分类号：TP393 单位代码：11232

密级：





专业硕士学位论文

**面向URLLC网络的性能评价与资源分配**

学 院： 计算机学院

学科(专业)： 计算机应用技术

学 号： 2011020222

作 者： 刘宗奇

指导教师： 陈昕 教授

完成日期： 二○一三年十二月三日

学 院： 计算机学院

学科(专业)： 计算机应用技术

学 号： 2011020222

作 者： 刘宗奇

指导教师： 陈昕 教授

完成日期： 二○一三年十二月三日

学 院： 计算机学院

学科(专业)： 计算机应用技术

学 号： 2011020222

作 者： 刘宗奇

指导教师： 陈昕 教授

完成日期： 二○一三年十二月三日

学 院： 计算机学院

学科(专业)： 计算机技术

学 号： 2016020291

作 者： 马圣程

学校指导教师：陈昕 教授

企业指导教师：魏丫丫 技术总监

完成日期： 二○一九年四月十五日

**学位论文版权使用授权书**

本人完全了解北京信息科技大学关于收集、保存、使用学位论文的规定，按照学校要求提交学位论文的印刷本和电子版本。学校有权保留学位论文并向中国科学技术信息研究所等国家主管部门或其指定机构送交论文的电子版和纸质版，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。学校有权适当复制、公布论文的全部或部分内容。学校有权将本人的学位论文加入《中国优秀硕士学位论文全文数据库》和编入《中国知识资源总库》。

学位论文作者签名：

年 月 日

□公开 □保密（\_\_\_\_年\_\_\_\_月） (保密的学位论文在解密后应遵守此协议)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 指导教师签名： | |  | 学位论文作者签名： | |  |
|  | 年 月 日 | |  | 年 月 日 | |

**硕士学位论文原创性声明**

本人郑重声明：所呈交的论文题目为《面向URLLC网络的性能评价与资源分配》学位论文，是本人在导师指导下，进行研究工作所取得的成果。尽我所知，除了文中特别加以标注的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明并表示了谢意。本学位论文原创性声明的法律责任由本人承担。

作者签字：

年 月 日

# 摘 要

自动驾驶、无人机控制和智慧医疗等应用技术的快速发展，离不开5G（The Fifth Generation）通信技术的核心支撑。作为移动通信的国际标准3GPP规定，高可靠低时延通信URLLC ( Ultra Reliable and Low Latency Communicaiton )作为5G三大特性之一，具有高于99.999%的传输可靠性和1毫秒以内时延限制等特点。URLLC是确保5G部署与应用的重要特性，直接影响着所支撑系统的实时性、有效性和安全性。为满足不同应用领域的时延、带宽和计算量的多样化需求，5G网络通过网络切片技术为应用的资源分配提供支撑。

首先，围绕5G网络架构的特点，运用随机网络演算理论，分析了URLLC通信中的时延和可靠性等性能指标，构建了5G网络的串联通信模型，给出了用于计算接入网时延的随机边界。同时，通过对时延边界的性能分析，以及对接入网时延边界分析方法的扩展，分析了时延与数据到达速率、服务速率以及串联服务节点之间的关系，给出了核心网以及整个URLLC网络的时延边界。

其次，针对5G网络的带宽、存储和计算等系统资源的分配问题，为了保证高可靠低时延通信服务，通过分析网络、计算和存储资源的特点，设置任务的切片参数，运用马尔可夫决策方法，构建了具有5G网络切片特点的系统资源分配优化模型，并采用深度强化学习DQN（Deep Q Network）求解，得到了满足网络性能要求的系统资源分配策略。

最后，围绕5G接入网的URLLC时延边界特性和网络带宽、存储、计算等系统资源的分配策略，运用Python和MATLAB工具，构建了5G接入网的URLLC和网络切片分配策略的仿真平台。通过该仿真平台，对URLLC网络的时延边界，包括时延与数据到达速率、服务速率等关系，以及5G网络的带宽、存储和计算等系统资源分配的关键性能指标进行验证。实验结果表明，边缘部署方式可以减少系统中的串联服务节点，从而降低时延；在保证URLLC任务优先处理的情况下，采用深度强化学习DQN的资源分配策略能够提高系统资源的利用率。

**关键词**：URLLC；性能评价；资源分配；随机网络演算；强化学习

ABSTRACT

5G communication technology is an important support for the rapid development of automatic drive, UAV control, and intelligent medical technology. As the international standard of mobile communication, 3GPP claims that URLLC, as one of the three scenarios of 5G, has the characteristics of higher transmission reliability than 99.999% and less than 1-millisecond delay limit. URLLC is a significant feature to ensure 5G deployment and application, which directly affects the real-time, effectiveness and security of the supported system. Network slicing technology supports the 5G network by SDN/NFV. It can help the system to meet the diversified demands of delay, bandwidth, and computation in different application areas.

Firstly, according to the characteristics of 5G network architecture, using stochastic network calculus theory, we analyze the performance criteria of delay and reliability in URLLC communication and construct a tandem communication model of the 5G network. We calculate a stochastic boundary for radio access network delay. By analyzing the performance of delay boundary and extending the analysis method of radio access network delay, we find out the relationship between delay and data arrival rate, service rate, and serial service nodes. And we provide the delay boundaries of the core network and the whole URLLC network.

Secondly, aiming at system resources allocation problems, system resources are abstracted to network resources, computing resources, and storage resources. In order to guarantee the service of URLLC and improve the utilization of system resources, the network slicing parameters are set according to the requirements of network, computing and storage resources. Using the Markov decision process method, an optimization model of system resource allocation with the characteristics of 5G network slicing is constructed and solved by deep reinforcement learning DQN. The system resource allocation strategy is obtained which meets the system performance requirements.

Finally, based on the characteristics of URLLC delay boundary of 5G network, a simulation platform of URLLC is constructed with MATLAB. Through this simulation platform, we confirm the delay boundaries of URLLC network, including the relationship between delay and data arrival rate and service rate. Using Python and Matplotlib liberary tools, a simulation platform for network slice allocation strategy of the 5G network is built. Experimental data shows that the deployment of edge computing can reduce the delay by decreasing the number of tandem servers in the system. In the simulation experiment of resource allocation, the strategy with the DQN method can improve the utilization of system resources while ensuring priority processing of URLLC tasks.

**KEY WORDS:** URLLC, Performance Evaluation, Resource Allocation, Stochastic Network Calculus, Reinforcement Learning

**目 录**

[摘 要 I](#_Toc6474107)

[ABSTRACT II](#_Toc6474108)

[英文缩略词表 VI](#_Toc6474109)

[第1章 绪论 1](#_Toc6474110)

[1.1 研究背景及意义 1](#_Toc6474111)

[1.2 国内外研究现状 3](#_Toc6474112)

[1.2.1 高可靠低时延通信技术研究现状 3](#_Toc6474113)

[1.2.2 网络性能评价问题研究现状 4](#_Toc6474114)

[1.2.3 网络资源分配问题研究现状 5](#_Toc6474115)

[1.3 研究内容与创新点 7](#_Toc6474116)

[1.3.1 主要研究内容 7](#_Toc6474117)

[1.3.2 主要创新点 8](#_Toc6474118)

[1.4 论文组织结构 8](#_Toc6474119)

[第2章 URLLC网络性能分析与资源分配关键技术研究 10](#_Toc6474120)

[2.1 高可靠低时延网络的架构及性能指标研究 10](#_Toc6474121)

[2.2 基于随机网络演算的高可靠低时延网络性能评价研究 11](#_Toc6474122)

[2.2.1 随机网络演算理论基础 11](#_Toc6474123)

[2.2.1 关于时延分析的研究 13](#_Toc6474124)

[2.2.2 关于可靠性分析的研究 15](#_Toc6474125)

[2.3 高可靠低延时网络资源分配问题研究 15](#_Toc6474126)

[2.4 本章小结 18](#_Toc6474127)

[第3章 URLLC网络性能评价 19](#_Toc6474128)

[3.1 高可靠低时延网络的时延分析 19](#_Toc6474129)

[3.1.1 高可靠低时延网络模型 19](#_Toc6474130)

[3.1.2 高可靠低时延网络时延边界研究 21](#_Toc6474131)

[3.2 高可靠低时延网络的可靠性分析 26](#_Toc6474132)

[3.3 本章小结 27](#_Toc6474133)

[第4章 URLLC网络资源分配策略研究 28](#_Toc6474134)

[4.1 高可靠低时延网络切片模型 28](#_Toc6474135)

[4.2 基于强化学习方法的网络切片优化策略 29](#_Toc6474136)

[4.3 网络切片算法实现 31](#_Toc6474137)

[4.4 性能分析 32](#_Toc6474138)

[4.5 本章小结 33](#_Toc6474139)

[第5章 网络性能评价与算法仿真 34](#_Toc6474140)

[5.1 随机网络演算性能分析 34](#_Toc6474141)

[5.2 时延分析器 39](#_Toc6474142)

[5.3 基于网络切片的资源分配算法仿真 40](#_Toc6474143)

[5.4 本章小结 41](#_Toc6474144)

[第6章 结论与展望 42](#_Toc6474145)

[6.1 结论 42](#_Toc6474146)

[6.2 展望 42](#_Toc6474147)

[致谢 44](#_Toc6474148)

[参考文献 45](#_Toc6474149)

[附录A 图目录 48](#_Toc6474150)

[附录B 表目录 49](#_Toc6474151)

[个人简历、在学期间发表的学术论文及研究成果 50](#_Toc6474152)

# 英文缩略词表

| 缩略词 | 英文全称 | 中文全称 |
| --- | --- | --- |
| 5G | The Fifth Generation | 第五代移动通信系统 |
| ITU | International Telecommunication Union | 国际电信联盟 |
| 3GPP | The Third Generation Partnership Project | 第三代合作伙伴计划 |
| eMBB | Enhance Mobile Broadband | 增强型移动宽带 |
| mMTC | Massive Machine Type Communication | 海量积极类型通信 |
| URLLC | Ultra Reliable and Low Latency Communication | 高可靠低时延通信 |
| LTE-A | Long Term Evolution Advanced | 高级第四代移动通信系统 |
| SNC | Stochastic Network Calculus | 随机网络演算 |
| SDN | Software Defined Network | 软件定义网络 |
| NFV | Network Function Virtualized | 网络功能虚拟化 |
| NS | Network Slicing | 网络切片 |
| API | Application Program Interface | 应用程序编程接口 |
| VNF | Virtual Network Function | 虚拟化网络功能 |
| RL | Reinforcement Learning | 强化学习 |
| PD | Packet Duplication | 数据包复制 |
| RAN | Radio Access Network | 无线接入网 |
| HARQ | Hybrid Automatic Repeat Request | 混合自动重传 |
| D2D | Device to Device | 设备对设备通信 |
| MIMO | Multiple-Input-Multiple-Output | 多入多出 |
| UDN | Ultra Dense Network | 超密集组网 |
| FDD | Frequency Division Duplexing | 频分双工 |
| TDD | Time Division Duplexing | 时分双工 |
| MGF | Moment Generating Functions | 矩母函数 |
| OFDMA | Orthogonal Frequency Division Multiple Access | 正交频分多址 |
| QoE | Quality of Experience | 用户体验质量 |
| E2E | End to End | 端到端 |
| EC | Edge Computing | 边缘计算 |
| NR | New Radio | 新型无线 |
| 4G | The fourth Generation | 第四代移动通信系统 |
| NSA | Non-Stand Alone | 非独立组网 |
| SA | Stand-Alone | 独立组网 |
| EPC | Evolved Packet Core | 演进型核心网 |
| NGC | Next Generation Core | 下一代核心网 |
| 5GC | The Fifth Generation Core | 5G核心网 |
| UP | User Plane | 用户平面 |
| CP | Control Plane | 控制平面 |
| UL | Uplink | 上联 |
| DL | Downlink | 下联 |
| BLER | Block Error Rate | 误块率 |
| gNB | gNodeB | 5G基站 |
| RRU | Radio Remote Unit | 远端射频单元 |
| BBU | Building Baseband Unit | 室内基带处理单元 |
| AAU | Active Antenna Unit | 有源天线单元 |
| DU | Distributed Unit | 分布式单元 |
| CU | Centralized Unit | 集中式单元 |
| MANO | Management And Network Orchestration | 管理与网络编排 |
| MDP | Markov Decision Process | 马尔科夫决策过程 |
| DQN | Deep Q Network | 深度Q学习网络 |

# 绪论

## 1.1 研究背景及意义

随着移动通信技术的快速发展，第五代移动通信技术已经渗透到生活中的多个领域，从视频直播到在线支付，从无人驾驶再到智慧城市。5G通信技术首次应用出现在2018年韩国平昌冬季奥运会上，主办方通过应用5G技术顺利地实现了体育赛事的在线直播，让观众可以通过网络流畅的观看比赛。根据国际电信联盟（ITU，International Telecommunication Union）公布的5G标准时间表，5G将在2020年开始全面商业化[1]。

5G网络从设计之初，就被考虑用来支持各种复杂的通信场景。第三代合作伙伴计划（3GPP，The Third Generation Partnership Project）将这些不同的场景分为三大类：增强型移动宽带（eMBB, enhance Mobile Broadband）、海量机器类型通信（mMTC, massive Machine Type Communication）和高可靠低时延通信（URLLC, Ultra Reliable and Low Latency Communication）[2]。高可靠低时延通信广泛应用于自动驱动、工业控制和其它一些延迟敏感系统。3GPP组织对URLLC通信标准的关键要求是，保证下行和上行用户平面数据通信的延迟小于1ms，同时保持数据接收可靠性在99.999%[3]。如此严格的延迟要求就需要新的5G技术来满足。

现有的LTE网络虽然能够实现可靠性目标，但是代价是要付出几十毫秒的时间延迟。这远远不能满足URLLC的低时延要求。因此，时延成为URLLC网络设计中的瓶颈，也是最亟需解决的问题。许多科研院校和科技公司已经提出了一些从工程上解决的方案以尽量降低时延，如HARQ重传和免授权传输技术。然而，如何从理论角度分析时延的产生，并提出有效的资源分配策略，从而在保证传输时延的前提下，提高资源利用率是一个很有价值的研究课题。

