

学校代号 10532

学 号 B141000244

分 类 号 TP391

密 级 普 通



湖南大学
HUNAN UNIVERSITY

博士学位论文

仲裁网络结构的 CPS 通信模型 优化与性能分析

学位申请人姓名 龚红仿

培 养 单 位 信息科学与工程学院

导师姓名及职称 李仁发 教授

学 科 专 业 计算机科学与技术

研 究 方 向 嵌入式计算系统

论文提交日期 2018 年 9 月 7 日

学校代号：10532

学 号：B141000244

密 级：普通

湖南大学博士学位论文

仲裁网络结构的 CPS 通信模型 优化与性能分析

学位申请人姓名：_____ 龚红仿 _____

导师姓名及职称：_____ 李仁发 教授 _____

培 养 单 位：_____ 信息科学与工程学院 _____

专 业 名 称：_____ 计算机科学与技术 _____

论文提交日期：_____ 2018 年 9 月 7 日 _____

论文答辩日期：_____ 2018 年 9 月 28 日 _____

答辩委员会主席：_____ 王志英 教授 _____

Communication Model Optimization and Performance Analysis of
Cyber-Physical Systems for Arbitration Network Structure

by

GONG Hongfang

B.S. (Changsha University of Electric Power) 1991

M.S. (Hunan University) 2004

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Engineering

in

Computer Science and Technology

in the

College of Computer Science and Electronic Engineering

of

Hunan University

Supervisor

Professor LI Renfa

September, 2018

湖南大学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权湖南大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

1、保密 ，在_____年解密后适用本授权书。

2、不保密 。

(请在以上相应方框内打“√”)

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

摘 要

信息物理系统(CPS)是由计算算法和物理组件的无缝集成所构建并依赖于这种集成的工程系统,是计算、通信、控制技术的有机与深度融合的系统。CPS的紧耦合性、资源受限、自适应性和异构性使得系统设计非常复杂。CPS描述了控制和通信方面的集成研究需求。虽然CPS在我们的日常生活中无处不在,但它们的设计和开发既困难又昂贵,而且没有合理的理论基础。传统上,控制器的设计及其稳定性和性能分析是独立于通信和计算架构的设计和分析而完成的,这反过来又加剧了所得到的CPS的验证。一方面,系统和控制领域专注于设计和开发诸如状态空间模型、系统辨识、过滤和鲁棒控制之类的方法,以便分析物理动机但抽象的实际过程模型。另一方面,计算机科学领域开发了用于实时计算、编程语言和编译器、嵌入式系统和信息安全的强大技术。这些方法使用非常强大的建模形式和验证工具。虽然这两个领域都有悠久的历史,并且几乎在所有技术设备中都发挥着重要作用,但没有共同的理论基础。因此,面向CPS的设计,需要通过整体CPS设计的理论和方法协同解决这两个领域的挑战。

汽车、列车、航空控制系统、工业自动化系统常常是分布式嵌入式系统,有大量的处理单元构成多个分布式控制应用,常常受到多个控制质量的约束。这种分布式特性导致CPS的设计需要考虑通信与控制协同设计问题。最近,研究人员设计了一类名为“仲裁网络控制系统(ANCS)”的单信道特殊的CPS体系结构,以强调所设计的控制系统中的网络调度(或仲裁),其显著特征是共享通信介质的消息仲裁机制。称这一ANCS体系结构为单信道ANCS系统,也称为单信道“仲裁网络结构”,用于CPS的整体建模。这一仲裁网络结构中的多个控制应用共享通信总线构成多个通信模型。本文研究仲裁网络结构中的CPS通信模型(或控制应用)优化问题,关注CPS中的网络参数与控制器参数优化。用排队系统建模ANCS,以改善系统性能、降低系统成本为目的,分别研究了ANCS中的多控制器调度、资源动态供应以及不确定的消息响应时间(MRT)等问题。主要的工作及创新点如下:

(1) 提出了一种新的双信道ANCS平台体系结构,基于分层建模理论和通信系统中的唤醒机制,建立了一种扁平半休眠多控制器(FSDMC)模型。在轻量级通信情况下,考虑系统成本最小和性能最优,允许FSDMC模型中的部分空闲控制器进入半休眠状态,以维护负载均衡,保持全局一致性。

(2) 针对FSDMC模型中控制器簇的调度问题,构建了一种新的N策略、部分服务台同步多重工作休假的有限容量的M/M/c排队系统(N/(d, c)-M/M/c/K/SMWV),建模FSDMC模型为一种拟生灭(QBD)过程,获得系统各种性能度量指标,构建系统成本和性能函数。基于这一排队模型,提出了一种多目标优化模型,以最小化非线性系统能耗函数和标称时延函数,并设计了一种基于多目标粒子群优化(MOPSO)算法的调度算法,求解该多目标优化问题,生成Pareto前沿和相应的非支配向量集。用最优停止算法获得半休眠控制器数的最优值,结合非支配向量集,生成了控制系统的各种参数的最优值,

配置 ANCS 系统。用基于事件触发协议的分层柔性时分多路复用/固定优先级(FTDMA/FP)总线调度策略配置 ANCS 中的总线。大量的数值实验表明了算法的有用性和最优策略下的控制系统设计效果。

(3) 针对资源约束下的资源动态供应问题,提出了滞弹性系统中的滞弹性计算模型和计算方法,给出了 FSDMC 模型中的滞弹性量化、正式、新的定义。首先建模 FSDMC 为一种新的 N 策略、部分服务台异步多重工作休假的有限容量的 M/M/c 排队系统(N/(d, c)-M/M/c/K/AMWV)。在这一排队模型的基础上,提出了一种最优化模型,最小化成本性能率(CPR)函数,获得系统在平稳状态下的各种参数的最优值。考虑不可预料的负载随时间的变化而随机改变,构建一种包含资源过度供应状态、正常状态和不足供应状态的 3-状态不可约马尔科夫过程,表示随机环境下的 M/M/1/K 队列。在这一队列中,分别生成了正常状态的条件平均队列长度和 3-状态概率,获得系统滞弹性精确值。当队列的平均到达率超过平均服务率时,设计了一种基于 PSO 的最优 CPR 不变自适应算法,动态调整控制器服务率。通过分析系统的高峰期行为,得出系统自适应调节参数能基本保持滞弹性不变。大量的数值实验表明提出的技术的有用性和有效性。

(4) 针对不确定时延问题,计算得到单信道 ANCS 中的不确定 MRT 的概率密度函数(pdf)和累积分布函数(cdf)的闭式表达式,并提出一种新的、命名为两参数概率变换灰色预测模型(PTGM(1,1)),预测每个消息的响应时间。首先,用一种基于事件触发协议的多层次柔性 TDMA/FP 调度策略配置 ANCS 的总线,再建立最优化模型获得总线的最优参数值。建立一种具有随机大小的批量到达 M/PH/1 队列模型,以获得每个控制应用的平均顾客数。基于这一排队模型,设计了一种约束优化问题,最小化 ANCS 的非线性成本函数,并生成每一消息的不确定响应时延的 pdf。基于 DFP 变尺度法,提出了一种最优 minislot 算法和 pdf 的参数估计算法,以分别获得总线 minislot 的最优值和相位型(PH)分布的矩阵参数最优值。设计了一种基于 DFP 方法的 PTGM(1,1)的两参数优化算法,最小化平均绝对百分误差,得到每一消息的两参数最优值,得到每个消息的响应时间的预测值。大量的数值实验给出了本文算法的有用性,表明了预测结果的精确性。

关键词: 信息物理系统; 仲裁网络结构; 通信模型优化; 资源动态供应; 不确定时延分析; 性能分析; 排队论

Abstract

A cyber-physical system (CPS) is an engineering system that is constructed by the seamless integration of computational algorithms and physical components and relies on this integration. It is an organic and deep integration system of computing, communication, and control technologies. The tight coupling, resource constraints, adaptability and heterogeneity of CPS lead to very complex system design. CPS describes the need for integrated research in control and communication. Although CPS are ubiquitous in our everyday life, their design and development are difficult, expensive, and without a sound theoretical foundation. Traditionally, the design of the controller and its stability and performance analysis is done independently from the design and analysis of the communication and computation architecture which in turn aggravates the verification of the resulting CPS. On the one hand, the systems and control domain has focused on designing and developing methods like state-space models, system-identification, filtering, and robust control in order to analyze physically motivated, yet abstract models of real processes. On the other hand, the computer science domain has developed powerful techniques for real-time computing, programming languages and compilers, embedded systems, and cyber security. These methods use very mighty modeling formalisms and verification tools. Although both areas have a long history and play an essential role in almost all technical devices, there is no common theoretical foundation. Therefore, CPS-oriented design needs to address the challenges of two domains using the theory and approach of holistic CPS design.

Automobiles, trains, avionics, and industrial automation systems are often distributed embedded systems. There are a large number of processing units in the system that form multiple distributed control applications, often subject to multiple quality-of-control (QoC) constraints. This distributed feature leads to the design of CPS that needs to consider communication and control co-design issues. Recently, a special CPS with single communication channel that named as arbitrated networked control system (ANCS) is proposed in order to emphasize that the control systems are to be designed for networks that are scheduled or arbitrated. We call this ANCS architecture a single-channel ANCS system, also known as a single-channel arbitration network structure, for the overall modeling of CPS. Multiple control applications in the arbitration network structure share communication bus to form multiple communication models. In this thesis, we study the optimization problem of the communication models (or control applications) of CPS in the arbitration network structure and focus on network parameters and controller parameters optimization in CPS. We use the queuing system to model the system dynamics in ANCS to improve system performance and reduce system cost. In ANCS, multi-controller scheduling problems, resource dynamic provisioning problems and uncertain message response time (MRT) problems are studied,

repectively.

The main works and innovations are summarized as follows.

(1) A novel ANCS platform architecture with dual communication channels is proposed. Based on the hierarchical modeling theory and the wake-up mechanism in the communication system, a flat semi-dormant multi-controllers (FSDMC) model is developed for the proposed ANCS. In lightweight communication, considering the minimum system cost and optimal performance, some idle controllers in the FSDMC model are allowed to enter a semi-dormant state to maintain load balancing and guarantee global consistency.

(2) Aiming at the scheduling problem of controller clusters in FSDMC model, a novel finite capacity M/M/c queuing system with N-policy and synchronized multiple working vacations of partial servers, N/(d, c)-M/M/c/K/SMWV for short, is proposed, and the FSDMC model is modeled as a quasi-birth-death (QBD) process. We obtain various performance metrics for the system and build system cost and performance functions. Based on the model a multi-objective optimization problem is then formulated to minimize the nonlinear energy consumption function and the nominal delay function presented in this study. To resolve the multi-objective optimization problem, a scheduling algorithm based on the MOPSO algorithm is proposed to generate the Pareto front and the corresponding nondominated vector sets. An optimal stopping algorithm is also designed to obtain the optimal value of the number of semi-dormant controllers. The optimal values of various parameters of the control system are obtained by using the above nondominated vector sets, and are applied to the proposed ANCS. The proposed ANCS uses a hierarchical Flexible TDMA/Fixed Priority (FTDMA/FP) scheduling policy that is based on the event trigger protocol. Extensive numerical results are provided to illustrate the usefulness of the proposed algorithms and the effects of the control system parameters on the optimal policy.

(3) Aiming at the problem of resource dynamic provisioning under resource constraints, an anelasticity computing model and calculation method for anelasticity systems are proposed. A new, quantitative, and formal definition of anelasticity for the FSDMC on the special CPS is proposed. We first model the FSDMC as a novel finite capacity M/M/c queuing system with N-policy and asynchronous multiple working vacations of partial servers, N/(d, c)-M/M/c/K/AMWV for short, and obtain various performance indices. Based on the model, an optimization problem is formulated to minimize the nonlinear cost-performance ratio (CPR) function in this study, and obtain the optimal values of the system parameters in a stationary state. Considering unexpected workload varies randomly over time, we construct an M/M/1/K queue in a randomly environment described by a 3-state (over-provisioning state, normal state, and under-provisioning state), irreducible Markov process. In this queue, the conditional average queue length in the normal state and 3-state probability are generated, respectively, and the anelasticity value of the system is precisely calculated. When the average arrival rate

exceeds the average service rate in the queueing system, an optimal CPR unchanged adaptive algorithm based on PSO is presented to dynamically adjust the controller service rate. By analyzing the rush-hour behavior of the system, we come to the conclusion that the system can basically maintain elastic invariance in adaptive adjustment parameters. Extensive numerical results show the usefulness and effectiveness of the proposed techniques.

(4) For the uncertain delay problem, the closed-form expressions of the probability density function (pdf) and the cumulative distribution function (cdf) of the uncertain MRT in the single-channel ANCS are calculated, and a novel grey model (1,1) (GM(1,1)) named as probability transformed GM(1,1) (PTGM(1,1)) with two parameters is proposed to predict the response delay for each message. We firstly configure the ANCS bus using a multi-hierarchical flexible TDMA/Fixed Priority scheduling policy based on event triggering protocol, and then establish an optimization model to obtain the optimal parameter values of the bus. An M/PH/1 queue model with random-sized batch arrivals is used to model the control application in the ANCS, and to obtain the average number of customers for each control application. Based on the model, a constrained optimization problem is formulated to minimize the nonlinear cost function of the ANCS, and the probability density function (pdf) of uncertain response delay for each message is generated. An optimal minislot algorithm based on DFP variable scale method and a parameter estimation algorithm of pdfs of the uncertain response delays are proposed to obtain optimal values of the bus minislot and matrix parameters of PH distribution, respectively. A two-parameters optimization algorithm of PTGM(1,1) based on DFP method is given to minimize mean absolute percentage error (MAPE), and then to obtain the optimal values of the two parameters and the predicted values of the response delay for each message. Extensive numerical results are provided to illustrate the usefulness of the proposed algorithms and to show the accuracy of the predicted results.

Key words: Cyber-Physical System; Arbitration Network Structure; Communication Model Optimization; Dynamic Resource Provisioning; Uncertain Delay Analysis; Performance Analysis; Queueing Theory

目 录

学位论文原创性声明和学位论文授权使用授权书	I
摘 要	II
Abstract	IV
目 录	VII
插图索引	X
附表索引	XII
符号列表	XIII
第 1 章 绪 论	1
1.1 选题背景	1
1.1.1 CPS 的特性	1
1.1.2 控制、NCS 和嵌入式控制系统	2
1.1.3 仲裁网络结构	4
1.2 选题意义	7
1.3 研究现状及挑战	8
1.3.1 研究现状	8
1.3.2 研究挑战	10
1.4 研究内容及主要贡献	12
1.4.1 研究内容	12
1.4.2 主要贡献	13
1.5 本文构想	14
第 2 章 相关工作	15
2.1 协同设计研究综述	15
2.1.1 控制与平台协同设计	15
2.1.2 资源竞争	16
2.1.3 时延问题	18
2.1.4 丢包问题	20
2.1.5 基于无源性的组合设计问题	20
2.2 QBD 过程	22
2.3 多目标优化问题求解	24
2.4 灰色预测模型	26
2.5 最优停止策略	28
2.6 总线协议	29

2.6.1 CAN	30
2.6.2 FlexRay	31
2.7 本章小结	32
第 3 章 扁平半休眠多控制器调度	33
3.1 引言	33
3.2 平台体系结构	33
3.3 排队模型	35
3.3.1 模型描述	35
3.3.2 QBD 过程模型	37
3.3.3 系统的平稳概率分布	39
3.3.4 系统性能度量	43
3.4 最优化分析	44
3.4.1 成本和性能分析	44
3.4.2 最优停止算法	45
3.4.3 FSDMC 调度算法	46
3.5 数值结果	48
3.5.1 系统性能度量灵敏度分析	48
3.5.2 期望的成本函数与标称时延函数的灵敏度分析	49
3.5.3 系统参数的最优值	53
3.5.4 比较和分析	56
3.6 本章小结	56
第 4 章 资源动态供应与滞弹性计算	58
4.1 引言	58
4.2 滞弹性度量	59
4.3 排队模型	60
4.3.1 QBD 过程模型	60
4.3.2 系统的平稳概率分布	64
4.3.3 系统性能度量	67
4.4 最优化和灵敏度分析	68
4.4.1 最小平均 CPR	68
4.4.2 系统性能度量灵敏度分析	69
4.4.3 CPR 灵敏度分析	70
4.5 滞弹性量化模型	73
4.5.1 马尔可夫环境下的 M/M/1/K 排队模型	73
4.5.2 弹性灵敏度分析	75

4.5.3 系统自适应分析.....	76
4.5.4 比较和分析.....	78
4.6 高峰期行为.....	80
4.7 本章小结.....	82
第 5 章 具有不确定时延的消息响应时间分析.....	83
5.1 引言.....	83
5.2 平台体系结构.....	84
5.3 排队模型.....	85
5.3.1 模型描述.....	85
5.3.2 QBD 过程模型.....	87
5.3.3 参考总线的消息队列.....	87
5.4 总线配置.....	90
5.4.1 最优化模型.....	90
5.4.2 总线配置算法.....	91
5.4.3 实验结果.....	94
5.5 不确定时延的概率分布.....	95
5.5.1 概率密度函数.....	95
5.5.2 每个消息的 WCRT.....	99
5.5.3 评估拟合精度.....	100
5.6 预测 MRT.....	102
5.6.1 GM(1,1)中的两参数背景值.....	102
5.6.2 概率变换 GM(1,1)模型(PTGM(1,1)).....	105
5.6.3 预测性能度量.....	106
5.7 本章小结.....	110
结 论.....	112
参考文献.....	115
致 谢.....	131
附录 A 攻读学位期间所发表的学术论文.....	132
附录 B 攻读学位期间所参与的科研课题.....	132

插图索引

图 1.1 闭环和开环系统.....	2
图 1.2 反馈控制系统.....	3
图 1.3 显式辨识与隐式辨识图.....	4
图 1.4 一般的 NCS.....	4
图 1.5 一个典型的分布式嵌入式系统(DES).....	5
图 1.6 具有三个分布式控制应用的系统体系结构.....	6
图 2.1 CAN 总线拓扑结构.....	30
图 2.2 单个信道的总线型拓扑结构.....	31
图 2.3 单个信道的星型拓扑结构.....	31
图 2.4 双信道的总线—星型混合拓扑结构.....	32
图 3.1 含有 5 个分布式控制应用的系统体系结构.....	34
图 3.2 控制应用 $i, i = 1, 2$ 的任务图.....	35
图 3.3 控制应用 $i, i = 3, 4, 5$ 的任务图.....	35
图 3.4 通信信道 A: 分层 FTDMA/FP 调度器, l 表示 FTDMA 周期长度.....	35
图 3.5 $N/(d, c)$ -M/M/c/K/SMWV 排队系统的状态转移率图.....	37
图 3.6 $F(10, d, 20, \mu_B, 1.0)$ 和 $E[T_q]$ 分别对应 μ_B 和 d 的变化图.....	50
图 3.7 $F(10, d, 20, 2.5, \mu_V)$ 和 $E[T_q]$ 分别对应 μ_V 和 d 的变化图.....	50
图 3.8 $F(c, 4, 20, \mu_B, 1.0)$ 和 $E[T_q]$ 分别对应 μ_B 和 c 的变化图.....	51
图 3.9 $F(c, 4, 20, 2.5, \mu_V)$ 和 $E[T_q]$ 分别对应 μ_V 和 c 的变化图.....	52
图 3.10 半休眠控制器数的频度图.....	53
图 3.11 四次调用 MOPSO 算法得到的 Pareto 前沿(a)~(d).....	55
图 4.1 $N/(d, c)$ -M/M/c/K/AMWV 排队系统的状态转移率图.....	60
图 4.2 λ 和 d 变化时 $CPR(d, 20, 3.4, 1.5, \lambda, 6.6)$ 变化图.....	71
图 4.3 λ 和 N 变化时 $CPR(4, N, 3.4, 1.5, \lambda, 6.6)$ 变化图.....	71
图 4.4 μ_B 和 d 变化时 $CPR(d, 20, \mu_B, 1.5, 16.9, 6.6)$ 变化图.....	71
图 4.5 μ_B 和 N 变化时 $CPR(4, N, \mu_B, 1.5, 16.9, 6.6)$ 变化图.....	72
图 4.6 μ_V 和 d 变化时 $CPR(d, 20, 3.4, \mu_V, 16.9, 6.6)$ 变化图.....	72
图 4.7 μ_V 和 N 变化时 $CPR(4, N, 3.4, \mu_V, 16.9, 6.6)$ 变化图.....	72
图 4.8 θ 和 d 变化时 $CPR(d, 20, 3.4, 1.5, 16.9, \theta)$ 变化图.....	73
图 4.9 θ 和 N 变化时 $CPR(4, N, 3.4, 1.5, 16.9, \theta)$ 变化图.....	73
图 4.10 当 $\rho < 1$ 时 p_{over} , p_{normal} 和 p_{under} 随 λ 变化的曲线图.....	75
图 4.11 当 $\rho < 1$ 时 p_{over} , p_{normal} 和 p_{under} 随 μ_B 变化的曲线图.....	76
图 4.12 当 $\rho < 1$ 时 p_{over} , p_{normal} 和 p_{under} 随 μ_V 变化的曲线图.....	76

图 4.13 参数 μ_B 与 λ 自适应关系曲线图	77
图 4.14 当 λ 变化同时 μ_B 变化时 p_{over} , p_{normal} 和 p_{under} 变化图.....	78
图 4.15 使用[39]中的模型绘制 p_{over} , p_{normal} 和 p_{under} 与 λ 的关系曲线一($\lambda \in [1, 34)$)	79
图 4.16 使用[39]中的模型绘制 p_{over} , p_{normal} 和 p_{under} 与 λ 的关系曲线二($\lambda \in [34, 64]$).....	79
图 4.17 当 $\lambda \in [1, 34) \cup [34, 64]$ 时系统在正常状态下 $AvgQL_{normal}$ 和 p_{normal} 关系曲线	81
图 5.1 具有六个控制应用的系统体系结构	84
图 5.2 通信总线: 多层次 FTDMA/FP 调度器, l 表示 FTDMA 周期长度	85
图 5.3 平台体系结构的多队列多服务台串并联混合排队系统	86
图 5.4 在一个超周期内第 3 个控制应用的消息队列	89
图 5.5 最小时隙与系统成本的关系	94
图 5.6 参考系统中的不确定时延	95
图 5.7 来自于不同控制应用的部分消息的响应时间的 pdf 曲线图.....	101
图 5.8 来自于第 3 个控制应用的部分消息的响应时间的 pdf 曲线图.....	101
图 5.9 消息 m_5 的仿真和拟合分布采样 Q-Q 图.....	101
图 5.10 消息 m_{69} 的仿真和拟合分布采样 Q-Q 图	102
图 5.11 消息 m_{34} 的实际值和预测值比较.....	108
图 5.12 消息 m_3 的实际值和预测值比较.....	108
图 5.13 消息 m_6 的实际值和预测值比较.....	109
图 5.14 消息 m_{34} 的响应时间的 cdf 的预测	109
图 5.15 消息 m_3 的响应时间的 cdf 的预测.....	110
图 5.16 消息 m_6 的响应时间的 cdf 的预测.....	110

附表索引

表 3.1 在 $\lambda=5.0, \theta=0.05, d=4, c=10, N=20$ 情况下(μ_B, μ_V)取不同值时系统性能度量	48
表 3.2 在 $\mu_B=2.5, \mu_V=1.0, d=4, c=10, N=20$ 情况下(λ, θ)取不同值时系统性能度量	48
表 3.3 在 $\lambda=5.0, \theta=0.05, \mu_B=2.5, \mu_V=1.0$ 情况下(d, c, N)取不同值时系统性能度量	49
表 3.4 在 $\lambda=5.0, \theta=0.05, \mu_V=1.0, c=10, d=4$ 情况下(N, μ_B)取不同值时的期望成本 函数和标称时延函数的变化	51
表 3.5 在 $\lambda=5.0, \theta=0.05, \mu_B=2.5, c=10, d=4$ 情况下(N, μ_V)取不同值时的期望成本 函数和标称时延函数的变化	51
表 3.6 在 $C_v=15, C_b=35, C_d=8, C_i=10000$ 情况下参数(C_h, C_i)对期望成本函数 $F(4, 20, 2.5, 1.0)$ 的影响	52
表 3.7 在 $C_h=1, C_i=10, C_d=8, C_i=10000$ 情况下参数(C_v, C_b)对期望成本函数 $F(4, 20, 2.5, 1.0)$ 的影响	52
表 3.8 在 $C_h=1, C_b=35, C_v=15, C_i=10$ 情况下参数(C_d, C_i)对期望成本函数 $F(4, 20, 2.5, 1.0)$ 的影响	53
表 3.9 分组消息到达的采样数据	53
表 3.10 图 3.11(a)中 Pareto 最优解和相应的非支配向量	55
表 3.11 Pareto 最优解和相应的非支配向量	56
表 4.1 在 $\lambda=16.9, \theta=6.6, d=4, c=10, N=20$ 情况下(μ_B, μ_V)取不同值时系统性能度量	69
表 4.2 在 $\mu_B=3.4, \mu_V=1.5, d=4, c=10, N=20$ 情况下(λ, θ)取不同值时系统性能度量	70
表 4.3 在 $\lambda=16.9, \theta=6.6, \mu_B=3.4, \mu_V=1.5$ 情况下(d, c, N)取不同值时系统性能度量	70
表 4.4 两个排队模型的成本函数 F 与标称时延函数 τ_{nom} 的比较结果	79
表 4.5 六个队列在各个位相上滞弹性和标准差	81
表 4.6 六个队列在各个位相上的空闲条件概率和队列长度分布的百分位数	81
表 5.1 汽车 CAN 系统的一个实例	88
表 5.2 在一个超周期内不同的子队列的消息平均到达率、消息数以及消息实例数	89
表 5.3 第 i 个消息相应参数 x_i 的估计值及其 WCRT	100
表 5.4 ANCS 中的消息拟合分布统计分析表	103
表 5.5 不同的 GMs(1,1)的最优参数值和性能度量	107

符号列表

符号	定义
CPS	信息物理系统 (Cyber Physical System)
ANCS	仲裁网络控制系统(arbitrated networked control system)
FSDMC	扁平半休眠多控制器(flat semi-dormant multi- controller)
QoC	控制质量(Quality of Control)
QoS	服务质量(Quality of Service)
DES	分布式嵌入式系统(distributed embedded systems)
PU	处理单元(processing unit)
ECU	电子控制单元(Electronic Control Unit)
MRT	消息响应时间(message response time)
RTA	响应时间分析(Response-Time Analysis)
WCRT	最坏情况响应时间(worst-case response time)
TT	时间触发(time-triggered)
ET	事件触发(event-triggered)
ST	静态段(static segment)
DYN	动态段(dynamic segment)
TDMA/FP	时分多路复用/固定优先级(Time Division Multiple Access/fixed priority)
LMI	线性矩阵不等式(linear matrix inequality)
PSO	粒子群优化(particle swarm optimization)
MOPSO	多目标粒子群优化(multi-objective particle swarm optimization)
QBD	拟生灭过程(quasi-birth-and-death)
FIFO	先进先出(first in first out)
FCFS	先来先服务(first-come first-served)
BD	总线驱动器(Bus Driver)
CPR	成本性能率(cost-performance ratio)
pdf	概率密度函数(probability density function)
cdf	累积分布函数(cumulative distribution function)
SMWV	同步多重工作休假(Synchronized Multiple Working Vacation)
AMWV	异步多重工作休假(asynchronous multiple working vacation)
GM(1,1)	灰色预测模型(grey model(1,1))
PTGM(1,1)	概率变换灰色预测模型(probability transformed GM(1,1))
ITGM (1,1)	改进的变换 GM(1,1)模型(improved transformed GM(1,1))

EDF	最早截止时间优先(earliest deadline first,)
DMR	截止时间错失率(Deadline Miss Ratio)
WCDFP	最坏情况截止时间错失概率(worst case deadline failure probability)
PJN	功率中继网络(power-junction- network)
AGO	累积生成运算(accumulated generating operation)
eCPS	弹性 CPS (elasticity CPS)
PH	相位型(phase-type)
LST	拉普拉斯-斯蒂尔切斯变换(Laplace-Stieltjes transform)
MAPE	平均绝对误差百分比(mean absolute percentage error)
RMSE	根均方误差(root mean squared error)
RMSPE	根均方百分比误差(root mean square percentage error)
λ	全局队列的到达率
μ_B	控制器忙期服务率
μ_V	控制器工作假期服务率
θ	控制器工作假期速率
λ_i	第 i 个子队列的平均到达率
μ_i	第 i 个子队列的平均服务率
ρ	系统利用率
F	期望成本函数
τ_{nom}	标称时延函数
l	总线环长度
R	率阵
\mathfrak{R}_{ij}	第 i 个控制应用的第 j 个服务位相
$MsgID_i$	第 i 个控制应用的消息集
$m_{i,j}$	消息 m_i 的第 j 个实例
P_i	消息 m_i 的固定优先级
ε_i	消息 m_i 的传输时间
T_i	消息 m_i 的周期
H	超周期
$MSNo$	每个总线环的最小时隙数
κ_i	第 i 个消息的 WCRT
$\hat{t}_{i,k,j}$	第 i 个控制应用中的第 k 个消息的第 j 个实例的采样值
$t_{h,i,j,s}^q$	第 h 个控制应用中的第 s 个总线环中传输的消息实例 $m_{i,j}$ 的等待时间
$t_{h,i,j,s}$	第 h 个控制应用中的第 s 个总线环中传输的消息实例 $m_{i,j}$ 的总时延

第1章 绪论

信息物理系统(cyber-physical system, CPS)是由计算算法和物理组件的无缝集成所构建并依赖于这种集成的工程系统。嵌入式计算、传感器监控、无线通信以及大数据处理等技术的发展使得物理过程、计算过程和通信过程高度集成与交互,从而产生集计算、通信和控制技术于一体的CPS系统,是时空多维异构的混杂自治系统。CPS通过感知环境,直接并大部分时间独立于人类操作员与其物理环境进行交互,计算所需行动并应用行动。所有这些都必须通过有限的资源实现,如有限的通信带宽、有限的功率、有限的计算能力等等。CPS中的控制动态与通信网络之间的集成研究近几年成为业界非常活跃的研究课题之一。这一集成研究问题常常归结为通信与控制协同设计^[1],其研究焦点主要集中在CPS中的控制与调度策略协同设计方面,以确保有效的控制质量(Quality of Control, QoC)和CPS的效用。

1.1 选题背景

1.1.1 CPS的特性

CPS无处不在,涵盖了小到纳米级生物机器人,大到全球能源协调与管理系统等涉及人类基础设施建设的复杂大系统。CPS的典型应用包括高可信医疗设备与系统、辅助生活系统、交通控制与安全、高级汽车系统、过程控制、能量保护、环境控制、航空航天、仪器仪表、关键基础设施控制(电力系统、水资源、通信系统等)、分布式机器人(远程监控、远程医疗)、防御系统、制造业、智能结构等各行各业,是构建人类未来智慧城市的基础^[2~4]领域。这表明CPS是未来的关键技术。CPS的特性可以归纳为如下四点,这些性质在各自研究中已经有所了解方案^[1]。

(1) 紧耦合性。CPS与物理世界紧耦合,通过传感器、控制器和执行器与外部环境直接交互作用。这是自动控制研究领域的研究内容。

(2) 热耗散与实时性。CPS设计存在某些限制,许多CPS用电池供电,允许简单的操作部署。有限的功率也就意味着有限的计算能力。此外,所能允许的热耗散可能是CPS设计中的一个限制因素。这种有限的热耗散问题是嵌入式系统领域的研究问题。CPS也是一种实时系统(Real-Time System),能够在确定的时间内执行计算或处理事务并对外部事件做出响应的计算机系统。实时系统把时间看成系统中的一个重要资源,任务必须在规定的截止时间(Deadline)之前完成,系统的正确性不仅依赖于计算结果的逻辑正确性而且与结果的完成时间有关。这种实时性问题也是嵌入式系统领域的研究问题。

