(u), \dots, F_L(u)\}$, 定义整数 $t \geq 1$ 表示全局冗余等级, $\Delta_l(u)$ 表示第 l 种资源的网元冗余增量, 计算方法为

$$\Delta_l(u) = \lambda_n \sum_{x_n \in \tilde{X}} \tilde{f}_l^n(u) + \mu_n \sum_{x_n \in (X^* - \tilde{X})} \tilde{f}_l^n(u) \quad (11)$$

表 2 核心网冗余层级划分

Tab. 2 Core network redundancy hierarchy division

层级	实现方式	运行指标
硬件冗余	双电源、RAID 存储、 N + 1 服务器配置	平均无故障时间 (MTBF)
虚拟化层	跨物理节点的虚拟机 HA 集群	故障检测时间
VNF 实例	多活实例部署 (Active-Active)	实例分布区域数量
业务层	SIP 会话复制、媒体流备份	终端恢复时间

式中, $\tilde{X} \subseteq X^*$ 表示单节点最简配置中的关键网元 (如 CSCF) 集合, $(X^* - \tilde{X})$ 表示非关键网元 (如 MRF) 集合, $\lambda_n > 0$ 表示关键网元冗余系数, $\mu_n > 0$ 表示非关键网元冗余系数, 且满足 $\lambda_n > \mu_n$, 则单节点冗余配置可表示为

$$Fr_l(u) = F_l(u) + t \cdot \Delta_l(u), l = 1, 2, \dots, L \quad (12)$$

为核心网单节点配置冗余的目的是, 在所选网元满足不超过 VNF 资源需求上限、系统冗余度满足安全运行率的条件下, 最小化算力资源部署成本, 相应的优化问题建立如下。

$$s. t. Fr_l(u) \leq \sum_{x_n \in X} \tilde{f}_l^n(u), l = 1, 2, \dots, L$$

$$P_{X^*, t}(u) \geq P^* \quad (13)$$

其中, 第一个约束条件表示资源配置上限约束, 第二个约束条件表示安全运行率约束; $P_{X^*, t}(u)$ 表示当单节点配置集合为 X^* 、冗余等级为 t 、并发用户数为 u 时系统的安全运行率; P^* 为根据系统运行环境及历史运行经验而人为设定的系统最低安全运行率阈值。

2 算法设计

2.1 单节点最简配置方案

贪心算法 (greedy algorithm, GA) 是一种基于局部最优决策的算法设计范式, 其核心思想是通过在每一步选择中采取当前状态下最优的局部解, 逐步构建问题的全局解^[22-23]。

根据优化目标 (10) 及其约束条件, 当网元节点数达到最小时, 各网元的资源需求值应相应最大, 则可基于贪心算法的思想逐步选取网元资源需求较大的节点, 并同时与剩余可用资源比较以确定最合适选取的节点。算法所用变量及其含义如表 3 所示。

表 3 贪心算法符号说明

Tab. 3 Description of greedy algorithm symbols

符号	含义	类型	详细描述
D	网元需求集合	集合	包含所有网元的资源需求函数值
S	资源池	对象集合	可用服务器资源集合, 每个对象包含剩余 CPU、内存信息
A	资源配置方案	列表	记录每个网元分配到的服务器 ID
d	单个网元需求	对象	包含 CPU、内存需求的网元实例
s	服务器实例	对象	物理服务器或虚拟机实例, 包含资源总量和已分配量

