

## 基于区域感知的通感算协同网络语义通信可靠性联合优化

陈津<sup>1</sup>, 窦高奇<sup>1</sup>, 蒋妍<sup>2</sup>, 朱尧<sup>2</sup>

(1. 海军工程大学, 湖北 武汉 430033; 2. 武汉大学电子信息学院, 湖北 武汉 430072)

**摘要:** 针对通感算协同 (ISCC) 网络提出了一种基于语义通信框架的联合优化方法。首先, 利用语义通信框架, 将 ISCC 网络的感知、通信和计算 3 个功能模块统一映射到语义编码、语义信道传输和语义解码过程, 建立了多目标感知与多服务器协作场景下的语义可靠性分析模型。其次, 从语义层面定量刻画系统端到端错误率, 在各模块错误率约束和总时延限制下, 构建服务器选择与时间资源分配的联合优化问题, 将原混合整数非线性问题转化为凸问题求解, 并基于拉格朗日方法和 KKT 条件揭示最优解的结构特性。仿真结果表明, 所提框架可有效指导 ISCC 网络资源配置, 在同构、异构服务器部署场景下, 相较传统方案显著提升了多功能协作可靠性, 验证了其在 ISCC 网络优化中的有效性。

**关键词:** 语义通信; 通感算协同; 多功能协作; 可靠性优化; 资源分配

**中图分类号:** TN92

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2026016

## Joint optimization of semantic communication reliability in integrated sensing, communication and computing networks based on regional perception

Chen Jin<sup>1</sup>, Dou Gaoqi<sup>1</sup>, Jiang Yan<sup>2</sup>, Zhu Yao<sup>2</sup>

1. Naval University of Engineering, Wuhan 430033, China

2. Electronic and Information School, Wuhan University, Wuhan 430072, China

**Abstract:** A joint optimization method was proposed based on the semantic communication framework for the ISCC (integrated sensing, communication, and computing) network. Firstly, leveraging the semantic communication theoretical framework, the sensing, communication, and computation functionalities of the ISCC system were unified and reconstructed with the semantic layer for multi-target sensing and multi-server collaboration scenarios. Secondly, the end-to-end error probability of the system was quantitatively characterized from the semantic level, and a joint optimization problem for server selection and time resource allocation was formulated. The original mixed-integer non-convex problem was transformed into a convex one. Moreover, the structure of its optimal solutions was analyzed based on the Lagrangian method and KKT conditions. Simulation results demonstrate that the semantic communication framework can effectively guide resource allocation in ISCC systems. Compared with traditional schemes, the proposed scheme significantly improves the reliability of the whole ISCC system in both homogeneous and heterogeneous server deployment scenarios, validating the effectiveness of the semantic communication framework in ISCC network optimization.

**Keywords:** semantic communication, ISCC, multi-functional cooperation, reliability optimization, resource allocation

收稿日期: 2025-10-23; 修回日期: 2026-01-08

通信作者: 窦高奇, 0909031022@nue.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61871473); 海军工程大学自主研发计划基金资助项目 (No.202509)

**Foundation Items:** The National Natural Science Foundation of China (No.61871473), The Independent Research and Development Program Project of the Naval University of Engineering (No.202509)

## 0 引言

随着 6G 技术的快速发展, 通感算协同 (integrated sensing, communication, and computation, ISCC) 已成为未来网络架构的关键技术趋势。在传统网络架构中, 感知、通信和计算功能相互独立, 导致资源利用效率低、系统响应时延高、能耗开销大等问题日益突出。ISCC 网络通过深度融合环境感知、信息传输和边缘计算功能, 实现了频谱、硬件、能量等资源的高效复用和系统性能的协同优化。在这种新型架构下, 感知设备不仅能获取环境信息, 同时还能进行数据传输; 通信链路在传递信息的同时可以提供目标检测和定位功能; 边缘计算节点能够就近处理感知数据, 大幅降低响应时延。这种多功能融合协同的架构在智慧城市、工业物联网、智能交通系统、无人机巡检、自动驾驶等场景中展现出巨大的应用潜力<sup>[1-2]</sup>。例如, 在输电线路巡检场景中, 无人机需要实时感知输电塔杆状态, 将检测数据传输至地面站, 并通过边缘计算快速识别潜在故障, ISCC 网络的集成设计能够显著提升巡检效率和准确性。特别是在需要实时环境感知和快速决策响应的关键任务中, ISCC 网络通过消除功能模块间的信息壁垒, 能够显著降低端到端时延, 提升系统整体效率和可靠性。然而, 在实际部署中, 多功能集成带来的系统复杂性急剧增加, 如何在资源受限和环境不确定的条件下保证系统可靠运行, 成为 ISCC 技术大规模应用的关键瓶颈。

不同于传统通信系统关注比特级准确传输, 语义通信强调传递信息的语义内容和实际意义, 能够在保证任务完成质量的前提下大幅提升传输效率<sup>[3]</sup>。在 ISCC 网络中引入语义通信理念, 可以从语义层面统一描述感知、通信和计算 3 个环节的性能, 为多功能协同优化提供新的视角, 从而有效表征 ISCC 网络中多任务的联合性能, 提升协同增益。然而, 在多目标感知和多服务器协作场景下, 语义通信面临诸多挑战: 首先, 多个感知目标的同时检测增加了感知阶段的复杂度, 任一目标的感知错误都可能导致整体语义失真; 其次, 多服务器环境下的服务器选择与任务分配存在复杂的耦合关系, 不同的服务器组合直接影响通信链路质量和计算处理能力; 最后, 在严格的时延约束下, 如何在感知、通信和计算 3 个阶段之间合理分配时间资源, 实现

各环节错误率的最优权衡, 是保障语义可靠性的核心难题。

为了完善 ISCC 网络的功能集成, 已有研究从网络架构设计与信号反射控制的角度展开探讨, 以提升各功能模块性能<sup>[4-6]</sup>。在感知功能优化方面, 文献[4]研究了一种无人机辅助的 ISCC 网络, 其中机载通感算协同 (integrated sensing and communication, ISAC) 设备能够在收集感知数据的同时保持与用户的通信, 并采用雷达互信息评估感知性能, 以保障雷达感知质量并提升系统整体性能; 文献[5]提出了利用边缘辅助节点的反射信号, 使 ISAC 设备具备被动感知能力, 并与主动感知相结合, 以提升感知数据处理效率。在感知与通信阶段联合优化方面, 为同时提升 ISAC 网络中的感知精度和通信传输速率, 文献[6]设计了一种自适应感知算法, 以实现感知和通信性能的同时优化。

语义通信旨在通过传输任务相关的特征级语义信息实现语义层面的有效还原, 同时降低带宽资源的占用, 并增强通信的鲁棒性和智能协同能力。文献[7]基于同义映射的思想, 引入了语义熵、语义互信息与语义容量等关键度量指标, 建立了一个系统的语义信息论框架, 进一步强化了语义通信的核心目标, 即接收端是否得到了与发送端相同的含义。在此基础上, 文献[8]针对带宽受限场景, 采用 Transformer 模型联合语义信道编码, 并基于信息瓶颈理论设计损失函数, 构建端到端的语义通信架构。为了降低带宽资源的占用和语义信息处理时延, 文献[9]提出将语义进一步划分, 从而传输语义中有信息的部分。文献[10]侧重于研究基于感算通融合的语义认知通信网络, 通过深度融合网络感知、计算与通信过程, 实现跨层协同与资源的高效利用。

多服务器协作能够基于并行处理降低整体服务时延, 提高任务完成的及时性。同时, 合理选择多台服务器能够避免因单节点处理能力不足导致的服务失败。文献[11]通过在密集异构网络中部署多台边缘计算服务器, 实现任务的并行卸载, 缓解小区边缘用户面临的高时延和低吞吐量问题。对于多服务器移动边缘计算 (mobile edge computing, MEC) 系统的任务调度优化, 文献[12]引入负载分配机制, 考虑了任务全分配和部分分配两种情况, 避免了额外的服务器能耗。文献[13]利用无源波束成形技术, 通过在邻近服务器超载时将部分移动设备重

定向至负载较轻的远端服务器, 实现系统性能的显著提升。

在上述研究中, ISCC相关工作多集中于单目标检测, 这在一定程度上简化了系统的优化设计, 实际场景中多目标构成的区域感知更具普遍性。此外, 尽管文献[10]强调ISCC与语义通信的深度融合可以改善网络性能, 但其对系统语义层面可靠性需求的探讨仍不完善。另一方面, 文献[11-13]主要致力于提升多服务器协同处理任务的效率, 优化的重点集中在任务本身的处理策略, 而忽略了计算时长分配对多服务器协作处理性能的影响。不同于已有文献的研究, 本文面向由多个感知目标构成的区域感知场景, 并同时考虑了多服务器协同计算下的最优服务器选择与任务负载分配策略, 基于语义错误率这一性能指标, 对多功能一体化ISCC网络中的语义通信可靠性进行了系统分析, 提出了一种联合优化资源分配的框架, 以实现整体语义错误率最小化。

