支持算网一体调度的可编程复合流水线架构

侯赛凤[1]，崔鹏帅[1,2,3]，胡宇翔[1,2,3]，董永吉[1,2,3]，刘义[4]
（1.信息工程大学信息技术研究所，河南 郑州 450002；2.先进通信网全国重点实验室，河南 郑州 450002；
3.网络空间教育部重点实验室，河南 郑州 450002；4.中国移动通信集团河南有限公司，河南 郑州 450000）

摘　要：为了应对算网一体（ICN）环境下多元化业务对差异化服务质量（QoS）保障和异构资源高效调度需求，针对传统网络架构面临的性能瓶颈，提出了一种支持算网一体资源调度和多业务QoS保障的可编程复合流水线架构设计，通过网络功能流水线部署系统支持可编程复合流水线架构中不同网络功能的QoS隔离保证和异构资源的分配。基于可编程队列的分层调度机制动态调整调度算法，适配个性化的业务QoS保障需求。基于资源定价和合作博弈求解异构资源分配策略，实现算网资源在数据转发过程中的灵活调度。实验表明，所提方法可有效隔离不同业务的性能干扰，满足多样化业务的定制化QoS需求，同时降低异构资源竞争导致的系统开销。

关键词：算网一体；可编程流水线；QoS保障；异构资源调度
中图分类号：TN934
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025126

Programmable composite pipeline structure supporting scheduling of integrated network and computing

HOU Saifeng[1], CUI Pengshuai[1,2,3], HU Yuxiang[1,2,3], DONG Yongji[1,2,3], LIU Yi[4]
1. Institute of Information Technology, Information Engineering University, Zhengzhou 450002, China
2. National Key Laboratory of Advanced Communication Networks, Zhengzhou 450002, China
3. Key Laboratory of Cyberspace Security, Ministry of Education of China, Zhengzhou 450002, China
4. China Mobile Communications Group Henan Co., Ltd., Zhengzhou 450000, China

Abstract: To meet the needs of differentiated QoS assurance and efficient heterogeneous resources scheduling of diversified services in the ICN environment, for the performance bottleneck faced by traditional network architecture, a programmable composite pipeline architecture that supported ICN resource scheduling and QoS guarantee was designed. The network function pipeline deployment system supported the QoS isolation guarantee of different network functions and the allocation of heterogeneous resources in the programmable composite pipeline architecture. The scheduling algorithm was dynamically adjusted based on the hierarchical scheduling mechanism of programmable queuing to adapt to the characteristic service QoS guarantee requirements. The heterogeneous resource allocation strategy was solved based on resource pricing and cooperative game to realize the flexible scheduling of computing and network resources in the process of data forwarding. Experiment results demonstrate that the proposed method can effectively isolate the performance interference, meet the customized QoS requirements of diversified services, and reduce the system overhead caused by heterogeneous resource competition.

Keywords: ICN, programmable pipeline, QoS assurance, heterogeneous resources scheduling

收稿日期：2025–04–28；**修回日期：**2025–07–02
通信作者：胡宇翔，chxachxa@126.com
基金项目：国家重点研发计划基金资助项目（No.2023YFB2903902）；中原科技创新领军人才基金资助项目（No.244200510038）
Foundation Items: The National Key Research and Development Program of China (No.2023YFB2903902), Science and Technology Innovation Leading Talents Subsidy Project of Central Plains (No.244200510038)

0 引言

数字经济已进入高速发展新阶段，作为关键经济动能对国民经济增长的贡献率持续攀升。据对无人驾驶、区块链、物联网及扩展现实四大前沿领域的测算，其算力需求至2030年将较2018年呈指数级跃升，综合增幅突破300倍，算力正跃升为数字经济时代的核心基础设施[1]。但是，当前算力设施建设面临着数据中心能耗高，算力资源利用效率不足，区域发展不协调、不均衡等挑战。随着摩尔定律趋于极限，单点计算设施（如大型计算中心）的建设，并不能完全满足计算能力需求[2]。将计算与网络融合，通过计算和网络设施来感知、连接与协同多元算力，使服务像用水和电一样即时可用，最终实现网络与算力融合为一体[3-4]。

算网一体（ICN, integrated computing and network）凭借无处不在的计算能力和网络智能突破单点计算能力的性能限制。ICN技术演进呈现两大特征[5-6]：其一，需处理海量差异化服务请求，这些请求不仅对计算资源提出峰值负载要求，更需网络侧提供低时延、高可靠性的传输保障；其二，系统需整合异构网络资源（包括但不限于节点的存储和转发资源等）和多样化算力节点，形成覆盖广泛的资源池。因此，需要通过动态调配云端、边缘及终端资源，实现计算能力、存储容量和网络带宽的按需分配。以全网资源协同为核心目标，着力解决多层级算力节点间的互联互通与智能调度难题，应对用户个性化需求带来的大规模资源动态调配挑战[7]。

在应对ICN的海量数据传输需求和千行百业数字化转型的双重挑战下，传统网络架构正面临前所未有的性能瓶颈[8]。基于分布式架构的运营管理模式难以实现异构资源的协同调度，僵化的协议栈和封闭式设备接口严重制约了网络资源的弹性扩展能力[9]。软件定义网络（SDN, software defined networking）通过构建“控制-数据”双平面架构，成功实现了网络智能的集中式管控与自动化编排[10-11]。新型SDN控制器采用微服务化设计架构，可动态感知应用层需求变化并自适应调整资源分配策略。SDN控制平面的智能化转型离不开数据平面的可编程特性支撑，支持可编程协议无关报文转发语言[12]的网络设备可以直接适配SDN架构，不仅能够实现数据包处理的深度定制[13]，更能通过标准化的RESTful API与控制器集群实时交互[14]，提升了网络功能的可扩展性。

面向ICN技术演进的两大特征，在可编程网络设备的数据转发过程中，针对算网服务的多元化业务请求，其服务质量（QoS, quality of service）需求呈现差异化特征。例如，全息媒体类业务依赖超大带宽传输能力实现信息同步和内容交互[15-16]。工业控制类业务则依赖微秒级端到端的时延保障[17-18]。因此，可编程网络设备不仅需要具备进一步细化区分业务流量的能力，而且还需要对多种业务、多种流量等传输对象进行统一的队列管理调度，从而为多业务提供精细化的QoS服务。针对多元化业务构建的网络功能流水线结构，其算网资源需求也呈现差异化特征。具体到单个可编程网络设备中，相同的算网资源可能在多个网络功能之间存在竞争关系[19]。对于任意软件功能实例，构建不同流水线的所需计算、存储、转发功能容量均不能超越片上资源总和。因此，给定资源总量，可编程网络设备需针对业务对异构算网资源的需求进行动态分配。

针对多元化的业务承载问题，研究人员开展了许多相关研究。HyperVDP[20]通过快速解析、控制流排序和动态阶段映射技术，使多个网络功能可以动态映射到不同的匹配动作阶段，实现了数据平面网络功能的完全虚拟化，但是并未实现性能隔离需求。文献[21]提出了一种基于可编程网络的算力调度方案，在网络路由和路径选择方面实现算力调度，但并未实现ICN调度。VirtP6[22]提出了一种支持并行流水线的虚拟化可编程数据平面结构，实现了数据包的高性能转发和网络功能的隔离，但缺乏异构资源调度能力。