对于资源池中的可用服务器资源, 需分别对网元

CPU 资源、内存资源分别进行叠加以计算总基础资源需求量, 计算公式为

$$F_{cpu}^*(u) = \sum_{x_n \in X} \tilde{f}_{cpu}^n(u) \quad (14)$$

$$F_{mem}^*(u) = \sum_{x_n \in X} \tilde{f}_{mem}^n(u) \quad (15)$$

通过贪心算法实现核心网单节点最简配置的流程如图 2 所示。

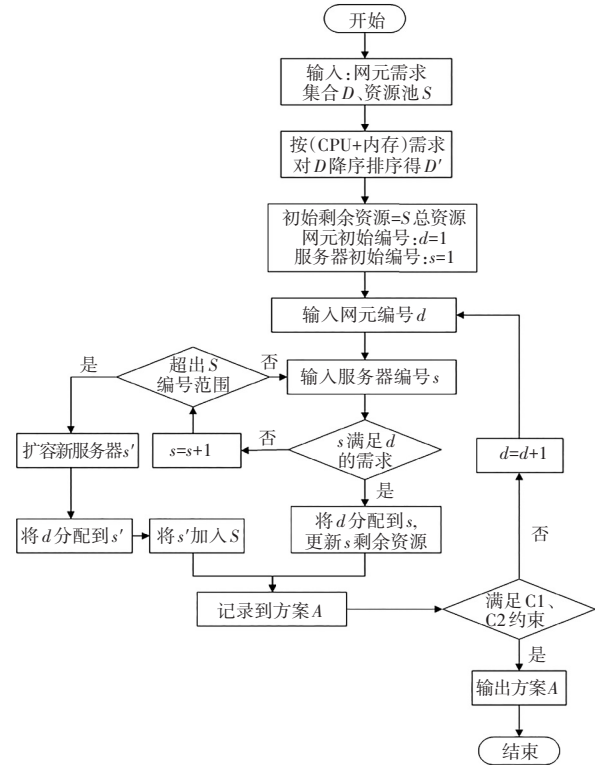


图 2 基于贪心算法的核心网单节点最简配置方法流程

Fig. 2 Flowchart of the simplest configuration method of a single node in the core network based on greedy algorithm

2.2 单/双节点冗余配置方案

本节基于递归算法的思想, 对优化问题 (13) 进行求解。根据式 (11) 可知, 本问题的冗余增量由关键网元冗余增量和非关键网元冗余增量组成, 同时根据 1.4 节中的核心网冗余层级划分可知, 非关键网元的冗余策略存在差异。为了便于算法设计, 先对通过 2.1 节算法中得到的单节点最简配置相关网元实施差异化管理策略, 如图 3 所示。

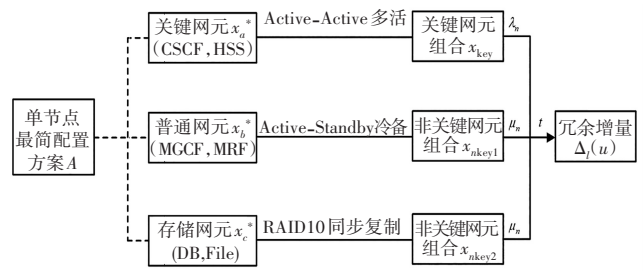


图 3 网元节点差异化管理策略

Fig. 3 Differentiated node management policy of NE

根据优化问题 (13), 通过递归算法实现核心网最优冗余配置的流程如图 4 所示。

得到核心网单节点冗余配置后, 为进一步针对 IMS

业务连续性提高所部署的冗余资源利用率,可结合多活实例同步技术和动态扩容技术配置双节点冗余。基于 1.1 节和 1.3 节所描述的混合组网架构建立同步机制,实例同步框图如图 5 所示。

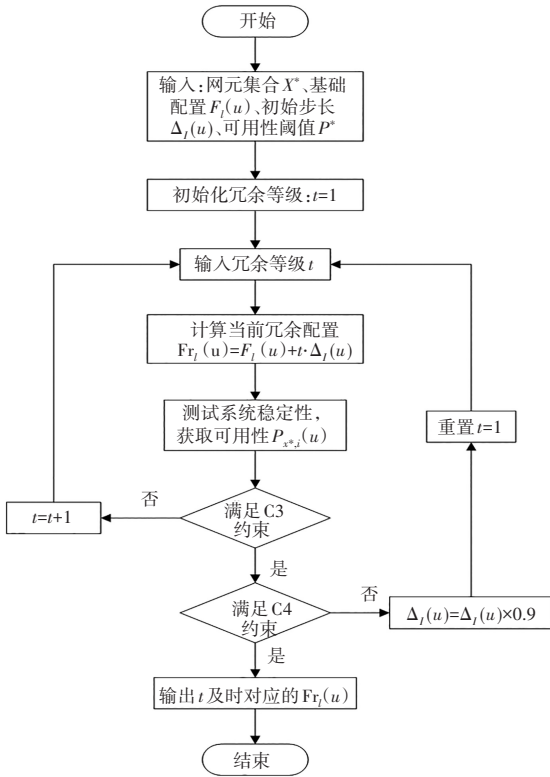


图 4 基于递归算法的最优冗余度搜索方法流程

Fig. 4 Flowchart of optimal redundancy search method based on recursive algorithm