本文主要贡献总结如下。

1) 建立了多目标多服务器ISCC网络的语义通信可靠性分析模型, 通过将语义通信的各组成部分映射到ISCC网络的感知、通信和计算模块, 定量刻画了系统端到端的语义错误率。

2) 提出了服务器选择与时间资源分配的联合优化方法, 在满足各阶段错误率约束和总时延限制的条件下, 最小化系统语义错误率。

3) 通过理论分析证明了优化问题的凸性, 并将原始混合整数非线性问题转化为可高效求解的混合整数凸优化形式。

4) 基于拉格朗日方法和卡罗需-库恩-塔克(KKT)条件, 揭示了最优解的结构特性, 为实际系统设计提供了理论指导, 并通过仿真验证了所提方案相比传统方案可显著提升语义可靠性。

## 1 系统模型

如图1所示, 考虑一个多功能ISCC网络系统, 需要基于区域状态执行行为判别任务, 系统包含一个区域内的 $I$ 个待感知目标、一个收发器、 $K$ 个可用服务器以及一个执行器, 其中 $\mathcal{I} = \{1, \dots, I\}$ 为感知目标集合,  $\mathcal{K} = \{1, \dots, K\}$ 为可用服务器集合。各服务器基于感知观测数据执行相应计算任务, 其计算结果用于控制执行器的行为。例如, 在工业物联

网中, 系统需要感知多个设备状态, 并上传至节点进行分析处理以实现生产线实时监控, 提供设备预测维护能力。设定收发器配置 $K_r$ 根接收天线和单根发射天线, 各服务器节点配备单根天线。系统采用时隙化结构运行, 时间被划分为帧周期, 整个执行过程逐帧进行且需在最大截止时间 $T_{\max}$ 内完成。在每帧的起始时刻, ISAC网络系统依次对 $I$ 个目标执行感知操作, 持续时间为 $t_{s,i}, i \in \mathcal{I}$ , 用于通过观测获取相应的目标特征信息 $m_i$ 。系统获取 $I$ 个目标的特征信息后, 对融合特征进行编码处理, 并以包含 $d$  bit的数据包形式向服务器端传输。通信持续时间为 $t_r$ , 服务器在接收并解码数据包后, 基于区域状态信息的估计 $\hat{m}$ 在计算时间 $t_c$ 执行计算操作, 获得计算结果 $f(\hat{m})$ , 并将该结果用于控制执行器的行为。

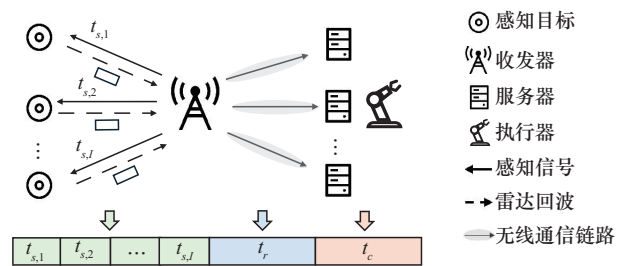


图1 多功能ISCC网络系统

在计算阶段, 使用 $c_0$ 表示系统总的工作负载,  $f_k$ 表示服务器 $k$ 的计算能力, 并且假设各服务器只要成功接收到输入的任务数据就立即开始处理计算任务。在多服务器系统中, 只有被选中的服务器才会被激活来处理来自收发器的任务。令二元决策变量 $a_k$ 表示对服务器 $k$ 的选择结果,  $a_k = 1$ 表示服务器被选中, 计算任务被卸载到该台服务器处理;  $a_k = 0$ 则表示该服务器不参与当前任务的处理。所以服务器选择结果的决策向量可以表示为

$$\mathbf{A} = \{a_1, \dots, a_K\} \quad (1)$$

所选服务器的集合用 $\hat{\mathcal{K}} = \{k | \forall a_k = 1\}$ 表示, 则集合 $\hat{\mathcal{K}}$ 的大小为 $\sum_{k=1}^K a_k$ 。

从系统请求的计算任务经过数据分割处理, 被分解为多个子任务, 这些子任务由被选中的服务器进行计算处理, 在平均分配的假设下, 每个服务器分配到的工作负载 $c$ 可以表示为

$$c = \frac{c_0}{\sum_{k=1}^K a_k}, \forall k \in \hat{\mathcal{K}} \quad (2)$$

考虑到系统运行环境中噪声的存在, 感知、通信与计算 3 个处理环节均无法保证完全可靠, 所以需要对各环节的错误概率特性进行深入分析。

### 1.1 感知模型

在语义通信中, 语义特征指能够反映任务相关含义的抽象信息表示。在多目标感知情形下, 将重点关注多个目标共同构成的区域状态这一“融合语义”而非单一目标的独立特征。本节通过检测目标在给定位置的存在性来刻画感知阶段错误率, 该检测过程本质上等效于对源环境的观测与特征提取。针对第  $i$  个目标的检测, 定义零假设  $H_0$  为目标不存在, 备择假设  $H_1$  为目标存在,  $y_d$  和  $y_s$  分别表示收发器从直达信道和监视信道接收到的信号, 则对第  $i$  个目标进行感知操作的接收信号可表示为

$$H_0^{(i)}: \begin{cases} y_{d,i} = \gamma_{d,i} s_i + a_{d,i} \\ y_{s,i} = a_{s,i} \end{cases}, H_1^{(i)}: \begin{cases} y_{d,i} = \gamma_{d,i} s_i + a_{d,i} \\ y_{s,i} = \gamma_{s,i} s_i + a_{s,i} \end{cases} \quad (3)$$

其中,  $s_i$  为执行第  $i$  个目标感知操作时 ISAC 发送的感知波形。由于感知系统对多个目标进行分时检测, 即在  $t_{s,i}$  的持续时间进行对第  $i$  个目标的感知操作, 因此  $\|s_i\|^2 = L_{s,i} P_{s,i}$ , 其中,  $P_{s,i}$  为感知功率, 时域符号  $L_{s,i}$  由  $L_{s,i} = \frac{t_{s,i}}{T_s}$  给出,  $T_s$  是单个符号的持续时间。当检测系统在没有违反给定的虚警概率阈值  $\delta_{\text{PFA}}$  的前提下仍无法检测到目标的存在, 则感知失败。根据文献[14]可知, 单个目标感知失败的概率为

$$\varepsilon_{s,i} = Q\left(\frac{P_{s,i} L_{s,i} \gamma_{s,i} - \sigma_s^2 \bar{\kappa}}{\sqrt{s P_{s,i} L_{s,i} \sigma_s^2 \gamma_{s,i}}}\right) \quad (4)$$

其中,  $\sigma_s$  为收发器收到的噪声功率。多目标感知系统从系统层面对多目标特征进行融合性建模, 并基于对多个目标的综合感知结果进行错误概率分析。执行器依据感知区域的整体状态做出行为决策, 考虑最坏情况下区域内任一目标的错误感知均可能导致该区域整体状态的错误判定, 因此区域感知的可靠性取决于所有目标的联合检测准确性。令  $P_s(Y=1)$  表示系统感知成功的概率, 由此得到的系统感知错误概率  $\varepsilon_s$  可表示为

$$\varepsilon_s = 1 - P_s(Y=1) = 1 - \prod_{i=1}^K (1 - \varepsilon_{s,i}) \quad (5)$$

### 1.2 通信模型

在通信阶段, 收发器将数据广播给多个服务器。考虑到不同传输链路信道状态不同, 设定收发器到服务器  $k$  的信道系数为  $h_{r,k}$ , 假设信道之间是独立同分布的且经历准静态瑞利衰落, 即  $h_k = \sqrt{\xi_k} \hat{h}_k$ , 其中,  $\sqrt{\xi_k}$  为大尺度路径损耗,  $\hat{h}_k$  为小尺度衰落, 则接收信号  $y_k = \sqrt{P_{r,k}} h_k x_k + a_k$ , 其中,  $P_{r,k}$  为发射功率,  $x_k$  为载有特征信息的信号,  $a_k \in \text{CN}(0, \sigma_r^2)$  为加性白高斯噪声。假设已知信道状态信息, 相应的信噪比 (SNR) 为  $\gamma_k = \frac{P_{r,k} h_k^2}{\phi_k \sigma_r^2}$ , 其中  $\phi_k$  是收发器到第  $k$  个服务器间的路径损耗。为满足严格的时延约束条件, 系统采用有限码长方案进行短包通信传输, 根据文献[15], 向服务器  $k$  传输的解码错误概率表示为

$$\varepsilon_{r,k} = P\left(\gamma_k, \frac{d}{L_r}, L_r\right) \approx Q\left(\sqrt{\frac{L_r}{V(\gamma_k)}} \left(C(\gamma_k) - \frac{d}{L_r}\right) \ln 2\right) \quad (6)$$

其中,  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$ ,  $C(\gamma) = \text{lb}(1 + \gamma)$  为香农信道容量,  $V = 1 - \frac{1}{(1 + \gamma)^2}$  为信道色散,  $L_r = \frac{t_r}{T_s}$  为码长。