随机网络演算（SNC，Stochastic Network Calculus）理论是延迟性能分析的有效工具。随机网络演算理论目前仍然在不断发展中，该理论的优点在于分析网络流量特性和评估系统通信性能[4]。与排队论不同，随机网络演算理论允许以很小的概率违背期望的指标限值，利用这一特性可以更好地获得统计复用增益。随机网络演算理论具有丰富的网络流量分析模型，可以针对数据积压、时延、可靠性等指标对网络系统进行性能分析与评价[5]。通过分析导致时延产生的原因，针对关键因素进行优化，调整网络资源部署，从而满足高可靠低延时场景下的性能的要求。

随着软件定义网络（SDN，Software Defined Network）和网络功能虚拟化（NFV,Network Function Virtualized）技术的快速发展，网络切片技术也应运而生。软件定义网络和网络功能虚拟化是实现可编程和灵活传输网络的一种新型网络架构。软件定义网络技术允许网络服务提供商通过应用程序编程接口（API，Application Program Interface）控制其资源来从而引入了网络的可编程性[6]。此外，软件定义网络技术采用OpenFlow协议来描述控制器与交换机之间的交互信息[7]，从而实现了网络控制层面与转发层面的分离，为网络切片技术提供了更广阔的设备兼容性。

网络功能虚拟化技术将原有的负责不同网络功能的物理设备转化为了虚拟化的网络功能（VNF，Virtual Network Function），可以在传输网络的不同部分动态创建网络功能虚拟化基础设施，为网络资源的部署和分配提供了灵活性[8]。软件定义网络和网络功能虚拟化为网络切片技术的实现提供了可能。网络切片技术可以将服务内容传递给定制的软件进行处理，同时可以根据特定的服务需求动态分配硬件资源[9]。因此，网络切片技术为5G通信的网络部署和资源分配提供了一种灵活的解决方案[10]。此外，网络切片技术可以根据网络服务需求动态分配网络中的资源，将基础设施映射成虚拟化的网络资源，为应用划分逻辑上独立的网络，实现网络硬件资源的复用。此功能可以提高服务质量和网络资源利用效率。同时，网络切片可以集成和分配一些相互独立的网络资源，为低延迟、高可靠性的应用场景提供服务[11]。

在5G网络应用中，如何根据不同场景的性能要求，为通信请求智能的提供所需的资源，目前仍是一个比较前沿的问题。随着人工智能技术的兴起，大量的采用机器学习方法对网络资源进行分配的研究不断涌现[12]。但目前5G网络仍未完全投入运营,对网络资源进行分配的网络切片策略还处于研究阶段。在这种情况下，缺乏大量的实际应用数据供有监督学习的智能算法进行训练。强化学习(RL，Reinforcement Learning)是一种不需要标记数据的学习方法，其算法思想是通过智能体与环境不断交互，获得反馈并调整动作以达到回报值最大化[13]。已经广泛的应用在棋牌博弈、游戏比赛、机器人、自动驾驶等领域。在最复杂的围棋比赛中战胜韩国李世石和中国柯洁的AlphaGo[14]其核心算法就是强化学习。当问题的复杂程度增加，环境的状态和动作的范围不断扩大，需要深度学习的方式来解决维度爆炸的问题。因此，深度Q学习结合了深度学习和强化学习的优点，成为了人工智能领域的新浪潮[15]。

为了将通信时延降低在标准要求以内，同时确保通信的可靠性，本文围绕URLLC场景下的性能分析与资源分配问题开展研究。首先，研究URLLC场景下的通信特点，以及用于分析性能的随机网络演算理论，解决资源分配问题的网络切片技术和强化学习技术。其次，研究了5G网络URLLC场景的网络架构，了解其性能指标，根据网络架构的特点建立分析模型，将网络模型中各个要素定义为随机网络演算理论分析所需要的随机过程。依据随机网络演算中时延的分析方法，结合随机网络演算中的一些特性，得出时延边界。再者，针对URLLC场景的应用特点，为满足通信需求而通过网络切片的方式进行资源分配。通过强化学习方法制定网络切片的分配策略，设置合理的切片分配的激励函数，利用贪心算法的思想选择激励值最大的行为，根据系统当前状态与行为值的选择建立贝尔曼方程，利用马尔科夫决策过程迭代更新完成系统的资源分配优化策略，有效的提高了系统的资源利用率。最后，通过实验分析，得到了影响时延产生的主要因素，分析了在满足URLLC时延要求情况下，各个影响因素变量的合理范围值。为URLLC通信的网络部署提供了有价值的指导和借鉴意义。通过仿真实验，验证了基于强化学习的网络切片策略的有效性，从而为实际环境下URLLC通信的资源分配提供了一种可选择的方案。

本课题依托于国家自然科学基金面上项目《5G超密集接入网智能动态资源分配及其优化方法研究》，项目编号：61872044，是5G超密集接入网智能动态资源分配及其优化方法问题中的关键问题之一，目的在于为通信时延的降低提供理论依据和优化方向，为智能动态资源分配策略进行前期预研，以提供一种既满足性能需求又提高资源利用率的解决方案。

## 1.2 国内外研究现状

随着自动驾驶、无人机控制等时延敏感通信应用的发展，URLLC通信越来越受行业关注。近年来，国内外学者围绕着URLLC的时延与可靠性等性能指标进行研究，提出了很多的性能评价方法与实施方案。针对网络部署与资源分配问题，以网络切片为技术手段的实现方式已在5G标准中敲定。

### 1.2.1 高可靠低时延通信技术研究现状

由于URLLC的标准尚未制定完成，许多研究者对URLLC的设计提出了不同的解决方案。总所周知，URLLC场景对于时延和可靠性的要求是极高的，所以许多学者将研究重点放在了如何设计和实现URLLC通信以满足性能要求。

Jimmy等提出了一种无需干预的物理层接口，将多种不同的协议接口整合为一个整体，提高了接口的兼容性。并且设计了一种分析架构。这种架构不但可以用传统的模型来分析可靠性，还可以分析特定技术的延迟概率分布,通过这种方式，能够根据延迟和可靠性来评价一个集成通信系统的性能[16]。

为了降低无线接口与互联网通信时的抖动，并且保证端到端通信的低时延，Delgado R A等人提出了一种新型多入多出级联控制器，该控制器的优点是解决了控制节点的数据流分割问题[17]。

Jaya Rao和Sophie Vrzic提出了一种数据包复制(PD, Packet Duplication)方法来满足延迟和可靠性的要求[18]。该数据包复制技术的原理是通过生成多个实例并在不相关的通道中同时发送，接收端再根据信道情况选择最佳的分组，以达到更好的传输可靠性。这种数据包复制技术可以在不增加无线接入网(RAN，Radio Access Network)复杂性的情况下提供一种经济有效的解决方案。

为了满足包括延迟和可靠性在内的关键需求，在文献[19-22]中讨论了目前主流的一些解决方案。这些技术包括HARQ快速重传(Hybrid Automatic Repeat request)、MIMO(Muiltiple-input-multiple-output)、波束形成、多协议接口整合、D2D(Device to Device)通信和超密集组网等。其中一些技术可以单独使用以提高性能，而有一些技术需要结合在一起，组合使用才能获得更好的性能。这些文章都提到了设计架构，需要在时域与频域上进行灵活的划分，这些精心的设计的目的，都是这是为了满足URLLC应用对低延迟和高可靠性的严格要求。

### 1.2.2 网络性能评价问题研究现状

针对URLLC场景的各项性能指标进行评价，是进一步优化改进URLLC通信，提高网络服务质量和用户服务体验的重要基础。由于URLLC对延迟和可靠性有着严格的要求，因此对URLLC的性能进行评估是非常有意义的。

**1）针对时延性能评估的研究**

Joachim等人提出一个可实现的延迟评估边界。比较了5G RAN传输中不同配置的延迟[23]。这些配置包括频分双工（FDD, Frequency Division Duplexing）、时分双工（TDD, Time Division Duplexing）、频段和时隙的使用。根据分析可知，利用高频带宽可以有效减少延迟，这也是毫米波技术得到广泛应用的原因。

随机网络演算理论是分析排队系统随机边界特性的有力理论工具，简化了性能指标的求解，在性能分析和理论边界计算中具有很好的实用效果。目前，针对无线网络边界性能评价已经取得了部分有效成果。Beck等人[24]设计了基于随机网络演算的DISCO网络性能分析器，分析了网络积压和延迟等边界特性，并获取网络随机边界性能，以适应新的网络需求。Li等人运用随机网络演算分析多基站网络时延，使用随机服务曲线和M/M/N模型推导排队延迟界限[25]。蒋玉明等人分析了无线电力通信系统的吞吐性能[26],通过将时延定义为约束条件，采用随机流量到达模型，推导出累计数据传输能力。

在端到端时延的分析工作中，M.Fidler等人使用随机服务过程来分析TCP的端到端延迟性能[27]。这篇文章采用随机服务过程实现了闭环流量的评价方法，并且同时分析了流量的积压和时延。K.Zheng等人分析了无线网络的有限状态马尔可夫信道的性能，基于矩母函数 (MGF，Moment Generating Functions) 推导出了马尔可夫信道的时延边界[28]。陈昕团队专注于LTE网络的端到端时延分析，研究通过优化资源分配以保证延迟性能[29]。在其研究中，证明了通信延迟受随机到达曲线和随机服务曲线的差值的约束。文献[30]是M.Fidler和Y.Jiang共同合作完成的，这两位作者是随机网络演算理论研究的代表性人物，他们在这篇文章中利用随机网络演算理论分析了多服务器系统的时延边界。

**2）针对可靠性评估的研究**

文献[31]提出了一种基于工作负载损失率的网络概率计算方法，该方法可通过对损失的计算来衡量网络服务的可靠性。文献[32]参考包络过程理论和矩母函数，得出了网络演算中的确定损失边界。文献[33]提出了一种新的网络服务模型，这种模型通过设置丢包参数来分析网络的丢包率。一篇来自华为公司的文章提出了一种免授权模式的上行链路传输机制[34]。该机制在没有调度请求的情况下动态授予传输权限，在保证安全性的前提下满足了上行传输中URLLC的可靠性要求。通过模拟不同数量活跃用户的随机到达请求，采用免授权模式，可以有效提高数据传输的可靠性。

### 1.2.3 网络资源分配问题研究现状

5G通信技术的设计初衷就是要满足不同类型的服务和应用。目前，很多学者在资源配置和能源效率方面做了很多研究。在URLLC场景中，如何将频谱资源合理分配以满足用户端设备的通信需求是一项长期引起关注的研究。

**1）结合能耗的资源分配问题研究**

Anand A和De Veciana G基于5G标准技术正交频分多址(OFDMA，Orthogonal Frequency Division Multiple Access)建立了一次性传输模型，该模型采用排队论对频谱资源进行分析的[35]。通过实验验证了在一次性传输系统中，长时间的小带宽比短时间的大带宽资源利用率更高。Mukherjee A等人[36]提出了一个节能方案。这种方案对一组相邻的基站进行协调，通过彼此共享一个睡眠时间表，让基站在休眠与唤醒的状态中切换。如果基站当前的流量负载较低，或只有较少的连接任务，则将其设置为休眠模式。在保证延时的前提下，这种切换方案比传统方式更加节能。

**2）基于网络切片的资源分配问题研究**

网络切片使网络运营商能够将逻辑上相互独立的网络资源分配给服务提供者、虚拟运营商和普通用户，并在网络基础设施上提供定制化的面向服务的功能，从而支持多种不同的服务和应用。通过对软件定义网络和网络功能虚拟化技术的应用，网络运营商可以提供自动化的、弹性的和可编程的网络切片。

随着网络应用数量增加，用户需求越来越复杂，应该考虑如何具体实现网络切片的算法，制定策略和机制来处理资源分配，特别需要应对高度动态化的资源共享需求。S. Vassilaras等学者关注于实时环境下的网络切片资源分配和控制算法，将网络资源管理和编排作为虚拟网络嵌入的一个实例 [37] 。在考虑到当前网络的动态特性基础上，应用图论将网络切片问题从网络拓扑的角度去描述，并将其归约为整数线性规划中的多商品流问题。

在IEEE Communication Magazine的5G网络切片专题研究中，S.Sharma等人阐述了面向服务的5G切片概念，分析了从实体网络到功能网络的范式转换[38]。充分考虑到了网络切片在整个生命周期内的设计、实现和编排，并提出了原生云网络切片方法。通过一个概念验证系统的演示，验证了端到端移动网络中的原生云网络切片方法在资源利用率方面的优越性。通过建立一种网络切片系统，并提出特定算法或者分配机制来解决问题的研究还有文献[39-43]，这些学者都对服务编排和切片操作进行了深入的研究，指出了满足所需定制和性能要求的关键支持技术。

**3）结合强化学习的资源分配研究**

随着人工智能的不断发展，很多技术都与人工智能相结合实现了效率的提升。此外，网络服务和应用需求的多样化、复杂化也需要通过智能的方法来提供支撑。很多学者建立的网络资源分配问题的求解，都借助于强化学习的方法来实现。文献[44-45]都是通过深度强化学习的方法分别解决了无线网和车联网下的资源分配问题。Yixue Hao等人利用网络切片技术来解决智能穿戴网络的资源分配问题，提出了基于超密集组网的新型的5G智能穿戴网络[46]。

面对时延敏感设备的海量接入、不同设备对用户体验质量QoE(Quality of Experience)要求不同和如何智能管理等挑战，结合机器学习和强化学习的方法实现了数据驱动的网络分配管理系统，很好的解决了资源分配的问题。文献[47]通过深度强化学习方法结合边缘计算（EC，Edge Computing）等技术手段，解决了满足URLLC通信性能要求的资源分配问题。

为了提高端到端（E2E，End to End）的平均可靠性，通过在边缘计算节点上设计了一个智能代理，开发了一个实时自适应的多用户卸载任务的计算资源分配策略。文献[48]从混合整数线性规划公式和启发式算法两方面给出了动态切片问题的求解方法。比较了动态切片与静态切片的优点，与静态切片相比动态切片可以将网络请求拒绝概率降低一个数量级。这可以帮助网络提供商接受更多的网络请求到其基础设施中，并可能增加其收入。