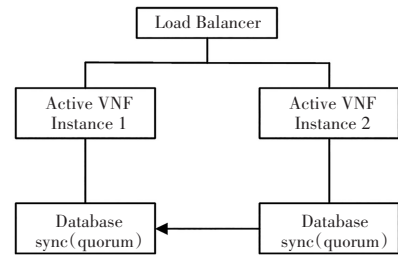


图 5 多活实例同步框图

Fig. 5 Multiactive instance synchronization block diagram

在该机制中,会话状态通过 ETCD 集群实现分布式锁,媒体流采用 RTP/RTCP 双路径传输,并基于 rsync 的增量同步进行数据配置^[24]。当单节点冗余配置中网元 $x_n \in X^*$ 发生故障,且剩余网元 $x_n \in (X - X^*)$ 存在可同步实例时,系统触发上述机制,采用图 4 流程向被同步的单个节点配置冗余,从而完成网元节点切换,实现双节点同步冗余配置。

为保证核心网具备足够同步实例,需结合动态扩容技术实时部署新 VNF 实例,调整业务边界。本节采用一种基于负载感知的弹性扩容算法进行设计,流程如图 6 所示。

在上述扩容流程中,当 VNF 容量占比低于 60% 时,执行 VNF 扩容,创建新 VNF 实例、迁移用户并验证迁移结果;当 VNF 容量占比不低于 60% 时,执行业务割接,调整 PNF-VNF 业务边界、在预设的维护窗口执行割接并验证割接结果。

此外,若因业务中断时间过长而导致扩容失败,则触发自动回滚机制:终止新建的 VNF 实例;将已迁移用户回切至原网元节点;向运维管理端发送告警通知;将相应故障代码记录到知识库中。

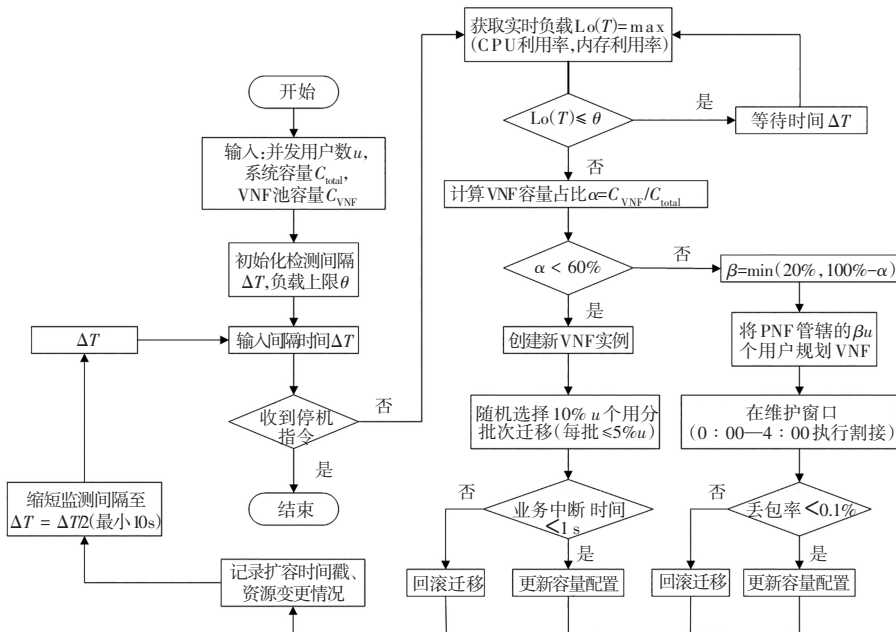


图 6 基于负载感知的弹性扩容算法流程

Fig. 6 Flowchart of elastic capacity expansion algorithm based on load awareness

3 仿真实验

3.1 实验环境配置和基础性能测试

考虑基于NFV的IMS行政交换核心网场景,实验环

境配置如表4所示。首先对核心网业务性能表现进行测试,以语音质量、视频峰值信噪比、会话接通率和SIP信令端到端时延为关键性能指标,采集稳态运行周期为30min的各项指标平均值数据,结果如图7~10所示。

表4 核心网测试环境参数配置

Tab. 4 Configure of the core network test environment parameters

参数	配置值	说明
云服务器架构	8 计算节点+3 控制节点	OpenStack Stein 云平台
单节点硬件配置	Intel Xeon Gold 6248 × 2	40 核/80 线程,384 GB DDR4
虚拟化平台	KVM +Ceph 存储集群	超分比 1 : 4(CPU), 1 : 1.5(内存)
测试工具	SIPp v3.6, Wireshark 3.6	流量生成与协议分析
业务模型	Passion 分布	0.25erl, 包含语音(80%)和视频通话(20%)