### 1.3 计算模型

为了表征系统计算错误概率, 用  $D_k$  表示服务器  $k$  的总计算时延,  $c_k$  表示服务器  $k$  的工作负载,  $f_k$  表示服务器 CPU 的工作频率, 并假设单帧期间 CPU 频率固定。根据数据分割模型, 假设各任务能够独立划分并分别在本地执行, 使任务间互不依赖且可并行计算<sup>[16]</sup>。各服务器接收来自多个用户的计算请求, 且遵循先到先服务的调度策略。因此  $D_k$  包含计算时延和排队时延, 具体表示为  $D_k = \frac{c_k}{f_k} + W_k$ , 其中  $W_k$  是当前帧中在本任务之前随机到达的任务造成的排队时延, 其在语义通信中可被视为影响接收端对消息语义内容正确解析的一类有效噪声因素。如果服务器在给定时间  $t_c$  内没有完成给定任务, 则服务器发生计算错误, 相应的错误概率为

$$\varepsilon_{c,k} = \Pr(D_k \geq t_c) = \Pr(W_k + \frac{c}{f_k} \geq t_c) = \Pr(W_k \geq t_c - \frac{c}{f_k}) \quad (7)$$

将计算阶段调整后的时延约束参数记为  $\hat{t}_c = \max\{t_c - \frac{c}{f_k}, 0\}$ , 为了满足系统对可靠性的较高要求即极低的计算错误率,  $W_k$  的互补累积分布函数满足  $\bar{F}_{W_k}(\hat{t}_c) = \varepsilon_{c,k} = \Pr(W_k \geq \hat{t}_c) \ll 1$ 。极值理论能够在高可靠场景下有效刻画任务时延等指标的尾部特征, 因此将其用于分析所考虑语义通信系统的计算错误率  $\varepsilon_{c,k}$  的尾部特征<sup>[16-17]</sup>, 当  $t_c$  超出阈值  $t_{th}$  时, 定义超出部分  $X'_k = \max\{\hat{t}_c - t_{th}, 0\}$ , 如果  $t_{th}$  接近  $F_{W_k}^{-1}(1)$ ,  $X'_k$  的条件累积分布函数可表示为

$$F_{X'_k|W_k > t_{th}}(x_k) = \Pr(W_k - t_{th} \leq x_k | W_k > t_{th}) \approx G(x_k; \sigma, \zeta) \quad (8)$$

$$G(x_k; \sigma, \zeta) = \begin{cases} e^{-\frac{x_k}{\sigma}}, & \zeta = 0 \\ 1 - \left(1 + \frac{\zeta x_k}{\sigma}\right)^{-\frac{1}{\zeta}}, & \zeta \neq 0 \end{cases} \quad (9)$$

其中,  $G(x; \sigma, \zeta)$  为广义帕累托分布;  $\sigma$  为尺度参数且大于 0;  $\zeta$  为形状参数, 其具体值受计算任务到达率和服务器计算能力的影响, 可以通过足够的历史数据获得, 并且其有效性不依赖于任何特定的任务分布<sup>[16]</sup>。由于  $\zeta$  的取值直接影响分布的尾部特征, 本文假设系统分析中仅考虑  $\zeta > -0.5$  的情形, 以确保互补累积分布函数存在有限上界。因此, 在该系统场景下, 给定计算时间  $t_c$  对应的错误概率为

$$\varepsilon_{c,k} = \left(1 - F_{W_k}(t_{th})\right) \left(1 - G\left(\max\left\{t_c - \frac{c}{f_k} - t_{th}, 0\right\}; \sigma, \zeta\right)\right) \quad (10)$$

由于通信和计算阶段涉及多服务器协同操作, 如前所述, 当多个服务器被选中时, 各服务器计算分配到的子任务, 原始任务的结果依赖于各子任务的计算输出。因此只有当前帧中所有选中服务器的链路的通信和计算过程均无错误发生时, 整个系统的通信和计算才被认定为可靠。令  $\varepsilon_k$  表示链路  $k$  的出错概率, 即服务器  $k$  解码失败或计算时延超过  $t_c$  的概率, 则

$$\varepsilon_k = a_k(\varepsilon_{r,k} + (1 - \varepsilon_{r,k})\varepsilon_{c,k}) \quad (11)$$

显然, 没有被选中的服务器不会对错误概率造成影响, 因为  $a_k = 0, \forall k \notin \mathcal{K}$ 。对于多服务器系统,

可将通信与计算阶段的联合错误概率记为  $\varepsilon_0$ , 结合上述分析可得, 当前帧中通信和计算阶段不可靠的概率表达式为  $\varepsilon_0 = 1 - \prod_k (1 - \varepsilon_k)$ 。在高可靠场景的需求下,  $\{\varepsilon_{r,k}, \varepsilon_{c,k}\} \leq \varepsilon_{th} \ll 0.1$ , 这意味着任意错误率的乘积项会远小于其累加项, 结合式(11),  $\varepsilon_0$  可被近似表示为

$$\varepsilon_0 = 1 - \prod_k (1 - \varepsilon_k) \approx \sum_{k=1}^K (a_k \varepsilon_{r,k} + a_k \varepsilon_{c,k}) \quad (12)$$

#### 1.4 语义通信可靠性描述

与传统通信侧重比特或符号级的准确传输不同, 语义通信强调的是基于环境的任务相关语义信息交换。其过程是在源端根据环境感知结果提取任务相关的语义特征, 通过语义信道传输后, 接收端通过语义译码器对特征信息进行恢复。并且当接收端基于失真观测  $\hat{s}$  获得的环境语义信息  $f(\hat{s})$  与发送端的语义信息  $f(s)$  一致时, 则认为通信在语义层面是可靠的。在语义信息理论框架中, 语义熵作为衡量任务相关语义不确定性的关键度量, 可反映语义任务被正确解释与使用的难度<sup>[7]</sup>。从系统可靠性的角度来看, 当发送端生成的语义任务与接收端重构的语义信息不一致时, 就会出现语义任务解释错误, 从而降低系统整体性能。本文重点关注多功能 ISCC 网络中的语义层, ISCC 网络将通信、感知与计算三大模块深度融合, 在实际运行过程中, 各模块均可能产生潜在错误。基于这一思想, 文献[18]从系统视角分析了这种语义任务不确定性在 ISCC 各功能组件间的传递与演化, 并据此构建了语义可靠性的度量框架, 从功能链路角度描述语义失真, 整体上刻画了多功能任务的系统性能。在 ISCC 网络中, 语义通信过程可视为源端基于环境的实时观测, 从原始物理信号中提取反映区域状态的特征信息。该特征提取过程本质上属于语义层抽象, 即将低维度的原始观测数据转化为能够表征区域状态的关键语义, 随后这些语义特征作为待传输数据, 在通信阶段完成编码与传输, 目的端接收后进行特征译码(计算阶段), 并据此进行任务决策。

语义通信系统对应 ISCC 网络中的 3 项功能, 即感知、通信和计算, 如图 2 所示。在感知阶段, 通过观测从源环境  $s$  中提取的语义特征  $m$ , 得到语义知识编码输出  $m = X_k$ ; 在通信阶段,  $m$  经过数据编码后得到可在物理信道上传输的底层数据表示

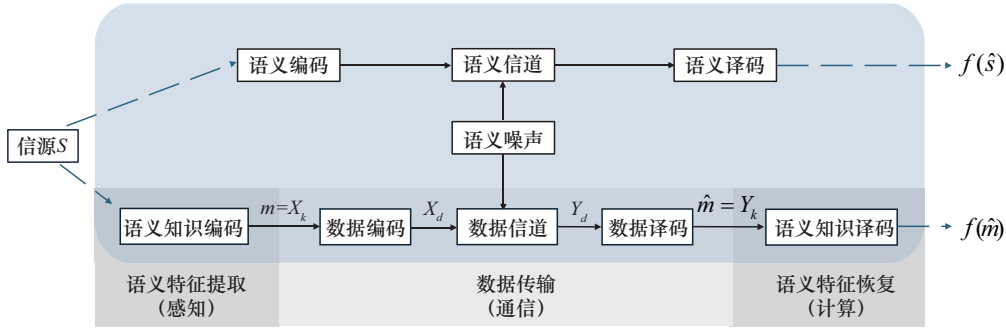


图2 语义通信系统各部分功能在 ISCC 网络中的对应组件

$X_d$ ，然后经过物理信道后得到  $Y_d$ ，并对其数据进行数据译码（解调、译码、去量化等）后得到  $\hat{m} = Y_k$ ；在计算阶段， $\hat{m}$  被映射到最终任务结果  $f(\hat{m})$ 。显然，如果接收端基于  $\hat{s}$  得到的  $f(\hat{s})$  与  $f(s)$  一致，则语义信息被认为正确传输，可以通过 ISCC 网络的各个组件描述任务的语义可靠性。由于后续传输错误率和优化问题分析需要，这里进行如下假设：①当语义传输速率  $R_{sem}$  小于语义容量  $C_{sem}$  时，传输码长趋于无穷大，语义通信可实现任意高的可靠性；②在给定的检测时刻，假设发送端从源环境中只能提取到唯一的语义特征  $m$ ，使基于该特征  $m$  做出的任务决策  $f(m)$  与  $f(s)$  完全一致。对于在真实环境状态中有可能提取到的其他可能的特征  $n$ ，根据  $n$  得到的决策  $f(n)$  与  $f(s)$  不同，即  $f(s) \neq f(n), \forall n \neq m$ 。