(3) 自适应性。既然CPS嵌入在物理系统中,很自然地,它们具备较长的生命期,其结果可能导致CPS的外部环境在整个生命期显著地改变,例如,物理系统可能出现故

障，或者诸如传感器之类的外部组件可能更换。那么，确保CPS继续有效地工作的方法属于自适应控制的研究范畴。

(4) 异构性。CPS由许多交互作用的异构组件组成，其关键特征是在数以千计节点的通信网络上的信息交互，而依附于网络的节点一般来说是各种各样的异构节点。例如，传感器以周期性的方式发送采样信息，但控制器在需要时仅仅发送控制信息。这种分布式异构控制系统的设计是网络控制系统(networked control system, NCS)的研究内容。

尽管这些特性的解决方案已经存在，但是用一种可靠的理论基础将这些解决方案组合成CPS集成方法是一个挑战。这一集成方法常常称为“通信与控制协同设计”。嵌入式控制器和通信网络的协同设计带来了系统分析与集成的挑战，包括网络协议、操作系统的兼容性、控制与实时计算系统的组合性能优化方法等，也带来了新的机遇。既然网络也需要设计，那么诸如时延、抖动和丢包等网络特征能够选择所期望的值，并能较好地利用它们设计控制器。而且，控制器设计需要适应体系结构级的约束。

1.1.2 控制、NCS 和嵌入式控制系统

反馈控制与自适应控制、NCS以及嵌入式控制系统等三种主要研究领域是CPS的关键技术。这三种技术的背景简要介绍如下^[1]：

(1) 反馈控制与自适应控制。

① 反馈控制。术语“反馈”与“控制”相互之间具有紧密关系。术语“反馈”描述了这样一种情况，一个动态系统的输出是另一个动态系统的输入，反之亦然。通过这种互联，使得两个系统的动态是强耦合的。术语“控制”涉及到一个动态系统生成的输出，主要是为了将另一个动态系统引导到所期望的行为。一般地，能区分开环和闭环系统(如图1.1)，一个闭环系统由反馈环连接，而一个开环系统没有反馈。

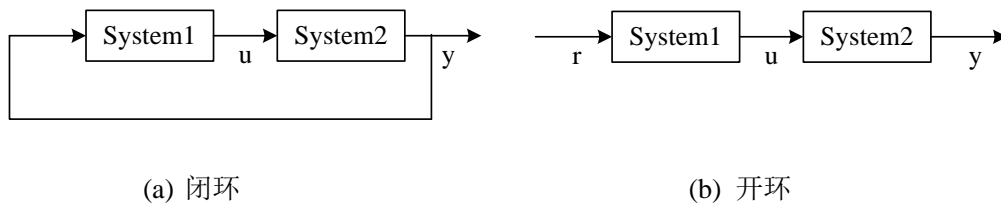


图 1.1 闭环和开环系统

典型地，一个反馈控制系统由监控系统性能的传感器、正确执行系统行为的执行器、指示期望行为的参考命令(reference command)以及计算这种正确性的时间和强度的控制器组成(如图1.2)。控制理论和控制工程致力于分析与合成给定系统控制结构的各种控制方法。

一个典型的设计过程开始于被控系统基本模型的确定。许多物理系统能用连续时间的普通的微分方程描述。某些情况下，差分方程用于描述离散时间系统，可能用基于系统辨识的方法生成该差分方程。一旦生成这些模型，控制器就用状态空间、频率域和几何方法进行设计。控制性能的成功不仅仅由系统模型的精度和使用的特定的控制方法确定，而且由实现体系结构的模型精度确定。实现体系结构本身将不同类错误引入系统，

这些实现错误可能与外部时延干扰和表示采样、调度、离散化以及量化错误的干扰一起建模。

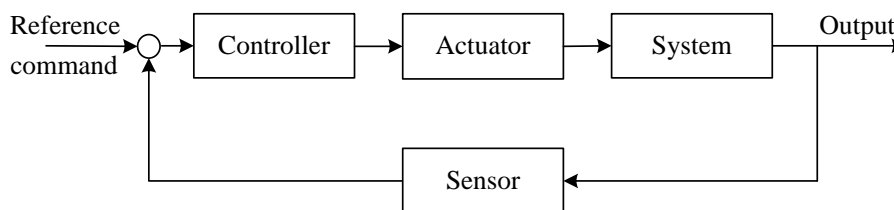


图 1.2 反馈控制系统

当讨论QoC时需要解决的两个主要性质是稳定性与性能。稳定性关心反馈控制中系统信号的好行为，其严格的度量是调整基本错误信号为0。闭环控制性能的目的典型地包括跟踪控制信号，除了拒绝或者最小化干扰、度量噪声和建模错误以外。特征化这些目标的现代方法是分析相关矩阵范数、优化各种性能目标等。

② 自适应控制。相比之下，经典的反馈控制假设系统参数、环境条件、干扰等所有参数完全已知，而自适应控制涉及到部分已知系统的控制。一个实际系统的特征很难足够精确地知晓，可能因各种因素随时间的推移而变化，诸如系统的老化、环境的改变、操作条件的变化、不可预见的干扰等都能影响系统的稳定性和性能。

一个自适应控制器是具有可调节的参数和调节参数机制的控制器^[5,6]。不同类型的自适应控制近几年有许多研究成果，主要集中在增益调度、模型参考自适应系统、自校正调节器、求极值控制等。如果一个系统仅仅是部分已知，或者作用于系统的干扰未知，那么辨识过程能用于搜集系统及其环境的信息。但是，如果目的是控制系统，那么单独的辨识并不充分。因此，一个自适应控制器同时解决两个问题：其一是必须确定系统特征；其二是必须确定导致期望目的的必要的控制行为。解决这一问题有两种不同的方法：一是在线估计参数，更新基于估计的控制参数(如图1.3(a))。这种方法称为间接控制或者显示辨识。二是并不明确地识别系统参数，但是为了提高性能指标而直接调整控制参数(如图1.3(b))。这种方法称为直接控制或者隐式辨识。

(2) NCS。当今 CPS 的基本特征之一是存在介于计算和物理实体之间的通信网络。因此，控制器和物理系统之间的交互可能会经历网络诱导延迟，甚至偶尔会丢失数据包。网络链路可以被视为遭受数据率限制的通信信道。因此，NCS 一些重要的基本问题如下：①网络引起的延迟、丢包和通信信道如何影响系统的稳定性？②在什么条件下 NCS 可以稳定，如何稳定它^[7]？

传统上，NCS 关注由某些组件共享通信网络的空间分布式动态系统的建模、设计与分析^[8](如图 1.4)。当前，NCS 成功地应用到了许多领域，如：汽车、远程手术、环境监测与监控、楼宇自动化、冶炼与能量系统、Internet 上的触觉协作等等。当组件增加或者移除时，NCS 不需要代价或者很少的代价，这增加了系统的可重配置性、整个控制系统的灵活性，简化了系统的诊断与维护。

文献[9]阐述了网络化实时系统中的控制与通信的研究挑战问题，关注控制空间分布式系统特定网络的构成、在数字反馈信道中系统相关数据率的需要、分布式异构传感器

数据的实时融合与注册、自治Agent网络协作控制理论等问题。尽管数字控制器很容易在空间上互联分布式组件，但是仍然需要硬连线进行连接。随着技术的进步，NCS在远程定位用低成本处理单元以及以安全和可靠的方式在共享总线上传送消息。

既然 NCS 结合了控制理论和通信理论两种不同的研究领域，那么其研究挑战不同于经典的控制系统，主要包括有限的网络资源、采样与时延问题、丢包以及同步问题等。

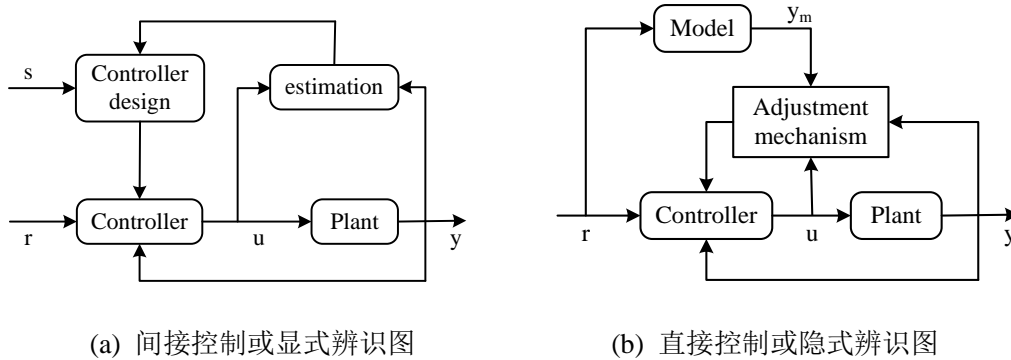


图 1.3 显式辨识与隐式辨识图

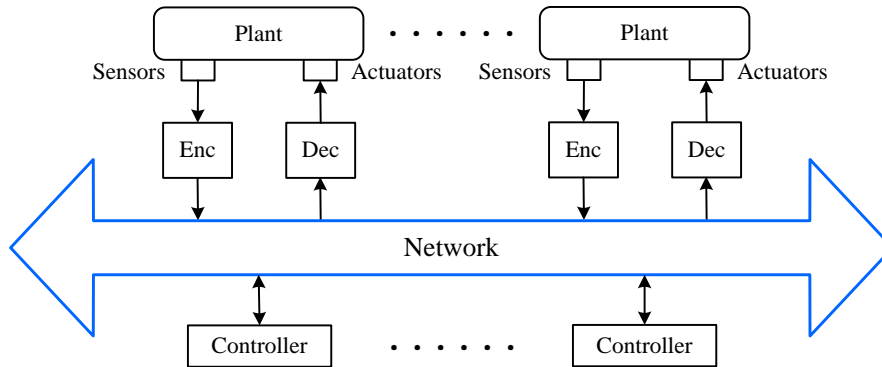


图 1.4 一般的 NCS

(3) 嵌入式控制系统。嵌入式控制系统无处不在，如：飞机、汽车、过程控制、楼宇系统、厨房电器等。一个嵌入式控制系统是一种在实时计算约束下完成专用功能的计算机系统，其典型特征是系统中互相通信的不同组件共享网络，这些组件包括系统中大量的分布式传感器和执行器。

第一个嵌入式系统之一是1961年由NASA设计的“阿波罗计划”^[10]。此后，关于嵌入式系统处理功率和功能性的能力显著地增加。第一个微处理器——Intel 4004，用于设计袖珍计算器。主要突破是，如果系统的作用发生改变，那么通过改变软件重用通用计算机硬件的能力。另外，许多复杂控制算法仅仅能用微处理器实现，这一点在汽车CPS中尤为明显，现代汽车使用多达100多个微处理器。嵌入式系统的特征不同于标准的PC系统，主要包括：复杂性、实时性、多级速率、制造成本、功耗等。嵌入式系统的设计挑战有：合适的硬件数量、截止时间问题、功耗问题、升级以及可靠性问题等。

1.1.3 仲裁网络结构

汽车、列车、航空控制系统、工业自动化系统等常常是分布式嵌入式系统(distributed embedded systems, DES)，有大量的处理单元(processing unit, PU)，如传感器和执行器

等，并通过诸如控制器局域网(Controller Area Network, CAN)、本地互连网(Local Interconnect Network, LIN)或FlexRay，最近也出现汽车以太网等进行通信。这种体系结构用于运行分布式控制应用，常常受到多个QoC的约束^[11]。这类CPS中内在的异构性和多样性使得系统非常复杂。给定这种CPS的分布式特性，采用NCS的研究技术进行研究是很自然的。传统上，在控制消息丢失、变延迟和抖动出现的情况下关注NCS领域的分布式控制器的建模与设计。这里，假设给出了网络的特征，并且主要聚焦于控制器设计。虽然这在无线网络控制上下文是有意义的，其中假设网络是预先设计的，但在许多其他域中，在设计网络本身时具有相当大的灵活性。例如，在汽车体系结构中，分布式控制器通过在FlexRay或者CAN总线通信的多个电子控制单元(Electronic Control Unit, ECU)来实现。在这种情况下，网络设计问题包括：①映射控制任务到ECU；②确定ECU的调度参数和通信总线的调度参数。如果给定这些参数，控制器的设计问题类似于NCS中的控制器分析和设计问题。控制器设计的总目标是选择合适的采样周期和相关的控制参数，使得：①控制功能是鲁棒的，且提供期望的性能；②在给定的资源约束下，所有的控制和其它任务与消息是可调度的。在这种情况下，网络与控制协同设计的问题包括依赖被控对象的网络参数选择、控制性能需求、控制器选择以及依赖于网络特征的控制参数选择等。这种协同设计带来了新的挑战 and 机遇，既没有在NCS中研究，又不是嵌入式系统的研究范围。

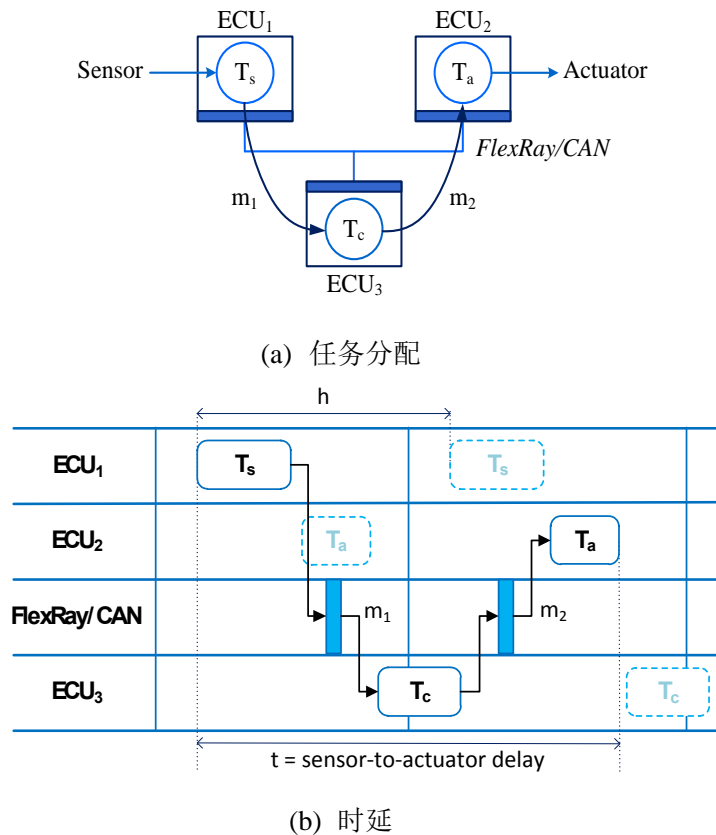


图 1.5 一个典型的分布式嵌入式系统(DES)^[11]

Annaswamy等设计了一类命名为仲裁网络控制系统(arbitrated networked control system, ANCS)的特殊CPS体系结构^[12]，以强调所设计的控制系统中的网络调度(或仲裁

裁), 其显著特征是共享通信介质的消息仲裁机制, 其中的通信网络及其附加信息用于设计控制器。称这一ANCS体系结构为单信道ANCS系统, 也称为单信道“仲裁网络结构”, 用于CPS的整体建模。这一仲裁网络结构中的多个控制应用共享通信总线构成多个通信模型。因此, 仲裁网络结构由共享总线的多个通信模型构成。为了举例说明一个ANCS, 考虑文献[11]中提出的典型的分布式嵌入式体系结构(如图1.5)。这里, 一个分布式控制应用(或通信模型)以及总的端到端的消息响应时间(即传感器到执行器延迟)分别如图1.5(a)和(b)^[11]。一个ECU(图中ECU₁)采集传感器数据(任务T_s)。然后, 通信总线(如FlexRay)传送消息m₁到第二个ECU(图中的ECU₃), 其内在的控制算法被调用(任务T_c)。控制器输出作为消息m₂经由总线发送到执行器(ECU₂), 激活执行器任务T_a。系统中的采样周期是h, 消息响应时间(message response time, MRT)是t(见图1.5(b))。

文献[13]设计了含有三个分布式控制应用的系统体系结构(如图 1.6)。每一个控制应用被划分成映射到不同 PU 的多个任务。这些 PU 通过共享通信总线进行通信, 并运行来自一个或者多个控制应用的不同任务。处理器的调度和共享总线的仲裁在各种控制信号中引入了时延, 这些时延依赖于 PU 和总线的调度/仲裁策略。采用分层时分多路复用/固定优先级(Time Division Multiple Access/fixed priority, TDMA/FP)调度算法仲裁总线, 采用固定优先级调度算法进行任务调度。文献[14~16]研究了该体系结构上控制与通信协同设计问题, 从分层调度、混合调度、多模态系统协同设计等三方面进行了研究, 解决了系统中的固定时延、有上界的丢包、不确定系统的自适应控制、自适应切换控制等网络通信与控制协同设计问题。文献[11]在假定给定网络特征的情况下, 重点关注 ANCS 中控制器设计问题, 运用“实时演算”^[17]分析方法设计网络, 时延感知的处理用于确定控制器, 并考虑传感器和执行器之间的端到端最坏情况时延上界设计控制器, 运用线性矩阵不等式(linear matrix inequality, LMI)方法讨论了系统的稳定性问题。文献[18]提出

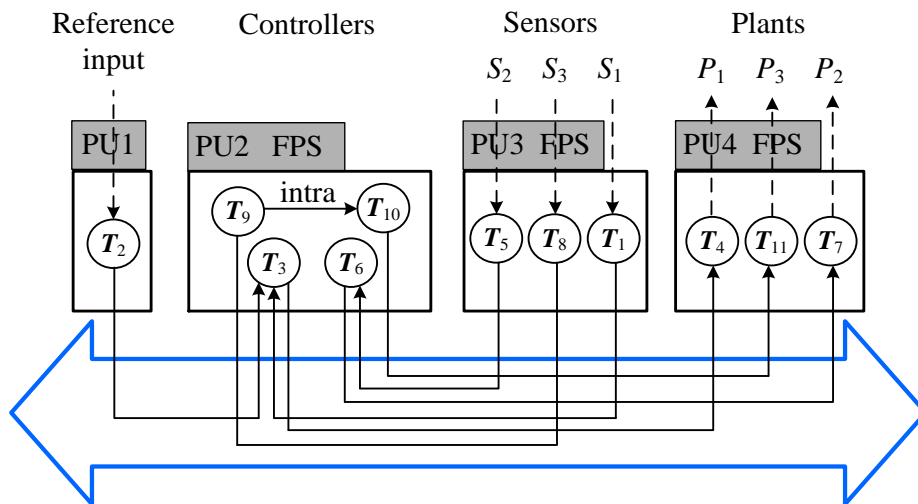


图 1.6 具有三个分布式控制应用的系统体系结构^[11]

了一个超限运行协同设计框架, 解决了在资源有限和共享而导致传输消息中的延迟变化的 NCS 中多个控制应用的实现问题, 以实现最优的控制性能和有效的资源利用率。但是, 在这一 ANCS 中, 存在以改善系统性能、减少系统成本为目的的多控制器调度问题; 资源约束下的资源动态供应问题; 不确定端到端的 MRT 问题等需要解决。本课题研究

仲裁网络结构中的 CPS 通信模型优化, 关注 CPS 网络通信与控制协同设计中的这三个问题, 考虑共享通信总线的参数和控制器的参数的协同计算。

1.2 选题意义

(1) CPS 协同设计是系统 QoC 与服务质量(Quality of Service, QoS)的需要。在 ANCS 中, 由于网络资源带宽的局限性, 当有多个控制节点接入网络的时候, 如何合理分配网络资源成为了 NCS 需要解决的问题。而且不同的通信协议会影响 NCS 信号传输的特征, 如节点驱动方式、数据包大小、网络诱导时延等影响整个系统的控制性能。因此网络仲裁和任务调度不仅仅会影响网络的利用率, 而且会影响网络的稳定性。NCS 的整体性能既受到控制算法的影响, 又受到其网络任务的调度策略和共享资源的仲裁策略的影响, 即系统 QoC 和网络的 QoS 共同决定 NCS 的全局性能, 而这两者的质量又都受到信息传输周期的影响^[19]。传输周期若超过上限值, 会导致控制系统性能不稳定, 因此从控制性能角度考虑, 需要较小的传输周期, 而过小的传输周期使得信息频繁地交换会导致网络冲突加剧, 增大了时延和丢包率, 使网络性能下降。又因为调度策略影响网络性能, 控制策略影响控制性能, 因此调度与控制相互约束, 网络性能与控制性能的优化需要通过调度与控制合理的协同设计来实现, 以获得系统最优的 QoC 和 QoS。

(2) CPS 协同设计是信息交互的实时性与准确性的需要。目前, 随着 CPS 的应用越来越广泛, 涵盖了小到纳米级生物机器人, 大到全球能源协调与管理系统等涉及人类基础设施建设的复杂大系统, 大量的 ECUs 加入到系统中, 导致在网络上传输的信息越来越多。而现有的网络通信带宽以及资源不能满足目前的复杂环境的需求, 所以需要对现有的网络总线协议以及系统内部分析方法做进一步改进。例如, 在汽车 CPS 中, 为了保证车载系统的安全可靠, 要求车载系统中信息交换具有很高的实时性和准确性, 即信息从控制单元发出后经过网络截止到接收节点并做出响应的需要在一定范围内, 即截止时间内, 否则较大的时延会导致系统的不稳定等问题。所以保证汽车的信息得到实时调度和控制具有十分重要的意义。

(3) CPS 协同设计是系统集成学科发展的需要。通信与控制协同设计的基本观点就是在 CPS 的设计中将实时调度理论和反馈控制理论结合起来, 考虑到网络仲裁和任务调度, 使可用的计算机资源得到最有效的利用并使总的系统性能达到最优^[20]。对 ANCS 这类特殊的 CPS 的网络仲裁与控制协同设计, 其优势在于将网络运行性能、带宽资源限制等更多的网络因素融入控制系统设计和性能优化中^[9], 得到行之有效的 ANCS 的分析、设计方法, 甚至获得 CPS 中的新的通信与控制一体化理论体系, 将是对系统集成学科的重大贡献, 具有重大的科学意义和实际应用意义。

(4) CPS 协同设计是 CPS 应用领域不断增长的复杂性的需要。CPS 的潜在能力正在稳步增长, 而成本正在以较高的速率下降。传统的计算机和 CPS 之间的不同在于与人的交互方式上。传统的计算机通过鼠标、键盘或者操纵杆输入, 而输出到显示屏, 但是, CPS 通过感知环境、计算需要的行为并应用该行为, 将物理环境直接和人进行交互, 并

且大部分时间独立地与人进行交互^[21]。所有的这一切都必须在有限的资源约束下进行，如：有限的通信带宽、有限的能量、有限的计算功率，等等。但是，CPS 因专用而设计，能完成专门任务，相对地，传统计算机因通用而设计。既然 CPS 被嵌入进物理设备，它们的生命期比其它的计算机系统的生命期要长，这使得在设计阶段验证 CPS 在整个生命期的可靠性和安全性成为必要。

1.3 研究现状及挑战

1.3.1 研究现状

(1) 控制器调度问题。在ANCS中，文献[1,11,12]研究了由时间触发(time-triggered, TT)协议产生的固定时延问题，时延感知的处理用于确定控制器，并考虑传感器和执行器之间的端到端最坏情况时延和最好情况时延的均值设计控制器，确保闭环系统的稳定性，而没有考虑时变时延问题。文献[1]提出了几种方法来处理特定类型的、采用分层和混合通信网络的ANCS的控制和通信集成协同设计问题。Zeng等从应用设计者的观点研究了ECU和FlexRay总线优化调度合成问题^[22]。Dvorak等提出了一种启发式算法，通过两个独立的通道与网关进行交互，将调度问题分解为ECU至信道分配子问题和信道调度子问题^[23]。Hu等研究了在TT协议的车载网络中处理实时应用的整体调度问题，其中捕获了系统设计和集成中的实际方面^[24]。Kang等^[25]提出了一种称为具有各种周期信号的带宽消耗最小化帧打包算法，将不同的周期信号打包成一个消息帧。但是，这些调度工作仅仅关注通信协议的静态段调度。

广泛相信，随着应用复杂性的持续增加，因此通信需求也随之增加，静态段(static segment, ST)带宽将不满足需求，纯粹的TT协议的实现可能过于昂贵。另一方面，优先级驱动的事件触发(event-triggered, ET)协议的实现遭受通常的时间不确定性而导致较差的控制性能^[26]。Zeng等^[27]在只有关于车辆的功能和结构的部分信息可用时，用统计分析的方法计算了CAN消息相应时间的概率分布。不断增加的数据在各种应用领域中带来了信息不均衡问题^[28]。这种在控制应用中的不均衡负载问题可能导致性能降低和能量损耗。为了改善系统性能，减少系统能耗，需要考虑新的控制设计范型，这类设计方法容纳体系结构级的约束，如计算和通信资源的可用性或成本^[1, 12]，以维护负载均衡，保持全局一致性，在系统实现成本和系统性能之间取得良好的折中。

(2) 资源供应问题。资源约束、成本敏感性和时间敏感性是这类系统的重要特征^[29,30]。在这类系统中，需要将CPU和网络资源分配给遭受物理现象强制约束的竞争应用^[31]。为了实现系统的最小成本和最优性能，有限的计算资源必须尽可能有效利用。因此，在这样的系统中，多个分布式控制应用需要按照应用需求动态供应资源。一个嵌入式计算平台依赖于按需分配/释放计算资源的机制和能力，获得更多的PUs来处理高峰负载，或者释放PUs以避免资源过度供应(over-provisioning)。这种动态资源供应和管理特征称

为“弹性(elasticity)”^[32]。Dustdar等^[33]介绍了弹性过程原理，定义了成本、性能和资源作为基本的弹性维数，形成了弹性系统的基础。

最近，关于CPS中的资源供应问题的研究已经引起了关注^[34]。在文献[35]中，弹性CPS (elastic CPS, 简称eCPS)被定义为一种自适应系统，能够在运行时从计算资源到物理设备进一步添加/删除组件，使得其成本、性能和资源使用情况与负载及自身需求相一致。Moldovan等^[36]将弹性引入到集成了计算过程、人和物的信息物理生态系统，并概述了一个新的计算领域的愿景，该领域称为“弹性计算(Elastic Computing)”，涉及弹性系统的研究。Candra等^[37]设计了一种称为“弹性剖面(Elasticity Profile)”的建模工具，该工具详述了建模混合系统的弹性行为的构造方法，以便在成本、性能和资源之间进行折中。Schmidt等^[30]在支持云计算和弹性基础设施的大规模CPS系统中辨识了研究挑战、机遇和效益。文献[38]开发了一种弹性计算中间件，联结计算资源与CPSs中的可穿戴设备、移动设备和连接设备，以卸载计算负载，增强计算结果。但是，这些工作既没有考虑CPSs中的弹性量化模型问题，也没有涉及到不同负载下系统成本和性能变化问题。

在云计算中，Ai 等^[32]用连续时间马尔科夫链提出了一种系统弹性度量模型。Li^[39]开发了一种云平台分析模型，将云平台建模为一种排队系统来研究弹性，综合分析研究云平台弹性、性能和成本问题。但是，由于系统资源受限而引起的时变时延，导致这一度量模型不能直接运用于 DESs 的弹性计算中。

(3) 不确定时延问题。现代高端汽车由多达100个、具有1亿行代码的ECU组成^[40]。广泛相信，这一数字在未来将持续增加^[26,29]。典型的例子包括引擎控制单元，车身控制子系统，底盘控制系统，以及类似于自适应巡航控制的安全功能。随着汽车电子系统复杂性的不断增加，在高级控制模型及其实际实现之间的间隙不可避免地变宽。产生这种间隙的主要问题是受控对象输出的传感信号与发送到控制执行器之间端到端的时延^[41]。控制系统在平台上的实际时延随时间而变化，可能与最坏情况下的时延显著不同并且小于最坏情况下的延迟。因此，在最坏情况假设下设计平台常常浪费许多资源。

在ANCS中，控制应用对MRT t 是敏感的。例如，在FlexRay协议中，每个通信总线环由ST段和动态段(dynamic segment, DYN)组成。在ST段，任务仅仅在预定的时隙访问总线，另一方面，为了访问DYN段的仲裁，任务被分配优先级。根据它们的段，控制应用中的这些段的时间效应是变化的^[12]。这些延迟，通常是随机的，可能显著降低系统性能，甚至引起系统不稳定^[42]。在ANCS中，由TT协议产生的固定MRT问题已经在文献[1,11,12]中研究。但是，这些方法基于MRT在最好情况延迟与最坏情况延迟之间的均匀变化的假设，而导致过度保守设计，因而无法确定切实可行的设计，但是，ANCS中的变时延问题还没有被考虑。

过去的研究工作中已经分析和讨论了车载网中的MRT问题。响应时间分析(Response-Time Analysis, RTA)^[43~45]是一种功能强大、成熟和完善的方法，用来在实时系统或者网络中分别计算任务或消息的响应时间上界。Tindell等^[46]开发了具有优先级队列的CAN总线RTA方法，这一分析提供了计算所有CAN消息的最坏情况响应时间(worst-case response times, WCRTs)，确保所有消息满足其截止时间。这项开创性的工作

导致了对车载网络的MRT分析的大量研究,已基本上成为了汽车通信总线的可调度性分析的基础。在这些研究工作中,运用于MRT分析技术主要包括:①优先级和先进先出(first in first out, FIFO)队列分析^[47];②基于偏移的分析^[48];③可中止和不可中止的传输请求^[49];④整数线性规划(ILP)^[50,51];⑤概率模型^[52,53];⑥基于有向无环图(DAG)的启发式算法^[54,55];⑦实时演算^[11,18]等等。但是,这些工作并没有从理论上考虑传感器上消息生成(到达)与网络上传输(服务)之间的关系,也没有考虑MRT的概率分布问题,更没有研究每一个消息的响应时间的精确预测问题。

1.3.2 研究挑战

面向CPS的设计中,传统的设计流是将控制器设计及其实现解耦合,导致:控制算法及其实现之间语义上的间隙;在许多场合,导致非常保守的设计。那么,传统的设计方法的主要缺点如下^[26]:

(1) 控制器与其实现平台间隙变大。传统上,控制工程师很大程度上忽略了控制器实现方面,而聚焦数学模型、分析以及高层仿真,这些工程师对控制系统实现的硬件知识往往认知有限。在这一过程中,当设计控制器时,需要许多简化的假设。当评价控制律时,他们忽略了计算时间和控制消息的通信时间。当实现平台变得更复杂、高层控制模型及其实现之间的语义间隙变宽时,控制理论研究人员开始研究NCS^[7,56]。大多数NCS的研究假设网络已给定,主要考虑控制器设计问题,而相反地,CPS设计者还需要考虑网络参数的处理。

(2) 资源利用和控制性能的保守性。控制应用在许多方面不同于实时应用。一般地,实时工程师负责实现控制系统,但这些工程师仅仅具备有限的控制技术知识。在实时系统中,实时应用常常由其截止时间特征化,而控制应用大多由其性能判断。控制应用在许多方面具有特殊性,在控制设计阶段忽略这种特殊性质,控制算法的嵌入式实现常常导致在资源利用和控制性能两方面的保守性。CPS设计范型专门用于开发这种控制应用,以改善设计方法。

CPS设计包含控制理论和嵌入式系统两方面的挑战:

(1) 控制设计挑战。一般地,动态系统(或被控对象)由微分方程建模:

$$\begin{aligned}\dot{x}(t) &= A_c x(t) + B_c u(t), \\ y(t) &= C_c x(t),\end{aligned}$$

其中, $x(t)$ 是状态向量, $u(t)$ 是控制输入, A_c 是系统矩阵, B_c 是输入向量。在状态反馈控制算法中,控制输入 $u(t)$ 由系统状态 $x(t)$ (反馈信号)计算,计算出的 $u(t)$ 用于被控对象,认为实现了期望的行为。整个闭环完成三种操作:

① 度量。动态系统输出 $x(t)$ 由一个或多个度量设备(或传感器)度量,度量信号就是反馈信号。

② 计算。反馈信号与期望的输出进行比较,在输入信号 $u(t)$ 必要的纠正由控制算法计算。

③ 执行。通过改变输入信号或者执行进行纠正。

典型地,这种反馈控制高层目标针对实现期望的行为,包括①快速,②精确,同时保证闭环稳定性。反馈控制系统的速度和精度量化了控制性能。性能概念大体上分为两类:稳态和暂态。暂态性能指标是驻留时间,即系统响应某些外部干扰有多快。另一方面,通常使用的稳态性能指标是二次代价函数,是跟踪精度和输入能量的组合,即:

$$J = \int_0^{\infty} [u(t)^2 + y(t)^T y(t)] dt,$$

其中, $u(t)$ 和 $y(t)$ 分别是系统的输入和输出。

在实现平台上连续执行这些操作,需要无穷的计算功率。因此,在这种反馈环数字实现平台上,这些操作仅仅在离散时间间隔(或采样时刻)执行,当两个连续采样时刻的时间间隔是常数时,连续时间动态系统能变换成如下的离散时间系统:

$$\begin{aligned} x[k+1] &= Ax[k] + Bu[k], \\ y[k] &= Cx[k], \end{aligned}$$

其中, $x[k]$ 和 $u[k]$ 分别是 $x(t)$ 和 $u(t)$ 在采样时刻 $t=h_k$ 的值。该系统的性能依赖于:①采样周期 h ,②传感器到执行器之间的时延 t ,如图1.5(b)。

因而,控制设计挑战依赖于:①较小的采样周期以改善控制性能;②较小的采样周期需要占用较高的计算和通信资源^[57,58]。那么,控制设计的总目标是选择合适的采样周期和相关的控制参数,使得:①控制功能是鲁棒的,且提供期望的性能;②在给定的资源约束下,所有的控制和其它任务与消息是可调度的。

(2) 平台设计挑战。在汽车CPS中, FlexRay总线的引入允许ECU和总线完全地同步调度。相对于占主导地位的CAN总线,在较好的配置情况下,同步TT调度明显地减少了端到端的时延,并在同样的时间避免了时间抖动。TT系统的配置需要所有ECU和总线的调度同时确定。在汽车CPS中,调度必须额外感知基础应用及其约束,被控对象及其特征的一种适当抽象有必要支持控制感知调度。CPS的性质必须被提取,并尊重一种模块调度框架,该框架从组件模型确定系统模型,最后确定整个系统的可行调度。

CPS的调度需要适当的应用模型。在汽车CPS中,所有的应用常常是周期性的,或由最小到达时间定义,调度目的是定义每一个任务的开始时间和到时隙的消息分配。当必须遵守所有应用的端到端时延的同时,ECU和总线的约束必须执行。依赖于低层的信息物理系统,应用程序的延迟也会影响应该考虑的控制质量。

随着系统灵活性的增加,汽车网络的优化正成为一个主要挑战。这里,以下组件必须确定:硬件资源及其配置(CPU、RAM等)的选择、总线的选择、任务和消息的分配允许可行的调度。设计空间探索任务通常分为:①资源分配;②任务映射;③消息路由。为了使汽车CPS能够使用设计空间探索方法,必须考虑控制理论和物理系统的特定约束。现有的体系结构设计空间探索方法和同步系统的调度方法的结合是一项具有挑战性的任务,需要整个系统的整体建模。开发整个系统的调度方法是一项具有挑战性的任务,通常基于ILP求解器。特别地,实现高效率和模块化是整个系统调度框架的重要设计目标,尽管当前的框架能处理100多个任务,但是它们的可扩展性一直备受关注。

本文研究仲裁网络结构中的 CPS 通信模型(或控制应用)优化问题,关注 CPS 中的网络参数与控制器参数优化。通过配置总线通信协议,从控制器调度、资源动态分配以及不确定 MRT 的分析问题等方面聚焦总线调度策略、总线协议参数以及 ECU 的调度参数,同时考虑系统成本最小和性能最优。

1.4 研究内容及主要贡献

1.4.1 研究内容

这种 ANCS 体系结构是一种由多个单输入单输出(single input single output, SISO)系统组成的多输入多输出(multiple input multiple output, MIMO)系统,每个控制应用相对独立,仅仅是共享通信总线。关注 CPS 中网络通信与控制协同设计中的若干问题,考虑共享通信总线的参数和控制器的参数的协同计算,而不考虑系统稳定性分析。研究如下三个问题:

(1) 控制器调度问题。为了系统实现成本和系统性能之间取得良好的折中,考虑新的控制设计范型。FlexRay 网络支持单信道配置和双信道配置,在双信道配置中,一些节点连接到两个信道,而另一些节点连接到一个信道^[59]。本文扩展文献[12]中提出的 ANCS 系统为一种具有双信道的混合总线-星型拓扑结构系统(见[60]中的图1-10),提出一种新的双信道 ANCS 体系结构。在这一系统中,基于分层建模理论和通信系统中的唤醒机制(见文献[59]中的图2),建立了一种扁平半休眠多控制器(flat semi-dormant multi-controller, FSDMC)模型,以维护负载均衡,保持全局一致性。本文中,为了在系统实现成本和系统性能之间取得良好的平衡,关注 FSDMC 中的所有控制器的同步调度问题。

(2) 资源约束下的资源动态供应问题。在 FSDMC 模型中,存在需要解决的两个主要问题:资源过度供应和系统超载。多个控制器的同步调度导致资源过度供应而招致不必要的成本浪费,而且,在某些情况下,由于不可预期的物理环境变化产生不期望的负载高峰。根据外部动态环境动态缩放 FSDMC 系统中的资源,以改善资源利用率。考虑在 FSDMC 模型中的控制器异步调度的情况下 ECU 计算资源的水平供应和垂直供应问题,以实现系统成本性能率(cost-performance ratio, CPR)最小化。

(3) 不确定时延问题。车载网络的 MRT 分析问题是一种关键和富有挑战的问题,已经在近十年引起了广泛的研究关注。在汽车电子系统中,例如,自适应巡航控制,主动安全功能通过雷达和摄像头采集 360° 环境视图,在产生驱动信号之前需要几个处理阶段,包括传感器融合、物体检测、控制和仲裁层^[53]。如果检测到物体,系统将降低巡航控制系统的目标速度,直至其与检测到的障碍物的速度相匹配。在这样的系统中,可以定义硬截止时间,并且 WCRT 可以防止碰撞。因此,必须尽可能准确地分析消息的 WCRT 以获得响应时间的安全上限。通过比较 WCRT 和相应的截止时间,可以验证消息的可调度性^[61]。但是,在许多时间关键型和非时间关键型应用程序中,控制性能和控制质量依赖于任务或消息的平均响应时间,这些时间必须进行分析 and 最小化。最坏情况事件的概

率可能小于硬件组件故障的概率,在这种情况下,针对最坏情况保证性能的设计可能是浪费的。因此,最坏情况的评估可能不充分,必须通过概率分析来补充^[27]。本文考虑最优最小时隙(minislot)下的总线配置,在单信道 ANCS中,分别计算具有不确定延迟的MRT的概率密度函数(probability density function, pdf)和累积分布函数(cumulative distribution function, cdf)的闭式表达式,并使用灰色系统理论预测延迟。

1.4.2 主要贡献

(1) 提出了一种新的双信道 ANCS 平台体系结构,基于分层建模理论和通信系统中的唤醒机制,建立了 FSDMC 模型。

(2) 构建了一种 N 策略、部分服务台同步多重工作休假(synchronized multiple working vacations, SMWV)的有限容量的 M/M/c 排队系统(N/(d, c)-M/M/c/K/SMWV),其中 c 表示控制器总数, d 表示半休眠控制器数, K 代表系统容量。在基于 ET 协议的柔性时分多路复用/固定优先级(flexible TDMA/Fixed Priority, FTDMA/FP)^[50]的总线调度策略下,用该排队系统建模 FSDMC 模型为拟生灭过程(quasi-birth-and-death, QBD),获得系统各种性能指标。通过求解多目标优化模型,并运用最优停止算法得到系统参数最优值以及成本、性能最优值。

(3) 提出了滞弹性系统中的滞弹性计算模型和计算方法,给出了 FSDMC 模型中的滞弹性量化、正式、新的定义,用随机环境下的 M/M/1/K 排队系统建模 FSDMC 模型,精确计算滞弹性值,衡量系统中计算资源动态供应的自适应缩放能力,以保持 CPR 最小值不变。

(4) 构建了一种 N 策略、部分服务台异步多重工作休假(asynchronous multiple working vacations, AMWV)的有限容量的 M/M/c 排队系统(N/(d, c)-M/M/c/K/AMWV),开发了一种最优化模型,最小化非线性 CPR 函数,获得统参数最优值。提出了一种基于 PSO 的最优 CPR 不变自适应算法,动态调整控制器服务率。

(5) 考虑单信道 ANCS 中的不确定时延问题,用具有随机大小批量到达的连续时间 M/PH/1 排队系统分析控制应用中消息的不确定响应延迟。将 ANCS 中的每个控制应用建模为 QBD 过程,在基于 ET 协议的多层柔性 TDMA/FP 总线调度策略下,构建约束优化模型,协同设计总线和控制器,获得 minslot 最优值 and 控制器服务率最优值,并计算每个消息响应时间的 pdf 和 cdf 的闭式表达式。

(6) 提出了一种两参数概率变换灰色预测模型(probability transformed grey model(1,1) with two parameters, PTGM(1,1)),即一种新的灰色预测模型(grey model(1,1), GM(1,1)),以精确预测每个消息的响应时间。提出了一种基于 Davidon-Fletcher-Powell (DFP)变尺度法^[63,64]的 PTGM(1,1)两参数优化算法,以最小化平均绝对误差百分比(mean absolute percentage error, MAPE)^[65],并获得每一消息的背景值的两参数最优值。

1.5 本文构想

本文致力于研究仲裁网络结构中的 CPS 通信模型(或控制应用)优化问题,关注 CPS 中的网络参数与控制器参数优化,以改善系统性能、减少系统成本为目的,考虑多控制器调度、资源动态供应以及不确定 MRT 的分析等问题。将排队论运用于系统建模与分析中,建立最优化模型,获得共享通信网络及控制器(ECU)的参数与系统性能指标。各章内容概括如下:

第二章从控制与平台协同设计、资源竞争问题、时延问题、丢包问题、基于无源性的组合设计问题等方面综述了 CPS 中的控制与通信协同设计的研究现状,指出了存在的问题,介绍了本文需要的基本概念、算法和总线协议等。

第三章提出了一种新的双信道 ANCS 体系结构,建立了一种 FSDMC 模型。构建了一种新的 $N/(d, c)-M/M/c/K/SMWV$ 排队系统,将 FSDMC 建模为 QBD 过程,求出排队系统的平稳概率分布,并获得系统各种性能指标。考虑系统成本和性能的折中,提出了一种多目标优化模型,并设计了一种基于多目标粒子群优化(multi-objective particle swarm optimization, MOPSO)算法的调度算法,求解该多目标优化问题,用最优化停止算法获得半休眠控制器数的最优值,生成了控制系统的各种参数的最优值。

第四章考虑双信道 ANCS 中资源约束下的资源动态供应问题,提出了在 FSDMC 模型上一种滞弹性系统中的滞弹性计算模型和计算方法,给出了滞弹性新的、量化的、正式定义。构建了一种新的 $N/(d,c)-M/M/c/K/AMWV$ 排队系统,并获得各种性能指标,建立最优化模型,最小化 CPR 函数,获得系统在稳定条件下的各种参数的最优值。用随机环境下的 $M/M/1/K$ 排队系统建模 FSDMC 模型,分别生成了条件平均队列长度和 3-状态概率,获得系统滞弹性精确值。通过分析系统的高峰期行为,系统自适应调节参数能基本保持滞弹性不变。

第五章考虑单信道 ANCS 中的不确定 MRT 问题,用一种具有随机大小的批量到达 $M/PH/1$ 队列模型建模 ANCS 中的每个控制应用,得到控制应用中顾客平均数,构建系统非线性成本函数。考虑基于 ET 协议的多层次柔性 FTDMA/FP 总线调度策略,设计一种约束优化模型,最小化系统成本函数,用 DFP 变尺度法,求解得到最优 minislot 值,配置总线。用最大似然估计法,得到每个控制应用和每个消息的响应时间的 pdf 和 cdf 闭式表达式。提出一种新的灰色预测模型 PTGM(1,1),精确预测每个消息的响应时间。

最后阐述了网络通信与控制协同设计在 CPS 设计中的地位和作用,进一步强调了研究 CPS 中的网络通信与控制协同设计问题的重要性和必要性,总结了全文的创新工作,指出了网络通信与控制协同设计在 CPS 设计中存在的问题,展望了进一步的研究方向。

第 2 章 相关工作

近几年，CPS中的网络通信与控制协同设计的研究成为业界非常活跃的课题之一。本章从控制与调度协同设计、资源竞争、时延、丢包以及基于无源性的组合设计等方面综述这一课题的研究进展，再介绍本文后面需要的基本概念、算法和总线协议。

2.1 协同设计研究综述

2.1.1 控制与平台协同设计

(1) 缩小控制模型的语义与实现平台的语义之间的间隙问题。这一问题近几年来已经引起了广泛关注^{[66],[67],[68],[69],[70],[71]}，例如，嵌入式系统建模语言 ESMoL^[66]，时间触发语言 Giotto^[67]，以及其它相关的形式化方法，如 PTIDES^[68]。在这方面，控制系统的研究热点集中在 NCS 的研究上^[7,72]，其主要方法是研究在网络上通信的分布式控制器，已经解决了诸如控制采样引起的时延、丢包和；时间抖动，以及它们对控制器设计、稳定性和性能的影响。研究人员已经尝试通过对控制器和平台系统化地协同设计来缩小这一间隙^[13,69,73,74]。一些协同设计经常导致一种内在的切换系统，其稳定性用诸如平均驻留时间(Average Dwell Time)方法^[75]和共同的二次李雅普诺夫函数^[12,76]进行分析。文献[69]通过指定采样周期和时延生成了稳定控制器，提出了协同设计实验方法，用 Jitterbug^[70]工具箱、开放式 T-Res 框架^[77]、基于 SystemC 的虚拟样机^[78]等的仿真实验结果表明所提出的算法产生了所期望的 QoC，并满足调度约束。文献[79]研究了超负载通信网络上分布式嵌入式系统的容错控制合成与验证。将系统体系结构建模为一种事件计数自动机，验证截止时间上界，用一种容错控制策略设计鲁棒稳定控制器，提高了系统的 QoC。但是，这种协同设计既没有提供分析保证，也没有生成来自调度器参数的消息时延的闭式表达式，更没有考虑混合协议。如果对协同设计作出任何改变，整个控制和调度设计过程需要重复进行。

(2) 控制与调度协同设计。文献[71,73]提出了一种类似的顺序协同设计方法，首先用时延脉冲建模方法设计控制系统，然后选择了一种基于最早截止时间优先(earliest deadline first, EDF)调度策略的物理体系结构，实施一种可调度性分析方法以确保所有的相关控制消息能够在有限时间内传送。如果可调度性不可行，设计过程需重新开始。值得注意的问题是设计方法没有提供分析上的保证，也不是增量式的，整个设计周期在系统将来升级的情况下需要重复进行。文献[73]依据采样时间和时延、映射控制环到计算/通信硬件组件以及调度分析方法，解决通信控制单元网络上的分布式控制系统的分析和实现问题。在文献[80]中，为了确定共享 WirelessHART^[81]网络的反馈控制环的最优采样率，研究了控制与调度协同设计问题，在确保所有数据流满足端到端的截止时间的同时，最小化整个系统的控制成本，并用多种方法求解了 WirelessHART 网络中时延有界的约

束优化问题。文献[82]采用接口变量集提出了 HVAC 系统的控制算法和嵌入式实现平台之间交互的协同设计分析方法,优化了满足用户舒适度的系统节能省钱问题。但是这些工作或者没有考虑映射控制任务到控制器的问题,或者没有关注如何确定控制器的调度参数和通信总线的调度参数问题,因而不能直接用于 CPS 设计。

(3) **控制和体系结构协同设计**。如何使 CPS 节点有效协作以完成更多计算任务是一个非常具有挑战性的问题^[83]。控制和体系结构协同设计可以使系统适应环境和负载的变化,甚至适应因重配置或者失败引起的系统体系结构的变化,其结果可以提高系统性能和更高的资源利用率^[26,84]。在过去的二十年中,人们已经探索了 CPS 中控制性能和体系结构设计之间的关系^[1,11,12,74~93],大多数文献的思想是用适当的方式重新设计体系结构,以实现更好的控制性能。文献[74,85]用一种基于闭环状态的策略确定非周期采样序列,并用相应的调度策略确保期望的控制性能。文献[86,87]考虑事件触发控制系统的闭环性能和通信负载之间的折中,提出了一种基于闭环输出的反馈策略保证系统的稳定性和期望的性能。另一些基于事件的策略用于文献[88,89]中,以确定采样序列保证系统的稳定性。文献[90,91]提出了基于事件的控制和调度协同设计方法,通过求解 LMI 最优化问题,最小化二次代价函数,提高具有有限带宽和计算资源的分布式嵌入式控制系统的控制性能。与文献[90]不同的是,文献[91]求解了基于事件的鲁棒协同设计和随机协同设计问题,确保系统的全局一致实际稳定性。在网络化嵌入式系统中,文献[92]设计了一种高层抽象的汽车分布式嵌入式系统体系结构,研究了基于高速输电线通信协议的系统性能评估方法,所提出的事务级建模方法能有效地获取实时图像处理体系结构特征并能正确地校正体系结构参数,以满足系统期望的性能。文献[79]研究了在故障或严重受限的通信网络上封闭分布式嵌入式控制系统的合成问题,建模分布式嵌入式体系结构为事件计数自动机网络以验证截止时间错失边界。为了改善 QoC,通过引入新的容错控制策略,设计了鲁棒稳定控制器。文献[93]用混合系统工具设计了一种非线性系统的事件触发稳定性框架,因为其能减少通信量和控制需要的计算资源而特别适合嵌入式系统和 NCS。文献[94]提出了一个联合优化框架,在网络化 CPS 中使用模拟退火算法,通过控制和通信协同设计在线优化控制时间间隔。但是,这些文献没有考虑到时延、丢包或时间抖动对系统性能的影响,也没有考虑共享通信资源的消息仲裁机制。

2.1.2 资源竞争

在CPS中,当组成一种新的服务时,由于资源的内部异构性因而服务的异构性,供应需要的资源是极具挑战性的。CPS中的与服务组合和资源供应相关的问题有待研究^[34]。资源管理和建模方面的资源供应问题很少被研究。Gunes等^[95]关于CPS的概念、应用和挑战给出了详细的综述,指出一个高度可缩放系统应该提供分散和聚集机制来平衡负载,提供有效的通信协议以改善性能。

(1) **资源优化供应**。Ravindran和Li^[96]提出了两种分别称为主动截止时间驱动和主动松弛驱动的主动资源分配算法,在异步实时分布式系统中主动分配资源以最大限度地满足在预期负载下未来时间间隔的总截止时间满足率。Marti等^[97]基于被控对象的反馈信

息, 为控制任务开发了一种命名为“Draco”的动态资源分配系统, 并将资源分配问题定义为优化问题, 在运行时重新分配计算资源, 最大化控制性能, 同时最小化计算开销。Tan等^[98]开发了无线网络中基于效用的资源配置算法, 在用户中分配有限资源以满足用户特定的服务质量(QoS)需求。为了解决自触发控制应用的计算资源的低效使用问题, Samii等^[99]提出了一种基于软件的中间件组件, 用于调度和优化单处理器平台上的多个自触发控制回路的控制性能和CPU利用率。Lozoya等^[100]提出了一种性能评估框架, 允许评估最新资源/性能感知策略是否能够在实际中实现。该框架的应用表明, 计算资源的在线管理是使嵌入式控制系统能够适应不同需求, 同时提供具有可调节资源利用率的特定控制性能的关键。Chantem等^[101]考虑弹性任务模型, 在出现不确定性时选择自适应周期实时系统任务周期。通过调整控制周期, 动态分配控制任务中的计算资源, 最大化负载管理性能, 同时仍然满足可调度性需求。但是, 这些工作仅仅考虑了资源分配性能优化问题, 没有涉及到成本有效性问题。

(2) **成本感知的 CPS 资源供应。**文献[102]研究了无容错异构嵌入式系统中可靠并行应用资源消耗成本最小化问题, 该问题包括两个子问题: 满足可靠性目标和最小化资源消耗成本。Chang 和 Chakraborty^[103]开发了汽车控制系统设计方法, 考虑包括存储、计算和能源资源的给定数量的资源, 旨在改善控制性能, 降低系统成本。为了避免服务台资源利用不足, 文献[104]用中间件实现并提供了基于外部协调器的弹性线程管理方案, 该中间件表明要更好地适合系统动态行为, 需要以低成本的方式支持变化数量的客户端。文献[105]研究了异构分布式多核嵌入式系统中的能量有效性和有效资源利用的联合优化问题。为了处理信息物理网络系统中快速能量均分问题, Zeng 等^[106]开发了三种新的混合稳定技术, 并建立了一种基于优化的网络拓扑设计框架。文献[107]综述了 CPSs 中的设计技术与应用问题。这些工作仅仅考虑系统成本问题, 忽略了资源利用不足的系统性能问题和系统资源供应的量化分析问题。

(3) **利用云平台的CPS资源供应。**以不同的方式利用云计算资源扩展CPS的功能提供了几种新的机遇。在2013年, 国家标准和技术研究所(NIST)定义了信息物理云计算体系结构框架为“一种能够快速构建、修改和提供自动扩展的信息物理融合系统的系统环境, 由一组基于云计算的传感器、处理器、控制和数据服务组成。”^[108]。Gai等^[109]关注CPS中的负载调度问题, 提出了一种解决因异构云计算瓶颈引起的限制问题, 所提出的机制称为基于云的智能优化负载模型, 用于预测云容量, 并考虑在动态网络环境中给异构云分配任务的可持续因素。

(4) **集成SDN与CPS的资源供应。**为了避免浪费资源, 或者避免不能及时传递正确的分析结果, Fu等^[110]提出了一种动态资源缩放框架, 包括一种基于Jackson开放式排队网络理论的精确性能模型, 能够处理基于云的实时流数据分析系统的任意算子拓扑结构。文献[111]综述了CPS与云计算集成研究进展。资源供应、调度和不同的服务是CPS云平台可能的研究实例。软件定义网络(Software-Defined Networking, SDN)范型被认为是克服CPS中的最优资源配置问题的关键^[112]。文献[113]提供了资源优化的SDN设计方法, 以实现CPS潜在的时间和空间域资源的最小资源利用率, 以及用于数据交换的所有

可能的共享无线/有线网络基础设施,并且进一步避免最坏情况分组拥塞,以便将非实时数据交换的性能追踪到最佳状态。但是,CPS云和软件定义信息物理网络的开发技术达到一定成熟度,还需要走较长的路。

(5) **考虑故障弹性(resilient)的 CPS 资源供应。**一种故障弹性系统被定义为一种自适应系统,该系统在内部或者外部发生异常时,能够通过修正其正常行为来维持和恢复其功能^[114,115]。针对分布式混合关键级 CPS 的意外负载高峰,Lakshmanany 等^[116]开发了一种称为“延展性(ductility)”的正式过载弹性度量,这种延展性包含一种称为“过载压缩打包”的延展性最大化打包算法,以最大化高关键级任务对过载的容忍度。Woo 等^[117]将正式的软件工程与一套反馈控制律相结合,为 CPS 可靠控制软件的设计和实现定义了一个协议,这些 CPS 系统在资源故障和/或可变资源配置的情况下必须具有弹性。Pereira 等^[118]从效率、鲁棒性和稳定性方面考虑分布式复杂系统的弹性问题,通过提出信任和风险管理机制来调节系统行为,增强多 Agent 系统固有的鲁棒性,从而为大规模 CPS 系统提供更具弹性的解决方案。但是,这些技术因“resilience”和“elasticity”的不同而不能运用于度量系统弹性(elasticity)。

2.1.3 时延问题

(1) **优先级和FIFO队列的消息时延分析。**在车载网的各种相关研究中,MRT分析是最重要的工作之一。Davis等^[119]发现文献[46]的开创性分析在某些情况下是有缺陷的,于是,他们驳斥和重新审视了原始分析,并提供了能用于计算CAN总线中的正确WCRT的修正分析方法。但是,这一工作并没有涉及到最优总线配置下的不确定MRT的分析。此后,Davis等和Mubeen等分别扩展了优先级和FIFO队列(见文献[47]中的图1)、基于偏移的分析(见文献[48]中的图1)、可中止和不可中止的传输请求(见文献[49]中的图1)。Mubeen等^[47]扩展了开创性的分析^[46,119],用优先级和FIFO排队模型支持CAN总线中的时间触发和事件触发同时发生的混合消息WCRT计算。Davis等^[120,121]扩展了CAN总线的分析,有些节点实现优先级队列,另一些节点实现先进先出(first in first out, FIFO)队列,支持具有任意截止时间的CAN消息分析。Schmidt等^[122]提出了一种改进算法,加速CAN总线中的具有优先级队列和FIFO队列的WCRT计算,并处理了通过新消息扩展现有的CAN消息集这一实际情况。Xie等^[123]为网关集成的CAN消息提出了一种探索性的WCRT计算方法。这些工作集中在优先级和FIFO队列问题上,并为CAN消息提供了响应时间分析和适当的优先级分配策略,且WCRT计算方法也得到了发展。然而,这些工作既没有从CPS的角度研究不确定性MRT分析问题,也没有从排队理论的观点对优先队列或FIFO队列问题进行建模,导致无法量化消息的动态行为。

(2) **具有偏移的消息时延分析。**为了避免由于高瞬时负载造成的截止时间错失,汽车嵌入式系统时常是具有偏移的调度^[124]。文献[125]研究了具有偏移的CAN消息WCRT分析。Mubeen等^[126]扩展了已有的CAN总线分析^[127],支持因限制有关消息抖动和截止时间的具有偏移的混合消息调度。Mubeen等^[128]忽略了这些限制,并通过对具有偏移量的CAN消息进行分析来扩展混合消息的分析。Mubeen等介绍了他们正在进行的工作,以

扩展CAN中具有可中止传输缓冲区的混合消息的RTA^[129]，并扩展这一工作，以支持CAN中具有偏移的混合消息的RTA^[48]。这一扩展的分析可应用于用来传输周期、零星以及混合消息的CAN的任何高层协议，其中周期和混合消息可以用偏移进行调度。具有偏移的WCRT的分析为架构师提供了一系列值，可用于将他的评估基于假设分析和集成过程。然而，最坏的情况评估是不够的，需要通过概率分析来补充。

(3) 时延概率分析。在实时系统研究中，很多研究人员进行了MRT的概率分析。一种任务响应时间分析的最初方法是实时排队论^[130]。任务集被建模为具有泊松到达和指数服务时间的Markov过程。给定一个调度算法和截止时间分布，计算直到作业截止时间的剩余时间分布。但是，这种方法是近似的，因为它基于非常繁忙的通信条件假设和有较高利用率($\rho \rightarrow 1$)的一组非常大的任务集合^[53]。DES中的流量很小，系统利用率低。Kim和Shin^[131]用连续时间Markov链建模一个应用为FIFO排队网络，考虑指数分布的任务执行时间。然而，这些工作没有明确地解决处理器负载的不确定性，也不容易扩展到网络调度或仲裁。

对于CAN总线的消息调度，Navet等^[132]因消息传输错误而引入了最坏情况截止时间错失概率(worst case deadline failure probability, WCDFP)的概念。WCDFP服从广义泊松过程，其随机参数值依据关键时刻(即WCRT场景)而获得。在文献[133]中，Nolte等当考虑给定数量的填充位的概率时，用不确定的消息传输时间扩展了WCRT分析，并在可靠性和时间线之间作出折中，但这一分析仍然在最坏情况场景执行。Broster等^[134]进行了在最坏情况场景下CAN消息延迟的随机网络错误影响的概率分析。Manolache等^[135]考虑具有给定的广义概率分布的可变执行时间的任务，提出了一种获得任务的期望截止时间错失率(Deadline Miss Ratio, DMR)的方法。但是，这一工作假设任务执行时间服从广义概率分布，任务执行截止时间是任意给定的，导致分析结果不精确。Tanasa等^[52]计算每个消息的DMR，这一工作依赖于混合整数线性规划以构建转移图，并用基于GPU的引擎计算DMR。但是，这些关于CAN总线消息调度分析的研究是基于最坏情况场景的。CAN中MRT的概率评估在这些技术中也没有得到充分的解决，特别是在存在不确定性的情况下。

(4) 随机分析与统计分析。在文献[136,137]中，Dáz等提出了一种随机分析框架，计算固定优先级系统的任务响应时间的概率分布，其中任务是周期性的、可抢占的，并用已知的偏移量激活。该文献计算了独立周期任务集的响应时间的概率质量函数，这些任务通过一种抢占式的基于优先级的调度器运行在单个处理器上。所有任务实例的激活时间是已知的，任务执行时间由概率质量函数定义。然而，这些分析假定了任务的(相对)激活时间的知识并且以悲观的方式估计由于没有抢占和释放抖动而造成的阻塞延迟。Zeng等^[53]扩展了先前关于CAN消息软件任务响应时间随机分析的工作，并将这一工作与采样时延比较，计算端到端延迟的概率分布。作者为每个远程节点使用一个特征干扰消息来建模由消息引起的干扰。这项工作将远程特征干扰消息的排队模型从具有随机初始偏移的确定性周期性改变为确定性偏移、周期性激活时间和随机排队抖动。特征干扰消息和随机抖动的引入导致系统建模和分析不精确。Zeng等^[27]在汽车中关于功能性和

体系结构仅仅部分消息可用时,用统计分析的方法计算CAN中MRT的概率分布。他们使用退化分布和伽马分布的概率混合模型来拟合CAN总线的MRT,并且使用回归分析来获得MRT的cdf的表达式。这种基于回归分析的概率分布拟合方法导致结果不精确。文献[138]给出了CAN总线中的MRT分析的综述。

2.1.4 丢包问题

文献[7,12,139,140]探讨了 NCS 中的丢包问题,共同的方法是用一种采样和保持框图分析系统的稳定性。但是,这些文献中很少处理协同设计问题。文献[12,139,140]在每一成功传送的数据包之后,采用了一种基于丢包最大数的丢包频率度量标准。但是,这些研究因每一成功传送的数据包需要补偿多个丢失的数据包而导致保守性分析。

实时系统研究人员研究了主动丢包问题^[141~147],设计了几种任务模型,使得丢包成为可能。文献[142]根据网络时延状况和采样值的偏差量,提出了一种基于变采样周期的无线网络多控制回路系统的控制策略,根据主动丢包特性对系统的离散化模型进行了稳定性分析。文献[143]提出一种基于时间片划分的异步采样法,研究多回路共享网络的MIMO网络控制系统,采用主动丢包法,将长时延问题转换成数据丢包问题,给出了系统的均方稳定条件以及控制器设计方法。文献[141]针对 FlexRay 网络控制系统的调度机制,提出了基于主动丢包的切换策略来减少网络时延对控制性能的影响。但是,这些文献没有考虑到连续多包丢失对系统稳定性的影响。文献[144]提出了(m;k)-firm 任务模型,对于任意 k 个连续数据包,多达 k-m 个包超过截止时间而被丢弃。文献[145~147]提出了几种分析(m;k)-firm 任务模型的方法。但是,他们仅关注给定的(m;k)-firm 任务调度,而没有分析丢包数的最大上界问题。

文献[139]考虑了在缓冲控制机制缺乏的情况下,如何计算最大丢包数问题。理论上,该方法能够扩展到捕获缓冲控制机制,但是这种基于自动机验证的方法用于大规模系统,需要搜索整个状态空间。文献[41]提出了 CPS 中具有信号丢失的控制与平台协同设计方法,在端到端时延大于时延阈值(threshold)时允许丢包。在时间窗口内的任意时间所出现的一定数量的丢包情况下,一种基于缓冲控制机制的平台分析方法确保丢包数量不超过控制设计所允许的最大数,多李雅普诺夫函数用于确保切换控制系统的稳定性。但是,没有给出求解连续丢包数上界的方法,也没有考虑在连续采样的滑动窗口中连续丢包和连续不丢包情况下的概率分布问题。