## 1.3 研究内容与创新点

### 1.3.1 主要研究内容

围绕高可靠低延时网络下的性能评价及资源分配问题，将重点研究的内容有：（1）高可靠低时延网络的通信特点及性能要求研究；（2）基于随机网络演算理论的高可靠低时延网络性能评价研究；（3）基于网络切片技术的高可靠低时延网络资源分配方法研究。

**（1）高可靠低时延网络的架构及性能指标研究**

分析5G高可靠低时延网络架构，研究网络通信过程中的核心组件，为分析网络性能提供模型依据。研究5G标准中关于高可靠低时延网络性能要求的规定，为后续分析时延和可靠性等指标提供数据支撑。

**（2）基于随机网络演算理论的高可靠低时延网络性能评价研究**

围绕5G高可靠低时延网络的性能要求，分别从时延和可靠性两个方面，系统构建分析模型。分析用户到达数据流量、系统服务速率以及网络部署结构，运用随机网络演算理论，建立时延边界分析模型。针对高可靠低时延网络性能指标要求的时延和可靠性等参数，给出其相应的服务速率和建议的网络结构，为网络资源分配的研究提供指导方向。

**（3）基于网络切片技术的高可靠低时延网络资源分配方法研究**

根据高可靠低时延网络的性能要求，为满足所需的服务速率，需要对网络的带宽、计算等资源进行分配。本文研究5G网络主流的网络切片技术，结合高可靠低时延通信的特点，建立随时间变化的资源动态分配模型。该模型可以反映当前网络的资源分配状态，以及切片任务对所分配资源的利用情况，通过对资源的利用情况设置效用函数，以实现最大化系统资源利用率的目标。运用深度强化学习方法，通过不断学习划分网络切片行为，自适应调整分配策略，在不需要标注数据的情形下，给出智能动态资源分配策略。

### 1.3.2 主要创新点

**（1）针对5G网络架构特点，建立网络性能评价模型**

针对5G高可靠低时延网络架构的特点，建立串联网络模型。该模型可以分析从用户端到云服务器的数据传输。根据随机网络演算理论中的随机服务过程和串联属性来对系统进行延迟分析。基于模型分析，得出时延与到达率和服务率之间的关系。计算了满足高可靠低延迟要求的最小系统服务速率。此外，通过分析显示，减少串联服务器的数量也可以减少系统延迟。这些分析结果为网络部署与资源分配提供了有价值的指导。

**（2）针对高可靠低时延网络的性能要求，得出系统时延边界**

为了满足低延迟高可靠的性能需要，提出了基于随机网络演算理论的分析方法，利用串联结构的性质，得到了一个适合于分析从用户端到基站的接入网时延随机边界。在此基础上进行扩展，推导得出了用于分析从用户端到云服务器整个网络的时延的随机边界。为了分析网络的可靠性，运用随机损失过程来表示系统的误块率。研究了错误块对延迟的影响，提出了一种延迟边界分析方法。基于分析结果，得到了时延与数据到达率、数据服务率之间的关系，并运用Python语言结合Matplotlib库设计并开发了时延模拟器，提高了时延分析效率。

**（3）运用强化学习去解决网络切片问题**

利用网络性能分析所得到的结果，指导网络资源分配的优化方向。为满足高可靠低时延的性能要求，设计基于网络切片的资源分配模型。综合考虑网络带宽、服务计算能力等资源，以及任务的时延要求等因素，提出了结合深度强化学习方法的资源分配策略。通过设置评价网络切片利用的效用函数，以最大化系统效用为目标，以是否划分切片为动作空间，利用马尔科夫决策理论和DQN方法，给出智能动态资源分配策略，有效的提高了网络资源利用率。

## 1.4 论文组织结构

针对高可靠低延时通信场景的特点，对网络性能中的时延和可靠性进行分析，构建了以网络切片技术为基础的资源分配模型，运用强化学习方法对分配策略进行优化。全文共分为六章，主要内容如下：

第一章，阐述了课题的研究背景及意义，结合问题的研究方向介绍国内外研究现状，简述了课题的主要研究内容、创新点和组织结构。

第二章，研究了5G网络架构以及高可靠低时延通信的特点；结合5G的架构特点和性能指标，研究了用于分析时延的随机网络演算理论。针对资源分配问题，研究了网络切片技术。为解决网络切片的策略优化问题，研究了强化学习优化方法。

第三章，研究了5G网络的架构特点并针对独立组网模式建立了性能评价模型。分别从时延和可靠性两个方面对5G高可靠低时延网络进行性能分析。在时延分析方面，先从接入网开始，分析了用户平面的时延边界，由此扩展到核心网以及云端的时延边界。在可靠性分析方面，通过引入随机差错过程求得在设定的违背概率情况下的误块率边界。

第四章，从带宽资源，计算资源和存储资源三方面对系统资源进行划分，分析了5G三大应用场景对这三类资源的需求差异，并建立了基于网络切片技术的资源分配优化模型。分析了强化学习对解决此类问题的合理性，并将网络切片资源分配优化问题转化为强化学习的序贯决策问题，并进行了算法实现。

第五章，对采用随机网络演算理论得出的时延边界进行仿真分析，将3GPP中对高可靠低时延网络的性能参数代入模型中，得到在设定违背概率情况下满足时延和可靠性要求的网络参数。在此基础上介绍了时延分析器，方便计算不同网络参数下的时延。进行了基于网络切片的资源分配仿真实验，验证了强化学习优化策略的优越性。

第六章，总结全文，在现有工作的基础上展望未来的改进方向。

# URLLC网络性能分析与资源分配关键技术研究

本章将重点研究高可靠低时延网络的结构特点，以及3GPP标准中关于性能的指标要求。根据网络结构的特点，研究随机网络演算理论中适合描述此类网络的模型和特性，从而为时延和可靠性的分析提供理论工具。深入研究了基于网络切片的技术的网络资源分配问题，同时对的强化学习方法进行了研究，为资源分配问题提供了解决方案。

## 2.1 高可靠低时延网络的架构及性能指标研究

**1) 5G网络架构**

高可靠低时延网络由国际电信联盟(ITU, nternational Telecommunications Union)提出，由第三代合作伙伴计划组织(3GPP, 3rd Generation Partnership Project)制定具体标准。在3GPP的TSG RAN WG1会议中，提出了5G NR(New Radio)的技术规范[49]。在NR中，将MIMO、HARQ、信道编码、双工通信模式等技术纳入到了5G接入网规范中，同时支持5G NR与4G LTE可以同时共存。这就引出了5G网络部署模式的问题。

根据3GPP标准Releas15的规范，5G的网络部署模式主要分为两大类: 独立组网(SA，Stand-Alone)和非独立组网(NSA, Non-Stand Alone)[50]。在非独立组网架构中，5G无线接入网络（RAN），也称为新无线（NR）与现有4G无线LTE和4G EPC基础设施核心网络结合使用，从而使新的5G无线技术无需更换原有4G的基础设施即可使用。在这种配置中，只支持4G服务，但可以享受5G网络提供的更高容量以及较低的延迟。在独立组网架构中，5G接入网NR连接到5G核心网NGC(Next Generation Core)，支持全套的5G服务，完全没有4G的存在[51-52]。关于非独立组网与独立组网的对比如表2.1所示。

**2) 性能指标研究**

3GPP对高可靠低时延网络的时延要求是，在用户平面(UP, User Plane)延迟的目标应为：上联(UL, Uplink)为0.5ms，下联(DL, Downlink)为0.5ms。此外，如果可能，延迟也应足够低，以支持将下一代接入技术用作可在下一代接入体系结构中使用的无线传输技术。

可靠性可以通过在一定延迟内传输指定字节的成功概率来评估，即在一定的信道质量下，从无线协议层入口点向无线接口的出口点传送小数据包所需的时间。对于URLLC可靠性要求是：在发送32字节小数据包的情况下，保证传输的误块率(BLER, Block Error Rate)在1\*10-5以内，同时满足用户平面延迟为1毫秒[2]。

表2.1 非独立组网与独立组网比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **组网方式** | NSA | NSA | NSA | SA |
| **接入网** | LTE,NR | eLTE | eLTE,NR | NR |
| **核心网** | EPC+ | 5GC | 5GC | 5GC |
| **新业务功能支持** | 提供新业务但功能受限 | 引入5G核心网 | 引入5G核心网 | 网络切片、新型QoS |
| **定位** | 过渡部署架构，适用于5G初期 | 过渡部署架构，适用于5G中期 | 准目标部署架构，适用于5G初期 | 5G目标部署架构，适用于5G全生命周期 |

如此严格的要求使得URLLC成为一项具有挑战性的服务，在技术实现方面，许多企业与学者都提出了不同的解决方案。然而，如何从理论上分析URLLC延迟的产生，并提出有效减少延迟的部署方案，仍是一项有挑战的工作。针对URLLC的性能保证问题，运用随机网络演算理论，构建分析时延和可靠性的模型将是研究的主要工作。

## 2.2 基于随机网络演算的高可靠低时延网络性能评价研究

### 2.2.1 随机网络演算理论基础

随机网络演算是一种用于分析通信网络的系统理论方法，其中使用最小加代数描述网络操作。网络服务节点被建模为最小加代数线性时变系统，输入输出关系由最小加代数中的线性时变系统理论控制。同时，将原有的确定性系统分析理论推广到具有概率特性的随机网络演算中，从而对随机网络进行建模和分析，将网络演算扩展到随机特性的方法是基于矩母函数和Chernoff边界理论[53]。

随机网络演算能够从理论方法上提供随机服务保证，从而更好地利用网络中的多路复用增益特性。在无线网络中，由于信道干扰、信号衰减等原因，无线信道的容量随时间随机变化，只能提供随机的服务保证。对于此类网络中性能分析和服务保证，随机网络演算能够很好的满足应用的需求。

在实际情况下，计算机网络由数据流和网络服务节点组成。与此对应的，随机网络演算理论也建立在流量过程和服务过程这两个基本概念之上。流量过程描述数据流的流量行为，服务过程描述网络服务节点的服务行为[54]。为了便于将随机网络演算理论应用于网络分析，其流量模型和服务器模型应满足一些基本性质。通常情况下，随机网络演算理论具有五个基本特性用来处理网络分析：

(1) **服务保证特性**

在特定的流量过程和特定的网络服务节点条件下，可以得到积压和延迟等单节点的随机服务保证。

**积压边界.** 如果一个系统，对于输入的网络数据流*A*提供服务下界曲线,假设数据流*A*有到达上界曲线，那么数据流的积压*B(t)*的边界为：

 (2.1)

其中=。

关于时延的边界将在2.2.1节具体描述。

(2) **输出特性**

流量模型的输入流可以表示服务器的输出流。

输出特性. 如果一个系统，对于输入的网络数据流*A*提供服务下界曲线，假设数据流*A*有到达上界曲线，那么输出的数据流*D*也是有边界的，对于所有的，有离开上界曲线，则离开数据流*D*的边界如下：

 (2.2)

(3) **串联特性**

多个网络服务器的串联部署的服务能力可以表示为网络服务节点卷积操作的形式。

**串联特性.** 在多个服务节点串联部署的网络*S*中，数据流依次经过各个节点。如果服务节点对数据的处理能力具有随机服务曲线，那么整个串联部署的网络*S*能提供随机服务曲线为，其中：

 (2.3)

 (2.4)

(4) **剩余服务特性**

对于具有多个竞争关系数据流的服务器，其承载的服务可以使用相同的服务器模型表示。

**剩余服务.** 如果一个系统，拥有一个网络数据流集合，集合中包含数据流*A1*和*A2*，系统对输入的数据流集合提供的服务能力有服务下界曲线，数据流*A2*有随机到达曲线，那么当时，对于数据流*A1*有：

 (2.5)

并且当为广义增函数时，为数据流*A1*的服务下降曲线。

(5) **叠加特性**

数据流的叠加可以用相同的流量模型表示。

**叠加特性.** 如果一个系统，拥有*n*个并行的网络数据流*Ai*，，假设每个数据流*Ai*有到达上界曲线，且为广义增函数，那么对于数据流集合*A*有，其到达上界曲线。

通过运用这五个基本特性，可以将网络模型简化，便于进行网络时延等性能指标的分析[4]。

### 2.2.1 关于时延分析的研究

随机网络演算是一种应用最小加代数来分析队列系统的理论。定义为非负增函数的结合，定义为非负减函数的集合。使用累积过程来表示数据流的总量。到达过程(Arrival Process)，离开过程(Departure Process)和服务过程分别表示为，，还有。对于任意的有，，且表示从*0*时刻到*t*时刻的累积达到数据量。和也表示各自的*0*到*t*时刻的累积数据量。通过到达过程、离开过程和服务过程的描述，就可以对随机到达曲线(Stochastic Arrival Curve)、随机服务曲线(Stochastic Service Curve)以及对时延(Latency)进行定义。

**定义2.1** **随机到达曲线**. 设是一个数据流，如果存在一个随机到达曲线和一个边界概率函数使得对于任意，满足：

 (2.6)

则称数据流具有随机到达曲线，并以为其边界概率函数。记为。

其中，作为随机到达曲线，其含义是数据流的上界，是违背概率(violation probability)函数。定义所要表达的含义是，数据流并不是严格的被限制在随机到达曲线内，有时可能突破边界，这种突破边界的可能性由边界概率函数来约束。

**定义 2.2** **随机服务曲线.** 设一个网络系统*S*，如果存在一个随机服务曲线 和一个边界概率函数，使得对于任意，满足：

 (2.7)

则称网络系统*S*具有随机服务曲线，并以为其边界概率函数。记为。

符号 表示累积最小加卷积操作，例如：

 (2.8)

其中是随机服务曲线，表示网络系统*S*在最坏情况下所能提供的的服务能力。与随机到达曲线类似，需要处理的数据可能比已离开的数据还要多，也就是说系统可能处理不过来到达的数据，但是这种可能性被边界概率函数约束。