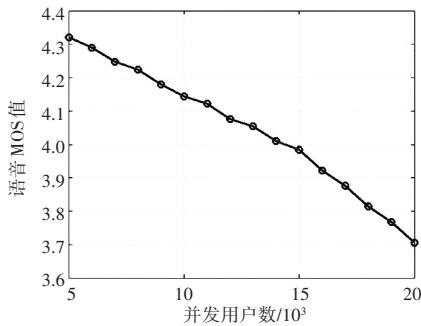


图7 语音质量测试结果

Fig. 7 Voice quality test results

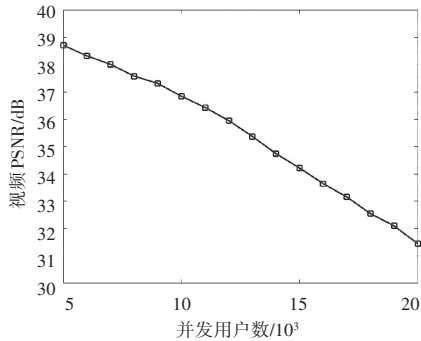


图8 视频峰值信噪比测试结果

Fig. 8 Video peak signal to noise ratio test results

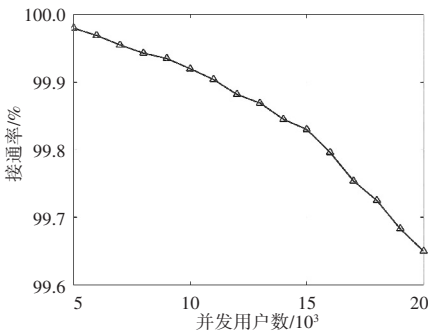


图9 会话接通率测试结果

Fig. 9 Test result of the session connection rate

保持并发用户数条件和数据采集方式不变,针对CSCF网元、HSS网元、MGCF网元的相应资源消耗以及虚机占用情况为指标进行测试,得到的结果如图11~14所示。

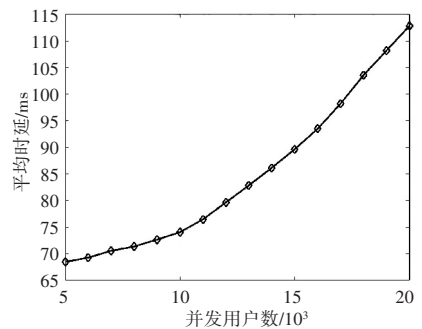


图10 SIP端到端时延测试结果

Fig. 10 SIP end-to-end delay test result

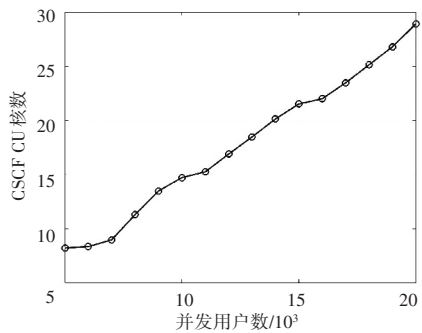


图11 CSCF CPU核数测试结果

Fig. 11 CSCF CPU core test result

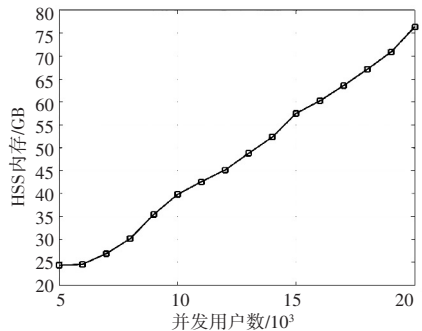


图12 HSS内存测试结果

Fig. 12 HSS memory test results

根据图7~14可知,当核心网并发用户数为5 000~20 000时,各业务性能指标以及资源消耗指标基本呈现线

性变化,其中,MOS 值在用户数增长至约 15 000 前保持在 4.0 以上,达到商用级质量;视频峰值信噪比始终保持在 30 dB 以上,属于 IMS 业务可接受范围。值得注意的是,在用户数为 15 000 时,会话接通率下降幅度进一步增大,同时存储 IOPS 需求出现拟合指数约为 1.2 的超线性增长,二者反映会话状态数据量累积效应。基于上述分析,15 000 为本测试环境的并发用户数关键阈值,需在后续实验步骤中关注该用户数下的系统负载情况,并利用该值做进一步研究。