2.1.5 基于无源性的组合设计问题

Sztipanovits^[148]等人将 CPS 中的控制与通信协同设计问题归结为系统集成学科问题,指出:系统集成在大规模 CPS 设计中宛如“一头装在陶瓷器中的大象”。系统集成是有效设计 CPS 系统的最大障碍,主要原因是缺乏稳固的科学理论基础。他还认为 CPS 集成学科的唯一挑战来自于 CPS 系统中的组件及其相互作用的异构性。

(1) 无源性的组合设计方法。无源性是大规模 CPS 系统组合设计的重要工具,特别是系统结构具备对称性时^[148]。无源动态系统理论对于系统组合框架稳定性提供了重要

的基础^[149]。无源性对于处理网络诱导时延、数据丢失以及量化提供了显著的优点。文献[150]证明了连续时间控制器怎样互联到连续时间被控对象,并对于任意固定时延保持系统稳定性。文献[151]研究了如何将离散时间控制器联接到连续时间“端口哈密尔顿”被控对象的方法,即使在数据丢失和时变延迟出现时仍能保持系统的无源性。文献[152]提出了一种基于模型的网络化多机器人实验系统的组合设计无源性方法,确保整个系统在出现网络时延和数据丢失的实现不确定性时依然保持稳定。无源性也用于多 Agent 系统的群组协调^[153~156]。文献[153]研究了一种基于无源性的控制非线性系统的设计框架,通过内在的反馈配置表达无源性,设计过程应用于群组协调和共识问题。这一框架已经应用于同步一组刚体^[154],将编队控制与路径跟踪相结合^[155],并解决在监控和护航跟踪应用中,由不确定的网络时延和数据丢失引起的实现不确定性而导致的不稳定问题^[156]等。但是,无源性仅仅是基于系统是否耗散了足够的能量的系统行为的二元表征,存在耗散的能量多于维持系统无源性所需能量的系统。这些基于无源性的组合设计方法没有考虑网络资源竞争的仲裁(或调度)问题。

(2) 单个控制器控制单个被控对象。文献[157]研究了网络上单个控制器控制单个连续无源被控对象的问题。这种 SISO 系统通过使用波变量(wave variables)^[158]在网络上传递控制和传感器数据,允许网络出现时延和丢包时保持系统 l_2^m 稳定性($l_2^m - stable$),并指出“允许时变延迟和数据丢失的 NCS 的设计方法依赖于基于无源性的网络控制基础”。这一设计方法典型地依赖于被限制在给定区域的系统或者控制器,以保持整个系统轨迹的稳定性^[159]。但是,并不是很清楚能否将这种 SISO 系统的设计方法直接运用到 MIMO 系统中。主要研究挑战是,如何开发一种正式的方法构建控制网络,在多个被控对象和控制器之间互联,使得整个系统保持稳定并能改变被控对象的行为? Kottenstette 等人设计了一种“功率中继网络”(power-junction- network, PJN)^[160]能够解决这一问题。

(3) 多个控制器控制多个被控对象。PJN 是另一类特殊类型的 CPS,互联多个控制器和被控对象的波变量,使得总的波功率输入(wave-power-input)总是大于等于总的波功率输出(wave-power-output)^[161]。这类“输入输出波变量网络”允许单个控制器准确地控制多个被控对象,哪怕是每个被控对象相应的动态不同。这种以“功率留存”(power preserving)的方式互联的波变量网络出现在遥控装置^[162]文献中,通过无源方式修改波变量的值来增大潜在的位置漂移;另一类互联的波变量的抽象出现在波数字滤波文献[158]中,主要关注连续时间参数滤波器的结构合成规则,以构造离散时间数字滤波器。文献[163]设计了两种 PJN—平均(averaging)PJN 和共识(consensus)PJN。通过采用无源采样器和无源保持^[164]连接的两种网络使得 m 个无源数字控制器控制多达 $n-m(n > m)$ 个连续时间无源被控对象成为可能。如果所有互连的被控对象和控制器是严格输出无源的,那么该网络是 L_2^m 稳定的($L_2^m - stable$)。文献[161]提出了一种无源性基本框架以合成 $l_2^m - stable$ 的数字控制网络,在被控对象和控制器之间的通信能够容许时变延迟和数据丢失。通过使用平均 PJN,允许单一控制器精确控制多个被控对象,即使每个被控对象的相应动态不同。引入了一种“无源向下采样器”和“无源向上采样器”,以便在保持稳定性和跟

$$\sum_{k=0}^{c-1} \pi_k e + \pi_c (I - R)^{-1} e = 1. \quad (2.7)$$

这些分析可归纳为如下定理：(证明见文献[192]定理 1.7.1)

定理 2.1^[166] QBD 过程(2.2)正常返当且仅当矩阵方程(2.4)的最小非负解 R 的谱半径 $\text{sp}(R) < 1$ ，并且齐次线性方程组(2.6)有正解。当过程正常返时，平稳分布可表示为方程(2.5)，且 $\pi_0, \pi_1, \dots, \pi_c$ 由方程组(2.6)和(2.7)唯一确定。

2.3 多目标优化问题求解

多目标优化问题也称为多标准优化问题，通常一个具有 n 维决策变量、 m 维子目标的多目标优化问题可以表述如下^[167,168]：

$$\begin{aligned} \min \quad & F(x) = (f_1(x), f_2(x), \dots, f_m(x)), \\ \text{s. t.} \quad & \begin{cases} g_i(x) \leq 0, & i = 1, 2, \dots, q, \\ h_j(x) = 0, & j = 1, 2, \dots, p, \\ x \in [x_{\min}, x_{\max}], \end{cases} \end{aligned} \quad (2.8)$$

其中， x 为 n 维决策变量，目标函数 $F(x)$ 包含 m 个子目标； $g_i(x) \leq 0 (i = 1, 2, \dots, q)$ 和 $h_j(x) = 0 (j = 1, 2, \dots, p)$ 分别定义了多目标优化问题的 q 个不等式约束条件和 p 个等式约束条件； x_{\min} 和 x_{\max} 分别为向量搜索的上下限。

多目标优化问题的 Pareto 前沿(Pareto front)定义如下^[169]：

定义 2.1 支配：一个决策向量 x_1 支配另一个决策向量 x_2 ，当且仅当 x_1 在所有目标上都不差于 x_2 ，也就是说 $f_k(x_1) \leq f_k(x_2)$ ， $k = 1, 2, \dots, m$ ；并且 x_1 至少在一个目标上严格好于 x_2 ，也就是说存在某个 $s (1 \leq s \leq m)$ ，有 $f_s(x_1) < f_s(x_2)$ 成立。类似地，一个目标向量 F_1 支配另一个目标向量 F_2 ，如果 F_1 在所有目标值上都不差于 F_2 ，并且至少在一个目标值上好于 F_2 。

定义 2.2 Pareto 最优：一个决策向量 x^* ，如果不存在一个决策向量 $x \neq x^*$ 支配它，即对于任意 $k (k = 1, 2, \dots, m)$ ，有 $f_k(x^*) \leq f_k(x)$ ，则称决策向量 x^* 是 Pareto 最优的。如果 x^* 是 Pareto 最优，则目标向量 $F^*(x^*)$ 也是 Pareto 最优。

这一定义指出，如果没有可行的决策向量 x 可以减少一些目标值而不会同时增加至少一个其他目标值，那么 x^* 是帕累托最优的。

定义 2.3 Pareto 最优集：所有 Pareto 最优决策向量组成的 Pareto 最优集合。

帕累托最优解也被称为非劣质、可接受或有效的解；它们的相应向量被称为非支配向量。

定义 2.4 Pareto 前沿(Pareto front):所有的 Pareto 最优目标向量组成的曲面称为 Pareto 前沿。

在一般情况下，不可能找到包含这些点的线或曲面的闭式表达式。生成 Pareto 前沿的正常过程是计算可行点及其相应的点。当这些点的数量足够时，可以确定非支配点并产生帕累托前沿。

用MOPSO算法求解上述多目标优化问题(2.8)。PSO是一种启发式算法[62]，其灵感来源于鸟群的觅食。PSO似乎特别适用于多目标优化，主要是因为该算法为单目标优化提供了高速的收敛速度。Carlos等[170]提出了一种基于自适应档案网格的MOPSO算法，其基本思想是将目标空间分割成几个超立方体，判断每个超立方体中所包含的非劣解个数来维护外部档案。自适应网格实际上是由超立方体形成的空间。作者提出算法考虑了约束处理机制和变异算子，大大提高了PSO算法的探索能力。MOPSO算法步骤如算法 2.1。

算法 2.1 基于自适应档案网格的 MOPSO 算法

输入：种群中粒子数、超立方体数、外部档案大小

输出：外部档案REP中的粒子位置

(1) 初始化种群POP:

FOR $i = 0$ TO MAX //MAX表示粒子个数

初始化POP[i];

END FOR

(2) 初始化每个粒子的速度 VEL:

FOR $i = 0$ TO MAX

VEL[i] = 0;

END FOR

(3) 评估 POP 中的每个粒子;

(4) 将代表非支配向量的粒子的位置存储在外部档案 REP 中;

(5) 生成已经探索的搜索空间的超立方体，并使用这些超立方体作为坐标系定位粒子，其中每个粒子的坐标根据其目标函数的值进行定义;

(6) 初始化每个粒子的历史最优位置PBests:

FOR $i = 0$ TO MAX

PBests[i] = POP[i];

END FOR

(7) WHILE 最大循环数尚未达到 DO

(8) 使用以下表达式计算每个粒子的速度:

$$VEL[i] = \omega \times VEL[i] + r_1 \times (PBests[i] - POP[i]) + r_2 \times (REP[h] - POP[i]) ;$$

其中惯量权重 ω 取值为0.4; r_1 和 r_2 是区间[0, 1]内的随机数; $REP[h]$ 表示从外部档案中取一个值; 索引 h 按以下方式选择: 包含多于一个粒子的超立方体的适应度等于用任何数 $x > 1$ (例如 $x=10$)除以该超立方体包含的粒子数的高, 这旨在降低包含更多粒子的那些超立方体的适应度, 并且可以将其视为一种共享适应度的形式。然后, 使用这些适应值并应用轮盘赌来选择需要获取相应粒子的超立方体。一旦选择了超立方体, 就从这一超立方体中随机选择一个粒子 h 。POP[i]是粒子 i 的当前值;

(9) 更新粒子的位置:

$$POP[i] = POP[i] + VEL[i];$$

(10) 在搜索空间内维护粒子以防粒子超出边界。当决策向量超出其边界时, 该决策向量取其相应边界的值(下界或上界), 并且它的速度乘以-1, 以便在相反的方向继续搜索;

(11) 评估 POP 中的每个粒子;

(12) 将 REP 的内容与超立方体内的粒子位置一起更新。这一更新包括将所有当前非支配粒子插入到外部档案中。在这个过程中, 外部档案中的任何支配粒子都会被删除。既然外部档案的大小是有限的, 只要它满了, 采用一种自适应网格方法来维护外部档案: 考虑超立方体所含非劣解的密度, 从密度高的超立方体中删除非劣解, 并加入到密度低的非劣解。如果插入到外部档案中的粒子位于网格的当前范围之外, 那么网格必须重新计算, 并且其中的每个粒子都必须重新定位;

(13) 当粒子的当前位置比其历史最优位置更好时, 使用下式更新粒子的位置:

$$PBests[i] = POP[i];$$

如果当前位置由历史位置支配, 则历史位置保持不变; 否则, 用当前位置替换这一历史位置; 如果它们中的任何一个都不被另一个支配, 那么 随机选择其中之一作为粒子的PBests;

(14) 增加循环计数器;

(15) END WHILE

2.4 灰色预测模型

由邓聚龙教授提出的灰色系统理论是一种用于处理具有数据较少和信息不完全特征的不确定性问题的理论^[171]。一种灰色预测模型 GM(1,1)已经用于许多预测问题(见文献[172]及其参考文献), 诸如电力负荷短期和长期预测、旅游流、航空业、汽车生产、波浪、风速和风能、农村人均年纯收入预测、燃料生产和碳排放、能源需求和经济增长等等。

在灰色系统理论中, 随机过程被认为是在一定范围内变化的灰色变量。在灰色模型的过程中, 通过累积生成运算(accumulated generating operation, AGO)将无序初始数据管理为规则生成数据序列。然后, 采用微分拟合方法建立微分方程来描述生成数据序列的规律。最后实现预测。GM(1,1)是最常见和简单的灰色模型, 操作方便, 预测精度高, 其具体建模过程如下^[172,173]:

(1) 假设 $X^{(0)}$ 是初始序列, 表示如下:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}.$$

(2) 由初始序列 $X^{(0)}$ 生成一次AGO序列 $X^{(1)}$ 为:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}, \quad (2.9)$$

其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$.

(3) 背景值数组 $Z^{(1)}$ 可以计算为:

$$Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}, \quad (2.10)$$

其中 $z^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)x^{(1)}(k-1), k = 2, 3, \dots, n$, 以及生成系数 α 通常为0.5。因

此, 背景值 $Z^{(1)}$ 表示为: $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$ 。

(4) GM(1,1)的一阶微分方程及其白化方程分别为:

$$x^{(0)}(k) + az^{(1)}(k) = b,$$

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b, \quad (2.11)$$

其中 a 和 b 分别表示发展系数和灰色输入。

(5) 由最小二乘法, 参数 a 和 b 可以计算为:

$$\hat{U} = \begin{pmatrix} a \\ b \end{pmatrix} = (B^T B)^{-1} B^T Y,$$

其中 $Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix}, B = \begin{pmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{pmatrix}$.

(6) 白化方程(2.11)的解为:

$$\hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-a(k-1)} + \frac{b}{a},$$

其中 $x^{(1)}(1) = x^{(0)}(1)$.

(7) 序列 $X^{(0)}$ 的预测值 $\hat{x}^{(0)}$ 可以通过逆累积生成运算获得, 表示如下:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) = (1-e^a) \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-a(k-1)},$$

其中 $k = 2, 3, \dots, n$, 且 $\hat{x}^{(0)}(1) = x^{(0)}(1)$.

2.5 最优停止策略

假定 P 是有限状态空间 S 的一个离散时间马尔可夫链 X_n 的转移矩阵。假设 f 为对应于每个状态的一个收益函数，它表示当该马尔可夫链到达这一状态后停止时的收益。一个**停止规则**或者**停时**是指一个随机变量 T ，它给出了马尔可夫链在何时停止。值得注意的是：玩家 n 时刻决定停止与否仅仅与 n 时刻以前发生的状态有关，而与未来无关。因此，对于时齐马尔可夫链，不难知道最合理的与未来无关的停止规则具有如下的形式：状态空间 S 被划分为两个子集 S_1 和 S_2 ，即 $S = S_1 \cup S_2$ ，且 $S_1 \cap S_2 = \phi$ ，其中 S_1 表示在最优策略下玩家继续游戏的状态集； S_2 表示在最优策略下玩家停止游戏的状态集。我们的目的是在所有的停止规则中选择能使期望收益最大化的停止规则^[174]。令 $v(x)$ 为状态 x 的价值，即在最优停止策略下的期望收益。 $v(x)$ 可以写成如下形式：

$$v(x) = \max_T E[f(X_T) | X_0 = x],$$

公式中最大值的选取基于所有合法的停止规则。

事实上， $v(x)$ 满足：

$$v(x) = \max \{f(x), Pv(x)\},$$

其中 $Pv(x) = \sum_{y \in S} p(x, y)v(y)$ ，且 $p(x, y) = \frac{1}{c}$ 。

令

$$T = \min \{j \geq 0 | X_j \in S_2\},$$

那么

$$v(x) = E[f(X_T) | X_0 = x].$$

称函数 $u(x)$ 为对应于 P 的**上调和函数**，如果它满足

$$u(x) \geq Pu(x).$$

假设 $u(x)$ 是上调和函数， T 是对应于如上定义的停止规则的时间。考虑时间 $T_n = \min\{T, n\}$ ，那么有

$$u(x) \geq E[u(X_{T_n}) | X_0 = x]. \quad (2.12)$$

结论： $v(x)$ 是对应于 P 且大于等于 $f(x)$ 的所有上调和函数 $u(x)$ 中的最小上调和函数，即

$$v(x) = \inf u(x).$$

因为根据式(2.12), 令 $n \rightarrow \infty$, 则得

$$u(x) \geq \lim_{n \rightarrow \infty} E[u(X_{T_n}) | X_0 = x] = E[u(X_T) | X_0 = x].$$

假定对于所有 x , 有 $u(x) \geq f(x)$, 那么

$$u(x) \geq E[u(X_T) | X_0 = x] \geq E[f(X_T) | X_0 = x] = v(x).$$

因此, 每个大于 $f(x)$ 的上调和函数 $u(x)$ 都大于等于收益函数 $v(x)$ 。若 $\{u_i(x)\}$ 是任意的上调和函数集, 那么 $u(x) = \inf_i u_i(x)$ 也是上调和的。给出最优停止策略的求解步骤。

初始化上调和函数 $u_1(x)$, 即

$$u_1(x) = \begin{cases} 0, & x \text{ 为吸收状态,} \\ \max_{x \in S} \{f(x)\}, & \text{其它.} \end{cases}$$

令

$$u_2(x) = \max \{Pu_1(x), f(x)\}.$$

那么 $u_2(x) \leq u_1(x)$, 且 $u_2(x) > f(x)$, 因此, $u_2(x)$ 也是一个大于 $f(x)$ 但小于 $u_1(x)$ 的上调和函数。

定义

$$u_n(x) = \max \{Pu_{n-1}(x), f(x)\},$$

这里 $u_n(x)$ 也是一个大于 $f(x)$ 但小于 $u_{n-1}(x)$ 的上调和函数。

得到期望收益函数为

$$v(x) = \lim_{n \rightarrow \infty} u_n(x).$$

得到最优停止策略 T 为

$$\text{停止集合} \quad S_2 = \{x : v(x) = f(x)\},$$

$$\text{继续游戏的集合} \quad S_1 = \{x : v(x) > f(x)\}.$$

2.6 总线协议

车载网络技术的运用具有简化线路、降低故障频率以及提高信息传输的速度和可靠性的功能, 可以实现对车辆本身的控制。相对于传统的点到点的布线方式, 车载网络技术的运用具有更多的优点, 具有更广泛的发展空间。目前, CAN和FlexRay总线协议是两种主要的车载网络协议。

2.6.1 CAN

CAN 是德国 Bosch 公司为解决现代汽车中众多的电控单元之间的数据交换而开发的一种串行通信协议。CAN 使用带有冲突避免的载波侦听多路访问控制(CSMA/CA)机制支持事件触发的数据交互, 其因结构简单、传输高效、设计灵活、鲁棒性强等特点, 已成为应用最为广泛的一种车载网络。其典型的应用是为车辆动力控制系统中的电子控制装置提供最高传输带宽为 500kbit/s 的网络连接, 以及为车身控制系统中的电子控制装置提供最高传输带宽为 125kbit/s 的网络连接。CAN 是一个基于微控制器互连的总线型拓扑网络, 典型的总线结构如图 2.1。

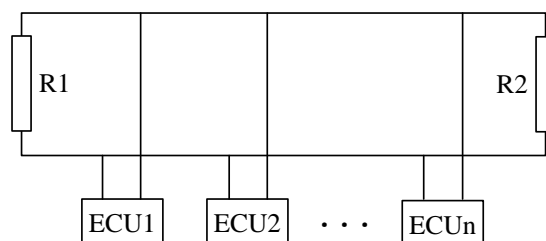


图 2.1 CAN 总线拓扑结构

CAN 总线的节点功能上主要有三部分组成, 分别是发送/接收驱动器、CAN 控制器、具体实现控制功能的 ECU(执行器)。CAN 协议遵从 ISO/OSI 网络开放系统模型, 按照这个标准模型, CAN 协议可划分为两层(国际标准 ISO11898), 即物理层和数据链路层, 数据链路层又细分为逻辑链路控制子层 LLC 和媒体访问控制子层 MAC。CAN 的其他高层协议则交给用户完成, 使得 CAN 的开发具有更大的灵活性。

CAN 是一种基于生产/消费模型的协议, 通过使用一种面向消息的传输协议来实现。在这种协议下, 网络中不对站点定义站点地址, 只定义消息。产生消息的站点称为生产者, 接收消息的站点称为消费者。消息的发送和接收通过消息标示符(ID)来管理。消息标示符在整个网络中是唯一的, 它不但定义了消息的内容, 也定义了消息的优先级。CAN 消息的具体发送机制是基于优先级竞争非破坏性仲裁机制, 即非破坏性 CSMA 机制。由于使用了生产/消费模型, 通信直接面向数据的内容而与节点地址无关, 基于 CAN 的系统往往在组合配置上具有较高的灵活性, 表现在: CAN 网络中节点的增加与去除较容易, 很容易实现模块标准化; 数个节点同时需要同一消息时可以一次就实现传输, 即一发多收; 节点通过接收过滤可以接收自己想要的信息, 可以实现点对点和多点对一点通信; 数据传输不依靠特定类型节点, 很容易实现网络的升级^[175]。

CAN 总线的时延是由其协议本身引起的, 其根本原因是 CAN 通讯的信道资源有限。由于只有一个信道, 因此高优先级的消息要优先发送, 低优先级的消息必须等待; 如果有消息正在使用 CAN 总线进行消息传输, 其它任何优先级的消息都必须等待。

目前, 解决 CAN 总线时延问题的主要方法之一是基于时分调度策略的 CAN 总线协议。利用时分调度原理来避免总线的相互碰撞, 从而使系统的实时性得到提高。这一般应用于传送周期性的实时信息。最典型的是 TTCAN 协议, 是 CAN 协议的扩展, 它在现有的数据链路层和物理层的顶端有个会话层, 该协议执行一种混合的、时间触发的

TDMA 流程，它也适用于事件触发通信。TDMA 将一个信道分成连续的周期性时隙，这些时隙在时间域里是完全正交的，使得不同的信号分配到不同的时隙里。

TTCAN 是基于主从结构的。在系统中，至少要有一个主节点作为整个系统的调度器，安排系统内所有节点的发送时间。这一策略包含两种可行的调度方法，即静态调度和动态调度。静态调度中所有节点的发送顺序的安排都是事先设计好的，在系统运行的时候不能更改，所以灵活性受到了限制；而动态调度则是在系统运行的时候动态产生调度顺序，以提高系统的灵活性，但也使系统的额外开销增大，占用较多的系统资源。但是，TTCAN 把时间片分配给每个节点，但如果存在节点通信量很少，或者在一段时间内没有通信产生时，这个时间片就被浪费。TTCAN 网络结构并不复杂，但要求极高精度的时钟同步，而且不方便于节点的随时加入和删除，在灵活性上还有待提高^[176]。

2.6.2 FlexRay

FlexRay协议^[39,40]被考虑成为将来汽车车载网事实上的标准。FlexRay是一种确定的、容错的、高速汽车总线系统。FlexRay有10Mbit/s的高速数据速率、冗余的通信信道、全局同步时间基准、可靠的消息传递、灵活地增减节点、TT和ET通信以及容错等重要性质，覆盖物理层和数据链路层。

FlexRay协议支持总线型(如图2.2)和星型拓扑结构(如图2.3)，或者两者的结合。FlexRay能够处理多达2条信道，而并不需要有相同的物理拓扑结构(如图2.4)。因此，FlexRay协议是非常灵活的协议。

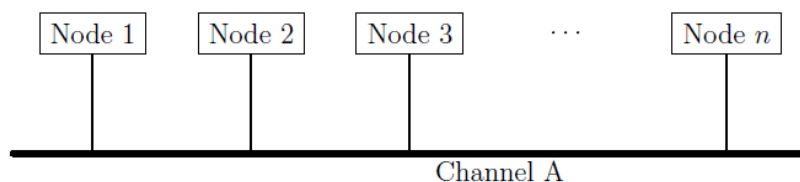


图 2.2 单个信道的总线型拓扑结构

FlexRay协议由一系列固定长度的通信循环组成。而一个这样的循环被分成静态段(ST)、动态段(DYN)、符号窗口(symbol window)和网络空闲时间(NIT)。仅仅ST和NIT是强制性的。前一个循环结束后，下一个循环立即开始，每一个循环的长度 l 在设计过程中设置，每一个循环有一个循环编号，从0到63增加，再回到0。FlexRay中的TT通信提供了高可预见性的时间行为，而ET通信提供了有效的带宽。

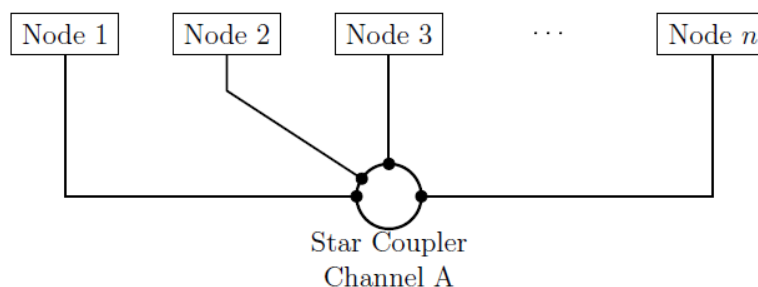


图 2.3 单个信道的星型拓扑结构

静态段的访问由TDMA方法组织。静态段被分成称为时隙(slots)的固定的、等长的时间窗口，每一个处理单元被分派一个或者多个用时隙号表示的时隙。动态段被分成比静态段时隙更小时隙的最小时隙，采用FTDMA协议。类似于静态段，最小时隙由允许消息传送的时隙号指示。但是，动态时隙依赖于在某个时隙所传送的消息大小而改变。如果在一个特定的时隙，没有消息传送，仅仅消耗一个小的最小时隙，时隙号被增加到下一个最小时隙。但是，如果在某个时隙传送消息，那么等消息完全传送完毕后，时隙号才能增加。因此，如果消息在总线上实际传送，总线资源才能被利用，否则，仅仅一个很小的最小时隙被浪费。如果很多节点在动态段发送消息，低优先级的消息可能经历非常大的时延，甚至可能在当前循环不能发送，或者被丢弃。因而，ET通信能产生随机时延，甚至丢包，极端情况下会导致连续丢包。这是值得研究的课题。

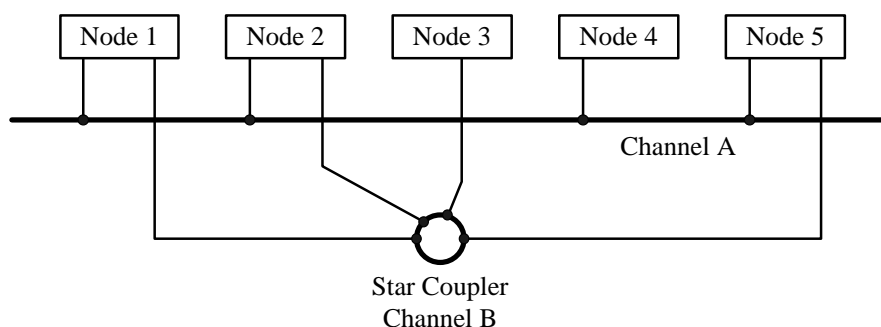


图 2.4 双信道的总线—星型混合拓扑结构

2.7 本章小结

本章从控制与调度协同设计、资源竞争、时延、丢包以及基于无源性的组合设计等方面综述了通信与控制协同设计的研究进展。但是，在这些工作中，没有考虑如下问题：ET 协议下的控制器调度问题，以达到系统成本和性能之间的折中；CPS 中时间相关的弹性量化建模问题；不同负载下系统成本和性能优化问题；消息生成与消息服务之间的关系；MRT 的概率分布；每个消息的响应时间精确预测等等，本文考虑解决这些问题。本章最后介绍了本文涉及到的基本概念、算法和总线协议。

第 3 章 扁平半休眠多控制器调度

本章考虑一种包含多个分布式控制应用的新的双信道 ANCS 体系结构(图 3.1), 设计总线调度策略, 计算控制器服务强度。运用排队论建模带有 ET 协议的 FlexRay 网络的 ANCS 系统, 同步调度 FSDMC 中的控制器, 计算系统的平稳概率分布和系统性能指标, 通过求解成本和性能函数的多目标优化问题并利用最优停止算法得出 ANCS 系统的参数最优值。

3.1 引言

随着 CPS 复杂性的持续增加, 系统中的不均衡负载问题越来越凸显, 甚至导致系统性能降低和能量损耗。为了改善系统性能, 减少系统能耗, 需要考虑体系结构级约束的新的控制设计范型。

文献[177]提出了一种 SDN 中的休眠多控制器模型, 将网络控制从转发业务中分离^[178]。在 ANCS 这一类特殊 CPS 中, 基于分层建模理论和通信系统中的唤醒机制(见文献[59]中的图 2), 设计了一种 FSDMC 模型。FlexRay 支持 ECU 的总线驱动器(Bus Driver, BD)的待机和休眠模式。唤醒符号通过发送引起 BD 唤醒本地 ECU 的模式, 将多节点集群带出低功耗状态。分布在网络中的唤醒信号由通信控制器和 BD 设备支持^[60]。考虑图 1.5 中的一种典型的分布式嵌入式体系结构。

FlexRay 网络支持单信道配置和双信道配置, 在双信道配置中, 一些节点连接到两个信道, 而另一些节点连接到一个信道^[59]。本文扩展文献[12]中提出的单信道 ANCS 系统(如图 1.6)为一种具有双信道的混合总线-星型拓扑结构系统(如图 2.4)。在这种拓扑结构中, 信道 A 是总线拓扑, 而信道 B 是一种环型拓扑。所有节点连接到信道 A, 而只有控制器节点连接到信道 B, 这些控制器构成了具有平等模式(服务具有等可能性)的扁平多控制器结构, 以维护负载均衡, 保持全局一致性。称这一 ANCS 系统为双信道 ANCS。本章中, 为了在系统实现成本和系统性能之间取得良好的平衡, 本文关注 FSDMC 中的控制器的调度问题, 应用排队论建模 ANCS 中的 FSDMC 模型, 构建一种 $N/(d, c)-M/M/c/K/SMWV$ 队列, 得到系统的各种性能指标。建立一种多目标优化模型, 最小化系统的非线性成本函数和标称时延函数, 用最优停止策略^[174]和 MOPSO 算法^[170]求解该多目标优化模型, 根据总线的平均传输速率, 协同设计控制器服务率, 动态配置系统参数。

3.2 平台体系结构

本文设计的包含多个分布式控制应用的双信道 ANCS 平台体系结构如图 3.1 所示。控制应用被分为映射到不同 PUs 的大量任务, 所有 PUs 通过共享的通信信道 A 通信,

并在一个或者多个控制应用中运行不同的任务。但是，仅仅 PUs 中的所有控制器节点通过信道 B 通信，以保持负载均衡，保持全局一致性。在这一结构中，PU1 承载负责从用户读取参考命令的任务，PU2 承载所有计算控制命令的任务，PU3 负责读取传感器的任务，负责提供被控对象命令的任务被映射到 PU4 上。这种设置将构成本文研究的技术基础。本文称这一控制器簇为 FSDMC 模型，并关注这一模型中的多控制器调度问题。

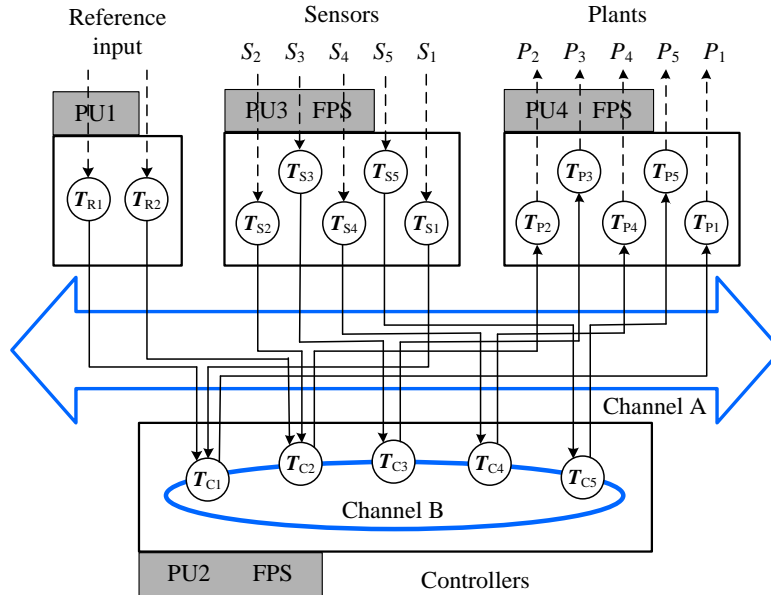


图 3.1 含有 5 个分布式控制应用的系统体系结构

在这一平台体系结构中，每一个控制器都与一个传感器和执行器相连，处理由相应的传感器经由共享通信信道 A 传送的任务。当一个控制器过载时，它将用令牌策略经由信道 B 转发部分任务给其它半休眠控制器进行处理。任务由半休眠控制器处理完毕后，将再经由信道 B 传送给相应的过载控制器，该过载控制器将经由信道 A 传送这一处理后的任务到相应的执行器。