在式(2.7)中出现的表示离开过程，随机离开过程与随机到达过程和随机服务过程相关，当时，它被描述为：

|  |  |  |
| --- | --- | --- |
|  |  | (2.9) |

通过式(2.9)可以更好地表示达到过程、离开过程和服务过程之间的关系。利用这些基本过程和随机曲线，就可以讨论延迟边界的定义。

**定义 2.3 随机时延过程.** 设和分别表示随机到达过程和随机离开过程。当时，时延过程描述为：

|  |  |  |
| --- | --- | --- |
|  |  | (2.10) |

式(2.10)表示时延过程的最小时值是，当满足条件使得*t*时刻到达的数据量小于或等于*t + d*时刻离开的数据量。这意味着数据不是立即离开服务器的，数据在服务器中的持续时间就是时延。到达过程小于等于离开过程，这意味着在*t*时刻到达服务器的数据在*t+d*时刻离开了服务器。如果大于离开过程，则表示*t*时刻到达的数据经过*d*时间,服务器没有处理完成。因此，使用小于或等于来描述服务器处理数据所需的最短时间，这就是时延。

根据延迟过程定义，利用随机到达过程和随机服务过程，就可以得出随机延迟边界。

**定理 2.4 随机时延边界.** 在一个网络系统*S*中，设是一个随机到达过程，其随机到达曲线为，边界概率函数为，记为；系统*S*其随机服务曲线为，边界概率函数为，记为；那么对于所有，有：

|  |  |  |
| --- | --- | --- |
|  |  | (2.11) |

其中函数表示与之间的最大水平距离，如图2.1所示；表达式表示边界概率函数与的累积最小加卷积运算。

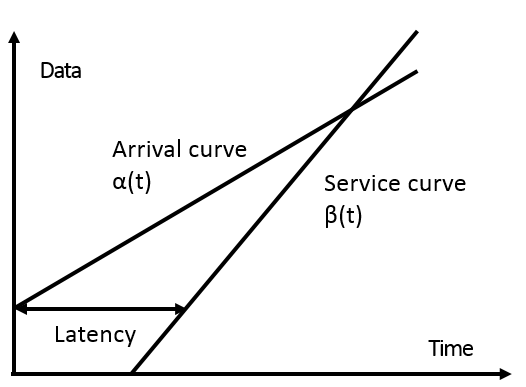


图2.1 随机网络演算中时延的定义

### 2.2.2 关于可靠性分析的研究

随机网络演算目前仍是一个处于不断发展完善的理论，并没有专门的随机过程用来分析可靠性。然而，有一种用于描述服务器出现服务错误的障碍过程，通过与随机服务过程做减法运算，可以用来表示数据处理的可靠性[4]。此外，还可以将随机网络演算理论进行发展，定义一种新的随机过程以及相应的边界曲线，来实现对可靠性的问题。

**定义 2.5 随机差错曲线.** 设一个网络系统的随机差错过程 具有以为边界概率函数的随机差错曲线，记为，对于任意,满足：

 (2.12)

式(2.12)表示网络系统在接受和处理数据时，有一定的几率会产生错误，过程表示时刻s到时刻t这段时间的误块率。随机差错曲线是过程的上界，但是这个上界有一定的概率会被突破，这个概率由边界概率函数来约束。

## 2.3 高可靠低延时网络资源分配问题研究

5G网络要支持多样化的服务需求，就需要将网络资源按照需求制定相应的分配策略和部署方案，但是由于资源有限，不能为每种应用场景都搭建一种专用的网络。因此，特别需要一种能够根据应用场景的需要灵活配置资源的解决方案，网络切片技术可以有效解决此类的资源分配问题。网络切片作为5G网络架构设计的关键技术之一，是支持多样化通信场景的重要手段。在3GPP的标准TS 22.261中有对于网络切片明确的要求[55]。网络切片能灵活调度及合理分配5G网络资源，满足不同应用场景的差异化需求。

根据3GPP制定的标准，5G网络的应用场景主要包括增强移动宽带（eMBB）、海量机器通信（mMTC）和高可靠低时延通信（URLLC）。相应的，5G网络切片的类型也要与应用场景的性能要求相匹配[56]，包括eMBB类型的切片，如网络监控切片、宽带视频切片和虚拟现实切片；mMTC类型的切片，如智能家居切片和工业控制切片；URLLC类型切片，如车联网切片和无人机控制切片。如图2.2所示。在进行网络切片的时候，在接入网切分时域或频域的资源，在传输网络切分网络带宽资源，在核心网编排网元功能，并且基于不同业务需求，可将计算资源或存储资源部署在边缘计算服务器上，实现灵活的资源分配。

图2.2 网络切片示意图

基于SDN/NFV技术的应用，网络提供商将物理网络基础设施划分为若干个不同的网络切片，为不同行业的不同通信业务提供服务[57]。NFV技术实现了将物理网络的各个实体功能映射为虚拟化的功能，SDN实现了各个功能组件的互联互通。要实现一个完整的网络切片功能，不但要划分资源，还要让资源合理的按照需求组合在一起，从前端到后端，从接入网到核心网，将不同功能组件有效的组合在一起，这就需要编排器（MANO, Management And Network Orchestration）实时的调配和管理。

网络切片的生命周期主要分为一下几个阶段：建立阶段、激活阶段，维护阶段和删除阶段[58]。在建立阶段，根据业务需求设计切片，调用网络资源，进行功能配置；在激化阶段，用设计好的方案生成网络切片实例，并投入运行；在维护阶段，对网络进行监控，汇报网络运行状态，对标网络指标，实时更新、调整、配置网络切片，满足业务需求变化；在删除阶段，根据调整需要，对网络切片进行删除，释放资源[59]。

通过对5G网络中网络切片的全过程管理，实现切片在网络中根据不同业务需求的灵活部署，可知网络切片的管理和编排是动态化的调整过程，这就需要一种能够动态优化资源分配的方法[60]。强化学习算法能够很好的解决动态优化的问题。

强化学习算法是一种非监督学习算法，它不关心当前输入的状态有什么特征，只关心对当前的输入应该采取什么样的动作，才能使长期的总体效果最优或者达到目的。这就需要智能体不断的与环境进行交互，智能体采用了一种动作，相应的环境会给智能体一个汇报，智能体根据汇报来评估采取的动作，有利于实现目标的动作被保留。不利于实现目标的动作被衰减。这一框架就是马尔科夫决策过程（MDP, Markov Decision Process）。

马尔科夫决策过程是通过智能体Agent不断的采取动作Action来和环境Environment进行交互，从而一步步的改变状态State并获得相应的Reward的过程，目的是找到最优策略。该过程具有马尔可夫性，即采取动作的策略只与当前状态有关，与历史状态无关。其形式化表述为：*M=<S,A,P,R,γ>*，其中：

*S*：有限的状态集合；

*A*：有限的动作集和；

*P*：状态转移概率，包含动作，即；

*R*：汇报函数；

*γ*：为折扣因子，用来计算累计汇报。

马尔科夫决策过程的求解就是找到最优的策略来使得累积汇报最大化，其中策略是根据特定状态选择相应的动作，通常使用*π*来表示策略，即：

|  |  |  |
| --- | --- | --- |
|  |  | (2.13) |

其含义是在状态为*s*的条件下，选择*a*作为动作的概率。

在一个给定的策略*π*下，累积回报为：

 (2.14)

累积回报在状态*s*处的期望值定义为状态值函数：

 (2.15)

相应的状态动作值函数为：

 (2.16)

状态值函数与状态动作值函数可以相互转化，其关系为：

 (2.17)

 (2.18)

从而得出最优状态值函数：

 (2.19)

最优状态动作值函数：

 (2.20)

最优策略可以通过最大化来决定，式(2.19,2.20)就是贝尔曼最优方程。

常用的求解贝尔曼最优方程的方法主要有Q-Learning，Sarsa，以及DeepMind公司提出的结合了深度神经网络的DQN（Deep Q Network）方法。

## 2.4 本章小结

本章首先研究了5G网络结构的特点，以及3GPP组织对高可靠低延时网络提出的性能要求。为了更好的满足性能的要求，研究性能评价理论工具随机网络演算，总结了随机网络演算理论对网络分析的基本特性，并分别从时延和可靠性两个反面对随机网络演算分析方法进行研究。针对网络资源分配问题，从3GPP标准入手，研究了5G网络资源分配的主流方法网络切片，从实现技术，切片类型，生命周期等几个方面对网络切片技术进行了分析。最后，对用于解决动态划分切片问题的强化学习方法进行了剖析，为后续的研究提供了理论基础和技术方案。

# URLLC网络性能评价

结合第二章对高可靠低时延网络架构的研究，本章将根据高可靠低时延网络的特点进行建模。利用建立的模型，代入3GPP对高可靠低延时网络的性能要求，从时延和可靠性两个方面进行分析，得出影响时延的关键因素，为性能优化以及网络部署提出有价值的指导意见。

## 3.1 高可靠低时延网络的时延分析

### 3.1.1 高可靠低时延网络模型

为了描述无线接入网（RAN）中的通信，本文利用随机到达曲线来描述用户设备UE数据发送到基站(gNB, gNodeB)端的过程。在此基础上，根据5G网络拓扑结构和随机网络演算的串联特性，推导出基站到云服务器其余阶段的数据传输。随机到达曲线的每一阶段都具有延迟特性，因此，URLLC系统的整个时延是从UE到云服务器各端通信的时延共同组成的。

由第二章的URLLC通信技术研究现状可知，5G网络包括独立网络(SA)和非独立网络(NSA)。 SA模式是建立全新的5G网络，而NSA模式是5G和4G LTE的结合，根据接入网和核心网所采用的具体实现技术，还可以进一步划分不同的组合方案，然而NSA模式只是为了节省部署成本而采取的一种过渡方案。为了简化系统模型，适应将来的5G全新网络，本文只讨论SA模式组网的情况。

4G LTE网络由UE、远端射频单元（RRU, Radio Remote Unit）、室内基带处理单元（BBU, Building Baseband Unit）、EPC组成，EPC是4G LTE的核心网，远端为云服务器。与4G 网络不同，5G网络主要由UE、gNB、NGC和云服务器组成。其中，5G基站gNB由有源天线单元（AAU, Active Antenna Unit）、分布式单元（DU, Distributed Unit）和集中式单元（CU, Centralized Unit）三部分组成[61]。

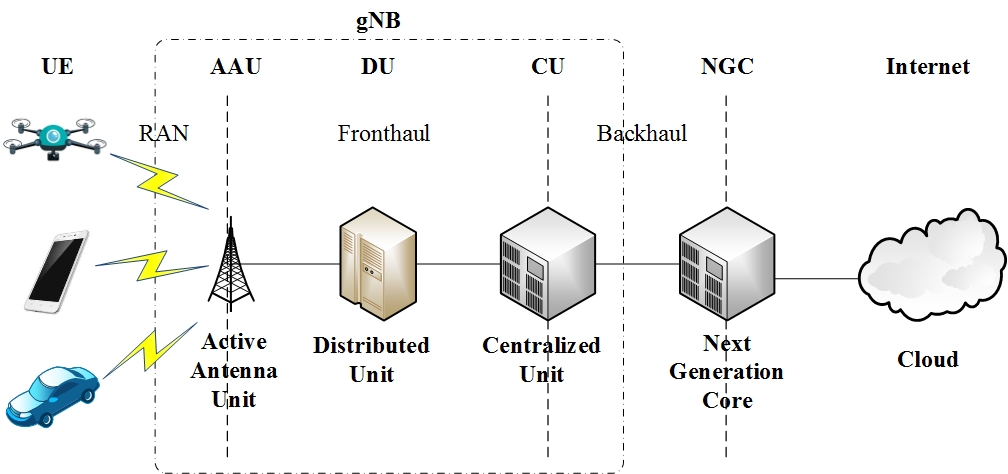
在接入网部分，AAU取代了原来4G中的RRU，还结合了一些BBU在物理层上的功能。4G中的BBU功能将在5G中重建为DU和CU，其中CU负责接入端的服务汇聚，专注于处理低实时性的协议而采用集中式部署；DU主要为终端设备提供访问接入，包括射频和部分信号处理，DU专注于实时性高的传输，适合采用分布式部署；DU和CU根据特定需要可以部署在一起也可以分开部署。NGC作为5G的核心网取代了4G中的EPC，5G核心网基于SDN/NFV技术而设计实现，旨在更好地适应多样化的云平台应用。总体网络架构如图3.1所示。

图3.1 5G网络架构

由于3GPP中对于用户平面的时延要求远远高于控制平面，所以主要讨论用户平面（UP）上的时延，而不是控制平面（CP）上的时延。用户平面的时延主要是用户终端与网络节点在IP层传输和接收数据产生的通信时间，而控制平面时延则是无线资源分配以及从空闲状态切换到活动状态所花费的时间[36]。

在实际通信中，UE设备首先接入AAU，AAU实际上是基站的一部分，这部分通信属于RAN。UE的数据将被AAU接受并转发给DU，如果DU和CU是联合部署的，那么数据可以直接到达CU，否则DU再将数据转发给CU。从AAU到CU这段通信属于前传(fronthaul)。数据离开CU继续向核心网NGC传输，这部分的通信属于回传(backhaul)。核心网处理数据并将其发送到云端，数据最终到达其目的地，整个通信过程结束。因此，5G系统的整体时延是由以下几个部分组成的：RAN接入网的处理时延，前传时延，回传时延，核心网处理时延，云端的时延。

时延的构成可由式(3.1)描述：

|  |  |  |
| --- | --- | --- |
|  |  | (3.1) |

其中，表示UE到AAU这段通信的时延，表示AAU到CU这段通信的时延，表示gNB到核心网这段通信的时延，表示核心网处理数据消耗的时延，表示数据从核心网到云端这段通信的时延，对于URLLC网络的时延分析，主要研究。

数据从用户终端UE经过各个网络节点最终到达云服务器端，根据随机网络演算理论，对通信的时延分析可以描述为：

 (3.2)

根据3GPP标准的要求，URLLC的用户平面时延要求在一毫秒以内，那么就需要式(3.2)中的*d*小于1，单位是毫秒（ms, millisecond），作为概率边界是一个很小的值，式(3.2)表示URLLC网络成功传输数据并且满足时延要求，不满足时延要求的情况其概率是非常小的[5]。