第一个假设保证了当语义传输的码长无穷大时，语义错误率可以任意小。在 ISCC 网络中，特征的提取由目标感知检测表示，因此第二个假设表明基于真实环境状态得到的决策和基于特征得到的决策是一一映射关系，即在雷达感知过程中只能提取一种可能的反映区域环境状态的特征。另外，本文的研究范围不具体涉及语义速率和语义信道容量。为了刻画语义错误率，结合上述假设，ISCC 网络中的语义可靠性如定义 1 所述。

**定义 1** 当且仅当服务器基于接收到的观测值  $\hat{m}$  计算出的决策  $f(\hat{m})$  与基于原始环境  $s$  的决策相同时，即  $f(s) = f(\hat{m})$ ，通信在语义上是可靠的。

显然，只有当语义层中的每个系统组件均可靠运行时，才能实现语义可靠性。需要强调的是，只要满足上述两个假设条件，定义 1 就不依赖于具体的语义编码器和解码器<sup>[18]</sup>。因此，本文采用语义错误率刻画语义特征传输的可靠性，而其与任务级连续指标属于链路不同层级，可通过有效语义速率

等概念实现关联。语义错误概率  $\epsilon_{sem}$  与语义可靠性  $P_{sem}$  的关系可表示为

$$\epsilon_{sem} = 1 - P_{sem} = 1 - (1 - \epsilon_s)(1 - \epsilon_o) \quad (13)$$

注意，对  $\epsilon_{s,i}$  同样有  $\epsilon_{s,i} \leq \epsilon_{th} \ll 0.1$ ，对式(5)、式(11)和式(12)进行化简，语义错误概率  $\epsilon_{sem}$  可近似表示为

$$\epsilon_{sem} \approx \sum_{i=1}^I \epsilon_{s,i} + \sum_{k=1}^K (a_k \epsilon_{r,k} + a_k \epsilon_{c,k}) \quad (14)$$

式(14)的近似成立分析与式(12)类似。需要说明的是，实际系统中感知阶段的语义错误可能影响通信编码效率及计算决策，但准确刻画这种错误传播需要获得感知错误的确定性结果，这将依赖于大量先验信息，在实际 ISCC 系统中难以获取。第 2 节将通过所提出的优化问题，基于目标函数  $\epsilon_{sem}$  将各阶段性能统一整合，从而在系统层面考虑阶段间的相互影响，使得在保持模型可解性的同时，实现对各阶段联合性能的有效优化。

## 2 优化问题

### 2.1 问题提出与重构

为提升语义通信的端到端可靠性，需要在系统各阶段之间对有限的码长资源进行合理分配，并同时优化服务器选择策略，以有效降低语义错误率。码长资源在通信、感知与计算 3 个阶段中具有较强的关联性 & 相互制约性，对系统整体可靠性有更直接的影响。基于上述考虑，本文构建了一个以最小化语义错误率为目标的优化问题，在给定总时延约束  $T_{max}$  内，通过联合优化 3 个阶段的时间分配及服务器选择策略，最大化语义可靠性。具体优化问题如下

$$\text{OP: } \min_{\{t_{s,1}, \dots, t_{s,i}, t_r, t_c, A\}} \epsilon_{sem} \quad (15a)$$

$$\text{s.t. } \mathbf{A} \in \{0,1\}^K \quad (15b)$$

$$\varepsilon_{s,i} \leq \varepsilon_{\text{th}}, \forall i \in \mathcal{I} \quad (15c)$$

$$\varepsilon_k \leq \varepsilon_{\text{th}}, \forall k \in \mathcal{K} \quad (15d)$$

$$\sum_{i=1}^I t_{s,i} + t_r + t_c \leq T_{\max} \quad (15e)$$

$$c = \frac{c_0}{\sum_{k=1}^K a_k}, \forall k \in \mathcal{K} \quad (15f)$$

$$\sum_{k=1}^K a_k \geq 1 \quad (15g)$$

其中, 约束式(15c)和式(15d)通过要求各阶段的错误率满足给定上界来保证系统的可靠性。由于问题OP中使用了二元决策变量 $a_k$ 表示服务器的选择情况, 因此将问题OP分解为一系列子问题, 并通过推导得出的相关结论对子问题进行分析, 求解出最优结果。

对于 $K$ 个服务器的系统, 存在 $(2^K - 1)$ 种服务器选择组合, 因此问题OP可被分解为 $(2^K - 1)$ 个不同 $\hat{\mathcal{K}}$ 的子问题。对于一个给定的集合 $\hat{\mathcal{K}}$ , 子问题可以表示为

$$\text{P1: } \min_{\{t_{s,1}, \dots, t_{s,I}, t_r, t_c\}} \varepsilon_{\text{sem}} \quad (16a)$$

$$\text{s.t. } c = \frac{c_0}{\sum_{k=1}^K a_k}, a_k = 1, \forall k \in \hat{\mathcal{K}} \quad (16b)$$

式(15c), 式(15d), 式(15e)

为了降低问题P1的复杂度, 下面给出3个刻画该问题特性的定理。定理1给出了语义错误率关于各阶段时长变量的单调性。

**定理1**  $\varepsilon_{\text{sem}}$ 分别关于 $t_{s,i}$ 、 $t_r$ 和 $t_c$ 单调递减。

**证明**  $\varepsilon_{\text{sem}}$ 关于 $t_{s,i}$ 的一阶偏导数为

$$\frac{\partial \varepsilon_{\text{sem}}}{\partial t_{s,i}} = \frac{\partial \varepsilon_{s,i}}{\partial t_{s,i}} + 0 = \frac{\partial \varepsilon_{s,i}}{\partial \omega_{s,i}} \frac{\partial \omega_{s,i}}{\partial t_{s,i}} = \frac{\partial \varepsilon_{s,i}}{\partial \omega_{s,i}} \sqrt{\frac{P_s \gamma_{s,i}}{2t_{s,i}}} \leq 0 \quad (17)$$

其中,  $\omega_{s,i} = \sqrt{\frac{sL_s P_s \gamma_{s,i}}{\sigma^2}} - Q^{-1}(\delta_{\text{PFA}})$ , 因此 $\varepsilon_{\text{sem}}$ 关于 $t_{s,i}$ 单调递减, 同样, 对于 $t_r$ 有

$$\frac{\partial \varepsilon_{\text{sem}}}{\partial t_r} = \sum_{k=1}^K \left( -\frac{1}{\sqrt{2\pi T_S}} e^{-\frac{\omega_{r,k}^2}{2}} \frac{L_r}{V_{r,k}} \left( C^2(\gamma_k) - \left( \frac{d}{L_r} \right)^2 \right) \right) \leq 0 \quad (18)$$

其中,  $\omega_{r,k} = \sqrt{V(\gamma_k)} L_r (C(\gamma_k) - \frac{d}{L_r})$ , 当 $\varepsilon_r \leq \varepsilon_{\text{th}} \leq 0.5$ 时,  $C(\gamma_k) - \frac{d}{L_r} \geq 0$ 成立, 不等式(18)成立, 即

$\varepsilon_{\text{sem}}$ 关于 $t_r$ 单调递减。此外,  $\varepsilon_{\text{sem}}$ 关于 $t_c$ 的一阶偏导数为

$$\frac{\partial \varepsilon_{\text{sem}}}{\partial t_c} = \sum_{k=1}^K \frac{\partial \varepsilon_{c,k}}{\partial t_c} = \sum_{k=1}^K \left( (-F_W(t_{\text{th}})) \frac{(1+\xi)}{\sigma^2} \cdot (1 - G(t_c - \frac{c}{f_k} - t_{\text{th}}; \sigma, \xi))^{-\frac{1+\xi}{\xi}} \right) \leq 0 \quad (19)$$

综上,  $\varepsilon_{\text{sem}}$ 关于 $t_{s,i}$ 、 $t_r$ 和 $t_c$ 都是单调递减的。证毕。

定理2根据定理1描述了达到最优分配时各时长之间的关系。

**定理2** 对于最优解 $t_{s,i}^*$ 、 $t_r^*$ 和 $t_c^*$ ,  $\sum_{i=1}^I t_{s,i}^* + t_r^* + t_c^* = T_{\max}$ 成立。

**证明** 通过反证法证明。假设 $t'_{s,i}$ 、 $t'_r$ 和 $t'_c$ 为最优解且 $T_{\max} - (\sum_{i=1}^I t'_{s,i} + t'_r + t'_c) = \alpha > 0$ , 则对于任意可行解 $t_{s,i}$ 、 $t_r$ 和 $t_c$ ,  $\varepsilon'_{\text{sem}}(t'_{s,i}, t'_r, t'_c) \leq \varepsilon_{\text{sem}}(t_{s,i}, t_r, t_c)$ 成立。此外, 假设存在可行解 $t''_r = t'_r + \alpha$ 、 $t''_{s,i} = t'_{s,i}$ 、 $t''_c = t'_c$ , 则由定理1的单调性可得 $\varepsilon''_{\text{sem}}(t'_{s,i}, t''_r, t'_c) < \varepsilon'_{\text{sem}}(t'_{s,i}, t'_r, t'_c)$ , 与 $t'_{s,i}$ 、 $t'_r$ 和 $t'_c$ 为最优解的假设矛盾。证毕。