在此背景下，为了解决多元化业务在数据转发过程中对算网资源的差异化需求，本文基于SDN的可编程开放架构，在数据平面引入支持ICN的复合流水线结构设计，支持多样化业务流量的精细化调度，实现性能隔离。通过数据包存储、转发和计算等逻辑处理的形式描述，实现计算存储转发资源在数据转发过程中灵活组合、调用，适配不同应用场景下个性化的处理需求。具体来说，本文研究工作和主要创新点如下。

1) 结合SDN的开放架构，提出一种支持ICN的可编程复合流水线架构，通过网络功能流水线部署系统实现可编程复合流水线结构中不同网络功能

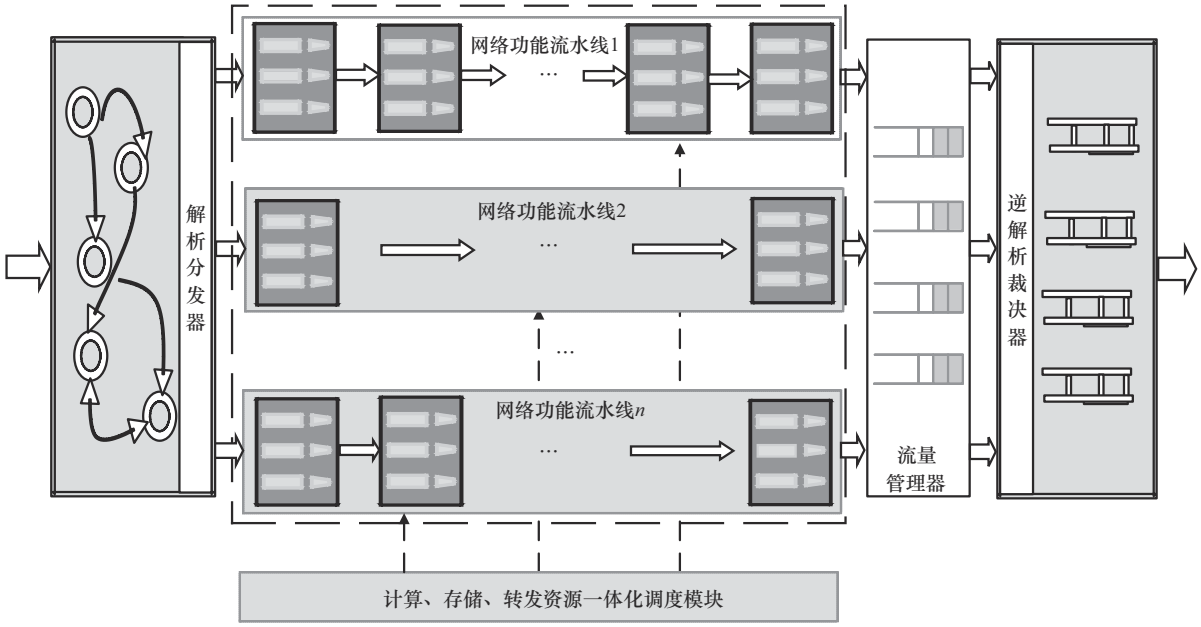


图1　支持ICN的可编程复合流水线架构

的QoS隔离保证和异构资源的分配，实现多要素的融合、感知、控制和管理。

2)针对数据转发过程中多业务QoS隔离保证问题，提出一种基于可编程队列的分层调度机制，通过动态调整调度策略适配不同应用场景下个性化的处理需求，实现差异化QoS隔离保证。

3)针对数据转发过程中网元设备内异构资源的竞争问题，设计了计算、存储、转发资源一体化调度模块，基于资源定价和合作博弈求解最优资源的分配策略，实现算网资源的高效利用。

4)在可编程网络设备中实现了支持ICN的复合流水线架构，并对业务QoS性能和算网资源利用率进行评估。仿真结果表明：本文可编程复合流水线架构能够保证不同网络功能之间的隔离性。同时，本文算法使流完成时间（FCT, flow complete time）减少了73%，有效降低了异构资源的资源开销，提高了资源利用率。

1　支持ICN的复合流水线架构

在算力网络中，可编程网络设备可通过动态编程、重构等方式，实现特定需求的网络功能。支持ICN的可编程复合流水线架构如图1所示，主要包括解析分发器，并行流水线结构，流量管理器，计算存储转发资源一体化调度，逆解析裁决器等模块。其中，并行的网络功能流水线在独立的运行环境中处理不同网络功能逻辑，维护特定的寄存器、流表等，且彼此之间不能互相访问；可编程流量管理器本身具有不可编程性，通过可编程队列调度可以解耦调度算法与硬件的强绑定关系；计算、存储、转发资源一体化调度模块将计算节点、存储设备、网络链路等异构资源抽象为统一的逻辑单元，为不同的网络功能流水线分配异构处理资源。

为了满足ICN演进的海量差异化服务承载和异构算网资源协同的需求，通过不同网络功能流水线承载差异化服务，不同业务流可按需分配独立处理通道。网络功能流水线组件采用融合异构处理资源的多级流水线设计，在支持类似P4可定义的转发流水线的基础上，将计算和存储功能挂载到并行的流水线上，再通过流水线的调用支持多种处理能力的融合，并采用协议无关的配置接口转发流表和动作信息，为不同的网络功能分配不同的流水线资源。复合流水线架构就是在并行流水线的基础上，结合可编程队列调度和异构资源调度支持ICN的海量差异化服务承载和异构算网资源协同。

然而，在可编程硬件流水线上同时部署不同网络功能是一项复杂且有挑战性的任务，需要考虑多个限制和目标。首先，为了实现不同网络功能的隔离，网络功能的部署需要保证QoS（包括时延和带宽等），提供资源和性能的细粒度控制。其次，可编程硬件平台中不同网络功能在不同异构资源之间

竞争协作，反映了网络运维成本和业务性能。因此，为了支持不同网络功能的部署，同时满足QoS保证和资源需求，需要实现以下2个目标。

1)性能的分层隔离。面对ICN的多样化和差异化业务需求，系统需要提供不同QoS保证的细粒度控制，并且支持不同网络功能之间的流量控制。本文提出一个多级分层调度器，以队列抽象结构实现不同网络功能的调度，控制不同网络功能流量的同时调度不同业务的QoS需求。

2)异构资源的统一分配。网络设备需要协同调度计算、存储和转发资源。通过对资源状态的细粒度感知，本文提出一种基于资源定价的异构资源分配策略，基于竞争协作关系利用短期博弈分配异构资源，由此构建支持不同网络功能资源需求的数据处理逻辑。

为实现上述2个目标，本文提出了灵活有效的网络功能流水线部署系统，如图2所示，支持可编程复合流水线中不同网络功能的QoS隔离保证和异构资源的分配。系统分为3个抽象层：业务层、网络设备层和资源层。在业务层，所有的业务被聚类为几个业务组，并根据聚类结果将业务分配到不同的网络功能。在网络设备层，不同网络功能实例基于可编程硬件流水线被抽象为队列调度结构。在资源层，所有可用的资源协调构成资源池，基于不同业务的QoS需求将资源分配给不同网络功能。