将图 3.1 中的五个控制应用分成两类。一类包含控制应用 1 和 2(如图 3.2, 见文献[11]中的图 3)，其中包含四类任务： T_{Ri} , T_{Si} , T_{Ci} 和 T_{Pi} , $i=1, 2$ ；另一类包含控制应用 3、4 和 5(如图 3.3, 见文献[11]中的图 4)，其中包含三类任务： T_{Sj} , T_{Cj} 和 T_{Pj} , $j=3, 4, 5$ 。任务 T_{Si} 经由信道 A 从传感器 S_i 读数据并发送传感器信号 m_{Si} 到任务 T_{Ci} 。类似地，任务 T_{Ri} 经由信道 A 从用户读参考命令 m_{Ri} 并发送到任务 T_{Ci} 。 T_{Ci} 用消息 m_{Ri} 和 m_{Si} 计算控制输入，并将输出经信道 A 发送到任务 T_{Pi} ，被控对象 P_i 从任务 T_{Pi} 接收控制输入，等等。

共享通信信道 A 服从事件驱动协议的分层 FTDMA/FP 总线调度策略(图 3.4)。FTDMA 技术的构思主要是为了克服 CAN 的限制，特别支持高性能实时通信^[179]。FTDMA 调度器运行在调度器顶层，即通信带宽被分成长度为 l 的相等周期(或环)，经由信道 A 在动态时段隙 i 传送的出入控制应用的消息服从固定优先级调度策略。在某个时段 k 期间如果没有消息发送，时段 k 有一个称为 minislot 的非常小的时间长度，否则时段 k 将有一个需要传送整个消息长度且等于 minislot 个数的时间长度。在图 3.4 中，当在第一个总线周期传送消息 m_{C5} 时，时段 5 占用两个 minislot (10 和 11)，但是在第二个总线周期时，时段 5 中没有消息发送，因此时段 5 仅仅占用一个 minislot(图中“MS”)。当一个控制器处于半休眠或者

空闲状态时，相应的时隙用“MS”表示，例如控制器2和3在第二个总线周期处于半休眠状态，其相应的时隙均为“MS”。

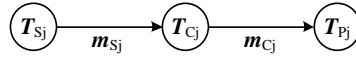


图 3.2 控制应用 $i, i = 1, 2$ 的任务图

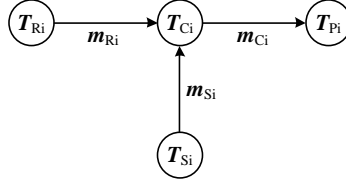


图 3.3 控制应用 $i, i = 3, 4, 5$ 的任务图

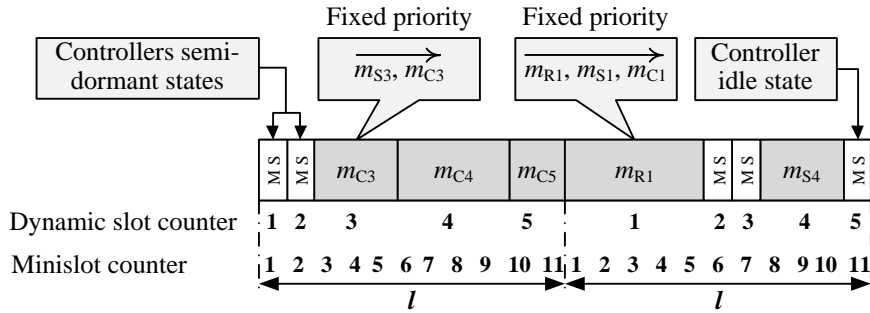


图 3.4 通信信道 A: 分层 FTDMA/FP 调度器, l 表示 FTDMA 周期长度

为了达到系统实现成本和系统性能之间良好平衡，本文构建 FSDMC 模型为一个排队模型，以计算系统的各种性能指标。从而提出一种多目标优化模型，最小化非线性成本函数和标称时延函数，并获得系统参数的最优值。

3.3 排队模型

3.3.1 模型描述

在排队系统中，N-策略是由Yadin和Naor引入的一种阈值(threshold)策略^[180]，用于控制系统的服务性能。文献[181]首先研究了具有可靠服务台的N-策略M/G/1排队系统，并证明了N-策略是一种在系统的有限或者无限范围内派生各种成本函数的最优策略。Chang等研究了一种具有单个可移动服务站的无限容量的N-策略M/G/1排队系统^[182]。一些研究工作还集中在具有单个或多重休假策略的多服务器排队系统。Zhang等^[183]和Xu等^[184]用“部分服务台休假策略”分析了M/M/C休假系统，其中一些服务台采取单个或多重休假策略。Ke等考虑一个有限缓冲的M/M/C排队系统，其中服务台不可靠并遵循(d, c)休假策略^[185]。但是，单个或者多重工作休假的排队模型在实际生活中更现实。

Servi等^[186]考虑一个在休假期间以不同的速率工作而不是完全停止的服务台，这样的休假被称为工作休假。Zhang等用QBD过程与矩阵几何解的方法研究了一种N-策略和多重工作休假的M/M/1排队系统^[187]。用类似的方法，Lin等考虑了一种单个工作休假和详尽服务的M/M/R队列^[188]。Yang等研究了带有多重工作休假和服务台故障的N-

策略 M/M/1 排队系统^[189]。Jain 等讨论了带有阻塞和同步工作休假策略的不可靠马尔可夫多服务台有限缓冲队列的建模和分析问题^[190]。但是，现有的研究工作，包括上面提及的工作，没有解决具有 N-策略和(d, c)工作休假机制的有限缓冲多服务台排队系统的优化问题。

本章构建一种 N/(d, c)-M/M/c/K/SMWV 排队系统，并详细描述这一系统。为了方便起见，称分组消息为“顾客”，控制器为“服务台”，以及半休眠状态为“工作休假状态”。在 FSDMC 模型中，每一个控制应用中的传感器采样数据和参考命令组成一个相应的子队列，c 个子队列分别由 c 个控制应用生成。CAN 总线的仲裁机制表明所有消息被传送，好像网络上的所有节点共享单个的全局优先级队列^[138]。所以，c 个子队列构成了共享总线的单个全局队列。假设从传感器到相应控制器的分组服从泊松分布，用户发送的分组也服从泊松分布(如控制应用 1 和 2)，而独立的泊松过程之和仍然是泊松过程。因此，假设到达每一个控制应用中相应控制器的分组服从速率为 $\lambda_i (\lambda_i > 0)$ 的泊松分布。每一个到达的分组在控制器是空闲的情况下直接进入控制器接受服务，否则加入到队列中等待服务。当控制器处理完成一个分组，立即传送给相应的执行器，如果队列不空，下一个分组进入控制器继续接受服务。显然，在全局队列中，分组到达的时间间隔服从参数为 λ 的指数分布，且 $\lambda = \sum_{i=1}^c \lambda_i$ 。

假设在正常忙期的服务时间是一个平均服务率为 μ_B 的指数随机变量，在工作休假期间的服务时间服从另一个平均服务率为 $\mu_v (\mu_v < \mu_B)$ 的指数分布。假设第 i 个子队列的阈值是 N_i ，容量是 K_i ，但是，一般来说， $N \neq \sum_{i=1}^c N_i$ ，并且 $K \neq \sum_{i=1}^c K_i$ ，这里 $i = 1, 2, \dots, c$ ，

$N_i = \left\lfloor \frac{N}{c-d} \right\rfloor$ ， $K_i = \left\lfloor \frac{K}{c-d} \right\rfloor$ 。假设第 i 个子队列的顾客数是 η_i ，并设 $\psi = \sum_{i=1}^c \eta_i$ 。如果

$\psi < c-d$ ，那么 d 个空闲的服务台开始同步工作休假，其中休假时间服从速率为 θ 的指数分布。当休假完成，并且 $\psi < N$ ，那么 d 个休假的服务台都开始另一个工作休假，否则，d 个休假的服务台从工作休假状态返回，继续以服务率 μ_B 服务等待的顾客，直到 $\psi \leq c-d$ 。如果 $\psi = K$ ，那么任何顾客不能进入任意的子队列。在第 i 个子系统中，如果 $\eta_i \geq N_i$ ，那么第 i 个服务台处于过载状态，它将通过信道 B 传送分组消息给处于休假状态的 d 个服务台，这 d 个服务台以速率 μ_v 处理这些消息。

在一般的周期性实时系统中，考虑周期任务为一系列相互独立的任务。每一个任务产生一个作业无限序列，假定每个作业都独立于同一任务的作业和其他任务的其他作业

$$B = \begin{pmatrix} \sigma_d & 0 \\ 0 & \delta_d \end{pmatrix}, \quad N \leq k \leq K,$$

$$C = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad N \leq k < K,$$

和

$$A_k = \begin{cases} -(\lambda + \delta_{k-c+d}), & 0 \leq k \leq c-d, \\ \begin{pmatrix} -(\lambda + \sigma_{k-c+d}) & 0 \\ 0 & -(\lambda + \delta_{k-c+d}) \end{pmatrix}, & c-d < k \leq c-1, \\ \begin{pmatrix} -(\lambda + \sigma_d) & 0 \\ 0 & -(\lambda + \delta_d) \end{pmatrix}, & c \leq k \leq N-1, \\ \begin{pmatrix} -(\sigma_d + \theta) & \theta \\ 0 & -\delta_d \end{pmatrix}, & k = K, \end{cases}$$

$$B_k = \begin{cases} \delta_{k-c+d}, & 1 \leq k \leq c-d, \\ \begin{pmatrix} \sigma_1 \\ \delta_1 \end{pmatrix}, & k = c-d+1, \\ \begin{pmatrix} \sigma_{k-c+d} & 0 \\ 0 & \delta_{k-c+d} \end{pmatrix}, & c-d+2 \leq k \leq c-1, \\ \begin{pmatrix} \sigma_d & 0 \\ 0 & \delta_d \end{pmatrix}, & c \leq k \leq N-1, \end{cases}$$

$$C_k = \begin{cases} \lambda, & 0 \leq k < c-d, \\ (\lambda \ 0), & k = c-d, \\ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, & c-d < k \leq N-1. \end{cases}$$

为了分析这一 QBD 过程, 需要找到式(3.1)所示的矩阵二次方程的最小非负解, 这一解称为率阵^[192]。

$$R^2 B + R A + C = 0 \quad (3.1)$$

定理 3.1 如果 $\rho = \lambda \delta_d^{-1} < 1$, 那么矩阵方程(3.1)有最小非负解

$$R = \begin{pmatrix} r & \frac{\theta r}{\delta_d(1-r)} \\ 0 & \rho \end{pmatrix}, \quad (3.2)$$

其中 r 是下面二次方程(3.3)的一个实根

$$\sigma_d z^2 - (\lambda + \sigma_d + \theta)z + \lambda = 0, \quad (3.3)$$

且 $r = \frac{1}{2\sigma_d}(\lambda + \sigma_d + \theta - \sqrt{\Delta})$, 这里 $\Delta = (\lambda + \sigma_d + \theta)^2 - 4\lambda\sigma_d$, $0 < r < 1$, 并且 r 满足下面关系式

$$\lambda + \theta + \sigma_d(1-r) = \sigma_d + \frac{\theta}{1-r} = \frac{\lambda}{r}. \quad (3.4)$$

证明. 假设二次方程(3.3)的另一实根是 r^* , 那么有

$$r, r^* = \frac{1}{2\sigma_d}(\lambda + \sigma_d + \theta \pm \sqrt{\Delta}).$$

显然, 下列不等式成立

$$\begin{aligned} (\lambda - \sigma_d + \theta)^2 < \Delta < (\lambda + \sigma_d + \theta)^2, \quad \lambda > \sigma_d, \\ (\sigma_d - \lambda + \theta)^2 < \Delta < (\lambda + \sigma_d + \theta)^2, \quad \lambda < \sigma_d. \end{aligned}$$

得到 $0 < r < 1$, $r^* > 1$, 以及 $r = \frac{1}{2\sigma_d}(\lambda + \sigma_d + \theta - \sqrt{\Delta})$.

将 r 代入方程(3.3), 则(3.4)成立。

矩阵 A 、 B 、 C 都是上三角阵, 令 $R = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}$,

将 R 代入矩阵方程(3.1), 则得到

$$\begin{cases} \sigma_d r_{11}^2 - (\lambda + \sigma_d + \theta)r_{11} + \lambda = 0, \\ \delta_d r_{22}^2 - (\lambda + \delta_d)r_{22} + \lambda = 0, \\ \delta_d r_{12}(r_{11} + r_{22}) + \theta r_{11} - (\lambda + \delta_d)r_{12} = 0. \end{cases} \quad (3.5)$$

为了获得方程(3.1)的最小非负解 R , 取 r_{11} 是方程(3.3)在区间(0, 1)上的根, 且令 $r_{11} = r$, $r_{22} = \rho$ 。将 r_{11}, r_{22} 代入方程组(3.5)中的第三个方程, 容易得到 r_{12} 。显然, 当且仅当 $\rho < 1$ 时, 有 $SP(R) = \max(r, \rho) < 1$ 。□

3.3.3 系统的平稳概率分布

如果 $\rho < 1$, 令 (L_v, J) 是该 QBD 过程 $\{L_v(t), J(t)\}$ 的平稳极限, 且

$$\pi_{kj} = \lim_{t \rightarrow \infty} P\{L_v(t) = k, J(t) = j\}, \quad (k, j) \in \Omega,$$

$$\pi_k = \pi_{k,0}, \quad 0 \leq k \leq c-d,$$

$$\pi_k = (\pi_{k,0}, \pi_{k,1}), \quad c-d < k \leq K.$$

定理 3.2 如果 $\rho < 1$ ，那么 (L, J) 的平稳概率分布为

$$\pi_{k,0} = \begin{cases} G \frac{1}{k!} \left(\frac{\lambda}{\mu_B} \right)^k, & 0 \leq k \leq c-d, \\ G \frac{1}{(c-d)!} \prod_{i=1}^{k-c+d} \frac{\lambda}{\sigma_i} - \frac{\theta r \pi_{c,0}}{\lambda(1-r)} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c} \sum_{i=1}^{k-c+d} \left(\prod_{j=i}^{k-c+d} \frac{\lambda}{\sigma_j} \right), & c-d+1 \leq k \leq c, \\ \pi_{c,0} \left(\frac{\lambda}{\sigma_d} \right)^{k-c}, & c < k < N, \\ r \pi_{c,0} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, & k = N. \end{cases} \quad (3.6)$$

$$\pi_{k,1} = \begin{cases} \pi_{c,0} \frac{\theta r}{\lambda(1-r)} \sum_{i=1}^{k-c+d} \left(\prod_{j=i}^{k-c+d} \frac{\lambda}{\delta_j} \right) \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, & c-d+1 \leq k \leq c, \\ \pi_{c,1} \left(\frac{\lambda}{\delta_d} \right)^{k-c}, & c < k < N, \\ \pi_{c,1} \left(\frac{\lambda}{\delta_d} \right)^{N-c} + \pi_{c,0} \frac{\theta r}{\delta_d(1-r)} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, & k = N, \end{cases} \quad (3.7)$$

其中

$$\pi_{c,0} = G \frac{\frac{1}{(c-d)!} \left(\frac{\lambda}{\mu_B} \right)^{c-d} \prod_{i=1}^d \frac{\lambda}{\sigma_i}}{1 + \frac{\theta r}{\lambda(1-r)} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c} \sum_{i=1}^d \left(\prod_{j=i}^d \frac{\lambda}{\sigma_j} \right)}, \quad (3.8)$$

$$\pi_{c,1} = G \frac{\frac{1}{\lambda(c-d)!} \left(\frac{\lambda}{\mu_B} \right)^{c-d} \prod_{i=1}^d \frac{\lambda}{\sigma_i} \sum_{i=1}^d \left(\prod_{j=i}^d \frac{\lambda}{\delta_j} \right)}{\frac{1-r}{\theta r} \left(\frac{\lambda}{\sigma_d} \right)^{c+1-N} + \frac{1}{\lambda} \sum_{i=1}^d \left(\prod_{j=i}^d \frac{\lambda}{\sigma_j} \right)}, \quad (3.9)$$

且常数因子 G 由归一化条件确定。

证明. $2N-c+d+1$ 维行向量 $(\pi_0, \pi_1, \pi_2, \dots, \pi_N)$ 满足方程

$$(\pi_0, \pi_1, \pi_2, \dots, \pi_N) B[R] = 0,$$

其中 $B[R]$ 是一个 $2N-c+d+1$ 阶方阵，且

$$\pi_{N,1} = \frac{\theta r}{\delta_d(1-r)} \pi_{N-1,0} + \frac{\lambda}{\delta_d} \pi_{N-1,1}. \quad (3.22)$$

将式(3.20)代入上面的方程(3.22)，得到(3.7)中的第三个表达式。

根据式(3.4)、(3.16)和(3.21)，得到

$$\sigma_d \pi_{c,0} - \lambda \pi_{c-1,0} = \sigma_d \pi_{N,0} - \lambda \pi_{N-1,0} = -\frac{\theta}{1-r} \pi_{N,0} \quad (3.23)$$

从式(3.13)和(3.23)，有

$$\sigma_{k-c+d} \pi_{k,0} - \lambda \pi_{k-1,0} = -\frac{\theta}{1-r} \pi_{N,0}, \quad c-d+1 \leq k \leq c \quad (3.24)$$

方程(3.24)迭代产生(3.6)中的第二个表达式。令 $k=c$ ，得到式(3.8)。

同理，根据式(3.14)、(3.15)、(3.17)和(3.19)，有

$$\delta_{k-c+d} \pi_{k,1} - \lambda \pi_{k-1,1} = \frac{\theta}{1-r} \pi_{N,0}, \quad c-d+2 \leq k \leq c, \quad \text{和} \quad \pi_{c-d+1,1} = \frac{\theta}{\delta_1(1-r)} \pi_{N,0}. \quad (3.25)$$

迭代上面的方程，得到(3.7)式中的第一个表达式。将(3.8)式代入，得到(3.9)式。

将式(3.24)和(3.25)代入(3.12)，并考虑式(3.11)，能得出式(3.12)恒成立。

最后，常数因子 G 由如下归一化条件确定：

$$\sum_{i=0}^K \pi_{i,0} + \sum_{i=c-d+1}^K \pi_{i,1} = 1. \quad \square$$

定理 3.3 如果 $k \geq N$ ，则 (L_v, J) 的平稳概率分布为

$$\pi_{k,0} = \begin{cases} \pi_{c,0} r^{k+1-N} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, & N \leq k < K, \\ \pi_{c,0} \frac{\lambda r^{K-N}}{\sigma_d + \theta} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, & k = K. \end{cases}$$

$$\pi_{k,1} = \begin{cases} \pi_{c,0} \frac{\theta r}{\delta_d(1-r)} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c} \left(\frac{\lambda}{\delta_d} \right)^{k-N} \left[1 + \sum_{j=1}^{k-N} \left(\frac{r\delta_d}{\lambda} \right)^j \right] \\ \quad + \pi_{c,1} \left(\frac{\lambda}{\delta_d} \right)^{k-c}, & N \leq k < K, \\ \pi_{c,0} \frac{r}{\delta_d} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c} \left[\frac{\lambda r^{K-1-N}}{\sigma_d + \theta} + \frac{\theta}{1-r} \left(\frac{\lambda}{\delta_d} \right)^{K-N} \right] \\ \quad \times \left[1 + \sum_{j=1}^{K-1-N} \left(\frac{r\delta_d}{\lambda} \right)^j \right] + \pi_{c,1} \left(\frac{\lambda}{\delta_d} \right)^{K-c}, & k = K. \end{cases}$$

证明. 当 $N \leq k < K$ 式，考虑矩阵几何解^[192]，得到

$$\pi_k = \pi_N R^{k-N}, \quad \text{i.e.} \quad (\pi_{k,0}, \pi_{k,1}) = (\pi_{N,0}, \pi_{N,1}) R^{k-N}. \quad (3.26)$$

根据式(3.2)，有

$$R^k = \begin{pmatrix} r^k & \frac{\theta r}{\delta_d(1-r)} \sum_{j=1}^k r^{j-1} \rho^{k-j} \\ 0 & \rho^k \end{pmatrix}.$$

将 R^k 代入(3.26)式, 得到

$$\pi_{k,0} = \pi_{N,0} r^{k-N} = \pi_{c,0} r^{k+1-N} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, \quad (3.27)$$

$$\pi_{k,1} = \pi_{N,0} \frac{\theta r}{\delta_d(1-r)} \sum_{j=1}^{k-N} r^{j-1} \rho^{k-N-j} + \pi_{N,1} \rho^{k-N}. \quad (3.28)$$

当 $k=K$ 时, 根据平衡方程, 得到

$$\lambda \pi_{K-1,0} - (\sigma_d + \theta) \pi_{K,0} = 0, \quad (3.29)$$

$$\lambda \pi_{K-1,1} + \theta \pi_{K,0} - \delta_d \pi_{K,1} = 0. \quad (3.30)$$

从(3.29)式, 有

$$\pi_{K,0} = \frac{\lambda}{\sigma_d + \theta} \pi_{K-1,0} \quad (3.31)$$

根据式(3.30)和(3.31), 得到

$$\pi_{K,1} = \frac{\lambda \theta}{\delta_d(\sigma_d + \theta)} \pi_{K-1,0} + \frac{\lambda}{\delta_d} \pi_{K-1,1}. \quad (3.32)$$

将式(3.27)和(3.28)代入式(3.31)和(3.32), 能得出 $\pi_{k,0}$ 和 $\pi_{k,1}$. □

3.3.4 系统性能度量

根据定理 3.2 和定理 3.3, 能数值估计 $N/(d, c)$ - $M/M/c/K/SMWV$ 排队系统的性能。该排队系统的各种性能度量如下:

(1) 系统中的平均顾客数:

$$E[L_s] = \sum_{i=0}^K i \pi_{i,0} + \sum_{i=c-d+1}^K i \pi_{i,1}. \quad (3.33)$$

(2) 队列中的平均顾客数:

$$E[L_q] = \sum_{i=c-d+1}^K (i - (c-d)) \pi_{i,0} + \sum_{i=c+1}^K (i-c) \pi_{i,1}. \quad (3.34)$$

(3) 处于正常忙期的服务台平均数:

$$E[N_B] = e \{E[W]\}. \quad (3.35)$$

(4) 处于工作休假期的服务台平均数

$$E[WV] = d \sum_{i=0}^K \pi_{i,0}. \quad (3.36)$$

(5) 处于空闲时期的服务台平均数:

$$E[I] = \sum_{i=0}^{c-d} (c-d-i)\pi_{i,0} + \sum_{i=c-d+1}^c (c-i)\pi_i \quad (3.37)$$

(6) 顾客在系统中停留的平均时延:

$$E[T] = \frac{E[L_s]}{\lambda(1-P_{loss})} \quad (3.38)$$

(7) 顾客丢失的概率:

$$P_{loss} = \pi_{c,0} + \pi_{c,1} \quad (3.39)$$

3.4 最优化分析

3.4.1 成本和性能分析

首先开发系统中每个顾客单位时间消耗的总的平均能量(或者成本)函数。定义如下的成本参数:

$C_h \equiv$ 驻留在系统中的每个顾客单位时间成本;

$C_b \equiv$ 服务台在正常忙期提供服务时单位时间诱发的成本;

$C_v \equiv$ 服务台在工作休假时期提供服务时单位时间诱发的成本;

$C_i \equiv$ 服务台处于空闲时期时单位时间诱发的成本;

$C_d \equiv$ 顾客在系统中延迟时单位时间的惩罚成本;

$C_l \equiv$ 系统丢失顾客时单位时间的成本。

利用上面的成本参数和 White 等^[193]提出的乘务员服务设备(crew-service equipment)的概念,在复杂和非线性成本结构下的单位时间的期望成本函数为

$$F(d, N, K, \mu_b, \mu_v) = C_h E[L_s] + C_b E[NB] + C_v E[WV] + C_i E[I] + C_d E[T] + C_l P_{loss} \quad (3.40)$$

其中 $E[L_s], E[NB], E[WV], E[I], E[T], P_{loss}$ 分别由方程(3.33)、(3.35)~(3.39)表示。在(3.40)式中,第一项和最后一项因顾客而产生的成本,其它各项由服务台产生的成本。

Annaswamy 等在文献[11]中用最好情况时延和最坏情况时延的平均值作为标称时延函数 τ_{nom} ,但是,他们假设最好时延是每一个采样周期的下界 nh ,其中 h 是采样周期, n 是采样数。这将引起系统实现偏差而导致系统不稳定。原则上,适当的 τ_{nom} 能确保系统的稳定性。选择顾客在队列中的平均等待时间 $E[T_q]$ 作为标称时延函数,即

$$\tau_{nom} = E[T_q] = \frac{E[L_q]}{\lambda(1-P_{loss})}$$

理论上, 处于休眠状态的控制器越多, 意味着系统消耗的成本越少, 那么顾客在队列中的平均等待时间 $E[T_q]$ 越大。考虑下面的多目标优化模型:

$$\begin{aligned} \min \quad & f = (F(d, N, K, \mu_B, \mu_V), \tau_{nom}), \\ \text{s. t.} \quad & \begin{cases} 0 < d < c < N < K < K', \mu_V < \mu_B, \\ \mu'_B \leq \mu_B \leq \mu''_B, \mu'_V \leq \mu_V \leq \mu''_V, \\ E[NB] \leq c - d, \end{cases} \end{aligned} \quad (3.41)$$

其中 K' 表示容量 K 的上界, μ'_B (或 μ''_B) 和 μ'_V (或 μ''_V) 分别表示 μ_B 和 μ_V 的下(或上)界。

本章的目的是确定半休眠控制器数 d 的最优值 d^* , 阈值 N 的最优值 N^* , 容量 K 的最优值 K^* , 服务台在正常忙期的服务率 μ_B 的最优值 μ_B^* , 以及服务台在工作休假时期的服务率 μ_V 的最优值 μ_V^* , 以最小化期望的成本函数和顾客在队列中的平均时延函数。为此, PSO 算法用于从数值上求解多目标优化问题。采用 MOPSO 算法生成多目标优化模型(3.41)的 Pareto 最优解。

3.4.2 最优停止算法

因为 Pareto 前沿相应的非支配向量不唯一, 需要确定离散决策变量 d 的唯一最优解, 同时得到离散决策变量 N, K 和连续决策变量 μ_B 和 μ_V 的最优值。用文献[174]中提出的马尔科夫链的最优停止策略确定 c 个控制器中的最优半休眠控制器数 d^* 。

假设 $P = (p_{ij})_{c \times c}$ 是离散时间马尔科夫链 X_n 的转移概率矩阵, 其状态空间为 $S = \{1, 2, \dots, c\}$ 。令 $p_{ij} = \frac{1}{c}$ 。假设第 i 个控制应用在 $[0, t]$ 上的分组消息数是 $\omega_i(t)$ 。设在区间 $[0, t]$ 上的收益函数列向量为

$$h_j(t) = \begin{cases} \omega_k(t) - \min \{ \omega_i(t), 1 \leq i \leq c \} = 0, & k = \arg \min_i \{ \omega_i(t) \}, \\ \frac{1}{c} \sum_{i=1}^c \omega_i(t) - (\omega_j(t) - \min \{ \omega_i(t), 1 \leq i \leq c \}), & \text{otherwise,} \end{cases}$$

其中 $k \in S$ 是吸收状态, 且 $1 \leq j \leq c$ 。最优停止算法描述如算法 3.1。

最优停止算法的目的是寻找最优停止策略最大化期望收益(见文献[174])。定义 $[0, t]$ 区间上的收益函数列向量 $H(t) = [h_1(t), h_2(t), \dots, h_c(t)]^T$, 定义所有控制应用中的最小分组消息数相应的状态 k 是一个吸收状态, 且收益为 0, 其它状态的收益值定义为所有控制应用的平均分组消息数与当前状态的分组消息数减去状态 k 的分组消息数之差。

$u_i(t)$ 表示第 $i-1$ 次迭代后得到的期望收益函数。在第(2)步首先初始化期望收益函数 $u_1(t)=[u_{11}(t), u_{12}(t), \dots, u_{1c}(t)]^T$ ，然后对第(3)步和第(4)步反复迭代获得最大期望函数 $u_k(t)=[u_{k1}(t), u_{k2}(t), \dots, u_{kc}(t)]^T$ 。Markov 链的状态空间被分成两个集合 S_1 和 S_2 ，如果 Markov 链的状态在集合 S_1 中，则继续；如果在状态集合 S_2 中，则停止。停止集 S_2 中的元素个数即为半休眠控制器数的最优值。

算法 3.1 最优停止算法

输入： $\omega_1(t), \omega_2(t), \dots, \omega_c(t)$ ，以及精度 $\varepsilon=10^{-7}$

输出： d^*

(1) 设 S_1 和 S_2 分别是继续和停止的状态集，且令 $S_1=S_2=\emptyset$ ，这里 \emptyset 表示空集；

(2) 令 $u_1(t)=[u_{11}(t), u_{12}(t), \dots, u_{1c}(t)]^T$ 表示初始期望收益函数，且

$$u_{1j}(t) = \begin{cases} 0, & h_j(t) = 0, \\ \max \{h_i(t), 1 \leq i \leq c\}, & h_j(t) \neq 0, \end{cases}$$

其中 $1 \leq j \leq c$ ；

(3) 设 $u_i(t) = \max \{Pu_{i-1}(t), H(t)\}$ ，其中期望收益函数 $u_i(t)=[u_{i1}(t), u_{i2}(t), \dots, u_{ic}(t)]^T$ ， $i=2, 3, \dots$ ；

(4) 对于每一个 $i=2, 3, \dots$ ，重复步骤(3)，直到 $|u_i(t) - u_{i-1}(t)| < \varepsilon e$ 为止，得到最大期望收益函数

$u_k(t)=[u_{k1}(t), u_{k2}(t), \dots, u_{kc}(t)]^T$ ，其中 e 是元素均为 1 的 c 维列向量；

(5) 对于每一个 $j \in [1, c]$ ，如果 $u_{kj}(t) = f_j(t)$ ，那么 $S_2 = S_2 + \{j\}$ ，否则 $S_1 = S_1 + \{j\}$ ；

(6) 令 $d^* = |S_2|$ ，其中 $|S_2|$ 是停止集 S_2 中的元素个数。输出 S_2 中的元素 j_1, j_2, \dots, j_{d^*} ，得到所有的半休眠控制器；

(7) 返回 d^* 。

3.4.3 FSDMC 调度算法

需要确定 $N/(d, c)$ -M/M/c/K/SMWV 排队系统中的参数 $\tilde{d}^*, \tilde{N}^*, \tilde{K}^*, \tilde{\mu}_B^*$ 和 μ_v^* 的最优值。

FSDMC 调度算法描述如算法 3.2。

算法 3.2 首先 M 次调用算法 3.1 统计半休眠控制器数的频率，选择其中的频率峰值作为半休眠控制器数的最优值。在第(5)步，调用 MOPSO 算法 M' 次生成多目标优化模型(3.41)的 M' 个 Pareto 最优解集。在每一个 Pareto 最优解集中，根据 $d_k^* = \tilde{d}^*$ 选择唯一非

支配向量 Φ_k^* 作为系统各参数的值。所有非支配向量 Φ_k^* ($1 \leq k \leq M'$) 的平均值 $\tilde{\Phi}$ 即是系统参数的最优值向量。算法 3.2 的时间复杂度是 MOPSO 算法的 M' 倍。

算法 3.2 FSDMC 调度算法

输入: $\omega_1(t), \omega_2(t), \dots, \omega_c(t)$, 和 θ, c .