URLLC的网络特性可以用前面介绍的随机过程描述为动态服务器模型。从UE发送的数据可以用随机到达过程*A(t)*来表示，网络服务节点提供的服务能力可以描述为随机服务过程*S(t)*。根据式(3.1)的假设，将URLLC网络描述为一个串联系统，因此，URLLC的时延适用于第二章中提到的随机网络演算串联特性，基于此特性，本章开始针对URLLC网络进行时延分析。

### 3.1.2 高可靠低时延网络时延边界研究

根据对网络架构以及通信过程的分析，UE的网络流以串联式的方式依次经过gNB、核心网和云端。每个网络节点 都提供了相应的随机服务曲线。先从无线接入网部分的时延开始分析，即UE到gNB部分，再对这部分的分析结果进行扩展，从而获得UE到云端的整个通信时延。

基站gNB包括AAU，DU和CU，相应的随机服务曲线为，，。数据是从UE发起的，最先到达AAU，所以随机到达过程用来描述数据的到达。数据从AAU离开再发向DU，以此类推，从DU到CU，其随机到达过程可以分布表示为，，也就是前一个网络节点的输出正是后一个节点的输入，前一个节点的随机离开过程等于后一个节点的随机到达过程。数据开始发起的是无线接入网系统gNB的到达过程，最后CU节点的随机离开过程也是gNB的随机离开过程。

考虑到gNB是一个串联部署的网络系统，用来表示有源天线单元AAU的数据处理能力，应用式(2.9)对于随机离开过程的定义可得：

|  |  |  |
| --- | --- | --- |
|  |  | (3.3) |

同理，可以得到DU的随机离开过程：

 (3.4)

又因为，用式(3.3)不等号的右端来替换式(3.4)中的可得：

 (3.5)

继续迭代，得到CU的随机离开过程：

 (3.6)

根据最小加卷积运算的结合律，可得：

 (3.7)

从gNB的无线接入网的子系统来看，是gNB的输入，是gNB的输出。因此，，，得到gNB子系统的离开过程：

 (3.8)

应用随机时延过程的定义和式(3.8)的结果，可以得到gNB系统的随机时延过程：



(3.9)

在这里要引入随机网络演算中的串联特性，以便于对随机时延边界的推导。如第2章2.2.1节的随机网络演算基础知识所介绍的那样：

**串联特性.** 在有多个服务节点串联部署的网络*S*中，数据流依次经过各个节点。如果服务节点能够提供的服务具有随机服务曲线，那么整个串联部署的网络*S*能提供随机服务曲线为，其中：





应用定理2.4和串联特性，结合式(3.9)，可以得出gNB系统的时延边界。

**推论3.1 gNB系统时延边界.** 在gNB子系统中，设为随机到达过程，其随机到达曲线为，边界概率函数为，记为；子系统中服务节点的随机服务过程分别为，和，并且，，记为，，；当，时有：



 (3.10)

其中，。

证明：因为时延过程的定义如式(2.10)：，就意味着，将不等式右边左移，结合式(3.9)得到：

 (3.11)

对于部分，结合式(3.8)展开得到：



 (3.12)

由于，，，，利用串联属性可以得出gNB随机服务过程等于, 其随机服务曲线等于，gNB能够提供的服务率小于，其边界概率函数由限制。因此，式(3.12)可以化为：







 (3.13)

根据式(2.8)，可将变换为，带入式(3.13)得：









  (3.14)

在式(3.14)中，为了构造式(2.6)的形式，增加了随机到达曲线，根据定义2.1可知，的概率是由来约束的，由随机服务曲线的定义可知，的概率是由违背概率函数来约束的。再利用定理2.4，将表示时延的变量*d*替换为，为和之间的最大水平距离，在时，其隐含的条件是：

 (3.15)

最后得出：













至此，推论3.1证明结束。

对于gNB子系统的时延分析，适用于接入网RAN的通信，即UE到CU这一阶段的网络通信。例如目前较为流行的移动边缘计算部署方式，将计算资源部署在基站，靠近用户侧。运用推论3.1可以尝试此类的场景的时延分析。对于整个网络通信，从UE到云端的时延分析，需要继续将推论3.1进行扩展，从而得到整个URLLC网络的时延边界。

**推论3.2 URLLC系统时延边界.** 在URLLC系统中，设为随机到达过程，具有随机到达曲线以及，记为；服务节点的随机服务过程分别为，和，随机服务曲线分别为，边界概率函数为，记为，，；那么，当，，系统整体时延满足：



 (3.16)

其中，

证明：gNB子系统是有AAU,DU和CU三部分组成的，而对于整个URLLC网络来说，不仅包含gNB，还包括核心网NGC和云服务，对推论3.1中的网络节点进行扩展就可以证明本推论。根据随机网络演算的串联特性，整个网络提供的服务能力具有随机服务曲线以及违背概率函数，记为，其中：

 (3.17)

将带入式(3.17)，得到：

 (3.18)

 (3.19)

将带入式(3.19)，得到：

 (3.20)

根据时延过程的定义2.3，URLLC网络的时延过程可以定义为：

 (3.21)

那么时延边界为：

 (3.22)

同样关注于这一部分，根据随机离开过程的定义可知：













 (3.23)

根据随机到达曲线式(2.6)可知，的概率边界由约束，由式(2.7)可知，的概率边界由来约束。对于长期运行的系统来说时，趋近于0。最后应用定理2.4得URLLC系统的时延边界：

 (3.24)

至此，推论3.3得证。

由推论3.3可知，有四个部分的因素需要讨论，这些要素分别是网络的随机到达曲线，随机到达过程的概率边界函数，系统总体的随机服务曲线和随机服务过程的概率边界函数，若需要获得更具体的结果，可以通过分解和来实现。

在高可靠低延时通信场景下，数据的传输需要更高的速率，从而使用更大的带宽，往往是利用非常小的时隙结合频率比较高的频段，毫米波技术就是这样一种技术。

假设UE发送的数据是服从泊松分布的，其平均到达速率为，那么随机到达曲线=，随机到达过程的概率边界函数为：

|  |  |  |
| --- | --- | --- |
|  |  | (3.25) |

将式(3.25)转换为矩母函数的形式和并利用Chernoff边界理论[62]可得：

|  |  |  |
| --- | --- | --- |
|  |  | (3.26) |

在设定了和之后，需要考虑的是URLLC系统随机服务过程的和。

为了简化问题，假设系统的各个服务节点能够提供相同的服务能力，假设每个服务器提供的服务能力，*C*为单位时间的服务速率，随机服务过程的概率边界函数为，根据串联特性定理3.1可以得出：



 (3.27)

应用文献[63]中的结论，可得：

 (3.28)

其中，，并且各个网络节点的服务边界概率函数为，，，，，其中和为服务边界参数，将式(3.26)和(3.28)的结果带入式(3.24), 并应用文献[64]中的定理3，可以得出：



 (3.29)

设, 设，那么*d*为时延边界，为时延边界的违背概率。

下面分别对*d*和的表达式求解来分析*d*和的关系。对于*d*，可知。对于，考虑长期运行情况下, 那么 ，将化简，可以得到：。将关于d和关于的式子合并可得：

 (3.30)

通过式(3.30)就可以分析时延*d*与服务速率*C*,以及到达速率还有违背概率之间的关系。

## 3.2 高可靠低时延网络的可靠性分析

对于评价URLLC网络的可靠性，目前的随机网络演算理论还没有完全吻合的评价方法。学术界常用的方法是从其它角度来考虑[31-33]，通过研究网络系统的丢包、错误和异常等的发生概率来反映系统的可靠性。本文尝试在随机服务过程中考虑发生差错的概率，以系统的服务率和误块率共同作用的效果作为系统的最终服务能力，从而对系统的可靠性进行分析。

为了研究URLLC的可靠性，运用随机差错过程来描述系统中的误块率。如定义2.5中的式(2.12)所示，用变量 来表示系统的误块率，如果的数值很小，表示误块率很低，说明系统运行非常稳定，出现错误的几率很小。在随机服务过程定义中，表示服务的数据量是没有考虑误码率的，那么现在引入了来约束误码率的发生上界，就是排除了错误发生只留下正确处理的数据的比率，设为不考虑误码率的系统服务率，那么为考虑了误码率以后系统正确处理数据的能力，可以通过设置，将其简化为。

在随机网络演算理论中，经常使用矩母函数来计算系统的性能边界。矩母函数由于其幂指数的形式，可以很方便的用乘法的运算来表达随机过程的求和关系，本文对于误码率和随机差错过程的分析将采用矩母函数的方法，结合式(2.12)可得出当时，随机差错过程的矩母函数为：

|  |  |  |
| --- | --- | --- |
|  |  | (3.31) |

应用Chernoff边界理论可知随机差错边界为：

|  |  |  |
| --- | --- | --- |
|  |  | (3.32) |

对于URLLC通信的可靠性分析，重点在于服务速率与误块率的关系。假设系统的服务速率会受到误块率的影响，应用文献[5]中包络模型的结论，将式(3.32)不等号右边转换为，从而随机差错过程变为：



 (3.33)

在考虑误块率与时延的关系时，可以通过对服务率加上误块率的影响来实现。设

未考虑误块率情况的服务率为*C*，误块率为，那么考虑了误块率情况的服务率为，将其带入随机时延边界式(3.29)替换里面的服务率*C*，就可以得到有误块率影响的时延边界：

|  |  |  |
| --- | --- | --- |
|  |  | (3.34) |

## 3.3 本章小结

本章针对高可靠低延时网络的性能评价问题从时延和可靠性两方面进行研究，分别求得这两个性能指标的概率边界。在时延分析方面，分析了5G网络URLLC通信的网络架构，分析其时延的构成，根据网络架构设计基于随机网络演算理论的分析模型。在此基础上，从gNB子系统和整体网络两个视角分析了数据到达与服务，利用随机网络演算的串联特性，获得了系统的随机时延概率边界。在可靠性分析方面，引入误块率这一性能指标，采用随机差错过程来表达系统出现错误的情况，并利用矩母函数及chernoff边界获得了随机差错过程的概率边界。

# URLLC网络资源分配策略研究

为了满足复杂多样的通信业务，5G网络通过网络切片技术来满足不同的需求。今年来随着人工智能技术的发展，5G网络也向智能化演进[65]。根据第3章性能分析的结果，本章提出了基于网络切片技术的网络资源分配策略，针对5G三大类应用场景，尤其是对高可靠低时延通信的需求，利用强化学习智能方法实现了动态分配切片的策略。

## 4.1 高可靠低时延网络切片模型

在网络切片模型中，假设系统分配切片的管理周期为*T*，当过了*T*时间以后，对已分配的切片按照需要进行重分配或回收。在*T*时间内将有m个待分配的切片任务，系统分配一个切片任务，就会得到相应的一个奖励，由于系统资源有限，系统从m个任务中选择n个任务进行分配(n<m)，使得系统总的收益最大。

在网络切片模型中，将资源抽象为三大类：带宽资源，存储资源和计算资源。设网络系统的总的带宽资源为*B*，总的存储资源为*S*，总的计算资源为*C*，那么一个切片任务对资源的需求定义为四元组*NSi=<bi,si,ci,di >*，其中：

*NSi*为第*i*个切片任务；

*bi*为第*i*个切片需要的带宽资源；

*si*为第*i*个切片需要的存储资源；

*ci*为第i个切片需要的计算资源；

*di*为第i个切片要满足的时延要求；

所有切片占用的带宽资源要小于系统的总带宽：

|  |  |  |
| --- | --- | --- |
|  |  | (4.1) |

所有切片占用的存储资源要小于系统能提供的总的存储空间：

|  |  |  |
| --- | --- | --- |
|  |  | (4.2) |

所有的切片占用的计算资源要小于系统能提供的总的计算能力：

|  |  |  |
| --- | --- | --- |
|  |  | (4.3) |

假设切片任务需要的存储资源*si*就是要处理的任务的数据量大小，那么时延要求*di*与切片的带宽资源有关。要求满足：

|  |  |  |
| --- | --- | --- |
|  |  | (4.4) |

设奖励函数为*R*(*NSi*)，参数为切片任务，奖励函数将与切片任务的参数相关，其函数表达式为：

|  |  |  |
| --- | --- | --- |
|  |  | (4.5) |

对于是否进行分配用变量来表示，为0,1变量，其定义为：

|  |  |  |
| --- | --- | --- |
|  |  | (4.6) |

则基于网络切片技术的资源分配模型可表示为：



s.t 





 (4.7)

在5G通信网络中，主要的应用场景有三大类，对于mMTC通信，其通信的特点是对带宽要求不高，传输数据量不大，时延要求不高，但是连接数量大，连接时间长；对于eMBB通信，其特点是对带宽要求大，对时延要求不高，传输的数据量大。对于URLLC通信，其特点是对时延要求特别高，对带宽的要求相对较高。

通过对切片任务中的参数大小的设置，就可以将切片分为定义的三类切片。对带宽参数b和存储参数s需求特别大的是eMBB通信，对时延参数d要求特别小的是URLLC通信，对这两个参数要求都不是特别严格，但是切片请求数量较多的，就可以归类为mMTC通信。根据切片任务的特点，就可以指定策略对优化问题进行求解。

## 4.2 基于强化学习方法的网络切片优化策略

通过4.1节对网络切片问题的描述，可知该问题是具有时序特点的序贯决策优化问题。即当前分配的网络切片会影响系统的资源剩余量，从而影响下一个切片的选择。如何在现有的资源总量不变的前提下，在任务队列中决定对到达的切片任务是否进行分配，使得已分配的切片任务所得到的总收益最大，从而达到网络资源充分利用的目的。

对于求解基于网络切片的资源分配序贯决策的优化问题，强化学习方法是非常适用的。首先，5G网络尚未大规模商用，对与网络切片技术在行业内并没有大量的实用数据，所以对于采用大量标注数据来进行训练的监督学习智能方法来说，并不是很适用，因为缺乏可用的训练数据，然而强化学习则没有受到训练数据限制的问题。其次，强化学习很适合处理这种具有马尔科夫性质的状态时变系统的优化问题，通过智能体和环境不断交互，选择动作，得到反馈并进行评估修正，进而不断优化，结合了神经网络的强化学习方法DQN更是能支持状态空间爆炸的复杂问题优化。因此选用强化学习方法来对基于网络切片的资源分配问题进行优化。