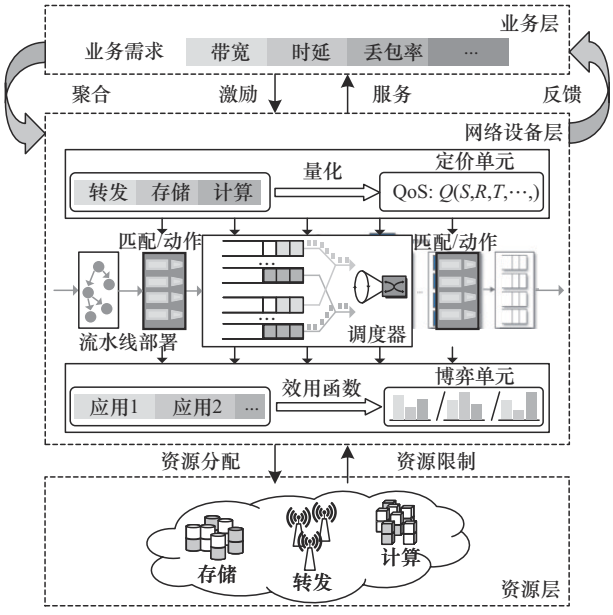


图2　网络功能流水线部署系统

首先，基于性能需求业务被聚类到合适的网络功能实例，网络功能的类型由业务聚类的输出结果决定。其次，基于数据平面流水线结构构建了网络功能的部署模型，可编程流量管理器提供针对不同网络功能实例的可定制的调度结构，实现区分网络功能实例的业务QoS隔离；基于可编程队列结构实现了各种网络功能实例的性能保证。最后，提供了一个定价单元和一个博弈单元以提高不同网络功能之间异构资源分配的有效性，结合资源限制，为每个网络功能流水线分配异构资源。

2　面向算网差异化需求的可编程调度器

面对算网应用的多样化和差异化业务需求，传统QoS技术难以区分多种网络功能承载的不同业务，或者保证不同应用的流量隔离。为了满足不同应用场景的优化目标，针对调度算法展开了许多研究。但是，这些算法大部分是硬件绑定的，一旦硬件设备设计完成，难以更改现有逻辑或添加一个全新的算法。另外，许多调度算法仅适用于固定场景，难以满足不同应用场景多样化的可定制QoS需求。HCSFQ[23]提出了能够以线速在商用可编程交换机上运行的分层公平队列调度算法。DeepQoS[24]提供了一个支持分层QoS的无核心状态包标记，可以在单个点上同时有效地标记不同HQoS层的资源共享策略。H-DRFQ[25]研究了支持业务数据流的细粒度QoS，提出了分层多资源共享公平队列算法，以保证期望的资源公平份额。

为了实现不同网络功能的QoS隔离，本文提出了支持自定义调度算法和分层QoS的可编程调度器结构。调度器能够以队列抽象结构的方式实现用户之间和用户内部业务的流调度，使用全新的调度算法支持多用户业务隔离，并且为用户提供可定制的确定性QoS服务。具体地，利用多级分层队列设计了一种支持多种级别流量聚合的分层可编程队列调度通用架构，设计层次化的准入控制策略，在可编程网络设备实现自定义调度算法。各层可采用不同的队列调度算法，实现不同用户间的加权最大-最小公平调度，以及内部业务的自定义可编程调度，达到多用户的流量隔离。

2.1　分层可编程调度器结构

分层可编程调度器的数据包处理过程如图3所示。可编程调度结构主要由两级可定制的调度器组

成，包括一个虚拟队列（VQ, virtual queuing）系统和一个令牌队列（TQ, token queuing）系统。VQ和TQ是由一组寄存器组成的虚拟队列。数据包携带的到达速率等信息被分类进入VQ，每个数据包的等级值被放入TQ。当任意数据包到达可编程设备后，首先数据包分类器识别数据包所属的业务类别，然后计算到达数据包的等级值，并将该值添加到令牌队列。一级调度器基于权重最大-最小公平算法调度不同用户，二级调度器基于等级值的准入控制算法执行不同业务之间的可定制调度。各组件的功能描述如下。

1) VQ系统

VQ存储数据包信息，包括到达速率估计和流权重，记为$Q(i,j)$，表示第j个用户的第i个数据流的数据包信息队列，基于数据包携带的信息计算每个用户的分配带宽。VQ之间采用权重最大-最小公平调度算法进行调度，根据相应的公平权重参数和到达速率估计实现多个用户之间的公平隔离。设置一个权重寄存器用于维持每个用户的权重值W_j。

2) TQ系统

根据业务分类设置了N个TQ，记为$R(i,j)$，表示第j个用户的第i个业务流的数据包令牌存储队列，基于存储在令牌中的等级值对TQ进行调度。数据包等级值代表数据包的优先级，表示数据包的调度顺序，其计算方法可以根据所选择的调度算法灵活定义。

3) 两级调度器

$S_{j,1}$表示第一级调度器，S_2表示第二级调度器。基于TQ中记录的等级值，$S_{j,1}$执行等级值的准入控制算法，决定如何调度某个用户业务类的下一个数据包。$S_{j,1}$控制数据包作为S_2所选用户的候选数据包，S_2基于存储在VQ中的到达速率估计和分配带宽调度该数据包。因此，当二级调度器决定调度下一个数据包时，会触发一级调度器。

由于受到P4语言和硬件资源的限制，实现队列结构和算法到可编程硬件交换机的映射面临着固有挑战。一方面，大多数调度算法是硬件绑定的，硬件设备一旦设计好，就难以修改现有逻辑或增加全新算法。考虑到服务需求的多样性和动态性，二层可编程调度器通过数据包的等级值计算实现不同的调度算法计算。例如，加权公平调度（WFQ, weighted fair queuing）算法是一种基于流的队列调度策略，其目的是实现网络中的流量公平和带宽共享。在WFQ算法中数据包的序列号可表示为等级值。随后，动态调整数据包的等级值并记录在窗口中。

另一方面，由于可编程设备难以实现多级队列结构，定义了滑动时间窗口实时记录到达数据包的等级值，仅使用单级队列实现分层结构。滑动时间窗口由n个寄存器组成，持续捕捉和更新最近n个到达数据包的等级值。数据包到达交换机后，系统对寄存器集进行复杂遍历以分析历史等级值的瞬时分布，该分布是智能队列准入控制决策的关键输入。

2.2　准入控制算法设计

当二级调度器准备调度下一个数据包时，它会触发基于业务流优先级和分配带宽的一级调度器。一级调度器基于等级值使数据包入队列，等级值基于可定制的调度算法计算，例如，WFQ根据加权值weight以及数据包的长度packet_length计算其等级值[26]，