基于定理1和定理2, 可以通过改写约束式(15e)缩减原问题的变量数。事实上, 在对式(14)的分析中, 优化目标函数已被近似为线性可加的形式。该形式有助于分析优化问题的凸性, 即若进一步证明问题满足凸性条件, 则可直接利用标准的凸优化工具求取全局最优解。接下来, 利用定理3对该问题的凸性进行进一步分析。

**定理3** 问题P1是凸问题。

**证明** 首先给出目标函数的Hessian矩阵

$$\mathbf{H} = \begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{b}} & \bar{\mathbf{c}} \\ \bar{\mathbf{b}}^T & \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_r^2} & \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_r \partial t_c} \\ \bar{\mathbf{c}}^T & \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_c \partial t_r} & \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_c^2} \end{bmatrix} \quad (20)$$

其中,  $\bar{\mathbf{A}}$ 是 $i \times i$ 维矩阵, 其元素为 $\bar{A}_{xy} = \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_{s,x} \partial t_{s,y}}$ ,  $x, y = 1, \dots, i$ ;  $\bar{\mathbf{b}}$ 和 $\bar{\mathbf{c}}$ 是 $i \times 1$ 维列向量, 分量分别为

$$\bar{b}_x = \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_{s,x} \partial t_r}, \quad \bar{c}_x = \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_{s,x} \partial t_c}. \quad \text{结合式 (14),} \quad \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_i^2} =$$

$$\frac{\partial^2 \varepsilon_i}{\partial t_i^2}, i \in \{s, r, c\} \text{ 和 } \frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_i \partial t_j} = 0, i \neq j \text{ 成立, 则可通过证明}$$

明对角线元素非负得到 Hessian 矩阵是半正定的。结合定理 1 可得

$$\frac{\partial^2 \varepsilon_{s,i}}{\partial t_{s,i}^2} = \underbrace{\frac{\partial^2 \varepsilon_{s,i}}{\partial \omega_{s,i}^2} \left( \frac{\partial \varepsilon_{s,i}}{\partial \omega_{s,i}} \right)^2}_{\geq 0} + \underbrace{\frac{\partial \varepsilon_{s,i}}{\partial \omega_{s,i}} \left( -\frac{3\sigma_s^2 \bar{\kappa} + L_{s,i}}{4 \sqrt{2P_s L_{s,i}^5 \sigma_s^2 \gamma_{s,i}}} \right)}_{\leq 0} \geq 0 \quad (21)$$

$\varepsilon_{\text{sem}}$  关于  $t_r$  的二阶偏导数为

$$\frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_r^2} = \sum_{k=1}^K \left( \frac{e^{-\frac{\omega_{r,k}^2}{2}}}{4T_s \sqrt{2\pi V_{r,k} t_r^3}} \left( (C_{r,k} + \frac{d}{L_r})^2 \omega_{r,k} + (C_{r,k} + 3\frac{d}{L_r}) \right) \right) \geq 0 \quad (22)$$

$\varepsilon_{\text{sem}}$  关于  $t_c$  的二阶偏导数为

$$\frac{\partial^2 \varepsilon_{\text{sem}}}{\partial t_c^2} = \sum_{k=1}^K F_W(t_{\text{th}}) \frac{(1 + \zeta)}{\sigma^2} \left( 1 - G\left(t_c - \frac{c}{f_k} - t_{\text{th}}; \sigma, \zeta\right) \right)^{\frac{(2 + \zeta)}{\zeta}} \geq 0 \quad (23)$$

所以目标函数的 Hessian 矩阵是半正定的, 此外, 矩阵中元素非负意味着  $\varepsilon_{s,i}$ 、 $\varepsilon_{r,k}$  和  $\varepsilon_{c,k}$  在每个变量上都是凸的, 从而式(15c)和式(15d)构成凸集, 而式(15e)是线性约束。因此问题 P1 是凸的。证毕。

根据以上对问题 P1 的分析, 可以将原问题 OP 转换为

$$\text{P2:} \quad \min_{\{t_{s,1}, \dots, t_{s,i}, t_r, t_c, \mathbf{A}\}} \varepsilon_{\text{sem}} \quad (24a)$$

$$\text{s.t.} \quad \sum_{i=1}^I t_{s,i} + t_r + t_c = T_{\text{max}} \quad (24b)$$

$$\varepsilon_k \leq a_k, \forall k \in \mathcal{K} \quad (24c)$$

$$a_k (\varepsilon_{1,k} + \varepsilon_{2,k} - \varepsilon_{1,k} \varepsilon_{2,k}) \leq \varepsilon_k, \forall k \in \mathcal{K} \quad (24d)$$

式(15b), 式(15c), 式(15d), 式(15f), 式(15g)

其中, 约束条件式(24c)和式(24d)有助于通过对目标函数进行线性化处理, 消除变量  $a_k$  与  $t_r$  的乘积项。问题 P2 的目标函数与约束条件均为仿射或凸形式, 经过重构后可变为混合整数凸优化问题。该方法的计算复杂度为  $\mathcal{O}(2^K - 1)$ , 因为其需要分别求解  $(2^K - 1)$  种服务器组合下对应的凸优化问题。该复杂度随服务器数量呈指数级增长, 其实际

可行性将在第 4 节中进一步讨论。

## 2.2 拉格朗日函数与 KKT 最优化条件

第 2.1 节已经通过对原始问题进行结构化分解并严格证明各子问题的凸性获得全局最优解, 在此基础上, 本节进一步构建原问题的拉格朗日函数并推导相应的 KKT 条件。注意, 由于原问题的凸性, KKT 条件既是最优性的必要条件也是充分条件, KKT 解不仅具有明确的数学形式, 而且对应原问题的全局最优解。这将有助于在确保系统语义通信可靠性提升的前提下, 进一步分析多服务器协同处理任务时的最优服务器选择策略及计算阶段时长分配的最优解结构, 从而更全面地理解多功能 ISCC 网络中的决策机制, 并揭示了在此类复杂系统中实现目标性能所需的有效资源选择与分配。在分析过程中, 将  $a_k$  松弛为连续变量且  $a'_k \in [0, 1]$ ,  $\mathbf{A}' \in [0, 1]^K$ , 然后将优化问题问

$$\text{P3:} \quad \min_{\{t_{s,1}, \dots, t_{s,i}, t_r, t_c, \mathbf{A}'\}} \varepsilon_{\text{sem}} \quad (25a)$$

$$\text{s.t.} \quad \mathbf{A}' \in \{0, 1\}^K \quad (25b)$$

$$c = \frac{c_0}{\sum_{k=1}^K a'_k}, \forall k \in \mathcal{K} \quad (25c)$$

$$\sum_{k=1}^K a'_k \geq 1 \quad (25d)$$

$$\varepsilon_k \leq a'_k, \forall k \in \mathcal{K} \quad (25e)$$

$$a'_k (\varepsilon_{1,k} + \varepsilon_{2,k} - \varepsilon_{1,k} \varepsilon_{2,k}) \leq \varepsilon_k, \forall k \in \mathcal{K} \quad (25f)$$

式(15c), 式(15d), 式(24b)

问题 P3 的拉格朗日函数为

$$\begin{aligned} \mathcal{L}(t_s, t_r, t_c, \mathbf{A}', \lambda, \mu) = & \sum_{i=1}^I \varepsilon_{s,i} + \sum_{k=1}^K \varepsilon_k + \sum_{i=1}^I \lambda_{s,i} (\varepsilon_{s,i} - \varepsilon_{\text{th}}) + \\ & \sum_{k=1}^K \lambda_k (\varepsilon_k - \varepsilon_{\text{th}}) + \mu \left( \sum_{i=1}^I t_{s,i} + t_r + t_c - T_{\text{max}} \right) \end{aligned} \quad (26)$$