为了使用强化学习的训练模式，需要对已定义的网络切片模型进行转化。强化学习方法需要明确状态空间，动作空间和回报函数等元素，结合网络切片资源分配模型分别对这这三种元素进行定义。

*State =* [*B, S, C, L*]

其中*State*为状态空间，其中包含四个变量，*B*为网络系统中当前可用的带宽资源(Bandwidth)，*S*为当前可用的存储资源(Storage)，*C*为当前可用的计算资源，*L*为当前任务队列的长度(Length of Queue)，就是还有多少待分配的切片任务。

*Task* = [*b, s, c, d*]

*Task*为切片任务，其中*b*为切片对带宽的需求，s为对存储空间的需求，c为对计算资源的需求，*d*是对时延的需求，其中时延d可以根据模型的中的约束条件来计算获得。

*Action =* [*0, 1*]

*Action*为动作空间，其中0表示不对当前切片任务进行分配，1表示对当前切片任务进行分配。

*Reward* = *b + s + c + 1/d*

*Reward*为回报函数，回报函数的设计主要体现两个思路，一个是消耗的资源越大，回报要越高；另一个思路是时延越小的任务，说明越紧迫，回报也相应越高。就是在时延上体现了高可靠低时延网络的特点，对于时延任务要求高的切片，将其回报设置较高，有利于模型优先分配此类高可靠低时延切片任务。

在算法中涉及到选取动作的策略，有基于贪心算法的策略，但这种策略容易陷入局部最优，所以本算法的选取动作的策略使用了ε-greedy策略，这种策略可以很好的平衡算法中利用(exploitaion)和探索(exploration)的关系，策略可以表示为：

 (4.8)

有了状态空间、动作空间，回报函数以及动词选择策略的定义，就可以写出资源分配问题的贝尔曼最优方程：

 (4.9)

(4.10)

通过不断的迭代训练，就可以得到相对优化的资源分配方案。

## **4.3 网络切片算法实现**

通过对基于网络切片的资源分配优化问题的分析，将其转化为适用与强化学习方法进行求解的问题。算法设计采用双层循环结构来解决问题，外层循环用于表示强化学习训练的回合，其循环次数为设定的训练回合数。内层循环用于处理切片任务，其循环次数为设定的切片任务队列长度。

内存循环的结束标志有两个，一个标志是任务队列长度为0，表示当前任务全部决策完成，另一个标志是系统资源为0，在网络资源B、计算资源C和存储资源S中的任何一种变量值为0时，表示不能再进行切片资源分配决策。内层循环完成一次，表示将队列中切片任务的全部进行决策一次，将生成一个决策序列D。每一个切片任务在决策完成后，都通过reward函数计算出一个收益值，内存循环结束后，收益值累计到总收益E中。

算法的核心思想在于，将当前系统资源状态state与决策动作action以及动作收益reward建立对应关系，并将这种关系保存在记忆池中，作为修正决策的历史数据。通过外层循环不断的迭代运行，记忆池中的数据将填满，这时触发DQN的学习过程。DQN学习就是利用保存的历史数据，代入贝尔曼最优方程q\*(s,a)，通过神经网络训练修正生成动作函数中的参数，从而使学习后的模型产生的决策动作更利于获得高的收益。

算法的具体实现如算法4.1：

|  |
| --- |
| 算法4.1 基于网络切片的资源分配算法 |
| **输入**：B, S, C, T,γ,Episode  **输出**：决策序列D, 总收益E  1：初始化B,S,C为系统能够提供最大资源，Episode为训练回合数  2：初始化T任务队列，确定切片任务的资源需求量  3：**Repeat**  4： E=0，设置结束标志flag=0  5： state = [B,S,C,len(T)]  6： **Repeat**  7： 根据生成的state选择动作action，利用策略π(action|state)  8： **if** B<=0 **or** S<=0 **or** C<=0 **or** T=∅ **then**  9： flag=1  10： **break**  11： **end** **if**  12： NS = T.pop() 获取当前任务  13： **if** action==0 **then**  14： 未分配，收益为0，系统资源保持不变  15： **else**  16： **if** B>=NS.b **and** C>=NS.c **and** S>=NS.s **then**  17： B=B- NS.b, C=C-NS.c, S=-NS.s 分配资源  18： reward(NS.b, NS.c, NS.s,NS.d)  19： state=[B,S,C,len(T)] 生成新的状态  20： **return** state,reward,flag  21： **else**  22： 系统剩余资源不满足当前切片的需求  23： reward=惩罚收益，负值  24： **return** state,reward,flag  25： **end** **if**  26： 保存state,action,reward,新的state  27： E=E+ reward，保存决策序列D  27： 检查已保存数据的数量，通过计算q\*(s,a)来调整参数  28： **if** 结束状态==True **then**  29： **break**  30： **end** **if**  31： **until** T=∅  32： **until** episode=0  33： 输出决策序列D，总收益E |

## 4.4 性能分析

影响算法4.1的时间复杂度的因素主要是算法中的两次循环，分别在第3行和第6行。第3行的循环为外层循环，其循环次数为训练的回合数。外层循环的执行次数没有变化，不受函数逻辑的影响，就是设定的目标训练回合数，若将训练回合数记为E，则时间复杂度为O(|E|)。

第6行的循环为内层循环，其循环次数为切片任务队列的长度。控制内层循环结束的条件有两个，一个是任务队列为空，另一个结束标志flag为1。当任务队列为空时，说明内存循环处理了队列中的所有切片任务，执行次数为队列的长度，记为L。当结束标志flag为1时，是因为出现了系统资源为空的情况。随着对切片任务的处理，系统资源不断分配而减少，当某一时刻系统资源减少为0的时候，不能再次进行分配，因而循环结束。这种情况的循环的执行次数小于队列长度，故内层的最大循环次数为L，其时间复杂度为O(|L|)。

综合内外层循环的时间复杂度，本算法的时间复杂度为O(|E||L|)。

算法4.1的空间复杂度主要受两方面影响，一方面是切片任务队列的长度，记为L，另一方面是保持训练数据的记忆库空间，记为M。无论是算法循环多少次，切片任务长度L的最大值都是固定，且随着内层循环中任务分配决策逐渐减少，在外层循环中再次恢复初值，其空间复杂度为O(|L|)。保存训练数据的记忆库的空间最大值也是固定的，每一次任务决策后的产生的数据都会保存在记忆库中。当记忆库填满的时候，会用新的数据覆盖旧的数据，所以记忆库的空间复杂度为O(|M|)。故算法总的空间复杂度为O(|L|+|M|)。

## 4.5 本章小结

本章围绕高可靠低时延网络资源分配问题开展研究，以最大化系统收益为目标。首先将5G通信系统资源抽象为网络资源、计算资源和存储资源，并根据对这三类资源需求量的不同，设计了eMBB，URLLC和mMTC三大类应用场景的网络切片任务。其次，定义了基于网络切片的资源分配模型，以最大化系统收益为目标，专门针对高可靠低时延网络设计了切片任务，以保证此类任务能够优先处理。最后，针对建立的优化问题，利用强化学习的方法，将网络切片问题的已知条件转化为强化学习的状态空间、动作空间和回报函数，通过不断训练，逐步优化模型，在兼顾高可靠低时延通信任务的情况下，最大化系统收益，从而提高了资源利用率。

# 网络性能评价与算法仿真

围绕着高可靠低时延网络通信的性能评价与资源分配问题，分别从高可靠低时延网络的时延和可靠性两方面对系统进行性能分析与评价，利用强化学习方法对基于网络切片的资源分配优化问题进行求解。在时延分析方面，联合考虑了数据到达速率、服务速率以及违背概率对时延的影响，在可靠性分析方面，研究了误块率和违背概率的关系，并分析了可靠性与时延的相互关系。针对资源分配问题，分配采用先来先服务、随机分配和强化学习等策略模拟了基于网络切片的资源分配场景，并对三种策略进行了比较，证明了强化学习方法在资源分配优化问题中的优越性。

## 5.1 随机网络演算性能分析

在这一节，根据第三章得到的随机时延边界进行数值上的分析，得到影响URLLC网络时延的参数。尽管3GPP尚未制定完成URLLC通信具体的实现标准，但仍然可以通过随机网络演算理论，对系统的时延做数值上的分析。假设URLLC通信网络是采用5G独立组网模式，用户端的数据到达服从泊松分布。同时，参考3GPP中对URLLC通信的性能指标作为时延分析的目标和依据，在3GPP的TR 38.913标准中规定，URLLC通信的用户平面要能够在传输32字节小数据包的情况下，将时延控制在1毫秒以内，同时保证误块率要低于1\*10-5。具体的参数设置见表5.1。

表5.1 性能评价参数设置

|  |  |
| --- | --- |
| **参数** | **数值** |
| 网络部署方式 | SA独立部署 |
| 流量模型 | 泊松分布 |
| 随机到达速率λ | 20(Gbit/s) |
| 随机服务速率C | 45,46,47,48(Gbit/s) |
| 服务边界参数a | 1 |
| 服务边界参数b | 3 |
| 串联服务节点数量 | 3,5 |
| 误块率BLER | 1\*10-5 |
| 目标时延 | 1(ms) |

以随机时延边界为依据，利用在第三章所得结论模拟了系统延迟与系统服务速率之间的关系。

图5.1显示了当违背概率在1\*10-5附近时，数据到达速率在20Gbit/s情况下，系统服务速率对系统时延的影响。在图5.1中，违背概率的值设定为从0.5\*10-5到1.5\*10-5, 在这个相对小的概率范围内来研究系统服务速率对时延的影响。为了使系统的时延能够满足3GPP标准的要求，也就是用户平面时延在1毫秒范围内，将系统服务速率设置在48Gbit/s才能让时延接近标准的要求。从图2中还可以看出，随着横坐标违背概率的增加，在固定系统服务速率下，系统时延会缓慢的降低，这验证了多路复用增益的效果，但同时也说明了这种效果并不明显，通过增大违背概率，并不能有效的降低系统时延。

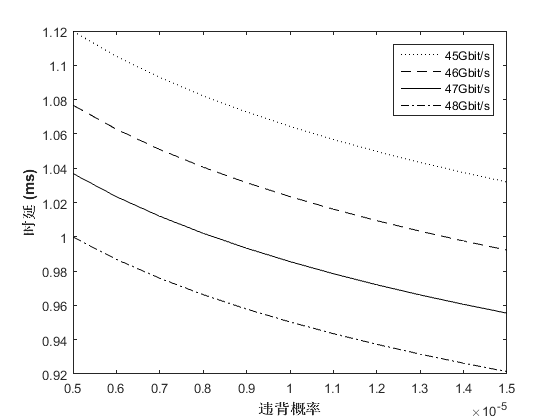


图5.1 系统服务速率对时延的影响

从图5.1中可以看出，随着服务速率的增加，系统时延在降低。在违背概率为1.5\*10-5时，系统服务速率大于46Gbit/s就可以使系统时延低于1毫秒。系统的服务速率从45Gbit/s增加到48Gbit/s才使得系统时延稳定的低于1毫秒，从而达到了3GPP标准的要求。

为了分析URLLC的可靠性，采用随机网络演算理论研究了误块率的随机边界。图5.2显示了在特定的违背概率范围内误块率的随机边界，根据3GPP中对可靠性的要求，设置了三个不同的误块率边界参数，从图3中可以看出误块率会随着违背概率的增加而降低。当系统的边界参数设置为0.00001时，系统的误块率也可以保证在1.5\*10-5以内，这也就意味着99.999%的传输正确性，也就是满足了可靠性要求的情况。

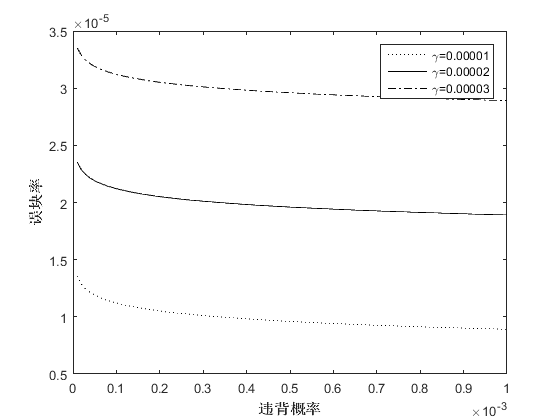


图5.2 误块率随机边界

基于URLLC的可靠性需求，研究了误块率对系统时延的影响。从图5.3可以看出，误块率对系统时延的影响并不明显。 当误块率的边界值从1\*10-5增加到 5\*10-2时，系统时延的变化很小。可靠性要求误块率不超过1\*10-5，如此小的值很难对系统服务速率产生大的影响，系统服务速率相对稳定，从而系统时延也相对稳定，因此系统的可靠性得到了保证。

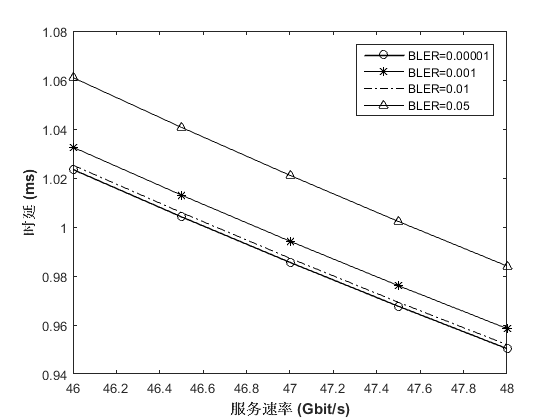


图5.3 误块率对时延的影响

图5.4阐释了系统时延和系统数据到达速率的关系。在图5.4中，系统时延随着数据到达率的增加而增加。URLLC的标准要求当系统的数据到达速率为20Gbit/s时，用户平面的系统时延要在1毫秒以内。随着系统服务速率的增加，系统时延会随之降低，而且会低于1毫秒。但是当数据到达速率也增加的时候，系统时延又会超出标准的要求。通过定性分析可以可知，系统的时延和服务速率成反比，与数据的到达速率成正比。由于系统资源的是有限的，不可能无限制的增加，所以系统的服务速率是有范围的，为了降低时延，应该制定资源分配策略，提高资源利用率，尽可能的提高系统服务速率，同时也应该安排好任务调度策略，降低数据到达速率的峰值，从而有效的降低系统时延。