输出: 最优值 $\tilde{F}(\tilde{d}^*, \tilde{N}^*, \tilde{K}^*, \tilde{\mu}_B^*, \tilde{\mu}_V^*)$, $\tilde{\tau}_{nom}$ 和 $\tilde{\Phi} = [\tilde{d}^*, \tilde{N}^*, \tilde{K}^*, \tilde{\mu}_B^*, \tilde{\mu}_V^*]$.

(1) 初始化决策变量的上、下界, 即 $N(c < N < K)$, $K(N < K < K')$, $\mu_B(\mu'_B \leq \mu_B \leq \mu''_B)$, 和 $\mu_V(\mu'_V \leq \mu_V \leq \mu''_V)$;

(2) 在时间间隔 Δt 内采样第 i 个控制应用中的第 j 次到达分组消息数 $\omega_{ij}(\Delta t)$, 其中 $i=1, 2, \dots, c$ 。假设采样次数是 M , 其中 M 表示一个较大的正整数, 例如 $M=10000$, 则总到达率为

$$\lambda = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^c \frac{\omega_{ij}(\Delta t)}{\Delta t}.$$

(3) 对于每一个 $j \in [1, M]$, 调用算法 3.1, 并分别统计半休眠控制器数 d^* ($d^* = 1, 2, \dots, c$) 的频度 $q(d^*)$;

(4) 令 $\tilde{d}^* = \arg \max_{d^*} \{q(d^*), 1 \leq d^* \leq c\}$, 则得到半休眠控制器数的最优值 \tilde{d}^* ;

(5) 对每一个 $k \in [1, M']$, 调用 MOPSO 算法, 并用定理 3.2 和定理 3.3 计算平稳概率分布 $\pi_{i,0}(0 \leq i \leq K)$ 和 $\pi_{j,1}(c-d+1 \leq j \leq K)$, 生成多目标优化模型(3.41)的 Pareto 前沿, 其中 M' 表示调用 MOPSO 算法的次数。然后, 根据最优值 \tilde{d}^* 选择模型(3.41)的唯一的 Pareto 最优解和相应的非支配向量 $\Phi_k^* = [d_k^*, N_k^*, K_k^*, \mu_{Bk}^*, \mu_{Vk}^*]$;

(6) 计算所有非支配向量 Φ_k^* ($1 \leq k \leq M'$) 的平均值 $\tilde{\Phi}$, 即 $\tilde{\Phi} = \frac{1}{M'} \sum_{k=1}^{M'} \Phi_k^*$, 得到所有决策变量的最优值向量 $\tilde{\Phi} = [\tilde{d}^*, \tilde{N}^*, \tilde{K}^*, \tilde{\mu}_B^*, \tilde{\mu}_V^*]$, 其中 $\tilde{N}^* = \text{INT}(\frac{1}{M'} \sum_{k=1}^{M'} N_k^*)$, $\tilde{K}^* = \text{INT}(\frac{1}{M'} \sum_{k=1}^{M'} K_k^*)$, 且 INT 表示取整函数;

(7) 将 $\lambda, \theta, c, \tilde{d}^*$ 和 $\tilde{\Phi}$ 分别代入 $F(d, N, K, \mu_B, \mu_V)$ 和 τ_{nom} , 得到最优值 $\tilde{F}(\tilde{d}^*, \tilde{N}^*, \tilde{K}^*, \tilde{\mu}_B^*, \tilde{\mu}_V^*)$ 和 $\tilde{\tau}_{nom}$;

(8) 返回 $\tilde{\Phi}, \tilde{F}, \tilde{\tau}_{nom}$.

3.5 数值结果

3.5.1 系统性能度量灵敏度分析

在稳定条件下，得到一些数值实验结果，表明系统参数的变化对系统性能的影响。用 Matlab 进行数值实验，固定系统最大容量 $K = 28$ ，并考虑下列三种情况下系统参数不同值的影响：

情况 1: $\lambda=5.0, \theta=0.05, d=4, c=10, N=20$ ，并改变 (μ_B, μ_V) 的值；

情况 2: $\mu_B=2.5, \mu_V=1.0, d=4, c=10, N=20$ ，并改变 (λ, θ) 的值；

情况 3: $\lambda=5.0, \theta=0.05, \mu_B=2.5, \mu_V=1.0$ ，并改变 (d, c, N) 的值。

上述三种情况下系统性能的数值结果分别见表 3.1~3.3。在表 3.1 中，随着 μ_B 或者 μ_V 的增加， $E[L_s], E[L_q]$ 和 $E[NB]$ 都减少，但 $E[I]$ 增加。在表 2 中表明：(i) 当 λ 增加时， $E[L_s]$ 和 $E[L_q]$ 明显增加，但是随着 θ 增加， $E[L_s]$ 和 $E[L_q]$ 缓慢减少；(ii) 当 λ 增加时， $E[NB]$ 增加，但是 $E[I]$ 减少；(iii) 随着 θ 的增加， $E[NB]$ 和 $E[I]$ 微弱减少。表 3.3 表明：(i) 当 d 增加时， $E[L_s]$ 和 $E[L_q]$ 增加，但 $E[NB]$ 和 $E[I]$ 减少；(ii) 当 c 增加时， $E[L_s], E[L_q]$ 和 $E[NB]$ 都增加，但 $E[I]$ 急剧增加；(iii) 与前面的(i)类似，当 N 增加时， $E[L_s]$ 和 $E[L_q]$ 微弱增加，但 $E[NB]$ 和 $E[I]$ 微弱减少。表 3.1~3.3 表明 $E[WW]$ 几乎等于 d ，无论其它参数如何变化。上述结果与 $N/(d, c)$ - $M/M/c/K/SMWV$ 队列的实际情况一致。

表 3.1 在 $\lambda=5.0, \theta=0.05, d=4, c=10, N=20$ 情况下 (μ_B, μ_V) 取不同值时系统性能度量

(μ_B, μ_V)	(2.5, 1.0)	(3.5, 1.0)	(4.5, 1.0)	(2.5, 0.5)	(2.5, 1.5)
$E[L_s]$	2.0617	1.4298	1.1113	2.0867	2.0450
$E[L_q]$	0.0372	0.0013	0.0002	0.0502	0.0287
$E[NB]$	2.0245	1.4285	1.1111	2.0366	2.0163
$E[WW]$	4.0000	4.0000	4.0000	4.0000	4.0000
$E[I]$	3.9755	4.5715	4.8889	3.9634	3.9837

表 3.2 在 $\mu_B=2.5, \mu_V=1.0, d=4, c=10, N=20$ 情况下 (λ, θ) 取不同值时系统性能度量

(λ, θ)	(4.0, 0.10)	(5.0, 0.10)	(6.0, 0.10)	(5.0, 0.05)	(5.0, 0.15)
$E[L_s]$	1.6034	2.06171701	3.0974	2.06171703	2.06171700
$E[L_q]$	0.0031	0.03719511	0.3986	0.03719514	0.03719509
$E[NB]$	1.6003	2.02452190	2.6988	2.02452189	2.02452191
$E[WW]$	4.0000	3.99999993	4.0000	3.99999995	3.99999991
$E[I]$	4.3997	3.97547817	3.3012	3.97547815	3.97547818

3.5.2 期望的成本函数与标称时延函数的灵敏度分析

作为 ANCS 的典型例子，高端汽车电子系统通常采用 FlexRay 总线来连接或组织大量的 ECU。例如，在汽车巡航子系统中，假设多个任务映射到 10 个 ECUs，这些 ECUs 以 5kB/s 的速率连接到双信道 FlexRay 总线。通过生成合适的系统参数来考虑系统实现

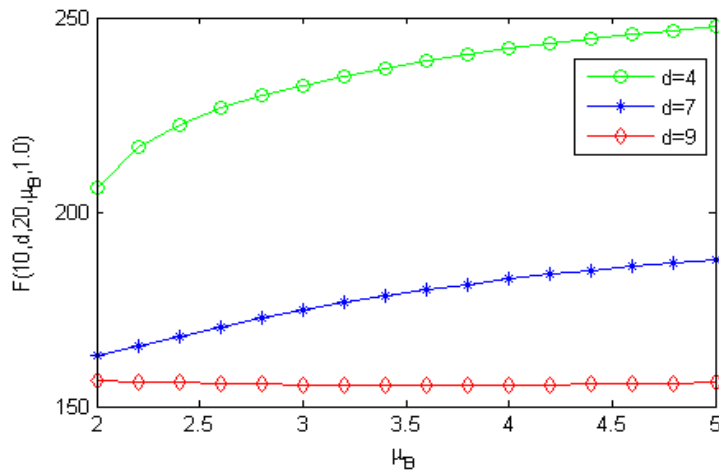
表 3.3 在 $\lambda=5.0, \theta=0.05, \mu_B=2.5, \mu_V=1.0$ 情况下(d, c, N)取不同值时系统性能度量

(d, c, N)	(4, 10, 20)	(7, 10, 20)	(4, 15, 20)	(4, 10, 15)
$E[L_s]$	2.06171 707	2.2657	2.0002	2.06171 667
$E[L_q]$	0.03719 520	0.4031	0.0001	0.03719 458
$E[NB]$	2.02452 187	1.8626	2.0001	2.02452 209
$E[WW]$	4.00000000	7.0000	4.0000	3.99999 945
$E[I]$	3.97547 813	1.1374	8.9999	3.97547 846

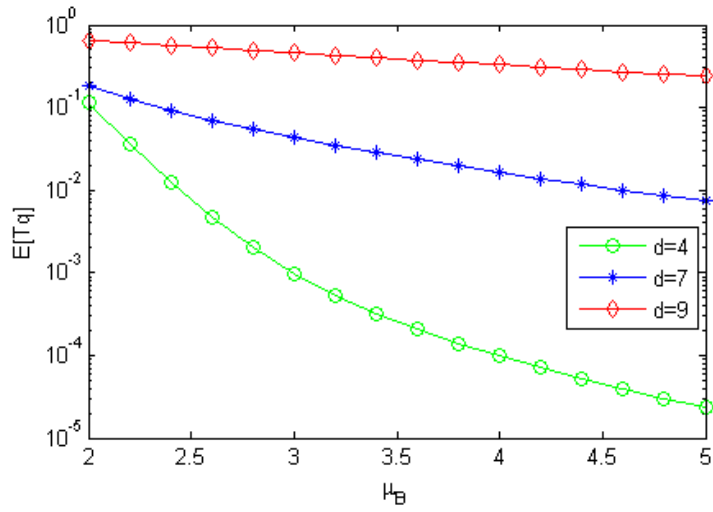
成本和系统性能之间的折中，使用的系统参数与文献[130,191]中的参数一致。首先分析关于决策变量的期望成本函数和标称时延函数的灵敏度，然后，通过求解多目标优化模型(3.41)，计算队列中的系统能耗和标称时延。固定成本参数值： $C_h = 1\text{mW}, C_b = 35\text{mW}, C_v = 15\text{mW}, C_i = 10\text{mW}, C_d = 8\text{mW}$ ，和 $C_l = 10000\text{mW}$ ，以及决策变量的值： $\lambda=5.0, \theta=0.05, K=28$ 。考虑如下六种情况：

- 情况 4: $c = 10, N = 20, \mu_V = 1.0$ ，并 μ_B 的值从 2.0 到 5.0 变化， $d = 4, 7, 9$;
- 情况 5: $c = 10, N = 20, \mu_B = 2.5$ ，并 μ_V 的值从 0.5 到 3.0 变化， $d = 4, 7, 9$;
- 情况 6: $d = 4, N = 20, \mu_V = 1.0$ ，并 μ_B 的值从 2.0 到 5.0 变化， $c = 6, 10, 15$;
- 情况 7: $d = 4, N = 20, \mu_B = 2.5$ ，并 μ_V 的值从 0.5 到 3.0 变化， $c = 6, 10, 15$;
- 情况 8: $d = 4, c = 10, \mu_V = 1.0$ ，并 μ_B 的值从 2.0 到 5.0 变化， $N = 15, 20$;
- 情况 9: $d = 4, c = 10, \mu_B = 2.5$ ，并 μ_V 的值从 0.5 到 3.0 变化， $N = 15, 20$ 。

对于期望成本函数 F 和标称时延函数 $E[T_q]$ 的数值结果，情况 4~7 分别显示如图 3.6~3.9，情况 8~9 分别表示如表 3.4~3.5。在图 3.6 中，随着 μ_B 的增加 F 增加，但是当 d 增加时 F 减少。图 3.7 表明当 μ_V 增加时 F 增加缓慢，当 d 增加时 F 也减少。但是，

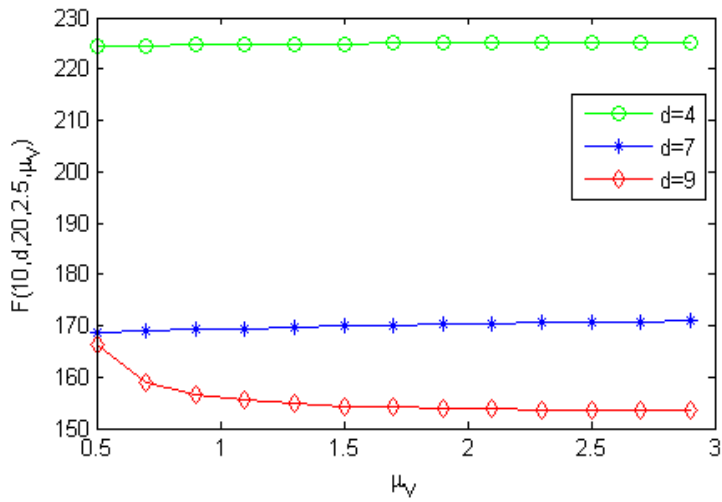


(a)

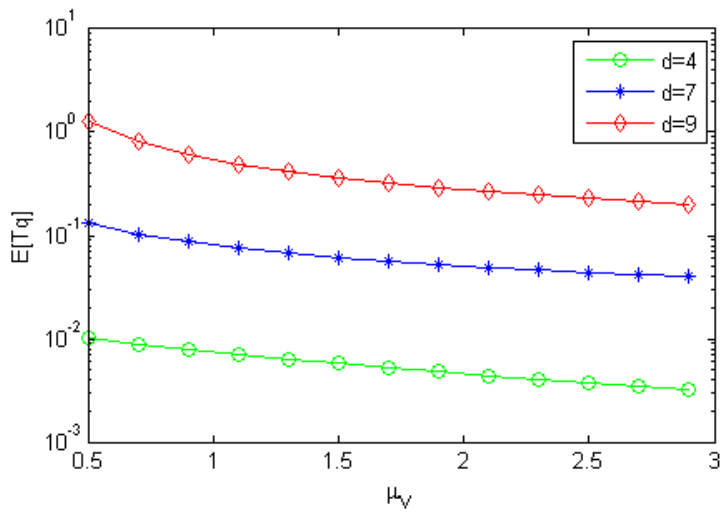


(b)

图 3.6 $F(10, d, 20, \mu_B, 1.0)$ 和 $E[T_q]$ 分别对应 μ_B 和 d 的变化图



(a)



(b)

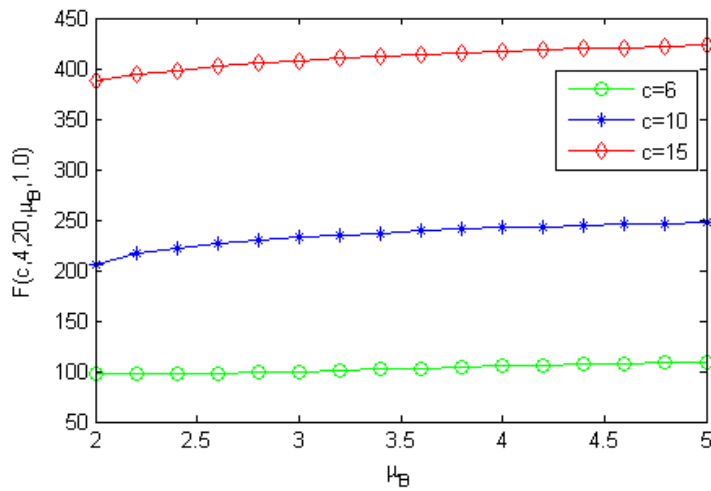
图 3.7 $F(10, d, 20, 2.5, \mu_v)$ 和 $E[T_q]$ 分别对应 μ_v 和 d 的变化图

表 3.4 在 $\lambda=5.0, \theta=0.05, \mu_V=1.0, c=10, d=4$ 情况下 (N, μ_B) 取不同值时的
期望成本函数和标称时延函数的变化

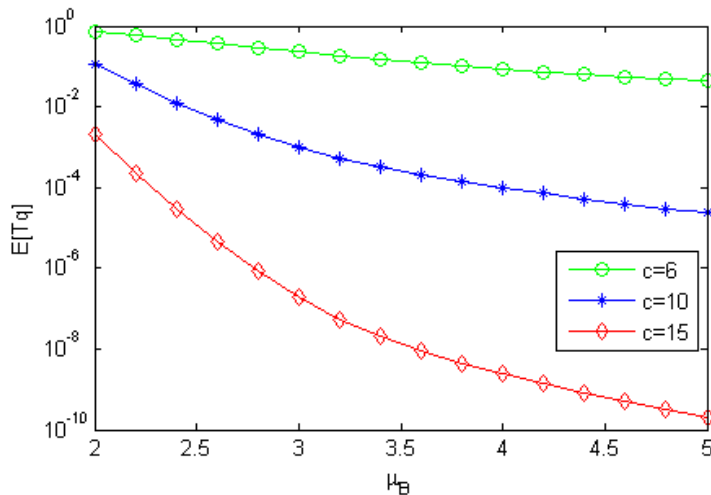
(N, μ_B)	(15, 2.0)	(15, 2.5)	(15, 3.0)	(20, 2.0)	(20, 2.5)	(20, 3.0)
$F(N, \mu_B)$	206.062959	224.747422	232.647279	206.062781	224.747418	232.647279
$E[T_q]$	0.11410504	0.00743892	0.00097276	0.11411172	0.00743904	0.00097277

表 3.5 在 $\lambda=5.0, \theta=0.05, \mu_B=2.5, c=10, d=4$ 情况下 (N, μ_V) 取不同值时的
期望成本函数和标称时延函数的变化

(N, μ_V)	(15, 0.5)	(15, 1.0)	(15, 1.5)	(20, 0.5)	(20, 1.0)	(20, 1.5)
$F(N, \mu_V)$	224.510822	224.747422	224.909647	224.510807	224.747418	224.909646
$E[T_q]$	0.01002990	0.00743892	0.00573672	0.01003027	0.00743904	0.00573677



(a)

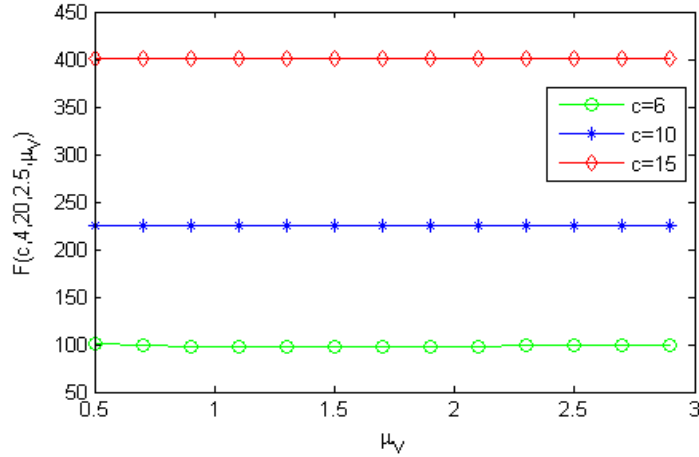


(b)

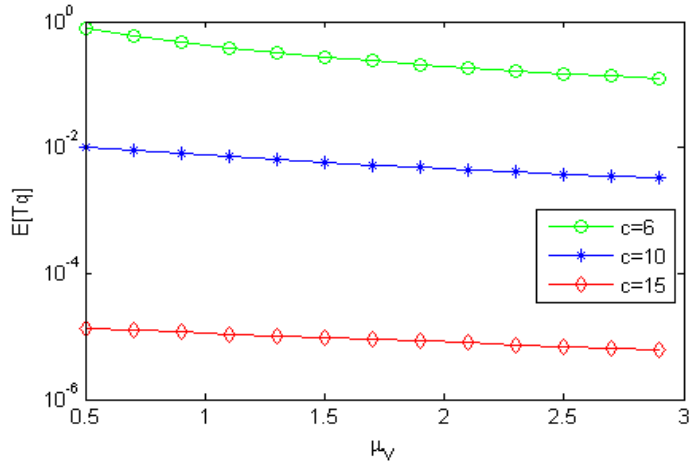
图 3.8 $F(c, 4, 20, \mu_B, 1.0)$ 和 $E[T_q]$ 分别对应 μ_B 和 c 的变化图

当 μ_B 、 μ_V 和 d 增加时, $E[T_q]$ 的变化分别与 F 的变化正好相反。图 3.8、图 3.9 和表 3.4、表 3.5 的情况分别与图 3.6 和图 3.7 类似。

接下来考虑在成本参数取不同值的情况下期望成本函数的行为。首先固定系统参数如下： $\lambda=5.0$, $\theta=0.05$, $\mu_B=2.5$, $\mu_V=1.0$, $d=4$, $c=10$, $N=20$, 和 $K=28$ 。表 3.6~3.8 表明了参数 (C_h, C_i) , (C_v, C_b) , 和 (C_d, C_l) 分别对期望成本函数的影响。当其它成本参数固定不变时, C_h 对 $F(4,20,2.5,1.0)$ 影响较大, 显然, 当 C_h 增加时 $F(4,20,2.5,1.0)$ 增加。这暗示当顾客在系统中的驻留成本增加时, 期望成本上升。系统中其它参数对期望成本的影响也被量化。但是, 需要强调的是, 当系统中的参数取不同的值时, 实验观察到的可能不同。



(a)



(b)

图 3.9 $F(c, 4, 20, 2.5, \mu_v)$ 和 $E[T_q]$ 分别对应 μ_v 和 c 的变化图

表 3.6 在 $C_v=15, C_b=35, C_d=8, C_l=10000$ 情况下参数 (C_h, C_i)

对期望成本函数 $F(4, 20, 2.5, 1.0)$ 的影响

(C_h, C_i)	(1, 10)	(3, 10)	(5, 10)	(1, 12)	(1, 14)
$F(4,20,2.5,1.0)$	224.7475	228.8710	232.9944	228.7966	232.8456

表 3.7 在 $C_h=1, C_i=10, C_d=8, C_l=10000$ 情况下参数 (C_v, C_b)

对期望成本函数 $F(4, 20, 2.5, 1.0)$ 的影响

(C_v, C_b)	(15, 35)	(17, 35)	(19, 35)	(15, 37)	(15, 39)
$F(4,20,2.5,1.0)$	224.7475	232.7475	240.7475	232.6985	240.6495

表 3.8 在 $C_h=1, C_b=35, C_v=15, C_i=10$ 情况下参数(C_d, C_l)
对期望成本函数 $F(4, 20, 2.5, 1.0)$ 的影响

(C_d, C_l)	(8, 10000)	(10, 10000)	(12, 10000)	(8, 20000)	(8, 30000)
$F(4,20,2.5,1.0)$	224.7475	225.5722	226.3969	224.7477	224.7478

3.5.3 系统参数的最优值

本节调用算法 3.2, 用数值实验得到 \tilde{d}^* , $\tilde{\phi}$, \tilde{F} , 和 $\tilde{\tau}_{nom}$ 的最优值。依然选择系统中的各种参数如下: $C_h=1mW, C_b=35mW, C_v=15mW, C_i=10mW, C_d=8mW, C_l=10000mW, \theta=0.05, c=10$ 。表 3.9 显示随机生成单位时间所有控制应用到达的分组消息数采样数据。

表 3.9 分组消息到达的采样数据

Ctrl. App. ID	1	2	3	4	5	6	7	8	9	10
Sample Data(B/s)	628	[361]	841	665	473	667	[266]	[256]	446	[397]

表 3.9 中, 通过调用算法 3.1, 在轻量级通信情况下, 方括号中的数据表示相应的控制器将要进入休眠状态。因此, 获得半休眠控制器数 $d = 4$ 。通过对采样数据求和, 得到总到达率为 $\lambda = \sum_{i=1}^c \omega_i(t) = 5000B = 5kB$ 。

$$\lambda = \sum_{i=1}^c \omega_i(t) = 5000B = 5kB$$

当调用算法 10000 次, 即 $M = 10000$, 得到半休眠控制器数的统计频度(如图 3.10), 表明半休眠控制器数服从正态分布。因此, 选择相应半休眠控制器数的最大频度作为最优值 \tilde{d}^* , 即 $\tilde{d}^* = 4$ 。值得注意的是, 图 3.10 中的采样数据并不唯一, 但是一般来说服从正态分布。

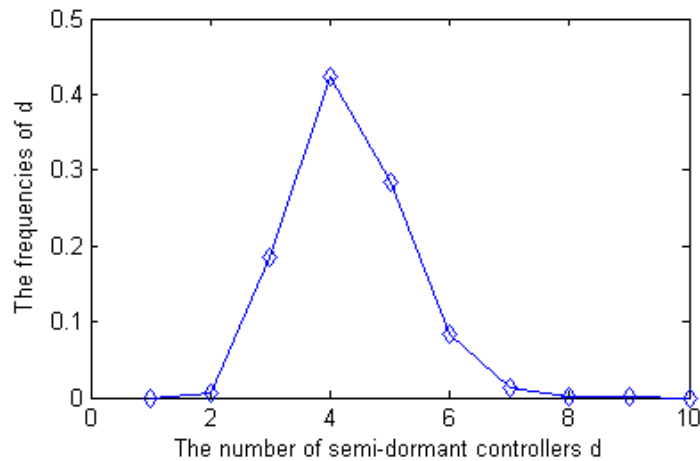
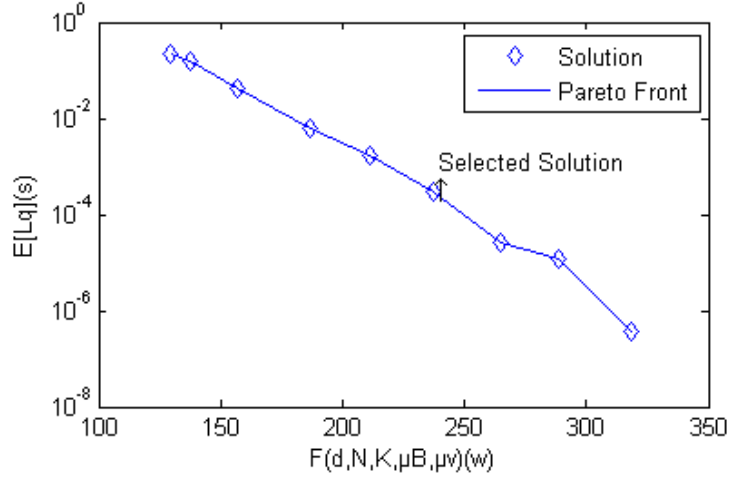


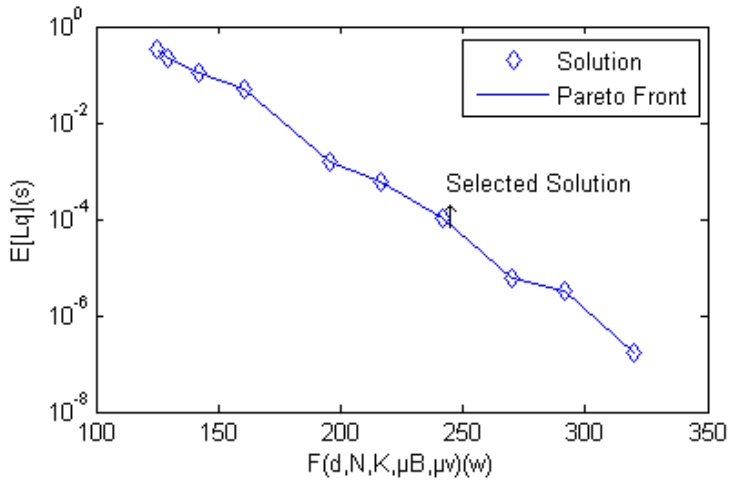
图 3.10 半休眠控制器数的频度图

下面调用算法 3.2 生成最优值 $\tilde{\phi}$, \tilde{F} , 和 $\tilde{\tau}_{nom}$ 。调用 MOPSO 算法的时, 用种群数为 100 个粒子, 档案大小为 40 个粒子, 变异率为 0.5, 迭代 3000 次, 以及 30 个自适应网格。而且, 令 $\lambda = 5.0$, 决策变量的取值范围 $d \in [1, c]$, $N \in [c+1, K-1]$, $K \in [N+1, K+10]$,

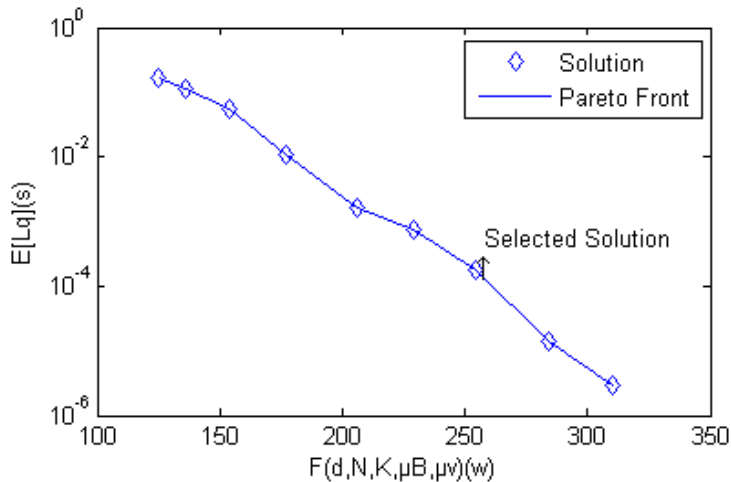
$\mu_B \in [2.0, 5.0]$, 和 $\mu_v \in [0.5, 3.0]$, 并且满足模型(3.41)的约束条件。图 3.11 表示了调用 MOPSO 算法四次分别获得的四种 Pareto 前沿。图 3.11(a)~(d)中, 当 $\tilde{d}^* = 4, 4, 3, 5$ 时, 分别选择了四组 Pareto 最优解。



(a)



(b)



(c)

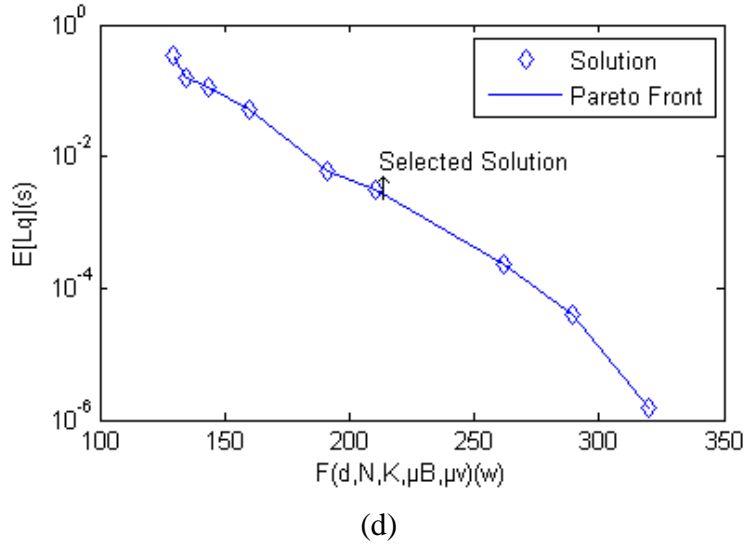


图 3.11 四次调用 MOPSO 算法得到的 Pareto 前沿(a)-(d)

图 3.11(a)中 Pareto 最优解和相应的非支配向量如表 3.10.