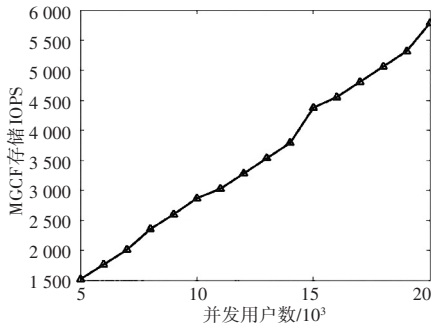


图 13 MGCf 存储 IOPS 测试结果
Fig. 13 MGCf stores IOPS test results

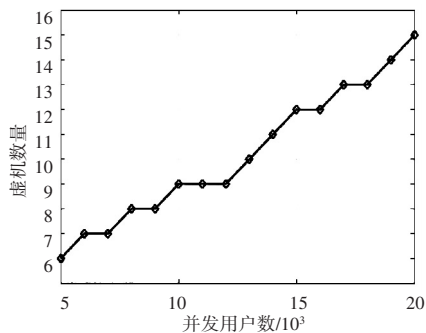


图 14 虚拟机占用测试结果

Fig. 14 Virtual machine occupancy test result

3.2 数据拟合与模型验证

由 3.1 节可知,核心网资源需求随用户数增长基本呈现现象变化,结合 1.2 节中的网元资源需求函数计算方法,应设置拟合函数的次数 $K = 1$ 。为进一步消除部分资源需求超线性变化的影响,此处通过加权最小二乘法(权重系数 $= \frac{1}{\sigma^2}$, σ 为资源需求测量值方差)确定最终的拟合函数^[25]。以 CSCF、HSS、MGCf 网元的 CPU 和内存资源为例,拟合结果如表 5 所示。

表 5 网元资源需求函数算例

Tab. 5 Example of NE resource demand function

网元	CPU 模型(核数)	内存模型/GB	拟合度 R^2
CSCF	$1.23u + 2.17$	$0.85u + 5.42$	0.992
HSS	$0.68u + 8.25$	$1.12u + 10.36$	0.987
MGCf	$0.15u + 1.05$	$0.21u + 3.14$	0.978

基于上述拟合函数,采用 3.1 节中的并发用户数条件和数据采集方式,以 CSCF 网元的 CPU 和内存需求为例进行模型预测误差分析,结果如图 15 所示。

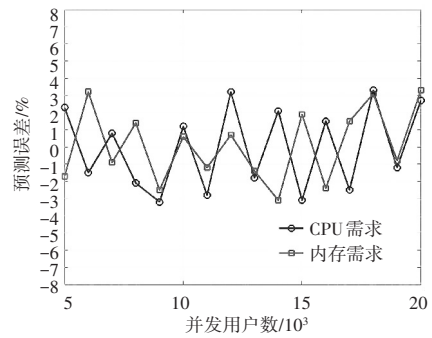


图 15 网元资源需求拟合函数误差分布

Fig. 15 Error distribution of NE resource demand fitting function

根据图 15,当用户数为 5 000~20 000 时,CSCF 网元的 CPU 和内存需求拟合值与实际测量值之间的误差绝对值均小于 4%,其中约 70% 的数据点误差绝对值不超过 3%,最大误差为 -3.3。

为进一步确认上述模型泛化能力,采用交叉验证(Cross-Validation)方法对上述结果进行分析^[26]。将实验结果按并发用户数规模划分为 5 个子集,其中每个子集包含 4 组用户规模(5 000、10 000、15 000、20 000)的完整测试数据。

验证步骤简要表述如下,设数据集为 D ,对每个子集 $D_i(i = 1, 2, \dots, 5)$ 。

1) 训练集: $D_{\text{train}} = \frac{D}{D_i}$ 。

2) 测试集: $D_{\text{test}} = D_i$ 。

3) 在 D_{train} 上拟合模型,在 D_{test} 上计算拟合度、平均绝对误差(mean absolute error, MAE)等性能指标。交叉验证结果如图 16~17 所示。