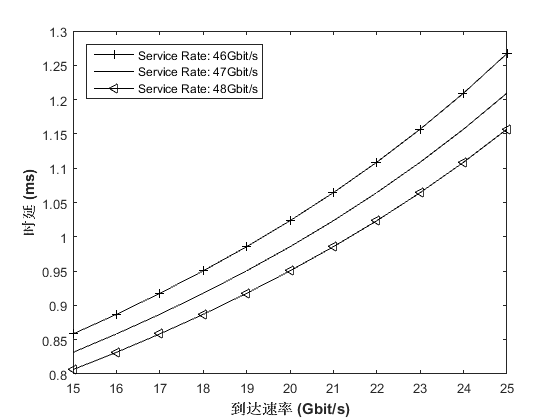


图5.4 时延与数据到达速率的关系

图5.5展示了系统时延、服务速率与串联服务节点数量之间的关系。图中将数据的到达速率恒定的设为20Gbit/s，时延的违背概率设置为1\*10-5，在此基础上可以得出，系统时延对串联服务节点的数量较为敏感。串联服务节点的数量多，说明网路部署中数据经过的服务器数量较多，路由次数多，串联服务节点的数量少则说明网络的路由次数较少，数据传输路径较短，要访问的目的地址靠近用户侧。为了保证低延迟的通信，有必要减少部署中串联服务器的数量。边缘计算通过将计算资源部署在靠近用户侧，减少了网络中的串联节点数量，从而有效降低了系统时延，这是通过改变部署方式来降低时延的一个典型例子。

从图5.5可以看出，由串联服务节点数量引起的时延斜率大于系统服务速率,图中灰色的平面是延时为1毫秒的平面，用这个平面来切割曲面，平面下方为满足系统时延要求的部署参数范围。这意味着当串联服务节点数量少于5台，服务速率超过40Gbit /s时，系统时延可以保持在1毫秒以下。增加系统的服务速率是减少延迟的另一种方法。当数据到达速率为20 Gbit/s时，服务速率应设置在40 Gbit/s以上。这样才可以保证URLLC的延迟要求。

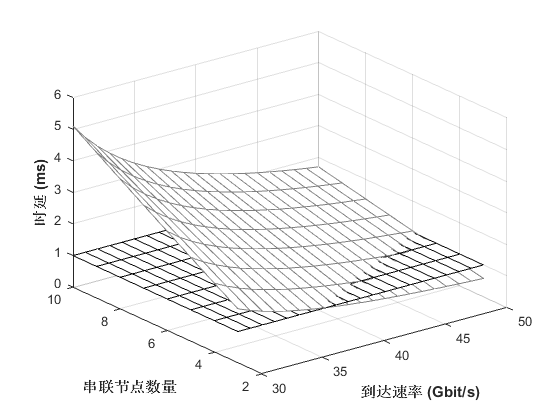


图5.5 串联服务节点数量对时延的影响

针对URLLC的性能要求做了一系列仿真实验，在延迟方面，URLLC要求用户平面传输时延在1毫秒以内。在可靠性方面，URLLC应该确保误块率为1\*10-5。围绕着这些性能指标，通过实验揭示了如何满足延迟和可靠性的要求。为了将网络的数据到达速率提高到20Gbit/s，则需要满足特定的网络建设条件，这些条件包括系统服务速率的范围，以及串联网络节点的数量。系统服务节点的服务速率应该提供40Gbit/s以上的处理速度，部署的串联服务节点应该尽可能的少，并且宜靠近用户侧，以减少路由次数。

## 5.2 时延分析器

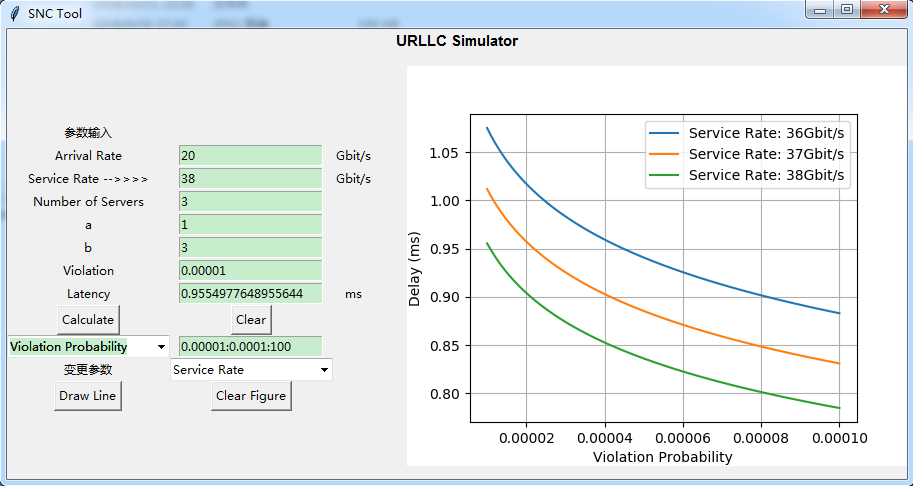
在3.1.2节得到的关于时延边界的结论基础上，采用Python语言结合Matplotlib函数功能库，实现了一款简洁实用的时延分析器。如图5.6所示。

图5.6 网络时延分析器

本时延分析器支持多种参数的输入，可以根据数据到达速率、系统服务速率、串联服务节点数量、服务边界参数和违背概率等输入变量来计算系统的时延。当输入框中添写了所需的参数之后，点击Calculate按钮，就可以在Latency对应的文本框中得到计算出的时延数据。

时延分析器还支持画图功能，通过下拉菜单选定序列化变量的名称，在输入框中添入序列化变量的起止范围以及变量数值变化的步长，就可以画出以选定的序列化变量为横坐标的时延变化曲线，该输入框中序列化输入参数的格式为linspace函数的输入格式。由于时延是本研究最关系的数据，所以图中的纵坐标都设定为时延。

时延分析器支持在横坐标变量固定的情况下，以追加模式画出选定参数不同数值对时延影响的曲线，这种曲线图可以直观的比较不同数据下时延的变化趋势。通过在变更参数的下拉菜单中选定一个变量作为要比较的参数，输入一次数值，点击Draw Line按钮，即可画出一条曲线，重复多次操作，就可以得到趋势变化图。当不再需要当前画图或者想要画新的曲线图的时候，可以点击Clear Figure清空当前画图。有了时延分析器可以很方便的计算时延，并得到不同参数对时延的影响，以及获得参数变化对时延影响的趋势，为时延分析提供便捷。

## 5.3 基于网络切片的资源分配算法仿真

在网络切片资源分配模型中，将5G通信网络资源抽象为三种：带宽资源、计算资源和存储资源，通过对这三种资源设置不同的需求量来区分不同的通信场景。其中，eMBB类型的网络切片任务需要较高的带宽资源和存储资源，时延要求适中；mMTC类型的网络切片任务对这三种资源的需求都不大，时延要求也不高，但是任务的数量较多；URLLC类型的网络切片任务对存储资源要求不高，但是对带宽资源和计算资源需求较高，对时延的要求也很高。

基于三种类型的特点生成了3组切片任务数据，任务类型的比率为URLLC:eMBB:mMTC=1:2:7，三组切片任务数量分别为20,40,80。如图5.7所示。

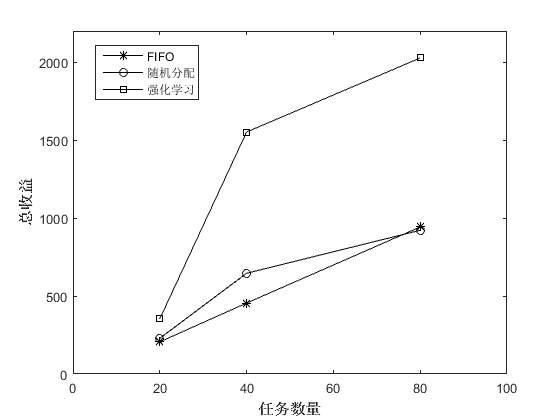


图5.7 网络切片策略总收益比较

比较了两种常见的方法，先来先服务(FIFO)和随机分配，通过获得的实验数据可知，采用强化学习的方法可以稳定的优于另外两种方法。采用随机的分配策略在任务数为20和40的情况下，都可以优于先来先服务，但是随机策略并不稳定，在任务数为80的情况下，其策略的总收益略差于先来先服务。

采用强化学习方法的分配策略其总收益并没有按任务数量的比例增长，这是由新增的任务的规则决定的。新增任务的生成要符合三种通信类型的比例关系，即URLLC:eMBB:mMTC=1:2:7，但任务对系统资源的占用参数是随机生成的，所以采用强化学习方法的分配策略其总收益的增长是合理的。

在分析强化学习方法训练出来的优化策略时，可以看到在获得总收益较高的策略中，多数情况都对URLLC类型的切片任务进行了分配的动作，这是符合实验的预期的，因为在设置回报函数的时候就将此类型的任务收益设置的高于其它两种类型，保证了高可靠低时延通信能够优先获得服务。

## 5.4 本章小结

本章利用第三章对高可靠低时延网络通信的性能分析结果，分别从时延和可靠性两方面进行了数值分析。将3GPP中对时延和可靠性要求的标准数据代入分析模型，得到了时延与数据到达速率、服务速率以及串联服务节点数量之间的关系，同时也分析了时延和可靠性之间的关系。通过性能分析，可以得出为减少时延而需要优化的方向，为网络部署提供有价值的指导意见。在此基础上，简单介绍了基于Python语言开发的时延分析器，能够更方便的计算时延与其他网络参数的相互关系。

在基于网络切片的资源分配策略研究中，分别采用先来先服务、随机分配和强化学习等策略对资源分配模型进行实验，通过仿真得出，采用了强化学习方法的资源分配策略能够在保证URLLC任务优先处理的情况下使系统资源得到高效利用。

# 结论与展望

## 6.1 结论

5G技术的快速发展和紧锣密鼓的商用步伐，使得5G网络正加紧部署。在车联网、无人机控制等技术日趋火热的情况下，高可靠低时延通信越来越受到业内的关注。目前，3GPP组织完成了对URLLC通信性能要求的标准制定，但对于网络部署的相关实施技术标准尚未制定完成。这一现状就催生出两个问题，如何满足URLLC的性能要求？如何对资源进行分配，来保证URLLC通信的顺利进行？本文围绕高可靠低时延通信的性能评价与资源分配问题进行研究，给出降低时延的网络部署意见，为资源分配提供优化策略。

首先，本文研究了5G网络的架构，根据架构特点利用随机网络演算理论建立了串联系统模型，用于分析时延和可靠性的性能边界。得出了时延与到达速率、服务速率、违背概率以及串联节点数量之间的关系。通过对3GPP标准的研究，将标准中数值参数带入模型，得出了满足时延和可靠性要求的网络参数。

其次，针对网络资源分配问题，以保证高可靠低时延通信优先服务，最大化系统收益为目标，构建了基于网络切片的资源分配优化问题模型。由于该优化问题的状态变化具有马尔可夫性，采用强化学习方法来解决切片任务分配的序贯决策问题。

最后，在网络性能分析方面，通过仿真得出了减少时延的优化方向。验证了边缘部署方式可以减少系统中的串联服务节点，从而降低时延。在网络切片的资源分配仿真实验中，分别采用先来先服务、随机分配和强化学习等策略对资源分配进行模拟，验证了采用强化学习方法分配策略的优越性。

## 6.2 展望

本文在5G高可靠低时延网络的性能分析研究上取得了一定的学术成果，但是仍存在很多需要深入研究的问题。在本文的研究基础上，未来计划进一步优化的问题主要有：

1) 可靠性分析

本在分析可靠性的时候通过随机差错过程来表达误块率，进而实现可靠性的分析。然而这种分析是基于一定违背概率假设的，在后面的工作中，希望能利用实际的网络测试仪器进行测量，提供更准确更直观的可靠性分析。

(2) 资源分配策略优化

在基于强化学习的资源分配策略中，通过训练得到的策略能够保证较优的系统总收益，但是并没有稳定的收敛到最优解。这需要后期不断完善模型的设计思路和调整训练参数，以达到稳定收敛的效果。

# 致谢

时光流逝，当初心心念念的研究生生活就要结束了。在计算机学院学习的三年，让我对计算机领域的了解从工程的角度转向到了学术的高度，我感到受益良多。在即将离开校园的时刻，我想向我的老师和同学表示感谢，感谢你们对我的培养和帮助，感谢你们对我的关心和陪伴。

首先，我要感谢的是我的导师陈昕教授。陈老师治学严谨，严格自律，无论多么忙，都会及时指导我们，甚至是占用了他自己的休息时间。在学术上，陈老师的视野开阔，格局很高，能准确的捕捉到前沿热门领域，尽全力为我们提供更高的科研平台。在工程上，陈老师与科研院所合作，让我们能够有足够的实践空间。在生活上，陈老师关心同学，经常组织团队活动，与往届的学长一起联谊，学习师哥师姐们的经验，多次组织体育运动，让同学们不但要增长知识，还要有健康的体魄，为团队营造了劳逸结合的氛围。我的成长与收获，都来自于陈老师言传身教和关心帮助。在此向我的导师表示真诚的感谢！

然后，非常感谢团队里的李卓老师，李老师对我的帮助非常具体。每当论文没有进展没有思路的时候，李老师都会给予指引和建议，让我瞬间醍醐灌顶，豁然开朗。还记得在南京参加会议的时候，李老师为了参加我的论文汇报，下了课立刻赶向机场，当我在会场上见到李老师的时候，瞬间就充满了自信。我的进步离不开李老师的鼓励和肯定，在此对李老师表示诚挚的谢意。

其次，感谢团队里的陈莹老师，陈莹老师卓越的学术素养和严格的自律态度时刻鼓舞鞭策着我，在论文的撰写上，感谢陈莹老师给与的意见和帮助。同时也要感谢同门的赵禹哲、王祥坤、唐超、张永超和杨英杰，在平时的讨论中，你们的思路让我茅塞顿开，受益匪浅，在你们的帮助下，让我少走了很多弯路，这段一起学习的经历让我难忘，是我人生中的宝贵财富。