其中,  $\lambda$  和  $\mu$  为拉格朗日乘子。对应的 KKT 条件为

$$\frac{\partial \mathcal{L}}{\partial t_{s,i}} = \frac{\partial \varepsilon_{s,i}}{\partial t_{s,i}} + \lambda_{s,i} \frac{\partial \varepsilon_{s,i}}{\partial t_{s,i}} + \mu^* = 0 \quad (27a)$$

$$\frac{\partial \mathcal{L}}{\partial t_r} = \sum_{k=1}^K a'_k \frac{\partial \varepsilon_{r,k}}{\partial t_r} + \sum_{k=1}^K \lambda_k^* \frac{\partial \varepsilon_{r,k}}{\partial t_r} + \mu^* = 0 \quad (27b)$$

$$\frac{\partial \mathcal{L}}{\partial t_c} = \sum_{k=1}^K a'_k \frac{\partial \varepsilon_{c,k}}{\partial t_c} + \sum_{k=1}^K \lambda_k^* \frac{\partial \varepsilon_{c,k}}{\partial t_c} + \mu^* = 0 \quad (27c)$$

$$\frac{\partial \mathcal{L}}{\partial a'_k} = (\lambda_k^* + 1)(\varepsilon_{r,k} + \varepsilon_{c,k}) + \sum_{i=1}^K a'_i \frac{\partial \varepsilon_{c,i}}{\partial a'_k} + \sum_{i=1}^K \lambda_i^* a'_i \frac{\partial \varepsilon_{c,i}}{\partial a'_k} = 0 \quad (27d)$$

$$\lambda_{s,i}^* \geq 0, \lambda_{s,i}^*(\varepsilon_{s,i} - \varepsilon_{th}) = 0, \forall i \in \mathcal{I} \quad (27e)$$

$$\lambda_k^* \geq 0, \lambda_k^*(\varepsilon_k - \varepsilon_{th}) = 0, \forall k \in \mathcal{K} \quad (27f)$$

消去式(27c)中 $\lambda$ , 可以得到

$$-\mu^* = \sum_{k=1}^K a'_k \frac{\partial \varepsilon_{c,k}}{\partial t_c} = \frac{F_W(t_{th}) - 1}{\sigma^2} \sum_{k=1}^K a'_k \left( \sigma + \zeta \left( t_c - t_{th} - \frac{c_0}{(f \sum_{k=1}^K a'_k)} \right)^{-\frac{1+\zeta}{\zeta}} \right) \quad (28)$$

在式(28)中, 因为 $\sum_{k=1}^K a'_k$ 对于每个项都相同, 所以它不包含在总和中, 则可以解得

$$t_c^*(a'_k) = t_{th} + \frac{c_0}{f_k \sum_{k=1}^K a'_k} + \frac{\sigma}{\zeta} \left( \left( \frac{-\mu^* \sigma}{(F_W(t_{th}) - 1) \sum_{k=1}^K a'_k} \right)^{-\frac{\zeta}{1+\zeta}} - 1 \right) \quad (29)$$

该计算时长的最优解表明, 如果服务器计算能力相同, 那么分配的最优计算时长与选择的服务器总数量有关, 与服务器的组合情况无关。此外, 工作负载的增加也会导致计算时长增大。

同样, 消去式(27d)中 $\lambda$ , 有

$$\varepsilon_{r,k} + \varepsilon_{c,k} + \sum_{i=1}^K a'_i \frac{\partial \varepsilon_{c,i}}{\partial a'_k} = 0 \quad (30)$$

对于不同服务器的选择概率 $a'_k$ , 可得

$$a'_k = \frac{-\varepsilon_{r,k} - \varepsilon_{c,k} + \frac{c_0}{(\sum_{k=1}^K a'_k)^2} \sum_{i \neq k}^K a'_i E(\sum_{k=1}^K a'_k f_i)}{-c_0 \frac{E(\sum_{k=1}^K a'_k f_k)}{(\sum_{k=1}^K a'_k)^2}} \quad (31)$$

$$\text{其中, } E\left(\sum_{k=1}^K a'_k f\right) = \frac{\partial \varepsilon_{c,k}}{\partial c} = \frac{(1 - F_W(t_{th}))}{\sigma f}.$$

$$\left( 1 + \frac{\zeta \left( t_c - t_{th} - \frac{c_0}{(f \sum_{k=1}^K a'_k)} \right)^{-\frac{1+\zeta}{\zeta}}}{\sigma} \right)^{-\frac{1+\zeta}{\zeta}}.$$

显然, 当 $\zeta \leq 0$ ,

$\sigma \geq 0$ 时,  $E$ 关于 $f$ 单调递减, 这意味着当其他计算节点的处理能力提升时, 相对较弱的服务器被选中的概率将降低; 相反, 服务器计算能力的增强将增大其在资源分配过程中的选择概率。此外, 当服务器的链路通信错误率和计算错误率较高时, 选择该服务器的概率也会降低。如果各服务器计算能力相同, 则有 $\frac{\partial \varepsilon_{c,1}}{\partial a'_k} = \frac{\partial \varepsilon_{c,2}}{\partial a'_k} = \dots = \frac{\partial \varepsilon_{c,i}}{\partial a'_k}$ , 令 $\frac{\partial \varepsilon_{c,i}}{\partial a'_k} = \mathcal{A}$ , 式(30)可写为

$$\varepsilon_{r,k} + \varepsilon_{c,k} + \mathcal{A} \sum_{i=1}^K a'_i = 0 \quad (32)$$

该形式下的最优服务器数量 $\sum_{k=1}^K a'_k$ 可表示为

$$\sum_{k=1}^K a'_k = \frac{c_0 E(\sum_{k=1}^K a'_k f)}{\varepsilon_{r,k} + \varepsilon_{c,k}} \quad (33)$$

在式(33)中, 系统总工作负载 $c_0$ 与最优服务器数量 $\sum_{k=1}^K a'_k$ 存在耦合, 无法将 $\sum_{k=1}^K a'_k$ 显式表示为 $c_0$ 的独立函数, 即使各服务器计算能力相同,  $c_0$ 的增加也不一定会导致选择的服务器总数增多, 这是因为当服务器总数固定时, 不同的服务器组合会使任务在各链路上的通信错误率发生变化, 进而对所选服务器链路的错误率产生影响。另一方面, 尽管无法在函数形式上分离最优解与其他参数, 但在各服务器计算能力相同的前提下, 可根据式(33)构建包含 $K$ 个未知量的方程组, 当系统各参数给定时, 能够通过数值求解该规模为 $K$ 的方程组得到每个服务器对应的最优选择概率 $a'_k$ , 这种考虑特殊情形下的数值求解方法复杂度更低, 在实际工程上也具有较广泛的适用性。

需要指出的是, 基于变量松弛并利用 KKT 条件所得的最优解  $a'_k$  对应于服务器选择概率的连续形式, 并非原问题 (OP) 中的最优整数解  $a_k$ 。为获得可行的服务器选择决策, 可对该连续解采取简单映射策略 (如四舍五入方法), 得到原问题的可行整数解 (即次优解)。这也意味着相较于直接求解原问题以获取全局最优解, 该基于松弛的求解方式在显著降低计算复杂度的同时牺牲了一定的最优性。

### 3 仿真结果

本节通过蒙特卡罗模拟得到数值仿真结果, 并对所提出的设计方案进行性能评估。除特殊说明外, 仿真中采用如下参数设置。收发器需要进行感知的目标数  $I=3$ , 收发器到各个目标的归一化反射系数分别为  $\gamma_{s,1}=0.8$ ,  $\gamma_{s,2}=0.9$ ,  $\gamma_{s,3}=1$ 。收发器对各目标感知功率相等且  $P_s=230$  mW, 通信阶段发射功率  $P_r=150$  mW。对于无线信道参数, 设定载波频率  $F=2.4$  GHz, 热噪声功率  $N_0=-154$  dBm/Hz。收发器发送到服务器的含有观测数据的数据包大小为  $d=1\ 600$  bit, 单个时域符号的持续时间为  $T_s=0.025$  ms。为了更全面地分析, 仿真中设置了同构服务器部署和异构服务器部署两种场景, 对于同构服务器, 各服务器到收发器的距离固定且相同, 距离范围为  $[5,20]$  m; 对于异构服务器, 设置其均匀分布在距收发器 5~20 m 处。使用文献[19]中提出的传输路径损耗模型, 表达式为  $\phi = 17.0 + 40\lg(r)$ ,  $r$  为收发器到服务器的距离。系统中总工作负载  $c_0=35$  Mcycle, 假设每个特定的服务器有相同的计算能力  $f$ , 且  $f=10$  GHz。为了满足 99.9% 的可靠性约束, 使用文献[17]中广义帕累托分布的相关参数, 即形状参数  $\zeta=-0.021\ 4$ , 尺度参数  $\sigma = 3.495\ 5 \times 10^6$ , 阈值  $d = 2.038\ 4 \times 10^7$ 。最后设定总时长资源  $T_{\max}=45$  ms, 错误概率阈值  $\varepsilon_{th}=0.01$ 。

将  $c_0$  设置为 90 Mcycle, 图 3 绘制了最小化语义错误率  $\varepsilon_{sem}^*$  与最大持续时间  $T_{\max}$  的关系, 并且引入穷举法 (用 “Exh.” 表示) 以验证所提方案的有效性。注意到, 在所提资源分配方案中,  $T_{\max}$  的增加会降低  $\varepsilon_{sem}$ , 因为更长的截止时间允许更有效地利用资源, 从而提升系统性能。与仅对通信阶段资源进行优化分配的传统方案相比, 所提方案在语义可靠性方面表现出显著优势, 这说明在