表 3.10 图 3.11(a)中 Pareto 最优解和相应的非支配向量

d	N	K	μ_B	μ_v	F	τ_{nom}
1	16	22	4.23	0.57	318.52	3.5453e-007
2	21	25	3.58	1.96	288.71	1.1431e-005
3	19	28	3.72	2.73	264.89	2.6012e-005
4	21	28	3.41	1.43	237.16	2.8473e-004
5	25	28	3.29	1.55	213.94	1.6183e-003
6	18	28	3.35	1.92	187.01	5.9476e-003
7	17	23	2.64	2.49	157.11	4.0771e-002
8	14	28	2.36	2.02	137.78	1.6100e-001
9	23	28	3.31	1.02	129.43	2.2617e-001

假设调用 MOPSO 算法 10 次，即 $M'=10$ ，获得 10 组 Pareto 最优解和相应的非支配向量(如表 3.11)。

在图 3.10 和表 3.11 中，获得设计上述 FSDMC 模型的参数值。当 $c=10, \theta=0.05$ 时，得到 $\tilde{d}^* = 4$ ， $\tilde{N}^* = 20$ ， $\tilde{K}^* = 28$ ， $\tilde{\mu}_B^* = 3.4$ ， $\tilde{\mu}_v^* = 1.5$ 。但是， λ 的值由 $\lambda = \sum_{i=1}^c \omega_i(t)$ 确定，这里， $\lambda = 5\text{kB/s}$ 。将上述参数值代入模型(3.41)中的目标函数，得到期望成本函数 F 和标称时延函数 τ_{nom} 的值，即 $\tilde{F} = 237.07\text{mW}$ 和 $\tilde{\tau}_{nom} = 2.9259\text{e}-004\text{s}$ 。这一结果基本上与表 3.10 和表 3.11 中的结果一致。在汽车巡航控制系统实例中，用本文的方法获得的控制器设计参数如下。在轻量级通信情况下，10 个控制器中允许 4 个控制器进入半休眠状态。在全局队列中，当分组消息数达到 20 个时，4 个半休眠控制器从休假状态返回。系统容量是 28。全局队列中总的泊松到达率是 5kB/s，每一个控制器在正常忙期和工作休假期的平均服务率分别是 3.4kB/s 和 1.5kB/s。因此，能获得相应的期望成本函数和标称时延函数最优值。

值得说明的是,如果 FlexRay 总线带宽是 10Mb/s^[194],或者 CAN 总线带宽是 500kb/s^[119],能相应地获得系统参数的最优值。如果 FSDMC 模型中控制器数 c 改变,半休眠控制器数 d 也可能变化。

表 3.11 Pareto 最优解和相应的非支配向量

No.	d	N	K	μ_B	μ_V	F	τ_{nom}
1	3	21	28	3.95	1.17	261.65	1.8834e-005
2	3	20	28	3.85	1.11	260.91	2.4267e-005
3	4	21	28	3.41	1.43	237.16	2.8473e-004
4	4	20	28	3.40	1.91	237.07	2.7046e-004
5	4	23	28	3.45	2.07	237.55	2.4130e-004
6	4	21	28	3.36	1.32	236.67	3.3089e-004
7	5	20	28	3.20	1.43	215.10	1.8553e-003
8	5	20	28	3.15	2.31	214.59	1.7335e-003
9	5	17	28	3.15	0.72	214.50	2.4159e-003
10	6	22	28	3.12	1.95	188.72	7.7024e-003
Average	4.3	20.5	28	3.404	1.542	—	—

3.5.4 比较和分析

在本章工作中,考虑系统实现成本和系统性能之间的折中。根据 ANCS 的特征,采用多种策略,如 N 策略、 (d, c) 休眠机制、有限容量 K 以及同步多重工作休眠等,建立 $M/M/c$ 排队模型。本章建立的排队模型比文献[177,185,188~190]中的排队模型更综合地考虑了 ANCS 的特征,并获得了排队系统平稳概率分布的闭式表达式,以计算系统各种性能度量。

本章建立了多目标优化模型。文献[177,185,188~190,194]针对最小化系统实现成本,仅仅考虑了单目标优化问题。文献[1,11,12,27]研究了系统时延问题,但没有考虑时延优化问题。文献[130,191]用排队论建模实时系统,但也没有涉及到时延优化问题。文献[53,136]用随机分析的方法研究了实时系统的时延问题,但没有讨论系统实现成本和系统性能折中的多目标优化问题。因此,本章构建的排队模型比上述文献中的排队模型更适合建模 ANCS。设计的最优停止算法生成半休眠控制器数的唯一最优值,这一算法是获得系统其它参数最优值的基础。本章的实验结果表明提出的算法的可行性和有用性。

3.6 本章小结

考虑一种新的具有双通信信道的仲裁网络控制系统(ANCS),并求解系统设计所需要的各种参数的最优值。设计了一种 FSDMC 模型,允许部分空闲控制器在轻量级通信条件下进入半休眠状态。考虑控制器簇的随机调度问题,减少系统成本并改善系统性能。通过分析系统特征,构建了一种 $N/(d, c)-M/M/c/K/SMWV$ 排队模型,分别构造成本和性

能函数。为了生成系统参数的最优值，本文开发了一种多目标优化模型最小化成本和性能函数。设计了 FSDMC 调度算法和最优停止算法，求解多目标优化问题，得到 Pareto 前沿和相应的非支配向量集，为系统设计提供服务。最后，数值实验验证了 FSDMC 模型的有效性。

第 4 章 资源动态供应与滞弹性计算

本章考虑图3.1所示的双信道ANCS平台的资源供应问题。第3章考虑了FSDMC模型中多个控制器同步休假问题，这种多个控制器同步休假模式导致资源过度供应而招致不必要的成本浪费。在某些情况下，由于不可预期的物理环境变化产生不期望的负载高峰。因此，根据外部动态环境动态缩放FSDMC系统中的计算资源，以改善资源利用率，是本章研究的焦点。依然运用排队论建模带这一ANCS系统，异步调度FSDMC中的控制器，计算系统的平稳概率分布和系统性能指标，通过求解最小化成本性能率函数优化问题，得出ANCS系统参数的最优配置，协同设计总线传输速率和控制器服务强度，动态供应系统资源。

4.1 引言

在双信道 ANCS 平台中，已经在第 3 章建模 FSDMC 模型为一种 $N/(d, c)-M/M/c/K/SMWV$ 排队系统，同步调度控制器。但是，在 FSDMC 模型中，存在需要解决的两个主要问题：资源过度供应和系统过载。在 $N/(d, c)-M/M/c/K/SMWV$ 模型中，处于空闲状态的某个控制器需等待其它控制器也处于空闲状态后同步进入工作休假状态，这种同步休假策略导致资源过度供应而招致不必要的成本浪费，而且，在某些情况下，由于不可预期的物理环境变化产生不期望的负载高峰。根据外部动态环境动态缩放 FSDMC 系统中的计算资源，以改善资源利用率。

在 ANCS 中，归因于控制模型及其实际实现之间的间隙的主要问题是受控对象输出的传感信号与发送到执行器的信号之间的端到端的时延^[41]。在弹性系统中，时间相关的弹性行为称为“滞弹性”，时间敏感的弹性系统称为滞弹性系统。当前的维基百科物理上定义滞弹性表述为“一个滞弹性材料是粘弹性材料的特殊情况。粘弹性材料有粘性和弹性特征，因此，展现了时间相关的应变。滞弹性材料在移除负载后将完全恢复到其原始状态”。当应用于计算时，滞弹性自然反射了 DES 系统中时间相关的资源供应的按需特性。称这一计算问题为“滞弹性计算”。资源供应意味着在需要时使用户可以无缝地使用资源^[34]。在 FSDMC 中，当外部应力(如分组消息请求数)作用于系统时，应变(如增加了资源利用率和平均任务响应时间)随时间非线性增加，即 FSDMC 平台变形。由于共享总线的资源约束，存在应力-应变滞后。这一滞后现象是弹性(elasticity)和滞弹性(anelasticity)系统的主要区别。

本章建模 FSDMC 为一种 $N/(d, c)-M/M/c/K/AMWV$ 队列，并用这一排队模型量化系统的各种性能指标，构建最优化模型，最小化非线性 CPR 函数。PSO 算法用于求解该最优化模型，获得系统在稳定条件下的参数最优值。在这一排队系统中，考虑系统自动缩放策略：控制器半休眠(资源水平缩放)和控制器服务强度的自适应变化(资源垂直缩放)，以保持 CPR 最优值不变。当系统负载增加时，系统成本增加，但系统性能降低，

因此, 应变(即 CPR)随着应力(即到达率)的增加而发生变化。当平均到达率超过平均服务率时, 控制器服务率自适应调整, 保持系统在稳定状态下的 CPR 不变。根据资源供应和负载之间的匹配情况, FSDMC 平台处于三种状态(下文中用“3-状态”表示), 即过度供应(over-provisioning)状态、正常状态(normal)和不足供应(under-provisioning)状态^[39]。用 3-状态 Markov 过程描述马尔可夫环境, 构建随机环境下的三阶段(3-位相) M/M/1/K 排队模型, 获得随时间变化而随机改变的突发负载下的系统滞弹性量化模型, 精确计算 FSDMC 平台的滞弹性值。

4.2 滞弹性度量

根据文献[35]中定义的弹性 CPS (elasticity CPS, eCPS)系统, 定义滞弹性 CPS 系统如下:

定义 1 一个滞弹性 CPS 系统是 eCPS, 其应变(如成本和性能)滞后于应力(如负载)。

文献[39]中, 具有动态可变负载的系统弹性被定义为系统处于正常状态的时间百分比(或者概率)。用 p_{over} 、 p_{normal} 和 p_{under} 分别表示系统处于过度供应状态、正常状态和不足供应状态的概率, 并且 $p_{over} + p_{normal} + p_{under} = 1$ 。在随机环境的 M/M/1/K 队列中, 条件队列长度密度常常运用于由 Markov 模型描述的时间相关的队列行为的研究中^[192]。当随机环境下的 M/M/1/K 队列处于正常状态时, 假设用 $AvgQL_{normal}$ 表示其条件平均队列长度。

定义 2 在 eCPS 系统中, 滞弹性计算如下:

$$Anelasticity = (AvgQL_{normal}, p_{normal}). \quad (4.1)$$

在 N/(d, c)-M/M/c/K/AMWV 队列中, 状态分成三类:

- 当 $0 \leq k < c$ 时, 一个状态是过度供应状态;
- 当 $c \leq k < N$ 时, 一个状态是正常状态;
- 当 $N \leq k \leq K$ 时, 一个状态是不足供应状态, 其中 k 表示系统中的分组消息数。

在随机环境下的 M/M/1/K 队列中, 用 (k, i) 表示一个状态, 其中 $k \geq 0$ 表示该排队系统中的分组消息数, i 从三个状态(或者三个位相、3-状态、3-位相, 下文同)取值, 即过度供应状态、正常状态和不足供应状态。为了简便起见, 这 3-状态分别用 1、2 和 3 表示。3-状态之间的状态转移描述如下:

- $(k-1, i) \xrightarrow{\lambda_i} (k, i+1)$, 当一个新的分组消息到达位相 i 时, 转移发生;
- $(k, i+1) \xrightarrow{\mu_i} (k-1, i)$, 当一个分组消息在位相 i 完成服务后, 转移发生, 其中 $k=c, i=1$; 或者 $k=N, i=2$, 并且 λ_i 和 μ_i 分别表示系统处于位相 i (或者状态 i)时的到达率和服务率。

在稳定条件下, 当系统中的分组消息数大于等于阈值 N , 并且系统当前处于不足供应状态时, d 个半休眠控制器从工作休假状态返回, 并继续为等待的分组消息提供服务; 当系统中的分组消息数小于控制器数 c , 并且系统当前处于过度供应状态时, 空闲的控制器立即开始工作休假, 直到工作休假的控制器数等于 d 为止。

为了获得系统实现成本和性能度量，本章建立 FSDMC 模型为一种排队模型，以计算系统的各种性能指标，设计最优化模型，最小化非线性 CPR 函数，并获得系统参数的最优值。

4.3 排队模型

4.3.1 QBD 过程模型

本章详细构建 $N/(d,c)-M/M/c/K/AMWV$ 排队系统，并分析 FSDMC 模型的实现成本和性能。为方便起见，依然称分组消息为“顾客”，控制器为“服务台”，以及半休眠状态为“工作休假状态”。在 FSDMC 中， c 个子队列构成共享总线的单个全局队列。在这个全局队列中，假设到达间隔时间、正常忙期服务时间和工作休假期的服务时间分别服从参数为 λ 、 μ_B 和 μ_V ($\mu_V < \mu_B$) 的指数分布。

在本节建立的排队系统中，除了工作休假的方式不同以外，其它特征与 3.3 节中提出的 $N/(d,c)-M/M/c/K/SMWV$ 排队系统的相应特征精确相同，前者是部分服务台同步多重工作休假，后者是部分服务台异步多重工作休假。在 $N/(d,c)-M/M/c/K/AMWV$ 排队模型中，一旦系统中的顾客数小于服务台数 c ，空闲服务台立即开始以指数速率 θ 工作休假，直到工作休假的服务台数等于 d 为止。当服务台休假完成，并且系统中的顾客数小于阈值 N ，那么休假服务台开始另一个独立同分布的工作休假；否则，休假服务台从工作休假状态返回，继续以服务率 μ_B 服务等待的顾客，直到顾客数小于等于 $c-d$ 为止。

在本章的工作中，假设所有到达间隔时间、服务时间和休假时间相互独立，并进一步假设顾客到达服从先来先服务(FCFS)规则。假设一个服务台一次仅仅能够服务一个顾客。

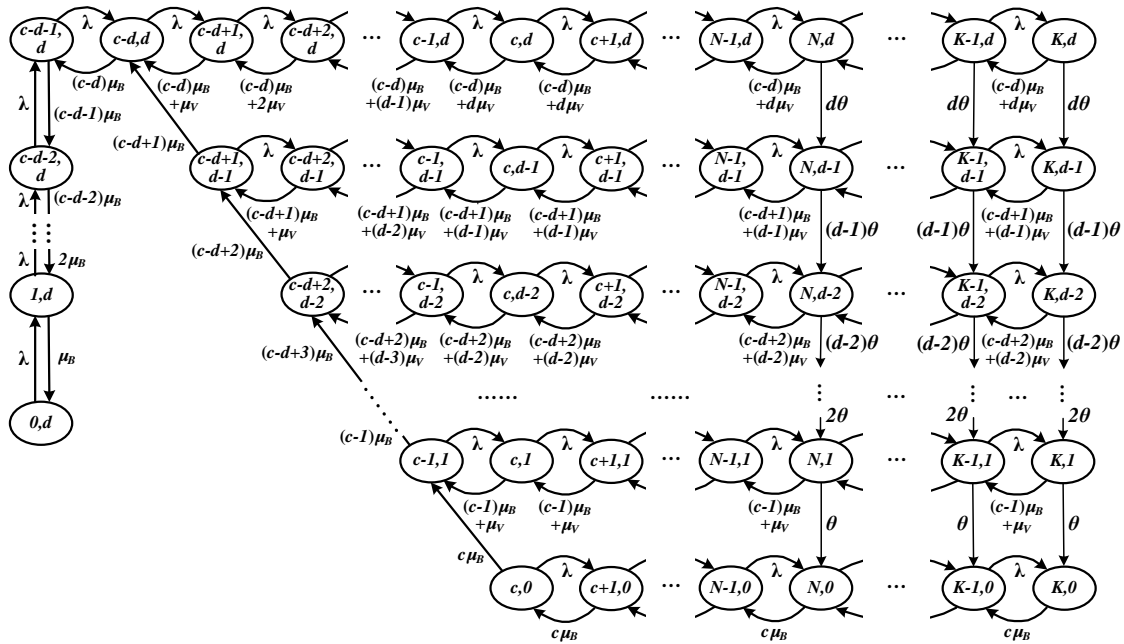


图 4.1 $N/(d,c)-M/M/c/K/AMWV$ 排队系统的状态转移率图

$$B_k = \begin{cases} \delta_{k,0}, & 1 \leq k \leq c-d, \\ \begin{pmatrix} \delta_{k,k-c+d} & & & \\ & \delta_{k,k-c+d-1} & & \\ & & \ddots & \\ & & & \delta_{k,1} \\ 0 & 0 & \cdots & \delta_{k,0} \end{pmatrix}_{(k-c+d+1) \times (k-c+d)}, & c-d < k \leq c, \\ B, & c < k \leq N-1, \end{cases}$$

$$C_k = \begin{cases} \lambda, & 0 \leq k < c-d, \\ \left(\lambda I_{(k-c+d+1) \times (k-c+d+1)} \quad 0_{(k-c+d+1) \times 1} \right), & c-d \leq k < c, \\ \lambda I_{(d+1) \times (d+1)}, & c \leq k \leq N-1, \end{cases}$$

以及

$$A = \begin{pmatrix} -(\lambda + \sigma_d + d\theta) & d\theta & & & & \\ & -(\lambda + \sigma_{d-1} + (d-1)\theta) & (d-1)\theta & & & \\ & & \ddots & \ddots & & \\ & & & -(\lambda + \sigma_1 + \theta) & \theta & \\ & & & & -(\lambda + \sigma_0) & \end{pmatrix}_{(d+1) \times (d+1)}, N \leq k < K,$$

$$B = \begin{pmatrix} \sigma_d & & & & \\ & \sigma_{d-1} & & & \\ & & \ddots & & \\ & & & \sigma_1 & \\ & & & & \sigma_0 \end{pmatrix}_{(d+1) \times (d+1)}, N \leq k \leq K,$$

$$C = \lambda I_{(d+1) \times (d+1)}, N \leq k < K,$$

其中 I 是相应阶的单位矩阵, $0_{(k-c+d+1) \times 1}$ 表示 $(k-c+d+1)$ -维 0 向量。

为了分析该 QBD 过程, 必须求解关于率阵 R 的矩阵二次方程的最小非负解

$$R^2 B + RA + C = 0. \quad (4.2)$$

定理 4.1 如果 $\rho = \lambda \sigma_0^{-1} < 1$, 那么矩阵方程(4.2)有最小非负解

$$R = \begin{pmatrix} r_d & r_{d,d-1} & r_{d,d-2} & \cdots & r_{d,1} & r_{d,0} \\ & r_{d-1} & r_{d-1,d-2} & \cdots & r_{d-1,1} & r_{d-1,0} \\ & & r_{d-2} & \cdots & r_{d-2,1} & r_{d-2,0} \\ & & & \ddots & \vdots & \vdots \\ & & & & r_1 & r_{1,0} \\ & & & & & r_0 \end{pmatrix}, \quad (4.3)$$

其中,

(i) 对角元素 r_k 是二次方程(4.4)的一个实根

$$\sigma_k z^2 - (\lambda + \sigma_k + k\theta)z + \lambda = 0, 0 \leq k \leq d, \quad (4.4)$$

且 $r_k = \frac{1}{2\sigma_k}(\lambda + \sigma_k + k\theta - \sqrt{\Delta})$, 其中 $\Delta = (\lambda + \sigma_k + k\theta)^2 - 4\lambda\sigma_k$, $0 < r_k < 1$, $r_0 = \rho$, 并且 r_k 满

足下面关系式(4.5).

$$\lambda + k\theta + \sigma_k(1 - r_k) = \sigma_k + \frac{k\theta}{1 - r_k} = \frac{\lambda}{r_k}, \quad (4.5)$$

(ii) 非对角元素满足

$$r_{u,v} = \begin{cases} \frac{(v+1)\theta r_{u,v+1}}{[(\lambda + \sigma_v + v\theta) - \sigma_v(r_v + r_u)]}, & u = v+1, \\ \frac{(v+1)\theta r_{u,v+1} + \sigma_v \sum_{g=v+1}^{u-1} r_{u,g} r_{g,v}}{[(\lambda + \sigma_v + v\theta) - \sigma_v(r_v + r_u)]}, & v+1 < u \leq d, \end{cases} \quad (4.6)$$

其中 $u = d - i, v = d - j, 0 \leq i < d, i < j \leq d$, 以及 $r_{k,k} = r_k, 0 \leq k \leq d$.

证明. 假设方程(4.4)的另一实根是 r_k^* , 那么

$$r_k, r_k^* = \frac{1}{2\sigma_k}(\lambda + \sigma_k + k\theta \pm \sqrt{\Delta}).$$

显然, 下列不等式成立:

$$\begin{aligned} (\lambda - \sigma_k + k\theta)^2 < \Delta < (\lambda + \sigma_k + k\theta)^2, \quad \lambda > \sigma_k, \\ (\sigma_k - \lambda + k\theta)^2 < \Delta < (\lambda + \sigma_k + k\theta)^2, \quad \lambda < \sigma_k. \end{aligned}$$

能得到 $0 < r_k < 1, r_k^* \geq 1$, 和 $r_k = \frac{1}{2\sigma_k}(\lambda + \sigma_k + k\theta - \sqrt{\Delta})$.

如果 $k = 0$, 那么 $r_0 = \rho$.

将 r_k 代入方程(4.4), 能得到方程(4.5)。

矩阵 A 、 B 和 C 都是上三角阵。在方程(4.2)中, 矩阵 R 也是一个上三角阵。因此, 矩阵 R^2 也能表示为一个上三角阵。设 R 的表达式如方程(4.3), 那么

$$(R^2)_{i,j} = \begin{cases} r_{d-j}^2, & i = j, 0 \leq j \leq d, \\ \sum_{h=i}^j r_{d-i,d-h} r_{d-h,d-j}, & 0 \leq i < d, i < j \leq d. \end{cases}$$

设 $u = d - i, v = d - j, g = d - h$, 以及 $r_{k,k} = r_k, 0 \leq k \leq d$ 。将 R^2 和 R 代入矩阵方程(4.2), 则得

$$\begin{cases} \sigma_v r_v^2 - (\lambda + \sigma_v + v\theta)r_v + \lambda = 0, 0 \leq v \leq d, \\ \sigma_v \sum_{g=v}^u r_{u,g} r_{g,v} + (v+1)\theta r_{u,v+1} - (\lambda + \sigma_v + v\theta)r_{u,v} = 0, 0 < u \leq d, 0 \leq v < u, \end{cases} \quad (4.7)$$

在方程组(4.7)中, 第一个方程就是方程(4.4)。取 r_v 作为方程(4.4)在区间(0, 1)中的根, 其中 $r_v = \frac{1}{2\sigma_v}(\lambda + \sigma_v + v\theta - \sqrt{\Delta}), \Delta = (\lambda + \sigma_v + v\theta)^2 - 4\lambda\sigma_v$ 。在第二个方程中, 如果 $u = v + 1$,

那么 $\sum_{g=v}^u r_{u,g} r_{g,v} = r_{u,v} r_{v,v} + r_{u,u} r_{u,v}$; 如果 $u > v + 1$, 那么 $\sum_{g=v}^u r_{u,g} r_{g,v} = r_{u,v} r_{v,v} + r_{u,u} r_{u,v} + \sum_{g=v+1}^{u-1} r_{u,g} r_{g,v}$ 。因此, 易得方程(4.6)。显然, 当且仅当 $\rho < 1$ 时, $SP(R) = \max_{0 \leq k \leq d} \{r_k\} < 1$ 。□

4.3.2 系统的平稳概率分布

如果 $\rho < 1$ ，那么设 (L_v, J) 是 QBD 过程 $\{L_v(t), J(t)\}$ 的平稳极限，且令

$$\pi_{k,j} = \lim_{t \rightarrow \infty} P\{L_v(t) = k, J(t) = j\}, \quad (k, j) \in \Omega,$$

$$\pi_k = \begin{cases} \pi_{k,d}, & 0 \leq k \leq c-d, \\ (\pi_{k,d}, \pi_{k,d-1}, \dots, \pi_{k,c-k+1}, \pi_{k,c-k})_{1 \times (k-c+d+1)}, & c-d < k \leq c, \\ (\pi_{k,d}, \pi_{k,d-1}, \dots, \pi_{k,1}, \pi_{k,0})_{1 \times (d+1)}, & c < k \leq K. \end{cases}$$

定理 4.2 如果 $\rho < 1$ ，那么 (L_v, J) 的平稳概率分布 $\pi_{k,j}$ 是

$$\pi_{k,j} = \begin{cases} G \frac{\lambda^k}{\prod_{i=1}^k \delta_{i,0}}, & 0 \leq k \leq c-d, j = d, \end{cases} \quad (4.8)$$

$$\pi_{k,j} = \begin{cases} \frac{\lambda^{k+j-c}}{\prod_{i=1}^{k+j-c} \delta_{c-j+i,i}} \pi_{c-j,j} + \sigma_j \left[\pi_{N,j} - \pi_{c,j} \left(\frac{\lambda}{\sigma_j} \right)^{N-c} \right] \sum_{h=0}^{k+j-c-1} \frac{\lambda^h}{\prod_{i=k-h}^k \delta_{i,i+j-c}}, & c-d+1 \leq k \leq c, c-k+1 \leq j \leq d, \end{cases} \quad (4.9)$$

$$\pi_{k,j} = \begin{cases} \frac{1}{\delta_{k,0}} \sum_{i=c-k+1}^d \sigma_i \left[\pi_{c,i} \left(\frac{\lambda}{\sigma_i} \right)^{N-c} - \pi_{N,i} \right], & c-d+1 \leq k < c, j = c-k, \end{cases} \quad (4.10)$$

$$\pi_{c,j} \left(\frac{\lambda}{\sigma_j} \right)^{k-c}, \quad c < k < N, 1 \leq j \leq d, \quad (4.11)$$

$$\pi_{c,0} \frac{1 - \rho^{k-c+1}}{1 - \rho}, \quad c < k < N, j = 0, \quad (4.12)$$

以及

$$\pi_{N,j} = \begin{cases} r_d \pi_{c,d} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}, & j = d, \end{cases} \quad (4.13)$$

$$\pi_{N,j} = \begin{cases} r_j \pi_{c,j} \left(\frac{\lambda}{\sigma_j} \right)^{N-1-c} + \frac{r_j}{\lambda} \left[\sigma_j \sum_{i=j+1}^d \pi_{N,i} r_{i,j} + (j+1)\theta \pi_{N,j+1} \right], & 1 \leq j \leq d-1, \end{cases} \quad (4.14)$$

$$\pi_{N,j} = \begin{cases} \pi_{c,0} \frac{1 - \rho^{N-c+1}}{1 - \rho}, & j = 0, \end{cases} \quad (4.15)$$

和

$$\lambda\pi_{N-1,d} + [r_d\sigma_d - (\lambda + \sigma_d + d\theta)]\pi_{N,d} = 0, \quad k = N, \quad j = d. \quad (4.28)$$

由式(4.20), 有

$$\delta_{k,0}\pi_{k,d} - \lambda\pi_{k-1,d} = \delta_{1,0}\pi_{1,d} - \lambda\pi_{0,d}, \quad 1 \leq k \leq c-d. \quad (4.29)$$

由式(4.19)和(4.29), 得到式(4.8)。

由(4.24), (4.25)和 $\delta_{c,0} = \sigma_0$, 则得

$$\pi_{k+1,0} = \rho\pi_{k,0} + \pi_{c,0}, \quad c < k < N. \quad (4.30)$$

反复迭代上述方程(4.30), 得到式(4.12)。

容易证明, 方程(4.11)满足式(4.26)。

由式(4.5), (4.11)和(4.28), 得到式(4.13), 即

$$\pi_{N,d} = r_d\pi_{N-1,d} = r_d\pi_{c,d} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c}.$$

将式(4.11)和(4.12)分别代入(4.27), 得到(4.14)和(4.15)。

从式(4.26), 得到

$$\sigma_j\pi_{N,j} - \lambda\pi_{N-1,j} = \sigma_j\pi_{c,j} - \lambda\pi_{c-1,j}, \quad 1 \leq j \leq d. \quad (4.31)$$

由式(4.23), (4.31)和 $\delta_{c,j} = \sigma_j$, 有

$$\delta_{k,k+j-c}\pi_{k,j} = \lambda\pi_{k-1,j} + \delta_{c,j}\pi_{N,j} - \lambda\pi_{N-1,j}, \quad c-d < k \leq c, \quad c-k+1 \leq j \leq d. \quad (4.32)$$

反复迭代方程(4.32), 得到式(4.9)。

在方程(4.9)中, 设 $k = c$, $j = d$, 并将式(4.8)和(4.13)代入(4.9)式, 得到式(4.16)。

由式(4.20)和(4.21), 得到

$$\pi_{c-d+1,d-1} = \lambda\pi_{N-1,d} - \delta_{c,d}\pi_{N,d} = \frac{\lambda - \delta_{c,d}r_d}{\delta_{c-d+1,0}} \left(\frac{\lambda}{\sigma_d} \right)^{N-1-c} \pi_{c,d}. \quad (4.33)$$

根据式(4.22), 有

$$\pi_{k,j} = \frac{\lambda + \delta_{k-1,0}}{\delta_{k,0}} \pi_{k-1,j+1} - \frac{\delta_{k,1}}{\delta_{k,0}} \pi_{k,j+1}, \quad c-d+2 \leq k \leq c-1, \quad j = c-k. \quad (4.34)$$

将式(4.9)代入(4.34)式, 得到下面表达式(4.35)。

$$\pi_{k,j} = \frac{\delta_{k-1,0}}{\delta_{k,0}} \pi_{k-1,j+1} + \frac{\sigma_{j+1}}{\delta_k} \left[\pi_{c,j+1} \left(\frac{\lambda}{\sigma_{j+1}} \right)^{N-c} - \pi_{N,j+1} \right]. \quad (4.35)$$

反复迭代上述方程, 得到

$$\pi_{k,j} = \frac{1}{\delta_{k,0}} \sum_{i=c-k+1}^d \sigma_i \left[\pi_{c,i} \left(\frac{\lambda}{\sigma_i} \right)^{N-c} - \pi_{N,i} \right], \quad c-d+2 \leq k \leq c-1, \quad j = c-k. \quad (4.36)$$

因此, 在式(4.36)中, 当 $k = c-d+1$ 以及 $j = d-1$ 时, 式(4.33)满足式(4.36), 故由式(4.33)和(4.36), 得到(4.10)式。

在式(4.9)中, 若 $1 \leq j \leq d-1$, 那么 $s = c-j$ 。根据式(4.9), (4.10), 以及式(4.14), 得到式(4.17)。由式(4.12)和(4.15), 得到式(4.18)。

最后, 待定常数 G 满足归一化条件如下:

$$\sum_{k=0}^{c-d-1} \pi_{k,d} + \sum_{j=0}^d \sum_{k=c-j}^K \pi_{k,j} = 1. \quad \square$$

定理 4.3 如果 $k > N$ ，那么 (L_v, J) 的平稳概率分布为：

$$\pi_{k,d} = \pi_{N,d} r_d^{k-N}, \quad N < k < K, \quad j = d, \quad (4.37)$$

$$\varphi_k^{(d)} = \pi_{N,d} \sum_{i=1}^{k-N} r_d^{i-1} \eta H^{k-N-i} + \varphi_N^{(d)} H^{k-N}, \quad N < k < K, \quad 0 \leq j \leq d-1, \quad (4.38)$$

$$\pi_{K,d} = \frac{\lambda r_d^{K-1-N}}{\sigma_d + d\theta} \pi_{N,d}, \quad k = K, \quad j = d, \quad (4.39)$$

$$\pi_{K,j} = \lambda \sum_{h=j}^{d-1} \left[\frac{1}{(h+1)\theta} \prod_{i=j}^h \frac{(i+1)\theta}{\sigma_i + i\theta} \pi_{K-1,h} \right] + \frac{\lambda r_d^{K-1-N}}{(d+1)\theta} \prod_{i=j}^d \frac{(i+1)\theta}{\sigma_i + i\theta} \pi_{N,d}, \quad (4.40)$$

$$k = K, \quad 0 \leq j \leq d-1,$$

其中 $\varphi_k^{(d)} = (\pi_{k,d-1}, \pi_{k,d-2}, \dots, \pi_{k,1}, \pi_{k,0})_{1 \times d}$ ， $R = \begin{pmatrix} r_d & \eta \\ 0 & H \end{pmatrix}$ ，且 $\eta = (r_{d,d-1}, r_{d,d-2}, \dots, r_{d,1}, r_{d,0})$ 。