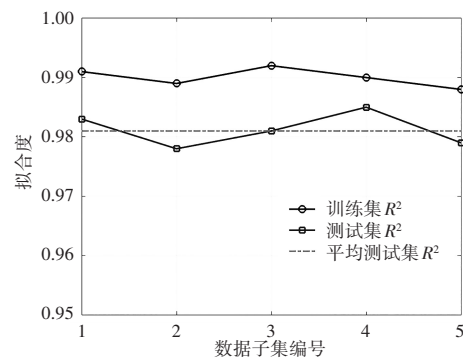


图 16 CSCF 资源需求拟合度交叉验证结果

Fig. 16 Cross verification results of CSCF resource demand

上述验证结果中,平均测试集拟合度为 0.981 ± 0.003 ,平均 CPU 预测 MAE 为 $2.12\% \pm 0.23\%$,平均内存预测 MAE 为 $1.90\% \pm 0.17\%$,训练集与测试集拟合度差异小于 0.01,可见模型在未见数据上表现稳定,未见明显过拟合;预测误差分布均匀,未出现系统性偏差。综上所述,本文所用核心网元资源需求函数拟合函数模型可有效反映核心网资源需求随并发用户数的变化特征,并可进一步用于后续网元资源配置。

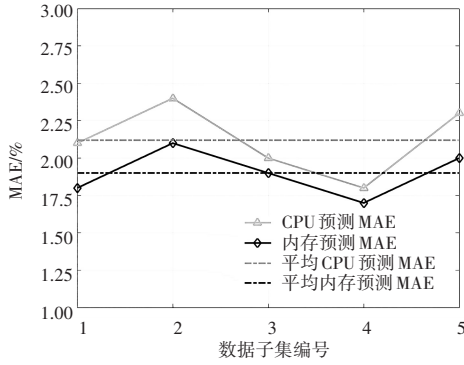


图 17 CSCF 资源需求平均绝对误差交叉验证结果

Fig. 17 Cross-validation results of CSCF resource demand average absolute error

3.3 核心网资源配置方案算例对比

本节基于 2.1~2.2 节的算法设计对比实验,以对 IMS 行政交换 NFV 核心网给出合理的工程建议。采用 3.2 节中的网元资源需求拟合函数,用户规模分别为 5 000、10 000、15 000 和 20 000,以 30 min 为单次测试周期,设置故障注入为随机关闭 30% 网元节点,负载模型为并发用户数阶跃式增长,每阶段步长为 5 000 用户数。在单节点最简配置、单节点冗余配置、双节点冗余配置的基础上增加一个基准方案作为对照组,具体方案描述如表 6 所示。

表 6 实验对比方案描述

Tab. 6 Description of experimental comparison scheme

方案	描述	优化维度
基准方案	传统静态配置(无 NFV 优化)	—
单节点最简配置	基于贪心算法	网元节点负载
单节点冗余配置	基于递归算法	系统鲁棒性
双节点冗余配置	基于多活实例同步和动态扩容技术	冗余资源利用率

在开始算例对比实验前,先确定冗余配置所需冗余度,设置该过程中用户数 $u = 15\ 000$,考虑单节点宕机,网络分区和存储三方面的故障,注入方式分别为强制关闭虚拟机、iptables DROP 规则以及移除 OSD 节点,相应的冗余度测试结果如表 7 所示。

表 7 冗余度测试性能指标对比

Tab. 7 Comparison of redundancy test performance indicators

冗余度	可用性 $P_x^*, t(u) / \%$	成本增幅 $\sum_{l=1}^L Fr_l(u) / \%$	故障恢复 时间/s
$t = 0$	99.92	0	无法恢复
$t = 1$	99.95	35	312
$t = 2$	99.99	58	107
$t = 3$	99.99	82	23

综合可用性、成本增幅和故障恢复 RTO 考虑,选择冗余度 $t = 2$ 作为初始冗余策略进行对比实验,并将冗余度调整范围设置为 $t \in [1, 3]$ 。在此基础上,以平均资源利用率为第一项参考指标,得到如图 18 所示的对比实验结果。

分析图 18 可知,除基准方案外的 3 种方案具有相似的变化趋势,整体变化幅度小于基准方案,其中单节点最

简配置方案的平均资源利用最高,双节点冗余配置其次,最后为单节点冗余配置,这是由于单节点冗余配置相对于最简配置额外部署了冗余资源,该部分资源用于提高系统可靠性但并不一定被核心网使用,而双节点冗余通过 VNF 实例同步进一步提高了冗余资源利用率,因此这 3 种方案资源利用率随用户数的增长变化关系符合本文中的模型设定,也符合 NFV 核心网场景实际。此外,图 18 中基准方案在并发用户数为 20 000 时因内存过载出现超售,而单节点最简方案也在该用户数条件下逼近临界值(100%),这体现出在高额并发数下冗余配置的重要意义。