ISCC 网络中, 仅针对物理层通信性能进行优化难以充分保障系统性能, 物理层优化只是系统设计中的一个组成部分, 无法全面刻画和反映语义通信系统的端到端可靠性。如果直接将时长资源对每个阶段进行均匀分配, 即使在截止时长较大时, 该方案下的  $\varepsilon_{sem}$  也会明显高于所提方案, 表明在多阶段任务执行系统中根据实际情况分配资源的重要性。如果给定感知错误率  $\varepsilon_s$  仅优化通算错误率  $\varepsilon_0$ , 会发现通过使感知时长满足感知错误率阈值而剩余时长全部用于通信与计算阶段, 随着  $T_{\max}$  的增加,  $\varepsilon_0$  能够持续降低。然而从系统整体语义可靠性的角度观察会发现, 即使  $T_{\max}$  较大, 语义可靠性的提高也非常受限, 说明语义错误率能够更准确地刻画系统整体性能的变化, 以及码长资源分配对系统可靠性的影响。此外, 在同构和异构部署下, 所提方案均与穷举法得到的全局最优解高度一致, 表明优化方案在显著降低复杂度的同时可得到 (近似) 全局最优解。在同构和异构两种场景下, 可用服务器总数的增加都会使  $\varepsilon_{sem}$  减小, 相关分析将在图 4 中继续深入讨论。

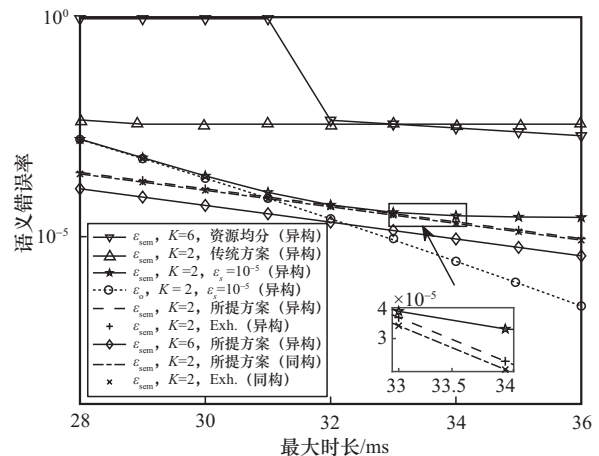


图 3 语义错误率与最大时长  $T_{\max}$  的关系

图 4 显示了服务器数量改变对优化选择和全部选择 (分别用 “Opt.” 和 “All” 表示) 两种策略的  $\varepsilon_{sem}$  的影响, 其中, 对于同构服务器仅展示最优方法下的结果。显然, 在同构和异构两种场景下的优化方案中, 增加服务器数量总体上都会提高系统语义可靠性, 这与图 3 的结果一致。然而, 随着服务器数量的持续增加, 其对语义可靠性的改善幅度会减少, 表明当服务器数量达到一定规模后,  $K$  不再是语义错误的主导因素, 因此,

需要考虑其他因素的影响。此外注意到,在异构场景下,若不经优化选择而是直接将所有服务器纳入系统,服务器数量的增加反而可能导致语义可靠性下降,这是因为当参与的服务器数量增加时,系统整体错误概率会因多条链路错误率的累积而增加,从而导致语义可靠性下降。在通信阶段,异构部署下的不同服务器对应的链路存在不同的路径损耗和信道条件,如果选择全部服务器,部分链路质量较差的服务器不可避免地加入系统,从而导致语义可靠性下降。这表明在大规模网络系统中,服务器选择优化策略具有重要意义。同时,仿真结果显示,在所考虑的 4 种策略中,多服务器配置相比单服务器配置在语义可靠性方面均表现出显著的性能提升。

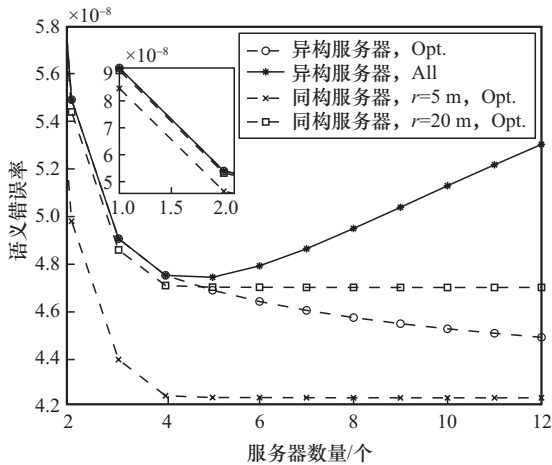


图 4 不同场景设置下服务器数量改变对语义错误率的影响

图 5 显示了总工作负载对语义错误率的影响及达到最小语义错误率时的最优服务器数量。与预期结果相符,随着总工作负载的增加,语义错误率增大,且无论是在  $K = 7$  还是在  $K = 3$  的配置下,服务器优化选择方案的性能始终优于非优化方案(直接选择所有服务器),该优势在服务器数量更大的情况下更加明显,这与图 4 的结果一致。另外,图 5 还表明,随着工作负载的增加,最优服务器数量呈现阶段性不变的特征,这与 3.2 节所得最优服务器数量表达式的特性相符,即工作负载的增加不一定要求系统选择更多的服务器。因为不同服务器组合对应不同的通信传输路径,即使是相同的最优服务器数量配置,链路错误率仍会因信道增益和路径损耗等因素的差异而发生变化。

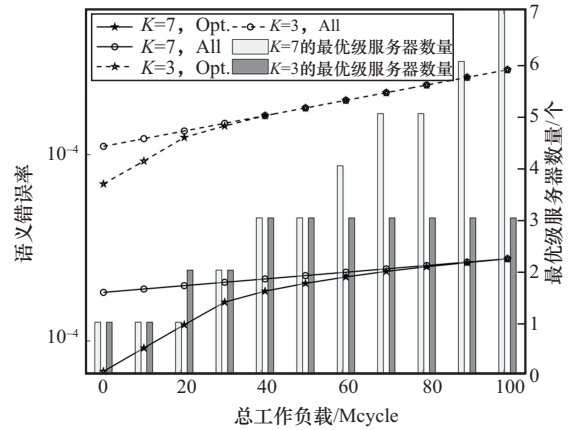


图 5 异构部署下工作负载与语义可靠性的关系

图 6 展示了  $\epsilon_{\text{sem}}$  随通信阶段发射功率的变化情况,并对比了不同服务器数量和虚警概率配置下的优化结果。结果表明,提高发射功率可有效降低语义错误率,这是由于  $P_t$  的提升改善了通信链路的信噪比,进而降低了通信错误率并提升了系统整体性能。与基准方案相比,所提方案仍具有更优的性能,并且这种改善随着服务器数量的增加变得更加显著。此外,当功率足够大时,优化选择与全选方案在语义错误率上接近相等,说明充足的功率配置能够有效缓解次优服务器选择策略带来的性能损失。最后,从图 6 中还可以看出,当系统要求较高的虚警概率阈值时,会导致较差的语义错误率,则需要分配更多的资源用于感知。这一观察表明,在感知部分,检测错误率和  $\delta_{\text{PFA}}$  之间存在着复杂的权衡关系。

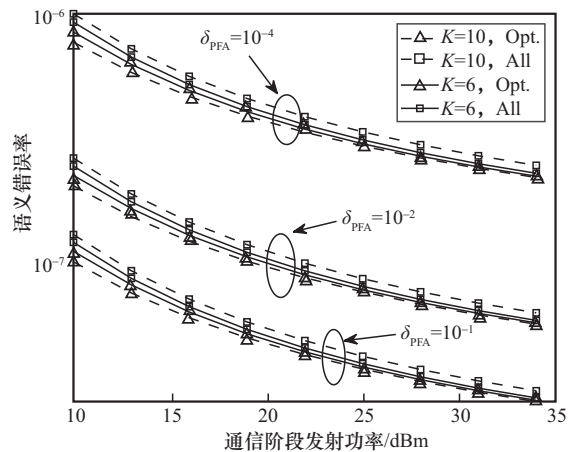


图 6 通信发射功率变化对语义错误率的影响

图 7 展示了异构场景下感知目标数量对语义可靠性的影响。由图 7 可知,感知目标数量增多,

语义错误率增大，此时各阶段对时长资源的竞争也更加激烈。如果系统语义可靠性不变，提高感知功率意味着可以检测更多的目标任务。尽管仿真中考虑的感知规模相对有限，所提方案在该范围内仍表现出良好的性能。图 8 主要分析了包含目标特征信息的数据包大小对语义可靠性的影响，同时也结合不同感知目标数量进行了对比观察。显然，数据包大小增加也会增大语义错误率，这是因为对于给定的码长，较大的输入数据需要更高的编码率，从而增加通信阶段的错误率，导致端到端的可靠性下降。在图 8 中，当可用服务器数量  $K = 2$  时，优化方案与非优化方案的优化结果相同，这是因为此时可选择的服务器组合方案较少，最优服务器组合恰好选择全部服务器的可能性显著增大。