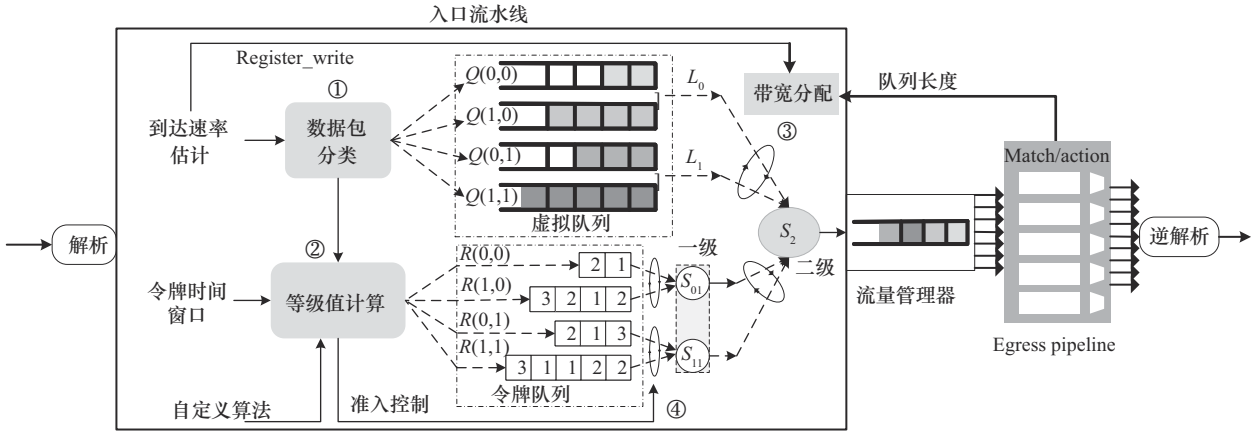


图3　分层可编程调度器的数据包处理过程

rank=Previous_rank+weight×new_packet_length，通过设置准入阈值控制数据包的转发或丢弃。为了记录最近时间周期T内到达数据包的等级值，维持令牌时间窗口实时捕捉到达数据包的等级值分布，根据数据包等级值生成累积分布函数（CDF, cumulative distribution function）。基于业务流的实际所需带宽和分配带宽的差值决定其优先准入百分比，再结合数据包等级CDF图决定业务流所属队列的准入控制阈值。

假设系统监测每个业务流的实际所需带宽为$[r_1,r_2,\cdots,r_i]$，分配带宽为$[F_1,F_2,\cdots,F_i]$，通过估计特定周期T内流达到速率实现加权最大-最小公平分配带宽，表示为

$$F_i = \min(f_i, r_i) \tag{1}$$

$$F_i = f_i + \sum(f_i - r_i)\frac{\omega_i}{\sum\omega_i} \tag{2}$$

其中，ω_i为业务流的权重；f_i为分配给每个业务流的加权份额，由业务流优先级和权重确定。

则任意用户业务流的分配带宽因子为

$$b_j = f(r_j - F_j) = \begin{cases} 1, & r_j \leqslant F_j \\ \dfrac{k(r_j - F_j)}{r_j}, & F_j < r_j \leqslant B \\ 0, & r_j > B \end{cases} \tag{3}$$

其中，$0 \leqslant b_j \leqslant 1$。

为了更好地描述业务流的带宽分配因子、数据包等级值和准入阈值，引入优先准入百分比A_j的定义，表示为

$$A_j = \begin{cases} \sum_{j=1} b_j, & \sum_{j=1} b_j \leqslant 1 \\ 1, & \sum_{j=1} b_j > 1 \end{cases} \tag{4}$$

基于每个用户业务流的优先准入百分比和等级值的累积分布函数，可以得出每个业务流的准入阈值为

$$C_j = \varphi_{\mathrm{CDF}}^{-1}(A_j) \tag{5}$$

准入控制过程基于带宽分配动态调整队列阈值，当带宽分配减少时，准入阈值相对“宽松”，但部分数据包可以入队；当带宽分配增加时，准入阈值相对“激烈”，只有小部分优先级较高的数据包可以入队。准入阈值的动态调整过程如下。

1)当某用户的业务流所需实际带宽满足$r_j \leqslant F_j$时，则$A_j = 1$，表示所有数据包均可进入租户所属队列，此时不需要数据包调度。

2)当某用户的业务流所需实际带宽满足$F < r_j \leqslant B$时，租户所属队列启动准入控制，随着所需带宽的增加，队列准入阈值越小，只有等级满足$v < C_j$的数据包才能进入该队列。

3)当某用户的业务流所需实际带宽满足$r_j > B$时，则$A_j = 0$，表示任何数据包均无法进入用户所属队列。

3 面向ICN的异构资源调度模块

计算、存储、转发资源一体化调度模块将计算节点、网络链路、存储设备等异构资源抽象为统一的逻辑单元，为不同的网络功能流水线分配异构处理资源。网络中的计算、存储和转发资源存在能力补偿关系。例如，缓存在不同网络节点或位置的部署策略，以及缓存的存取速率和容量，均会对相关节点的转发资源开销产生影响。因此，异构资源分配不仅因业务变化而调整，也因异构资源被替代而调整。DRF-CSN[27]研究了一种通用的多资源最大-最小公平分配算法，支持多种类型资源的调度，但该分配策略仅由每种网络功能的主导资源决定。2L-MRA[28]提出了一种基于效用的异构资源联合分配机制，该机制针对特定切片的固定资源分配，但没有考虑不同应用的差异化需求。

当底层资源被网络功能重利用时存在竞争协作关系，并且传统网络的资源属性与网络功能的能力强绑定，因此，建立一个动态的资源承载模型是具有挑战性的。为了实现网络功能到异构资源的动态适配，本文设计了计算、存储、转发资源一体化调度模块对异构资源进行分配调度，基于资源定价和合作博弈的算网资源调度框架如图4所示。通过资源定价机制可以在满足业务需求的前提下对资源进行预分配，从而避免网络服务突变造成的资源负载过重，进而提高网络运维效率，优化资源配比。同时，合作博弈理论模型可以激发各网络功能之间的竞争与协作，既可以满足不同业务对低时延、高可靠性、用户连续性等服务质量的需求，又可以达到网络性能最优化和资源利用率最大化的目的。

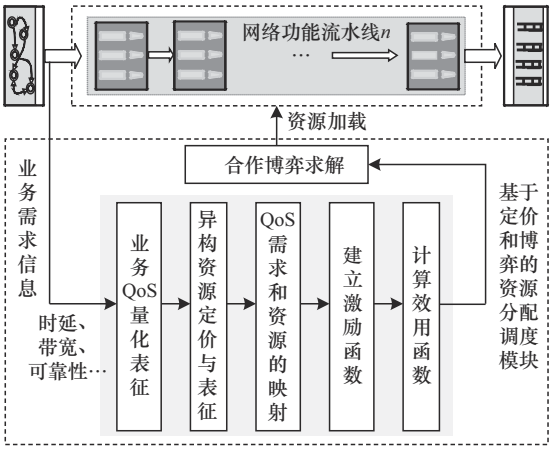


图4　基于资源定价和合作博弈的算网资源调度框架

3.1　资源定价模型

业务按照性能需求聚类到不同的网络功能流水线后，各网络功能基于资源定价的竞争与协作机制进行短期博弈分配算网资源。首先，制定异构资源的定价策略，对异构资源进行量化，确定不同类型和规模的网络资源的价格。其次，获取业务对服务质量的需求，进行需求预测和分析。然后，根据业务的需求和资源定价，进行资源分配决策。最后，对异构资源的使用情况进行监控，根据实际情况对定价策略和资源分配进行调整。面向网络功能优化运行的资源定价策略主要包括以下步骤。