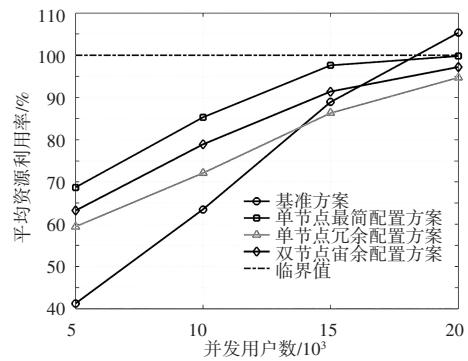


图 18 不同方案的平均资源利用率对比

Fig. 18 Comparison of average resource utilization of different schemes

综合图像和实验数据可得如下结论:基于贪心算法的单节点最简配置方案相对于其他方案可提升约 31%~45% 的平均资源利用率,但高负载时逼近硬件极限;基于递归算法的单节点冗余配置可牺牲 8%~12% 利用率换取可靠性,而双节点冗余配置方案折衷效果更显著,利用率损失仅 4%~6%。

接着,固定用户数为 15 000,以服务可用性为第 2 项参考指标,基准方案、单节点最简配置方案、单节点冗余配置方案、双节点冗余配置方案分别对应方案编号 1、2、3、4,得到如图 19、20 所示的对比实验结果。

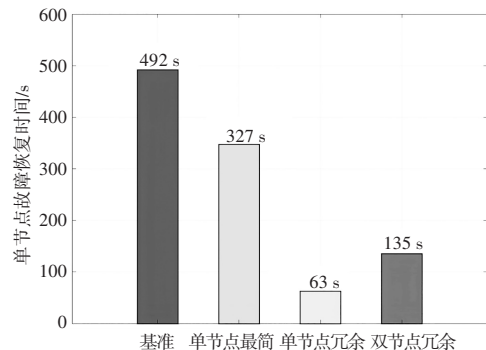


图 19 不同方案的单节点故障恢复时间对比

Fig. 19 Comparison of single-node fault recovery time in different solutions

根据图 19、20 和实验数据,核心网元冗余配置相对于无冗余的单节点配置在业务持续性上具有优势,其中单节点冗余配置方案的性能优于双节点冗余配置方案,可归因

于双节点冗余资源部署中多活实例同步过程和动态扩容过程一定程度上提高了系统总时延。具体而言,上述2级冗余度单节点配置将原故障恢复时间缩短约87%,双节点配置通过动态冗余保持中断率 $<0.15\%$,同时减小固定冗余的资源浪费。结合图18分析结果可知,双节点冗余配置的最终效益为在牺牲一定资源调度无感性的条件下,提高系统平均资源利用率,使得核心网在该资源配置方案下能实现较好的性能折衷。

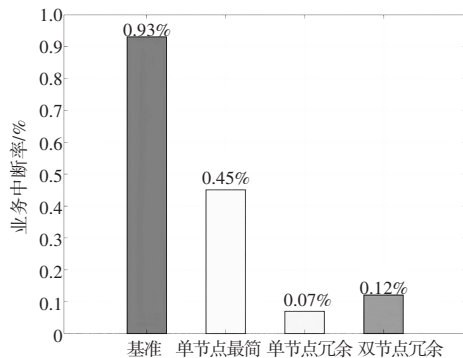


图20 不同方案的业务中断率对比

Fig. 20 Comparison of service interruption rates of different schemes

3.4 实验结论

基于3.1~3.3节的实验分析,可得出如下实验结论。

1) 本文所用网元资源需求函数拟合模型的误差主要来源于监控工具精度、采样间隔、线性假设与真实系统的偏差、背景流量波动以及硬件性能波动,经检验具备泛化能力,但当核心网并发用户数为15 000时,部分网元资源需求因会话状态数据量累积而开始出现超线性变化,此时应考虑对系统进行负载均衡优化,即针对部分网元节点配置一定冗余度。

2) 本文所用基于贪心算法的单节点最简配置方案和单/双节点冗余配置方案均明显优于传统无NFV优化的静态配置方案。单节点最简配置方案的平均资源利用率 $>30\%$,硬件成本降低35%,适用于预算有限、可靠性要求适中的场景;单节点冗余配置方案的故障恢复时间 $<2\text{ min}$,满足IMS行政交换网严苛SLA,适合核心节点、不可中断的关键业务;双节点冗余配置方案在成本增加 $<12\%$ 的前提下,在单节点冗余基础上提高资源利用率且保证99.99%可用性,能有效应对负载波动,适用于作为核心网默认方案。