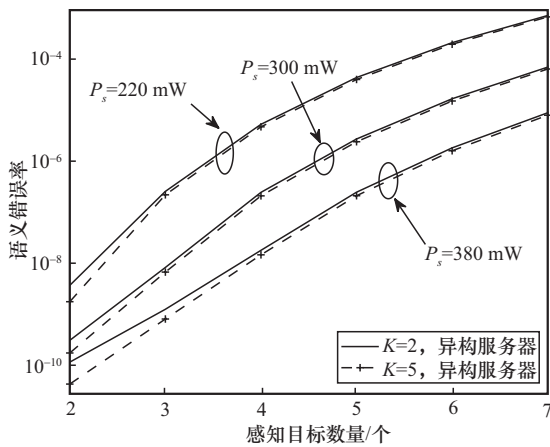


图 7 异构场景下感知目标数量对语义可靠性的影响

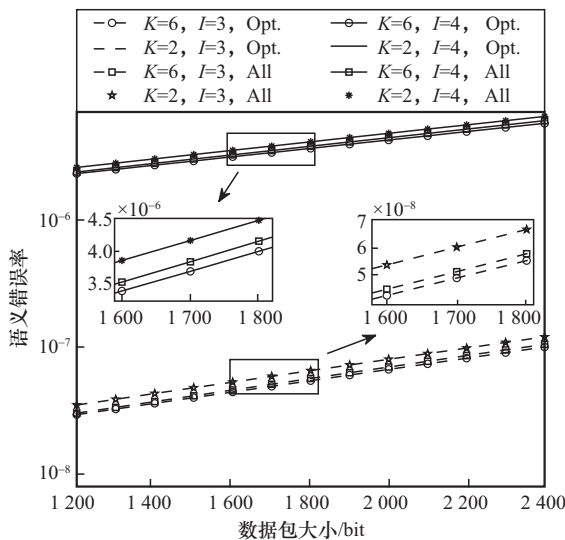


图 8 数据包大小对语义可靠性的影响

### 4 结束语

本文针对多功能 ISCC 网络中多目标感知与多服务器协作的任务需求，在语义层引入具体系统组件刻画 ISCC 框架中的语义可靠性，并在感知、通信与计算 3 个阶段的错误率阈值及最大截止时长约束下，提出了联合优化时长资源分配与服务器选择的语义可靠性最大化问题。通过问题分解，将原问题转化为可高效求解的混合整数凸优化形式，同时结合拉格朗日方法与 KKT 条件揭示了最优解的结构特性。仿真部分讨论了不同系统关键参数对语义可靠性的影响，验证了所提方案的有效性和规模适应性。未来工作中，笔者将进一步探讨该框架下的能耗开销问题，并考虑将强化学习及深度学习等智能资源分配策略引入所提出的语义通信可靠性分析模型中，系统性地评估数据驱动方法与结构化优化方法在综合性能方面的差异。

### 参考文献:

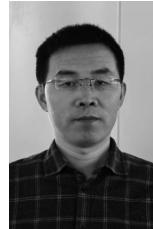
- [1] 刘聪聪, 赵军辉, 任瑞星, 等. 面向智慧交通的通感算一体化: 关键技术、挑战与展望[J]. 移动通信, 2025, 49(12): 155-162.  
Liu C C, Zhao J H, Ren R X, et al. Integrated sensing, communication, and computation for smart transportation: key technologies, challenges and prospects[J]. Mobile Communications, 2025, 49(12): 155-162.
- [2] 李瑞华, 裴郁杉, 黄蓉. 工业场景通感算一体化关键技术体系研究[J]. 移动通信, 2025, 49(12): 150-154.  
Li R H, Pei Y S, Huang R. Research on key technology system of integrated sensing, communication, and computing for industrial scenarios[J]. Mobile Communications, 2025, 49(12): 150-154.
- [3] 张平, 牛凯, 姚圣时, 等. 面向未来的语义通信: 基本原理与实现方法[J]. 通信学报, 2023, 44(5): 1-14.  
Zhang P, Niu K, Yao S S, et al. Semantic communications for future: basic principle and implementation methodology[J]. Journal on Communications, 2023, 44(5): 1-14.
- [4] Huang B Z, Chen N, Fan X W, et al. Joint resource allocation and task scheduling for UAV-assisted ISCC networks[C]//Proceedings of the 2024 13th International Conference on Communications, Circuits and Systems (ICCCAS). Piscataway: IEEE Press, 2024: 456-461.
- [5] Dou C L, Huang X M, Wu Y, et al. Multi-access edge computing empowered integrated hybrid sensing and communication: a computation-efficient design[C]//Proceedings of the 2024 International Conference on Ubiquitous Communication (Ucom). Piscataway: IEEE Press, 2024: 471-476.
- [6] 王晓云, 张小舟, 马良, 等. 6G 通信感知一体化网络的感知算法研究与优化[J]. 通信学报, 2023, 44(2): 219-230.  
Wang X Y, Zhang X Z, Ma L, et al. Research and optimization on the sensing algorithm for 6G integrated sensing and communication network[J]. Journal on Communications, 2023, 44(2): 219-230.
- [7] 牛凯, 张平. 语义通信的数学理论[J]. 通信学报, 2024, 45(6): 7-59.

- Niu K, Zhang P. A mathematical theory of semantic communication[J]. *Journal on Communications*, 2024, 45(6): 7-59.
- [8] 刘伟, 王孟洋, 白宝明. 面向带宽受限场景的高效语义通信方法[J]. *西安电子科技大学学报(自然科学版)*, 2024, 51(3): 9-18.  
Liu W, Wang M Y, Bai B M. Efficient semantic communication method for bandwidth constrained scenarios[J]. *Journal of Xidian University (Natural Science)*, 2024, 51(3): 9-18.
- [9] 石光明, 杨旻曦, 高大化, 等. 面向语义信息直传的通信架构[J]. *通信学报*, 2023, 44(5): 15-27.  
Shi G M, Yang M X, Gao D H, et al. Communication framework for directed transmission of informative semantic[J]. *Journal on Communications*, 2023, 44(5): 15-27.
- [10] 朱慧翔, 柴靖轩, 肖泳, 等. 基于感算通融合的语义认知通信网络[J]. *移动通信*, 2023, 47(3): 31-39.  
Zhu H X, Chai J X, Xiao Y, et al. Semantic-aware communication network based on integrated sensing, computation and communication[J]. *Mobile Communications*, 2023, 47(3): 31-39.
- [11] Khan M S, Jangsher S, Qadir J, et al. Cooperative offloading multi-access edge computing (COMEC) for cell-edge users in heterogeneous dense networks[J]. *Computer Networks*, 2025, 267: 111332.
- [12] Kan T Y, Yao C A, Wei H Y. QoS-aware mobile edge computing system: multi-server multi-user scenario[C]//*Proceedings of the 2018 IEEE Globecom Workshops (GC Wkshps)*. Piscataway: IEEE Press, 2019: 1-6.
- [13] He W, He D Z, Ma X Y, et al. Joint user association, resource allocation, and beamforming in RIS-assisted multi-server MEC systems[J]. *IEEE Transactions on Wireless Communications*, 2024, 23(4): 2917-2932.
- [14] An J C, Li H B, Ng D W K, et al. Fundamental detection probability vs. achievable rate tradeoff in integrated sensing and communication systems[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(12): 9835-9853.
- [15] Polyanskiy Y, Poor H V, Verdú S. Channel coding rate in the finite blocklength regime[J]. *IEEE Transactions on Information Theory*, 2010, 56(5): 2307-2359.
- [16] Liu C F, Bennis M, Debbah M, et al. Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing[J]. *IEEE Transactions on Communications*, 2019, 67(6): 4132-4150.
- [17] Zhu Y, Hu Y L, Yang T, et al. Reliability-optimal offloading in multi-server edge computing networks with transmissions carried by finite blocklength codes[C]//*Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops)*. Piscataway: IEEE Press, 2019: 1-6.
- [18] Zhu Y, Yuan X P, Hu Y L, et al. Semantic reliability maximization: a cooperative perspective in integrated sensing, communication and computation networks[C]//*Proceedings of the GLOBECOM 2023 - 2023 IEEE Global Communications Conference*. Piscataway: IEEE Press, 2024: 5073-5079.
- [19] Corre Y, Stephan J, Lostanlen Y. Indoor-to-outdoor path-loss models for femtocell predictions[C]//*Proceedings of the 2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*. Piscataway: IEEE Press, 2012: 824-828.

## [作者简介]



陈津 (1986-), 男, 湖北枝江人, 海军工程大学副教授、博士生, 主要研究方向为语义通信、智能网络、多媒体通信等。



窦高奇 (1981-), 男, 山西沁县人, 博士, 海军工程大学教授、博士生导师, 主要研究方向为智能压缩编码、语义通信等。



蒋妍 (2003-), 女, 重庆人, 武汉大学硕士生, 主要研究方向为语义通信、工业物联网等。



朱尧 (1986-), 男, 内蒙古乌兰察布人, 博士, 武汉大学研究员、博士生导师, 主要研究方向为语义通信、工业物联网、高可靠低时延通信、移动边缘计算等。