1)业务 QoS 量化表征。首先，定义一系列 QoS 指标，如时延、带宽、丢包率等，这些指标可以根据具体业务需求和网络应用来确定。其次，对需求进行量化，将业务的需求转化为具体的数值要求。例如，对于视频流业务，可以将时延要求定义为特定的毫秒数，带宽要求定义为特定的数据传输速率等。然后，将量化的需求映射到对应的 QoS 参数上。本文简单地使用加权和的方法将网络业务需求量化为 QoS 参数。在这种方法中，可以为每个 QoS 指标（如时延、带宽、可靠性）分配一个权重，假设 3 个 QoS 指标量化值分别为：D、B、R，对应的权重分别为 ω_1、ω_2、ω_3。则可以使用式(6)将业务的 QoS 需求量化为一个综合的 QoS 分数，即

$$Q = \omega_1 D + \omega_2 B + \omega_3 R \tag{6}$$

2)建立 QoS 需求和资源的映射关系。不同网络业务的性能差异表现在网络业务的服务质量与异构资源使用的映射情况不同，对各种异构资源的敏感度不同，例如部分应用可以利用较少的带宽和存储资源来满足业务的低时延需求，并且提供较高的服务质量，因此在定价模型中，需要对此进行量化与表征，将其转化为具体的数值要求，根据异构资源的参数设置建立映射函数。假设网元中共有 l 种网络功能，对应于业务的需求，各网络功能可以提供时延、带宽和可靠性这 3 种服务，每种服务的质量量化表征为 D_l、B_l、R_l，设计每种服务与异构资源 m 的映射关系为

$$\psi_l = \sum_{l,m} \alpha_{lm} \lg u_m \tag{7}$$

其中，$\psi \in \{D,B,R\}$，α 表示不同资源对每种服务的作用权值，u 表示某种资源的使用量。

3)异构资源定价。定价的依据之一是 QoS 需求与资源的映射关系。这意味着需要对不同类型的资源（如带宽、计算能力、存储空间等）进行量化，并根据其对业务 QoS 的贡献以及业务需求等因素来制定合理的定价策略，对每单位带宽资源、计算资源以及存储资源分别定价。每个网络功能流水线承载的业务量不同，所需资源也各不相同。为了保证网络功能流水线之间的公平性，可以根据当前资源剩余量和实时业务需求进行价格调整，定义根据需求导向的价格调整因子为

$$\varepsilon_m = \begin{cases} \dfrac{-2T_m}{\chi_m - 1}, & T_m \geqslant \dfrac{\chi_m}{2} \\ \dfrac{2T_m}{\chi_m + 1}, & T_m < \dfrac{\chi_m}{2} \end{cases} \tag{8}$$

其中，$T_m = \chi_m - u_m$ 为某种资源的剩余量，χ_m 为某种资源总量。

当某种资源的剩余量较少且对该资源的需求较大时，可以通过提高资源价格来获取奖励；当某种资源剩余量较多且对该资源的需求较小时，可以通过降低资源价格刺激资源消费，根据业务需求导向的资源定价可表示为

$$P_m = \varepsilon_m p' \tag{9}$$

其中，p'为价格调整前的资源定价。

4)建立激励函数。这是网络定价模型中的关键一步，在这个过程中，由于网络功能为业务提供服务，业务需要为各网络功能给予一定的激励。激励的设置依据是业务的 QoS 以及网络功能提供的服务质量。为了实现这一目标，需要为各个 QoS 指标分别建立与网络功能所提供服务的数值映射，即

$$M_l = \beta_l(\omega_1 D_l + \omega_2 B_l + \omega_3 R_l) \tag{10}$$

其中，M_l表示网络功能l通过提供单位QoS所得到的激励，β_l为不同网络功能对应的QoS权值。

5)计算效用函数。在激励函数以及异构资源定价的基础上，计算每个网络功能流水线的效用函数。该效用函数的具体构成分为两部分：一部分是业务为网络功能各类服务所提供的激励之和；另一部分是网络功能的成本，即使用异构资源所付出的总成本。两者之差构成了网络功能的效用函数，效用函数为

$$U_l = \beta_l(\omega_1 D_l + \omega_2 B_l + \omega_3 R_l) - \sum_m P_m \quad (11)$$

算法1描述了资源定价策略的设置过程。

算法1 资源定价策略伪代码描述

输入 QoS_weights #QoS权重，network_functions #网络功能，resource_types #资源类型，resource_capacity #资源容量，base_prices #基础定价，NF_beta #网络功能权值

输出 NF_utilities #效用函数

```
def main( )：
service_QoS = quantify_service_req(QoS_weights); #业务QoS量化分数
service_quality = {}; #服务质量量化表征
for NF in network_functions: #不同网络功能
    for service in ['D', 'B', 'R']: #QoS指标
        service_quality[(NF, service)] = calculate_service_quality(NF, service); #建立QoS需求和资源的映射关系
resource_prices = {}; #资源定价
for R in resource_types: #不同资源类型
    adjustment_factor = calculate_adjustment_factor(R); #资源价格调整因子
    resource_prices[R] = adjustment_factor * base_prices[R]; #资源定价表示
NF_incentives = {}; #激励函数
for NF in network_functions: #不同网络功能
    NF_incentives[NF] = calculate_incentive(NF, service_QoS); #计算网络功能激励
NF_utilities = {}; #效用函数
for NF in network_functions: #不同网络功能
    NF_utilities[NF] = calculate_utility(NF, NF_incentives, resource_prices); #计算网络功能效用
return NF_utilities。 #输出效用函数
```

尽管动态定价机制为资源分配提供了理论优势，但它们的实现带来了不小的计算复杂度挑战。在资源受限场景或高密集度请求环境下，要求对价格进行实时调整，以准确地反映瞬时资源的可用性。价格均衡计算过程中决策空间维度呈多项式增长$O(n^k)$，为了解决这一基本的权衡，提出了一个计算复杂性感知的定价框架。该框架通过2种协同机制动态调节适应粒度：首先，采用卡尔曼滤波的自适应定价调制方案，将瞬态波动与持续资源趋势解耦；其次，基于熵的状态采样，通过滑动窗口自适应将计算复杂度限制在$O(\log n)$。

3.2 合作博弈求解

由于资源的定价是一个动态优化过程，在完成对异构资源的统一定价后，再结合以业务需求为基准的约束条件，使网元中的各个网络功能互相协调，根据业务需求量和资源定价之间的实时关系进行合作博弈，对底层资源进行合理的配置以最低网络资源耗费满足用户的业务需求。