4 结论

本文研究了IMS行政交换NFV核心网的网元资源配置问题,其网络架构为包含VNF、PNF、二层交换机、三层路由器及独立设置OMC/OSS的混合组网。基于此,通过历史测试数据建立核心网元资源需求预测模型,并利用该模型所反映的数据分布特征研究用户规模与核心网元算力适配关系。为了得到合理的网元配置方案以优化核心

网业务可靠性和资源利用率,本文提出了基于贪心算法的单节点最简配置方案、基于递归算法的单节点冗余配置方案以及基于多活实例同步和动态扩容技术的双节点冗余配置方案,结合冗余层级划分和网元节点差异化管理策略进行优化设计。最后,仿真实验验证了基础性能的可实现性、所建立拟合模型的泛化能力、所提方案的性能及各自的优势区间,仿真结果与理论分析吻合,证明了本文方法的可行性,具备实际应用的参考价值。基于本论文所提出的网元资源配置方案,可根据实际场景参数来进一步确定系统负载区间、设定资源弹性系数,从而实现冗余度自适应配置。未来,该方案可延伸拓展至IMS行政交换网业务编排、迁移技术研究等领域,为IMS行政交换核心网NFV建设提供帮助。

参考文献

- [1] CAMARILLO G, GARCIA-MARTIN M A. The 3G IP multimedia subsystem (IMS): merging the Internet and the cellular worlds[M]. Hoboken, NJ: John Wiley & Sons, 2011.
- [2] KOVACIKOVA T, SEGEC P, KUBINA M. IMS in the next generation network[C]//Proceedings of the 11th WSEAS International Conference on Communications. Agios Nikolaos, Crete Island, Greece: WSEAS, 2007: 45-50.
- [3] FITE L. Leverage Advanced TCA to optimize IMS: move from proprietary to commercial-off-the-shelf components to build next-generation IP networks[J]. Electronic Design, 2007, 55(11): 59-62.
- [4] 魏娜. IMS互联互通技术在企业网络中的应用[J]. 电子世界, 2019(8): 128-129.
- [5] 田文锋, 毕庆刚, 李炳林, 等. 国家电网公司行政交换网迁移演进方案研究[J]. 电力信息与通信技术, 2015, 13(4): 1-5.
- [6] HERRERA J G, BOTERO J F. Resource allocation in NFV: a comprehensive survey[J]. IEEE Transactions on Network and Service Management, 2016, 13(3): 518-532.
- [7] CARELLA G, CORICI M, CROSTA P, et al. Cloudified IP multimedia subsystem (IMS) for network function virtualization (NFV)-based architectures [C]//Proceedings of 2014 IEEE Symposium on Computers and Communications (ISCC). Funchal: IEEE, 2014: 1-6.
- [8] 李磊, 周正, 陈家璘, 等. 基于云计算的电力智能信息数据网架构研究[J]. 自动化技术与应用, 2024, 43(6): 176-180.
- [9] 刘航, 吴言. 基于人工智能云计算的医院人岗匹配方法研究[J]. 自动化技术与应用, 2025, 44(7): 164-168.
- [10] 曾京文, 杨琦, 江淦, 等. 电力IMS交换网业务应用技术研究[J]. 电力信息与通信技术, 2021, 19(12): 75-81.
- [11] CARPIO F, DHAHRI S, JUKAN A. VNF placement with replication for load balancing in NFV networks[C]//2017 IEEE International Conference on Communications (ICC). Paris, France: IEEE, 2017: 1-6.
- [12] 梅增杨, 柳旭, 华伟韬, 等. 电力IMS虚拟化核心网单节点两阶段硬件配置研究[J]. 电力信息与通信技术, 2024, 22(12): 77-82.
- [13] 崔跃君, 杨甲明, 顾新桥, 等. 基于物联网技术的电网物资供应优化模型建立[J]. 自动化技术与应用, 2024, 43(1): 108-111.
- [14] 田文锋, 李炳林, 卜宪德. 国家电网公司IMS行政交换网网元扩容方案研究[J]. 电力信息与通信技术, 2015, 13(12): 134-138.
- [15] 潘远, 张华兵, 刘梓健, 等. 基于图注意力网络的算力资源调配优化研究[J/OL]. 自动化技术与应用, 1-6(2024-12-18). <https://link.cnki.net/urlid/23.1474.TP.20241218.0956.006>.
- [16] 李亮, 孙丽丽, 曹新智. IMS行政交换网平滑割接方案的研究与应用[J]. 电力信息与通信技术, 2015, 13(10): 75-78.

(下转第168页)