最后，感谢百忙之中抽出时间对我论文进行评审的各位专家学者表达我衷心的感谢！

# 参考文献

1. ITU-R M.2083-0. IMT Vision - Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond [S]. 2015
2. 3GPP. TR 38.913. Study on Scenarios and Requirements for Next Generation Access Technologies[S]. 2017
3. Soldani D, Guo Y J, Barani B, et al. 5G for Ultra-Reliable Low-Latency Communications [J]. IEEE Network, 2018, 32(2):6-7
4. Jiang Y, Liu Y. Stochastic Network Calculus [M]. London, Springer, 2009
5. Fidler M, Rizk A. A Guide to the Stochastic Network Calculus [J]. IEEE Communications Surveys & Tutorials, 2015, 17(1):92-105
6. 纳多. 软件定义网络[M]. 人民邮电出版社, 2014
7. 张笛, 李兴华, 刘海, 马建峰. SDN网络中面向服务的网络节点重要性排序方法[J]. 计算机学报, 2018, 41(11):206-218
8. 王进文,张晓丽,李琦,等.网络功能虚拟化技术研究进展[J].计算机学报,2019, 42(2):185-206
9. Chen Jing, He Kun, Yuan Quan, et al. Batch Identification Game Model for Invalid Signatures in Wireless Mobile Networks [J]. IEEE. Trans. Mobile Comp, 2017, 16(6):1530–1543
10. Chen M, et al. A 5G Cognitive System for Healthcare, Big Data and Cognitive Computing [J]. 2017, 1(1)
11. Retal S, Bagaa M, Taleb T, et al. Content delivery network slicing: QoE and cost awareness[C]. IEEE International Conference on Communications. IEEE, 2017
12. 廖晓闽,严少虎,石嘉,等.基于深度强化学习的蜂窝网资源分配算法[J]. 通信学报,2019, 40(2)
13. 郭宪,方勇纯. 深入浅出强化学习原理入门[M],电子工业出版社,2018
14. Silver D , Huang A , Maddison C J , et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587):484-489
15. 刘全,翟建伟,章宗长.深度强化学习综述[J].计算机学报,2018,1:1-27
16. Nielsen J.J, Liu R and Popovski P. Ultra-Reliable Low Latency Communication Using Interface Diversity [J]. IEEE Transactions on Communications, 2018, 66(3):1322-1334
17. Ramón A. Delgado, Lau K, Middleton R H, et al. Networked Delay Control for 5G Wireless Machine-Type Communications Using Multiconnectivity [J]. IEEE Transactions on Control Systems Technology, 2018, PP(99):1-16
18. Rao J, Vrzic S. Packet Duplication for URLLC in 5G: Architectural Enhancements and Performance Analysis [J]. IEEE Network, 2018, 32(2):32-40
19. Pocovi G, Shariatmadari H, Berardinelli G, et al. Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements [J]. IEEE Network, 2018, 32(2):8-15
20. Popovski P, Nielsen J J, Stefanovic C, et al. Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks [J]. IEEE Network, 2018, 32(2):16-23
21. Ji H , Park S , Yeo J , et al. Introduction to Ultra Reliable and Low Latency Communications in 5G[J]. 2017
22. Hyoungju J, Sunho P, Jeongho Y, et al. Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects [J]. IEEE Wireless Communications, 2018, 25(3):124-130
23. Sachs J , Wikstrom G , Dudda T , et al. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication[J]. IEEE Network, 2018, 32(2):24-31
24. Beck M A, Henningsen S, Xu Q, et al. Demo abstract: An integrated tool of applying stochastic network calculus for network performance analysis[C] Computer Communications Workshops. IEEE, 2017
25. Li. Z, Gao. Y, Salihu. B. A, et al. Network Calculus Delay Bounds in Multi-Server Queueing Networks with Stochastic Arrivals and Stochastic Services[C]. 2015 IEEE Global Communications Conference (GLOBECOM), 2015.1-7
26. Li Z, Jiang Y, Gao Y, et al. Delay and Delay-Constrained Throughput Performance of a Wireless-Powered Communication System [J]. IEEE Access, 2017, 5:21620-21631
27. Ralf Lübben, Fidler M. Estimation method for the delay performance of closed-loop flow control with application to TCP[C]. IEEE Infocom -the IEEE International Conference on Computer Communications. IEEE, 2016
28. Zheng K, Liu F, Lei L, et al. Stochastic Performance Analysis of a Wireless Finite-State Markov Channel [J]. IEEE Transactions on Wireless Communications, 2013, 12(2):782-793
29. Chen X, Si Y, Xiang X. Delay-bounded resource allocation for femtocells exploiting the statistical multiplexing gain [J]. The Journal of Supercomputing, 2015, 71(9):3217-3236
30. M.Fidler, B.Walker and Jiang. Y. Non-Asymptotic Delay Bounds for Multi-Server Systems with Synchronization Constraints [J]. IEEE Transactions on Parallel and Distributed Systems, 2018, 29(7):1545-1559
31. Gulyás, András, Bíró, József. A stochastic extension of network calculus for workload loss examinations [J]. Communications Letters IEEE, 2006, 10(5):399-401
32. Deng Y, Lin C. An Extended Stochastic Loss Bound with Moment Generating Function[C]. International Conference on Communications \& Mobile Computing IEEE, 2010
33. Sami Ayyorgun, et al. A Composable Service Model with Loss and a Scheduling Algorithm[C]. Infocom, Hong Kong. 2004
34. Wang C, Chen Y, Wu Y, et al. Performance Evaluation of Grant-Free Transmission for Uplink URLLC Services[C]. IEEE Vehicular Technology Conference, 2017.1-6
35. Anand A, De Veciana G. Resource Allocation and HARQ Optimization for URLLC Traffic in 5G Wireless Networks [J]. IEEE Journal on Selected Areas in Communications, 2018:1-1
36. Mukherjee A, Energy Efficiency and Delay in 5G Ultra-Reliable Low-Latency Communications System Architectures [J]. IEEE Network, 2018, 32(2):55--61
37. Vassilaras S, Gkatzikis L, Liakopoulos N, et al. The Algorithmic Aspects of Network Slicing [J]. IEEE Communications Magazine, 2017, 55(8):112-119
38. Sameerkumar Sharma. Raymond Miller. Andrea Francini. A Cloud-Native Approach to 5G Network Slicing[J], IEEE Communications Magazine, 2017, 55(8):120-127
39. Li Xi, Ramon Casellas, Giada Landi, et al. 5G-Crosshaul Network Slicing: Enabling Multi-Tenancy in Mobile Transport Networks [J] IEEE Communications Magazine, 2017, 55(8):128-137
40. Zhang Haijun, Liu Na, Chu Xiaoli, et al. Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges [J]. IEEE Communications Magazine, 2017, 55(8):138-145
41. Katsalis K, Nikaein N, Schiller E, et al. Network Slices toward 5G Communications: Slicing the LTE Network [J]. IEEE Communications Magazine, 2017, 55(8):146-154
42. Lanfranco Zanzi, Vincenzo Sciancalepore, Andres Garcia-Saavedra, et al. OVNES: Demonstrating 5G network slicing overbooking on real deployments. IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) [C]. 2018. 1-2
43. Dario Bega, Marco Gramaglia, Albert Banchs, et al. Optimising 5G infrastructure markets: The business of network slicing. IEEE INFOCOM 2017 - IEEE Conference on Computer Communications[C]. 2017. 1-9
44. Ying H, Nan Z, Yin H. Integrated Networking, Caching, and Computing for Connected Vehicles: A Deep Reinforcement Learning Approach [J]. IEEE Transactions on Vehicular Technology, 2018. 67(1):44-55
45. Ying H, Zheng Z, Yu F R, et al. Deep Reinforcement Learning-based Optimization for Cache-enabled Opportunistic Interference Alignment Wireless Networks [J]. IEEE Transactions on Vehicular Technology, 2017, 66(11):10433-10445
46. Hao Y , Tian D , Fortino G , et al. Network Slicing Technology in a 5G Wearable Network[J]. IEEE Communications Standards Magazine, 2018. 2(1):66-71
47. Yang Tianyu, Hu Yulin, Cenk Gursoy. M, et al. Deep Reinforcement Learning based Resource Allocation in Low Latency Edge Computing Networks[C]. 2018 15th International Symposium on Wireless Communication Systems (ISWCS), 2018. 1-5
48. Muhammad Rehan Raza, Matteo Fiorani, Ahmad Rostami, et al. Dynamic slicing approach for multi-tenant 5G transport networks [J]. IEEE/OSA Journal of Optical Communications and Networking, 2018, 10(1):77-90
49. 3GPP. Final Report of 3GPP TSG RAN WG1. Hangzhou, China[S]. 2017
50. 3GPP. TR 21.915 v0.6.0. Technical Specification Group Services and System Aspects, Release 15 Description[S]. 2019
51. 3GPP. TR 23.799 v14.0.0. Technical Specification Group Services and System Aspects, Study on Architecture for Next Generation System, Release 14[S]. 2016
52. 杨旭, 肖子玉, 邵永平, 等. 5G网络部署模式选择及演进策略[J]. 电信科学, 2018, 34(6):144-152
53. Yang G, Xiao M, Gross J, et al. Delay and Backlog Analysis for 60 GHz Wireless Networks [C]. 2016 IEEE Global Communications Conference (GLOBECOM). 2016
54. 吴璇, 朱晓荣. 基于随机网络演算的物联网业务流量建模与分析方法[J]. 南京邮电大学学报(自然科学版), 2018, 38(175):26-32
55. 3GPP. TS 22.261 16.6.0 Technical Specification Group Services and System Aspects; Service requirements for the 5G system[S]. 2018
56. 3GPP. TR 22.891.Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers[S]. 2016
57. Rost P, Breitbach M, Roreger H, et al. Customized Industrial Networks: Network Slicing Trial at Hamburg Seaport [J]. IEEE Wireless Communications, 2018,25(5):48-55
58. 3GPP. TR 28.530.V0.2.0. Management of Network Slicing in Mobile Networks; Concepts, Use Cases and Requirements[S].2017
59. Rost P, Breitbach M, Roreger H, et al. Customized Industrial Networks: Network Slicing Trial at Hamburg Seaport [J]. IEEE Wireless Communications, 2018, 25(5):48-55
60. 任语铮,谢人超,黄韬.基于增强学习的5G网络切片资源动态优化方案[J]. 中兴通讯技术,2018,24(1):31-36
61. 3GPP. TR 38.801 V14.0.0. Technical Specification Group Radio Access Network, Study on new radio access technology: Radio access architecture and interfaces, Release 14[S]. 2017
62. Wu K, Jiang Y, Li J. On the model transform in stochastic network calculus[C]. In: International Workshop on Quality of Service IEEE.2010. 1-9
63. Sun F, Li L, Jiang Y. Impact of duty cycle on end-to-end performance in a Wireless Sensor Network[C]. Wireless Communications and NETWORKING Conference IEEE, 2015.1906-1911
64. Beck M. Towards the analysis of transient phases with stochastic network calculus[C]. Telecommunications Network Strategy and Planning Symposium IEEE, 2016. 164-169
65. Li R , Zhao Z , Zhou X , et al. Intelligent 5G: When Cellular Networks Meet Artificial Intelligence[J]. IEEE Wireless Communications, 2017:2-10

# 附录A 图目录

[图2.1 随机网络演算中时延的定义 15](#_Toc6468812)

[图2.2 网络切片示意图 16](#_Toc6468813)

[图3.1 5G网络架构 20](#_Toc6468814)

[图5.1 系统服务速率对时延的影响 35](#_Toc6468815)

[图5.2 误块率随机边界 36](#_Toc6468816)

[图5.3 误块率对时延的影响 36](#_Toc6468817)

[图5.4 时延与数据到达速率的关系 37](#_Toc6468818)

[图5.5 串联服务节点数量对时延的影响 38](#_Toc6468819)

[图5.6 网络时延分析器 39](#_Toc6468820)

[图5.7 网络切片策略总收益比较 40](#_Toc6468821)

# 附录B 表目录

[表2.1 非独立组网与独立组网比较 11](#_Toc6435463)

[表5.1 性能评价参数设置 34](#_Toc6435464)

# 个人简历、在学期间发表的学术论文及研究成果

**个人简历：**

马圣程，男，1986年5月生，辽宁人。

2003.09 -- 2007.06，北京信息工程学院，获学士学位。

2007.07 -- 2010.08，北京中科网威信息技术有限公司，研发工程师。

2010.09 -- 2014.06，北京天融信软件公司，研发工程师。

2014.06 -- 2016.09，北京二六三企业通信有限公司，研发工程师。

2016.09 -- 至今，北京信息科技大学计算机学院计算机技术专业学习，攻读硕士学位。

**参与科研项目：**

1. 2019.01—2022.12，国家自然科学基金面上项目“5G 超密集接入网智能动态资源分及 其优化方法研究”（编号：61872044）
2. 2014.01—2017.12，国家自然科学基金面上项目“LTE-A飞蜂窝系统的动态资源分配与性能评价研究”（编号：61370065）
3. 2015.01—2017.12“十二五”国家科技支撑计划项目“城镇化进程中社会安全风险评估与应急处置技术研究与示范” 子课题“城镇社会安全多源大数据处理关键技术”（编号：2015BAK12B03-03）

**发表论文：**

1. **Shengcheng Ma**, Xin Chen, Zhuo Li, YingJie Yang. A Retrieval Optimized Surveillance Video Storage System for Campus Application Scenarios [J]. Journal of Electrical & Computer Engineering, 2018(8):1-10. (EI: 20181805117616)
2. **Shengcheng Ma**, Xin Chen, Zhuo Li, Ying Chen. Delay Analysis for URLLC in 5G Based on Stochastic Network Calculus[C]. 2nd International Conference on 5G for Ubiquitous Connectivity. Nanjing, China. 2018. (Best Paper Reward)

**专利：**

监控视频的存储方法和系统（申请号：201810128720.4）

**获奖情况：**

2018.11 获评“校级三等奖学金”

2017.11 获评“校级一等奖学金”