将网络功能间的合作博弈表示为：$G=[L,U]$，其中$L=\{1,2,\cdots,l\}$代表l种网络功能的集合，定义S为L的子集，即$S\subseteq L$，代表网络功能组成的联盟，U_i表示不同网络功能对应的效用函数，使用用户的体验质量值作为合作博弈效益值。在进行收益分配时，边际贡献往往是考虑的重要因素之一，Shapley值[29]即是运用边际贡献进行收益分配的解。边际贡献的计算需要考虑参与联盟的顺序，对于网络功能1和2组成的联盟$\{1,2\}$，先参与者网络功能1的边际贡献为$U(\{1\})-U(\varnothing)$，后参与者网络功能2的边际贡献为$U(\{1,2\})-U(2)$，为了解决参与顺序带来的前后边际贡献差距，Shapley值将所有的排列方式全部考虑，相加之后求平均，即全部边际贡献向量的算术平均值为

$$\mu_l = \sum_{S\in N}\frac{(|S|-1)!(|L|-|S|)!}{|L|!}\cdot [U(S)-U(\{S-l\})]$$

$$v_l = \frac{\mu_l}{\sum_l \mu_l} \quad (12)$$

其中，μ_l为网络功能l的Shapley值，$|S|$为联盟S中成员的数量，$|L|$为全部网络功能的数量，$U(S)$为联盟S的收益，$U(\{S-l\})$为联盟$\{S-l\}$的收益，$\{S-l\}$指S除去网络功能l形成的联盟，v_l为网络功能l的分配系数。由此可得出异构资源的分配结果。

通过先定价再博弈的资源分配架构，对抽象化的业务流对应的多种网络功能所需要的异构资源进行优化配比，异构资源分配结果则通过调用协议无关的接口配置转发流表和动作信息，为不同的网络功能分配不同的流水线资源。异构资源调度模块不仅实现了资源的高效利用，同时进一步保障了可编程流水线架构的隔离性能。

4　仿真分析

4.1　仿真参数设置

首先，为了评估本文可编程流水线结构的可行性，搭建了一个硬件实验拓扑：2台Intel Tofino交换机分别连接2台服务器，每个服务器配置8核CPU，64 GB RAM和一个40 G NIC（Intel XL710）。虚拟机为Ubuntu 16.04.6 LTS，运行4.10.0-28-generic Linux kernel，搭载i7-3 770 3.40 GHz处理器，8 GB内存。2台服务器分别发送2种业务流，在2个交换机上部署2条流水线以支持2种业务流的转发功能，链路带宽为20 Gbit/s。另外，为了模拟大规模网络环境，本文在软件交换机BMv2的基础上，自定义复合流水线的网络功能处理引擎。在Mininet网络仿真平台上模拟大规模真实网络环境：搭建leaf-spine拓扑，包含9个叶交换机、4个脊交换机和144个服务器。每个交换机安装BMv2软件交换机，连接16个服务器，接入和叶−脊链路带宽分别设置为10 Gbit/s和40 Gbit/s。

4.2　仿真结果

为验证本文算法中可编程流水线承载多元化业务的隔离性，在硬件实验拓扑上进行带宽隔离实验。隔离性是指多个网络功能在同一个可编程网络设备中相互独立运行，性能互不影响。2台服务器分别发送业务流A和业务流B，业务流A包含3 000条流，业务流B包含1 000条流，业务流的发送速率均为40 Gbit/s。业务流内部具有相同的优先级，本文分别评估不同权重下，即权重A:权重B=1:1和权重A:权重B=2:1，业务流A和业务流B的公平性。

可编程流水线架构的隔离性验证如图5所示。当2个业务具有相同的优先级时，即权重A:权重B=1:1，带宽均等地分给每种业务；此时，业务流A的总吞吐量约为18.2 Gbit/s，业务流B的总吞吐量约为17.36 Gbit/s，2种业务流的吞吐量近似相等。同样地，当2种业务具有不同的优先级时，即权重A:权重B=2:1，带宽按照业务优先级分给每种业务；此时，业务流A的总吞吐量约为25.8 Gbit/s，业务流B的总吞吐量约为13.2 Gbit/s，业务流A的吞吐量是业务流B的2倍。

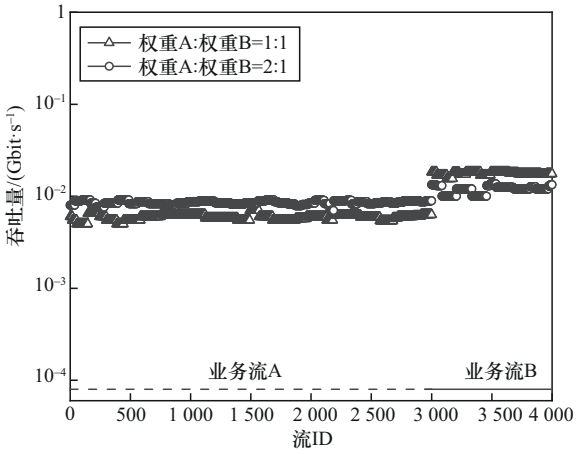


图5　可编程流水线架构的隔离性验证

为了模拟大规模真实环境下的业务性能和资源需求，在Mininet网络仿真平台上进行模拟仿真实验。根据多样化应用场景的性能需求，生成不同性能需求和资源需求的服务请求数据，可编程交换机内部支持3种网络功能数据处理逻辑：音视频类业务、工业互联网业务及全息媒体业务。每种类型业务的数据处理逻辑参数设置如下：计算资源需求为10~70 MIPS，存储资源需求为1~80 GB。为了展现高效调度资源的能力，业务请求强度从200~2 200个随机产生，业务量按比例3:2:1注入节点。每种类型业务请求的存储资源在0.2~10 GB服从均匀分布，请求的计算资源在5~15 MIPS服从均匀分布，请求的带宽资源在0.2~1 Mbit/s服从均匀分布。

将本文算法与HCSFQ、DeepQoS和H-DRFQ进行对比，验证本文算法的性能保障能力。FCT是衡量网络性能的重要参数，测量业务的FCT用来验证多业务QoS性能保证效果。不同业务请求强度下的平均FCT对比如图6所示，随着业务请求强度的增加，不同方案的平均FCT是递增的；与Deep-

QoS和H-DRFQ相比，本文算法能够实现类似性能，平均FCT减少了约40%，具有微弱优势；与HCSFQ相比，本文算法在性能方面具有较大优势，平均FCT减少了约67%~73%。

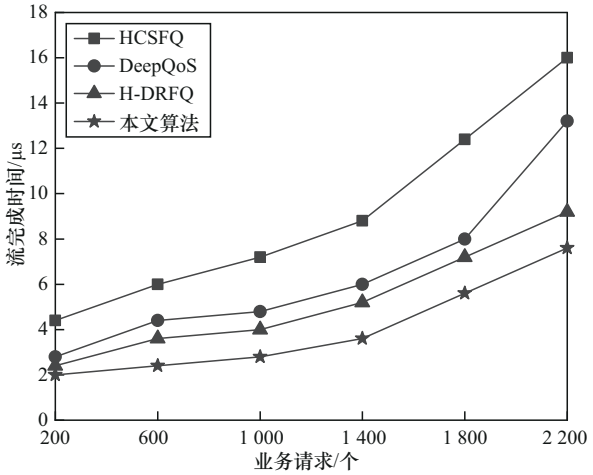


图6　不同业务请求强度下的平均FCT对比

相同业务请求强度下各个时钟周期T下的平均FCT对比如图7所示。在实验过程中，本文将可编程设备的业务请求总数设置为1 000个，但3种业务类型的数据量比例随机变化。从图7可以看出，在相同的业务请求强度下，本文算法的FCT最短，H-DRFQ方案次之，其次是DeepQoS，HCSFQ方案的FCT最长；并且各对比方案的抖动较大。这是由于对比方案无法支持区分差异化业务性能的可定制调度，难以最大限度地满足业务性能。本文算法能够根据不同类型业务的差异化性能需求定制调度策略，实现不同网络功能的性能需求。

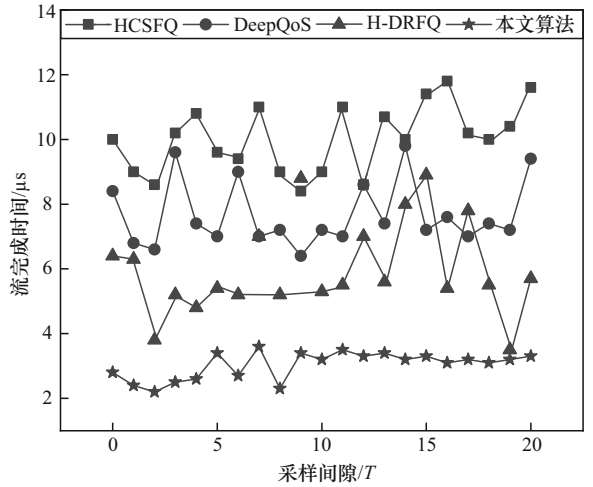


图7　相同业务请求强度下的平均FCT对比

本文针对ICN环境中面向多样化应用场景的资源分配方法进行对比，记录不同业务请求强度下的资源开销对比。业务的QoS指标包括时延、带宽、可靠性，资源定价模型中的参数设置为：音视频类业务的指标权重为1:1:2，工业互联网业务的指标权重为2:1:1，全息媒体业务的指标权重为1:2:1。不同资源对各种业务的作用权值与指标权值设置相同。通过与DRF-CSN和2L-MRA进行对比，验证本文算法的资源利用率，测量可编程交换机的资源开销，并进行归一化。不同业务请求强度下的归一化资源开销对比如图8所示，总体来说，所有方案的资源开销都随着业务请求强度的增加而增加；本文算法的资源开销相较于2L-MRA平均减少了22%，相较于DRF-CSN平均减少了15%。2L-MRA和DRF-CSN难以适应不同业务的服务密集度，仅以固定的方式部署资源。本文算法可以根据业务需求动态调整资源分配，实现计算、存储和转发资源的高效利用。因此，本文算法的资源利用率最高，在相同流量下的资源开销最低。

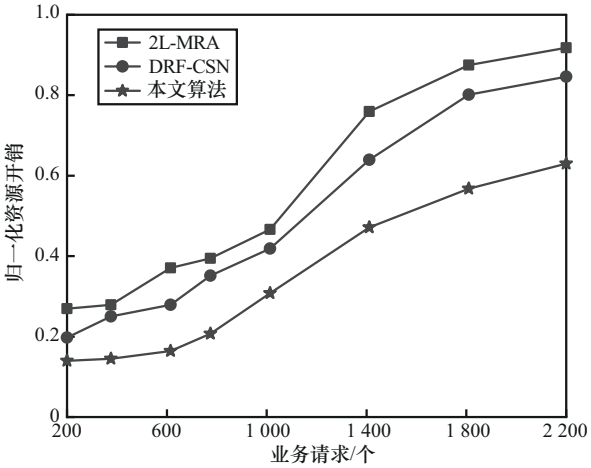


图8　不同业务请求强度下的归一化资源开销对比

相同业务强度下各种算法的归一化资源开销对比如图9所示。同样地，在实验过程中，本文将可编程设备的业务请求总数设置为1 000个，同时3种业务类型的数据量比例随机变化。从图9中可以得出，相同的业务请求强度下，本文算法的资源开销相较于2L-MRA平均减少了13%，相较于DRF-CSN平均减少了8%；并且对比方案的抖动也较大。这是由于2L-MRA采用针对切片的固定资源分配方案，没有考虑不同应用的差异化需求，不能满足不同类型业务的资源需求，资源利用率较低。

DRF-CSN 提出的分配策略仅由每种网络功能的主导资源决定，当某一种类型的业务占比较大时，该主导资源将给其他类型应用带来瓶颈效应。本文算法可以根据业务请求和实时资源剩余量联合调度各种资源，能够最大化资源利用率，以降低资源开销。

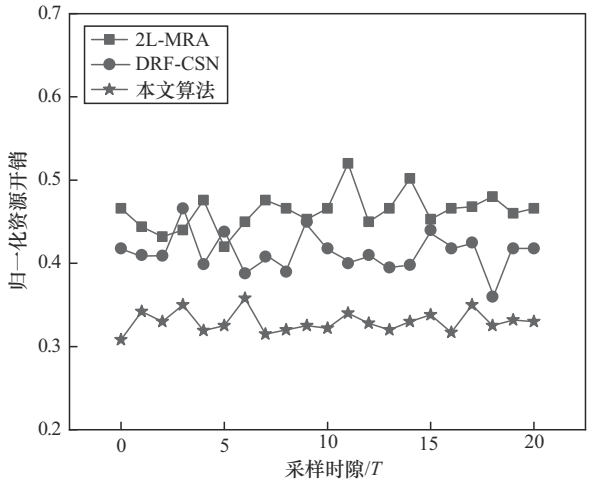


图9　相同业务请求强度下的归一化资源开销对比

5　结束语

本文对基于 ICN 的差异化服务质量保障和异构资源高效调度需求进行研究，考虑业务 QoS 保障和算网异构资源调度，构建可编程数据平面的复合流水线结构，实现 ICN 环境下多业务流量的精细化调度。针对数据转发过程中多业务 QoS 隔离保证问题，通过不同的网络功能流水线承载多元化业务，基于可编程队列的分层调度机制动态调整调度算法，适配个性化的业务 QoS 保障需求；针对数据转发过程中网元设备内异构资源的竞争问题，将计算节点、网络链路、存储设备等异构资源抽象为统一的逻辑单元，基于资源定价和合作博弈求解异构资源分配策略，实现计算存储转发资源在数据转发过程中灵活调度。仿真结果表明，本文复合流水线架构能够支持 ICN 调度，同时，本文算法能够实现较好的性能保障和较高的资源利用率。

参考文献：

[1] 中国信息通信研究院 . 中国算力发展指数白皮书[R]. 2023.

[2] 朱海龙, 杨帆, 蒋如一, 等 . 算网自智: 研究进展与展望[J]. 通信学报, 2024, 45(10): 191-206.
ZHU H L, YANG F, JIANG R Y, et al. Computing-network intelligence: research progress and prospects[J]. Journal on Communications, 2024, 45(10): 191-206.

[3] WANG X Y, DUAN X D, YAO K H, et al. Computing-aware network (CAN): a systematic design of computing and network convergence[J]. Frontiers of Information Technology & Electronic Engineering, 2024, 25 (5): 633-644.

[4] 张宏科, 权伟, 刘康 . 算力网络研究与探索[J]. 中兴通讯技术, 2023, 29 (1): 1-5.
ZHANG H K, QUAN W, LIU K. Research and exploration of computing power network[J]. ZTE Technology Journal, 2023, 29(1): 1-5.

[5] 杨帆, 宋闻萱, 许方敏, 等 . 工业互联网算网一体技术研究[J]. 无线电通信技术, 2023, 49(1): 63-71.
YANG F, SONG W X, XU F M, et al. Research on the application of computing force network technology in industrial Internet of Things[J]. Radio Communications Technology, 2023, 49(1): 63-71.

[6] 张悦 . 云网融合向算网融合演进的关键技术研究[J]. 中国宽带, 2024 (11): 1-3.
ZHANG Y. Research on key technologies of cloud network convergence evolving to computing network convergence[J]. China BroadBand, 2024(11): 1-3.

[7] 张宏科, 于成晓, 权伟, 等 . 融算网络体系基础研究[J]. 电子学报, 2022, 50(12): 2928-2934.
ZHANG H K, YU C X, QUAN W, et al. Fundamental research on computing integration networking[J]. Acta Electronica Sinica, 2022, 50(12): 2928-2934.

[8] 中国移动 . 算网一体: 网络架构及技术体系展望白皮书[R]. 2022.

[9] 胡宇翔, 伊鹏, 孙鹏浩, 等 . 全维可定义的多模态智慧网络体系研究[J]. 通信学报, 2019, 40(8): 1-12.
HU Y X, YI P, SUN P H, et al. Research on the full-dimensional defined polymorphic smart network[J]. Journal on Communications, 2019, 40 (8): 1-12.

[10] FEAMSTER N, REXFORD J, ZEGURA E. The road to SDN: an intellectual history of programmable networks [J]. ACM SIGCOMM Computer Communication Review, 2014, 44(2): 87-98.

[11] KOBAYASHI M, SEETHARAMAN S, PARULKAR G, et al. Maturing of OpenFlow and software-defined networking through deployments[J]. Computer Networks, 2014, 61: 151-175.

[12] BOSSHART P, DALY D, GIBB G, et al. P4: programming protocol-independent packet processors[J]. ACM SIGCOMM Computer Communication Review, 2014, 44(3): 87-95.

[13] BIFULCO R, RÉTVÁRI G. A survey on the programmable data plane: abstractions, architectures, and open problems[C]//Proceedings of the 2018 IEEE 19th International Conference on High Performance Switching and Routing (HPSR). Piscataway: IEEE Press, 2018: 1-7.

[14] LARSEN J K, GUANCIALE R, HALLER P, et al. P4R-type: a verified API for P4 control plane programs[J]. Proceedings of the ACM on Programming Languages, 2023, 7(OOPSLA2): 1935-1963.

[15] KWON W, CHEON S, CHOI K, et al. Real-time holographic media system utilizing HBM-based holography processor[C]// SIGGRAPH Asia 2024 Posters. New York: ACM Press, 2024: 1-3.

[16] LIU Y L, WU S, XU Q, et al. Holographic projection technology in the field of digital media art[J]. Wireless Communications and Mobile Computing, 2021, 2021(1): 1-12.

[17] YU H B, ZENG P, XU C. Industrial wireless control networks: from WIA to the future[J]. Engineering, 2022, 8: 18-24.

[18] HOSSAIN M M, PURDY G. Role of industry 4.0 in zero-defect manufacturing: a systematic literature review and a conceptual framework for future research directions[J]. Manufacturing Letters, 2024, 41: 1696-1707.

[19] 胡宇翔, 崔子熙, 李子勇, 等. 基于领域专用软硬件协同的多模态网络环境构造技术[J]. 通信学报, 2022, 43(4): 3-13.
HU Y X, CUI Z X, LI Z Y, et al. Construction technologies of polymorphic network environment based on codesign of domain-specific software/hardware[J]. Journal on Communications, 2022, 43(4): 3-13.

[20] ZHANG C, BI J, ZHOU Y, et al. HyperVDP: high-performance virtualization of the programmable data plane[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(3): 556-569.

[21] 李铭轩, 曹畅, 杨建军. 基于可编程网络的算力调度机制研究[J]. 中兴通讯技术, 2021, 27(3): 18-22, 61.
LI M X, CAO C, YANG J J. Computing power scheduling mechanism based on programmable network[J]. ZTE Technology Journal, 2021, 27(3): 18-22, 61.

[22] 李子勇, 胡宇翔, 田乐, 等. 支持多个网络功能共存的可编程数据平面虚拟化[J]. 电子与信息学报, 2023, 45(10): 3667-3675.
LI Z Y, HU Y X, TIAN L, et al. Virtualization of the programmable data plane for supporting coexistence of multiple network functions[J]. Journal of Electronics & Information Technology, 2023, 45(10): 3667-3675.

[23] YU Z L, WU J F, BRAVERMAN V, et al. Twenty years after: hierarchical core-stateless fair queueing[C]//18th USENIX Symposium on Networked Systems Design and Implementation. Berkeley: USENIX Association, 2021: 49-65.

[24] FEJES F, NÁDAS S, GOMBOS G, et al. DeepQoS: core-stateless hierarchical QoS in programmable switches[J]. IEEE Transactions on Network and Service Management, 2022, 19(2): 1842-1861.

[25] YOU C Q, ZHAO Y M, FENG G, et al. Hierarchical multiresource fair queueing for packet processing[J]. IEEE Transactions on Network and Service Management, 2022, 20(1): 726-740.

[26] LI Z Y, HU Y X, TIAN L, et al. Packet rank-aware active queue management for programmable flow scheduling[J]. Computer Networks, 2023, 225: 109632.

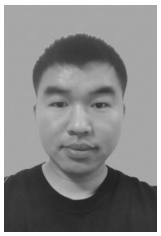
[27] MISHRA P K, CHATURVEDI A K. State-of- the-art and research challenges in task scheduling and resource allocation methods for cloud-fog environment[C]/ 2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT). Piscataway: IEEE Press, 2023: 1-5.

[28] MOHAMMED T, JEDARI B, DI FRANCESCO M. Efficient and fair multi-resource allocation in dynamic fog radio access network slicing [J]. IEEE Internet of Things Journal, 2022, 9(24): 24600-24614.

[29] SUN Q H, ZHANG J Y, LIU J F, et al. Shapley value approximation based on complementary contribution[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(12): 9263-9281.

[作者简介]

侯赛凤（1996– ），女，安徽宿州人，信息工程大学博士生，主要研究方向为新型网络体系架构、可编程网络数据平面。



崔鹏帅（1990– ），男，河南安阳人，博士，信息工程大学副研究员，主要研究方向为可编程网络数据平面、协议无关转发技术。

胡宇翔（1982– ），男，河南周口人，博士，信息工程大学教授、博士生导师，主要研究方向为新型网络体系架构、路由与交换技术。

董永吉（1983– ），男，吉林汪清人，博士，信息工程大学研究员，主要研究方向为可编程网络数据平面、协议无关转发技术。

刘义（1979– ），男，山东临沂人，中国移动通信集团河南有限公司高级工程师，主要研究方向为数据通信网、云计算和算力网络。