证明. 当 $N < k < K$ ，根据矩阵几何解^[192]，有

$$\pi_k = \pi_N R^{k-N}, \text{ i.e., } (\pi_{k,d}, \varphi_k^{(d)}) = (\pi_{N,d}, \varphi_N^{(d)}) R^{k-N}. \quad (4.41)$$

分块矩阵 R 的 k 次幂如下：

$$R^k = \begin{pmatrix} r_d^k & \sum_{i=1}^k r_d^{i-1} \eta H^{k-i} \\ 0 & H^k \end{pmatrix}. \quad (4.42)$$

将式(4.42)代入式(4.41)，式(4.37)和式(4.38)分别产生。

当 $k = K$ 时，基于平衡方程，有

$$\lambda \pi_{K-1,d} - (\sigma_d + d\theta) \pi_{K,d} = 0, \quad j = d, \quad (4.43)$$

$$\lambda \pi_{K-1,j} - (\sigma_j + j\theta) \pi_{K,j} + (j+1)\theta \pi_{K,j+1} = 0, \quad 0 \leq j \leq d-1. \quad (4.44)$$

由式(4.43)，得到(4.39)式，即

$$\pi_{K,d} = \frac{\lambda}{\sigma_d + d\theta} \pi_{K-1,d} = \frac{\lambda r_d^{K-1-N}}{\sigma_d + d\theta} \pi_{N,d}.$$

由式(4.44)，得到

$$\pi_{K,0} = \lambda \sum_{h=0}^{j-1} \frac{h! \theta^h}{\prod_{i=0}^h (\sigma_i + i\theta)} \pi_{K-1,h} + \frac{j! \theta^j}{\prod_{i=0}^{j-1} (\sigma_i + i\theta)} \pi_{K,j}, \quad 1 \leq j \leq d. \quad (4.45)$$

在式(4.45)中，令 $j = d$ ，则得

$$\pi_{K,j} = \lambda \sum_{h=j}^{d-1} \left[\frac{1}{(h+1)\theta} \prod_{i=j}^h \frac{(i+1)\theta}{\sigma_i + i\theta} \pi_{K-1,h} \right] + \prod_{i=j}^{d-1} \frac{(i+1)\theta}{\sigma_i + i\theta} \pi_{K,d}, \quad 0 \leq j \leq d-1. \quad (4.46)$$

将式(4.39)代入式(4.46)，则得到式(4.40)。□

4.3.3 系统性能度量

用上述得到的平稳分布，数值上评估 $N/(d, c)\text{-M/M/c/K/AMWV}$ 排队系统的性能，该系统的各种性能指标表示如下：

(1) 系统中顾客期望数是

$$E[L_s] = \sum_{i=0}^{c-d-1} i \pi_{i,d} + \sum_{j=0}^d \sum_{i=c-j}^K i \pi_{i,j}. \quad (4.47)$$

(2) 队列中顾客期望数是

$$E[L_q] = \sum_{j=0}^d \sum_{i=c-j+1}^K (i-c+j)\pi_{i,j}. \quad (4.48)$$

(3) 正常忙期服务台平均数是

$$E[NB] = e[E]WV. \quad (4.49)$$

(4) 工作假期服务台平均数是

$$E[WV] = d \sum_{i=0}^{c-d-1} \pi_{i,d} + \sum_{j=1}^d \sum_{i=c-j}^K j\pi_{i,j}. \quad (4.50)$$

(5) 空闲期服务台平均数是

$$E[I] = \sum_{i=0}^{c-d} (c-d-i)\pi_{i,d}. \quad (4.51)$$

(6) 顾客在队列中驻留的平均等待时间是

$$E[T_q] = \frac{E[L_q]}{\lambda(1-P_{loss})}. \quad (4.52)$$

(7) 系统中顾客损失概率是

$$P_{loss} = \sum_{j=0}^d \pi_{j,d}. \quad (4.53)$$

4.4 最优化和灵敏度分析

4.4.1 最小平均 CPR

(1) **成本**: 利用文献[193]中的船员服务(crew-service)设备的概念, 单位时间的平均成本函数表示为

$$Cost(d, N, K, \mu_B, \mu_V, \lambda, \theta) = C_h E[L_s] + C_b E[NB] + C_v E[WV] + C_i E[I] + C_d E[L_q] + C_l P_{loss}, \quad (4.54)$$

其中, 成本参数 C_h, C_b, C_v, C_i, C_d , 以及 C_l 的定义与 3.5.2 节中的定义相同, 且 $E[L_s], E[L_q], E[NB], E[WV], E[I], P_{loss}$ 分别由方程(4.47)~(4.51)和(4.53)计算。

(2) **性能**: 用顾客在系统中驻留的平均时间的倒数作为系统性能指标(见文献[39]中的第5节): $P = \frac{1}{E[T]}$, 其中 $E[T] = \frac{E[L_s]}{\lambda(1-P_{loss})}$ 。

(3) **CPR**: 根据文献[39]中的(4.16)式和本文中的(4.54)式, 定义FSDMC模型的CPR为 $CPR = \frac{F}{P} = Cost * E[T]$ 。考虑如下最优化模型:

$$\begin{aligned} & \min \quad CPR(d, N, K, \mu_B, \mu_V, \lambda, \theta), \\ & \text{s. t.} \quad \begin{cases} 0 < d < c < N < K < K', \mu_V < \mu_B, \\ \mu'_B \leq \mu_B \leq \mu''_B, \mu'_V \leq \mu_V \leq \mu''_V, \\ \lambda' \leq \lambda \leq \lambda'', \theta' \leq \theta \leq \theta'', \\ E[NB] \leq c - d, \end{cases} \end{aligned} \quad (4.55)$$

其中, 参数的上下界与多目标优化模型(3.41)中的参数上下界相同。

本章的目的是系统中的参数 $d, N, K, \mu_B, \mu_V, \lambda, \theta$ 的最优值, 给出系统在稳定状态下的配置。因此, PSO 算法[62]数值上实现, 以处理该最优化问题。

(4) 最优系统配置: 用PSO算法求解最优化模型(4.55), 得到系统CPR的最小值。采用3.5.2节中的系统参数值为: $C_h=1\text{mW}, C_b=35\text{mW}, C_v=15\text{mW}, C_i=10\text{mW}, C_d=8\text{mW}, C_l=10000\text{mW}, c=10$ 。决策变量的取值范围为: $d \in [1, c], N \in [c+1, 25], K \in [N+1, 60], \mu_B \in [1.6, 3.4], \mu_V \in [0.5, 1.5], \lambda \in [4, 30]$, 和 $\theta \in [0.01, 10]$, 并满足模型(4.55)中的约束条件。

在本章的实验中, 获得的决策变量的最优值为: $d=4, N=20, K=42, \mu_B=3.4, \mu_V=1.5, \lambda=16.9, \theta=6.6$, CPR 的最小值为 $CPR_0 = 53.1134$ 。用这些参数值配置系统, 并称此时的系统为“稳定系统”, 或者称此时的系统处于“稳定状态”。值得说明的是, 在求解过程中, μ_B 和 μ_V 的最优值总是取值为各自的约束条件的上界 μ_B'' 和 μ_V'' , 这是因为随着 μ_B 或者 μ_V 的增加, CPR 随之减少(见图 4.4~图 4.7)。因此, 采用 3.5.3 节中 μ_B 和 μ_V 的值作为系统参数 μ_B 和 μ_V 各自的值。

4.4.2 系统性能度量灵敏度分析

在稳定条件下, 得到一些数值实验结果, 表明了系统参数的改变对系统性能产生的影响。用 Matlab 进行数值实验, 固定系统最大容量为 $K = 42$, 并考虑在下面三种情况下系统参数的不同值:

情况 1: $\lambda=16.9, \theta=6.6, d=4, c=10, N=20$, 且 (μ_B, μ_V) 的值变化;

情况 2: $\mu_B=3.4, \mu_V=1.5, d=4, c=10, N=20$, 且 (λ, θ) 的值变化;

情况 3: $\lambda=16.9, \theta=6.6, \mu_B=3.4, \mu_V=1.5$, 且 (d, c, N) 的值变化;

上面三种情况下系统参数的数值实验结果分别如表 4.1~4.3 所示。在表 4.1 中, 发现随着 μ_B 或者 μ_V 的增加, $E[L_s]$ 、 $E[L_q]$ 以及 $E[NB]$ 都减少, $E[WW]$ 稍微减少, 但 $E[I]$ 增加。表 4.2 表明(i)当 λ 增加时, $E[L_s]$ 和 $E[NB]$ 增加, 但当 θ 增加时, $E[L_s]$ 和 $E[NB]$ 略微增加; (ii)当 λ 增加时, $E[L_q]$ 和 $E[WW]$ 增加, 但当 θ 增加时, $E[L_q]$ 和 $E[WW]$ 稍微减少; (iii) 当 λ 增加时, $E[I]$ 减少, 但当 θ 增加时, $E[I]$ 无变化。表 4.3 表明(i)当 d 增加时, $E[L_s]$ 、 $E[L_q]$ 和 $E[WW]$ 显著增加, 但 $[NB]$ 和 $E[I]$ 显著减少; (ii)随着 c 的增加, $E[L_s]$ 和 $E[WW]$ 减少, 且 $E[L_q]$ 显著减少, 但 $E[NB]$ 增加, 且 $E[I]$ 明显增加; (iii) 与前述(i)类似, 当 N 增加时, $E[L_s]$ 、 $E[L_q]$ 和 $E[WW]$ 略微增加, 但 $E[NB]$ 和 $E[I]$ 略微减少。

表 4.1 在 $\lambda=16.9, \theta=6.6, d=4, c=10, N=20$ 情况下 (μ_B, μ_V) 取不同值时系统性能度量

(μ_B, μ_V)	(2.4, 1.5)	(3.4, 1.5)	(4.4, 1.5)	(3.4, 0.5)	(3.4, 2.5)
$E[L_s]$	9.2830	5.5363	4.0266	6.2840	5.1866
$E[L_q]$	3.3622	0.8914	0.2710	1.4613	0.6358
$E[NB]$	5.1571	4.2040	3.5380	4.3220	4.1467
$E[WW]$	4.4462	4.4362	4.2175	4.4683	4.4033
$E[I]$	0.3967	1.3598	2.2444	1.2097	1.4500

表 4.2 在 $\mu_B=3.4, \mu_V=1.5, d=4, c=10, N=20$ 情况下(λ, θ)取不同值时系统性能度量

(λ, θ)	(15.9, 6.6)	(16.9, 6.6)	(17.9, 6.6)	(16.9, 5.6)	(16.9, 7.6)
$E[L_s]$	5.0840	5.536 280	6.0379	5.536 249	5.536 423
$E[L_q]$	0.6585	0.891 352	1.1843	0.891 786	0.891 037
$E[NB]$	4.0486	4.204 032	4.3480	4.203 491	4.204 562
$E[WW]$	4.3752	4.436 159	4.4940	4.436 711	4.435 634
$E[I]$	1.5763	1.359 808	1.1580	1.359 798	1.359 803

表 4.3 在 $\lambda=16.9, \theta=6.6, \mu_B=3.4, \mu_V=1.5$ 情况下(d, c, N)取不同值时系统性能度量

(d, c, N)	(4, 10, 20)	(7, 10, 20)	(4, 15, 20)	(4, 10, 15)
$E[L_s]$	5.5363	8.0472	4.9762	5.4805 01
$E[L_q]$	0.8914	4.7934	0.0098	0.8135 01
$E[NB]$	4.2040	2.3067	4.9526	4.2314 00
$E[WW]$	4.4362	7.6232	4.0138	4.3995 01
$E[I]$	1.3598	0.0700	6.0336	1.3691 00

4.4.3 CPR 灵敏度分析

在稳定条件下，得到一些实验结果，表明系统参数值的变化对 CPR 的影响。固定成本参数值： $Ch=1mW, Cb=35mW, Cv=15mW, Ci=10mW, Cd=8mW, Cl=10000mW$ ，以及控制器数 $c = 10$ 和最大容量 $K = 42$ ，并考虑系统参数取不同值时的下列八种情况：

- 情况 4： $N=20, \mu_B=3.4, \mu_V=1.5, \theta=6.6$ ，且 λ 的值从 1.0 到 30.0 变化， $d=4, 7, 9$ ；
- 情况 5： $d=4, \mu_B=3.4, \mu_V=1.5, \theta=6.6$ ，且 λ 的值从 1.0 到 30.0 变化， $N=15, 20, 25$ ；
- 情况 6： $N=20, \mu_V=1.5, \lambda=16.9, \theta=6.6$ ，且 μ_B 的值从 2.0 到 20.0 变化， $d=4, 7, 9$ ；
- 情况 7： $d=4, \mu_V=1.5, \lambda=16.9, \theta=6.6$ ，且 μ_B 的值从 2.0 到 20.0 变化， $N=15, 20, 25$ ；
- 情况 8： $N=20, \mu_B=3.4, \lambda=16.9, \theta=6.6$ ，且 μ_V 的值从 0 到 10.0 变化， $d=4, 7, 9$ ；
- 情况 9： $d=4, \mu_B=3.4, \lambda=16.9, \theta=6.6$ ，且 μ_V 的值从 0 到 10.0 变化， $N=15, 20, 25$ ；
- 情况 10： $N=20, \mu_B=3.4, \mu_V=1.5, \lambda=16.9$ ，且 θ 的值从 2.0 到 20.0 变化， $d=4, 7, 9$ ；
- 情况 11： $d=4, \mu_B=3.4, \mu_V=1.5, \lambda=16.9$ ，且 θ 的值从 2.0 到 20.0 变化， $N=15, 20, 25$ ；

上述八种情况下的 CPR 数值结果分别如图 4.2~4.9。图 4.2 表明，当 λ 和 d 增加时，CPR 明显波动。在图 4.3 中，(i)当 λ 增加时，CPR 首先减少，然后显著增加；(ii)当 N 增加时，CPR 首先无变化，但随后明显增加。在图 4.4~4.7 中，当 μ_B 和 μ_V 分别增加时，CPR 明显减少，但是随着 d 的变化，CPR 明显波动，而当 N 增加时，CPR 略微增加。图 4.8 和图 4.9 表明当 d 和 N 增加时，CPR 增加，但当 N 增加时，CPR 细微改变。在 4.5 节，将考虑应力(到达率)对应变(CPR)的影响。

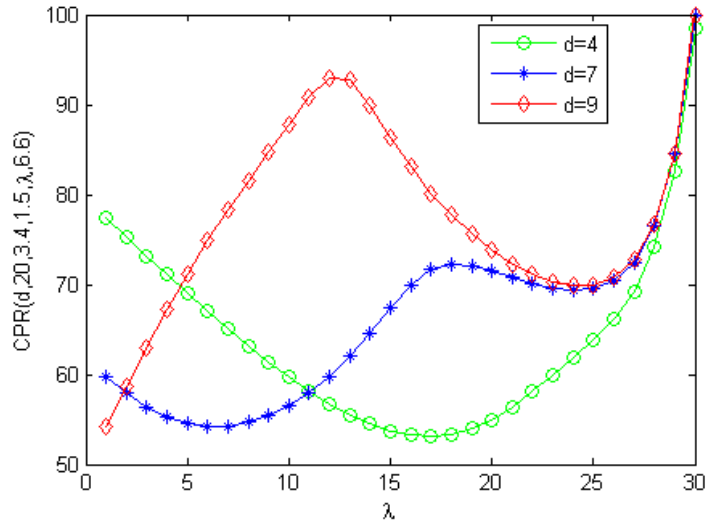


图 4.2 λ 和 d 变化时 $CPR(d, 20, 3.4, 1.5, \lambda, 6.6)$ 变化图

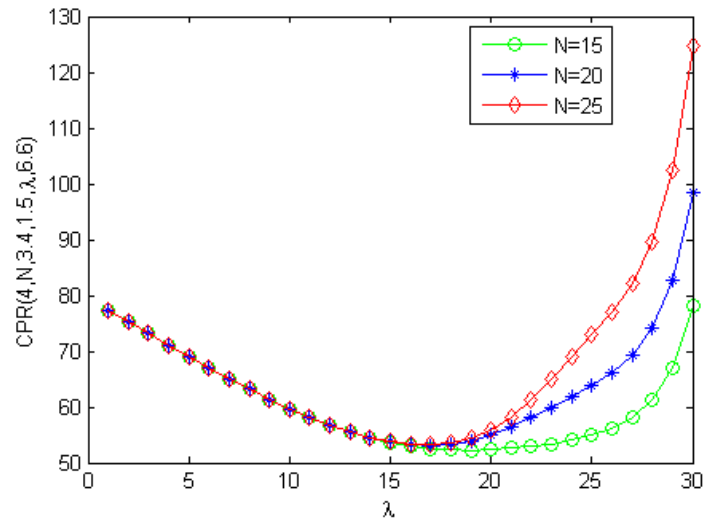


图 4.3 λ 和 N 变化时 $CPR(4, N, 3.4, 1.5, \lambda, 6.6)$ 变化图

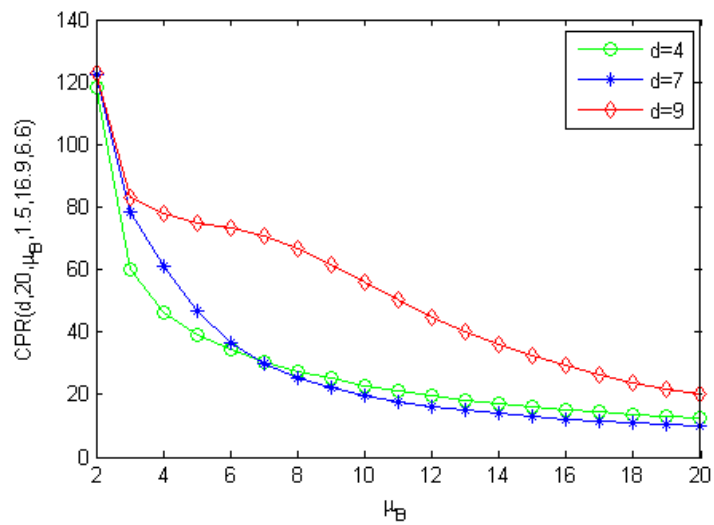


图 4.4 μ_B 和 d 变化时 $CPR(d, 20, \mu_B, 1.5, 16.9, 6.6)$ 变化图

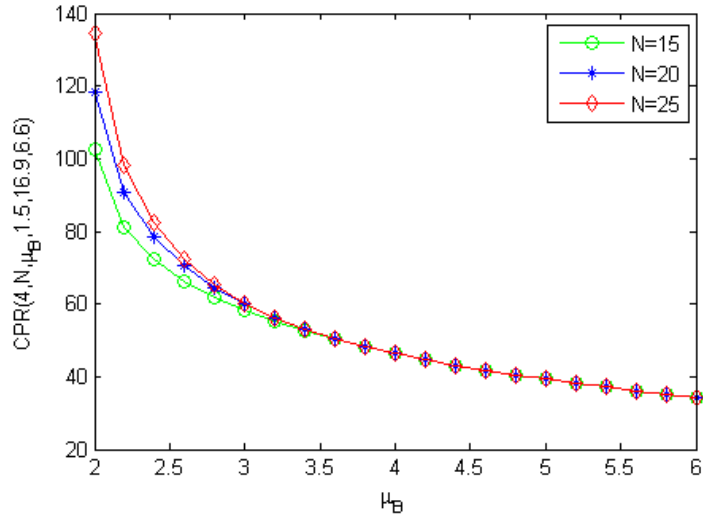


图 4.5 μ_B 和 N 变化时 $CPR(4, N, \mu_B, 1.5, 16.9, 6.6)$ 变化图

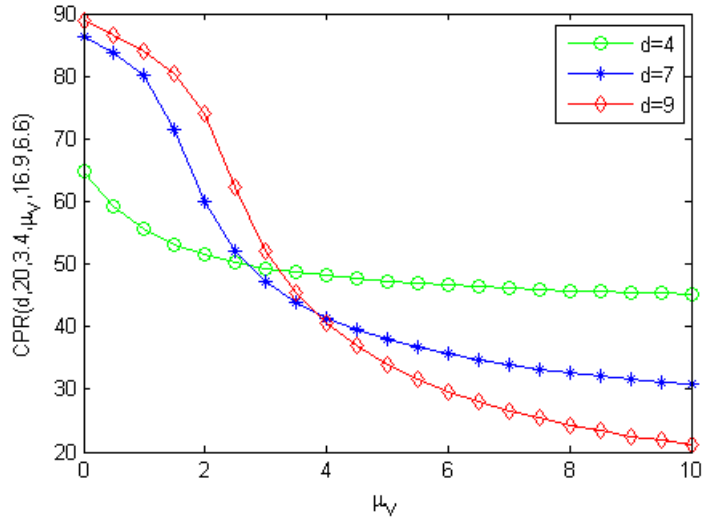


图 4.6 μ_V 和 d 变化时 $CPR(d, 20, 3.4, \mu_V, 16.9, 6.6)$ 变化图

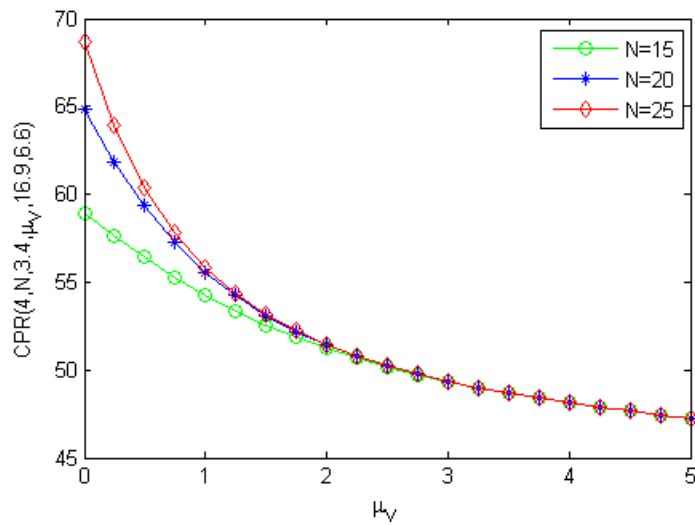


图 4.7 μ_V 和 N 变化时 $CPR(4, N, 3.4, \mu_V, 16.9, 6.6)$ 变化图

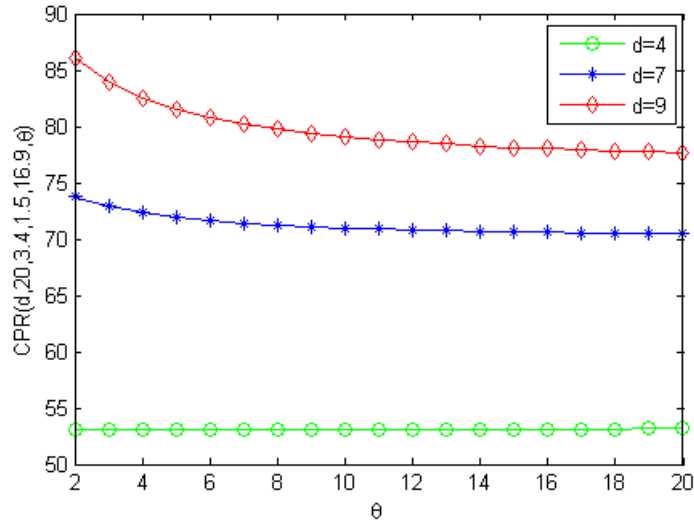


图 4.8 θ 和 d 变化时 $CPR(d, 20, 3.4, 1.5, 16.9, \theta)$ 变化图

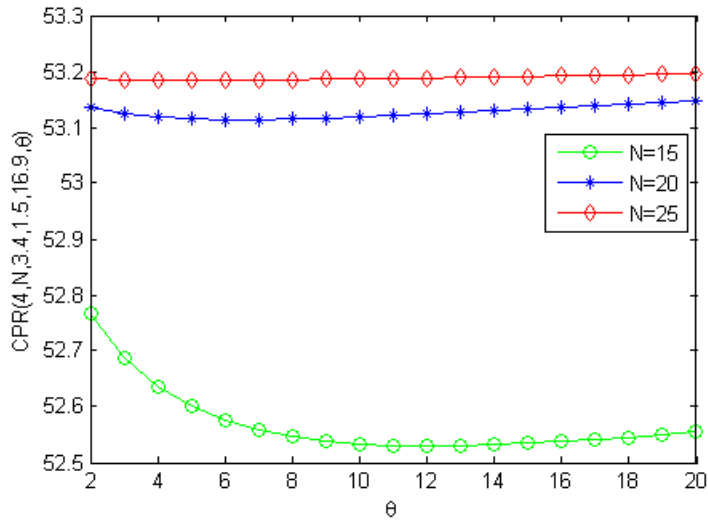


图 4.9 θ 和 N 变化时 $CPR(4, N, 3.4, 1.5, 16.9, \theta)$ 变化图

在 $N/(d, c)$ -M/M/c/K/AMWV 排队系统中，到达率可能随时间的变化而随机变化，导致可能的高峰期(rush-hour)行为。因此，队列可能表现长时间严重恶化。本章建立一种马尔可夫环境^[192]下的 M/M/1/K 排队模型，构建滞弹性量化模型，并分析系统高峰期行为。

4.5 滞弹性量化模型

4.5.1 马尔可夫环境下的 M/M/1/K 排队模型

在这一部分，考虑一种马尔可夫环境下的 M/M/1/K 队列模型，其中 K 表示系统容量。马尔可夫环境由具有生成器 \bar{Q} 的 3-状态、不可约马尔可夫更新过程描述。为简便起见，3-状态分别由状态 1, 2 和 3 表示。在状态 $i, 1 \leq i \leq 3$ 期间，顾客依据速率为 λ_i 的泊松过程到达一个单服务台队列。在状态 i 的时间间隔期间，平均服务率为 μ_i 。当在过程 \bar{Q} 中，状态 i 转移到状态 j 时，平均到达率和平均服务率同时变化到新的值 λ_j 和 μ_j 。

在图 4.1 中, 根据 3-状态的定义, 应该注意到仅仅转移在状态 1 和 2 以及状态 2 和 3 之间发生。因此, 矩阵 \bar{Q} 可以表示为

$$\bar{Q} = \begin{pmatrix} -\mu_1 & \mu_1 & 0 \\ \mu_2 & -2\mu_2 & \mu_2 \\ 0 & \mu_3 & -\mu_3 \end{pmatrix},$$

其中 $\mu_i, 1 \leq i \leq 3$ 表示 3-状态中的第 i 个状态平均服务率。

在矩阵 \bar{Q} 中, μ_i 由下式计算

$$\mu_i = E[NB_i] \mu_B + E[WV_i] \mu_V, \quad i = 1, 2, 3, \quad (4.56)$$

其中, $E[NB_i]$ 和 $E[WV_i]$ 分别表示系统处于状态 i 时的正常忙期和工作休假期的平均服务台数, 其值可由式(4.49)、(4.50)和(4.51)计算而得。

设 $\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3)$ 是马尔可夫过程 \bar{Q} 的平稳概率向量, 那么

$$\tilde{\pi} \bar{Q} = 0, \quad \tilde{\pi} e = 1. \quad (4.57)$$

由(4.57)式, 得到 $\tilde{\pi} = (\mu_2 \mu_3, \mu_3 \mu_1, \mu_1 \mu_2) (\mu_1 \mu_2 + \mu_2 \mu_3 + \mu_3 \mu_1)^{-1}$. 经过标准构建, M/M/1/K 队列可以由具有生成元 \tilde{Q} 的 QBD 过程进行研究, 即

$$\tilde{Q} = \begin{pmatrix} \bar{Q} - \Delta(\lambda) & \Delta(\lambda) & & & & \\ \Delta(\mu) & \bar{Q} - \Delta(\lambda + \mu) & \Delta(\lambda) & & & \\ & & \ddots & \ddots & \ddots & \\ & & & \Delta(\mu) & \bar{Q} - \Delta(\lambda + \mu) & \Delta(\lambda) \\ & & & & \Delta(\mu) & \bar{Q} - \Delta(\mu) \end{pmatrix},$$

其中 $\Delta(\lambda)$ 和 $\Delta(\mu)$ 分别表示对角矩阵 $diag(\lambda_1, \lambda_2, \lambda_3)$ 和 $diag(\mu_1, \mu_2, \mu_3)$ (见文献[192]中 6.2 节)。

矩阵 \tilde{R} 是下面矩阵二次方程的最小非负解

$$\tilde{R}^2 \Delta(\mu) + \tilde{R} [\bar{Q} - \Delta(\lambda + \mu)] + \Delta(\lambda) = 0, \quad (4.58)$$

其中 \tilde{R} 可以用快速 Newton-Shamanskii 迭代算法求解^[195]。

设 $X = (x_0, x_1, x_2, \dots, x_K)$ 是稳定队列的平稳概率向量, 其中 $x_k = (x_{k,1}, x_{k,2}, x_{k,3}), 0 \leq k \leq K$ 。在 $x_{k,j}$ 中, k 和 j 分别表示队列长度和环境状态, 且 $j = 1, 2, 3$ 。

定理 4.4 如果 $\tilde{\rho} = \tilde{\pi} \tilde{\lambda} (\tilde{\pi} \tilde{\mu})^{-1} < 1$, 那么 x_k 的平稳概率分布为

$$x_k = \begin{cases} \tilde{G} \tilde{\pi} (I - \tilde{R}) \tilde{R}^k, & 0 \leq k < K, \\ \tilde{G} \tilde{\pi} (I - \tilde{R}) \tilde{R}^{K-1} \Delta(\lambda) [\Delta(\mu) - \bar{Q}]^{-1}, & k = K, \end{cases} \quad (4.59)$$

其中 $\tilde{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^T$, $\tilde{\mu} = (\mu_1, \mu_2, \mu_3)^T$, 且

$$\tilde{G} = [\tilde{\pi} (I - \tilde{R}^K) e + \tilde{\pi} (I - \tilde{R}) \tilde{R}^{K-1} \Delta(\lambda) (\Delta(\mu) - \bar{Q})^{-1} e]^{-1}. \quad (4.60)$$

证明. 根据文献[192]中的定理 6.2.1, 有

$$x_k = \tilde{G} \tilde{\pi} (I - \tilde{R}) \tilde{R}^k, \quad 0 \leq k < K, \quad (4.61)$$

其中 \tilde{G} 表示待定常数。

由 $x_K \tilde{Q} = 0$, 得到平稳方程如下:

$$x_{K-1} \Delta(\lambda) + x_K [\bar{Q} - \Delta(\mu)] = 0. \quad (4.62)$$

由式(4.61)和(4.62), 得到方程组(4.59)中的第二个表达式。

待定常数 \tilde{G} 由归一化条件得到

$$\sum_{k=0}^{K-1} \tilde{G} \tilde{\pi} (I - \tilde{R}) \tilde{R}^k e + \tilde{G} \tilde{\pi} (I - \tilde{R}) \tilde{R}^{K-1} \Delta(\lambda) [\Delta(\mu) - \bar{Q}]^{-1} e = 1. \quad (4.63)$$

由式(4.63)，能得到式(4.60). □

4.5.2 弹性灵敏度分析

根据方程组(4.59)，能计算弹性度量如下。系统在 3-状态下的概率分别表示为：

$$p_{over} = \sum_{j=1}^3 \sum_{i=0}^{c-1} x_{i,j}, \quad (4.64)$$

$$p_{normal} = \sum_{j=1}^3 \sum_{i=c}^{N-1} x_{i,j}, \quad (4.65)$$

$$p_{under} = \sum_{j=1}^3 \sum_{i=N}^K x_{i,j}. \quad (4.66)$$

在通信强度 $\rho = (\tilde{\pi}\lambda)(\tilde{\pi}\mu)^{-1} < 1$ 的情况下，通过数值实验证明系统基本参数的变化对弹性的影响。固定控制器数 $c=10$ ，半休眠控制器数 $d=4$ ，阈值 $N=20$ ，以及最大容量 $K = 42$ ，并考虑系统参数取不同值时下列四种情况：

情况 12: $\mu_B=3.4, \mu_V=1.5, \theta=6.6$ ，且 λ 的值从 5.0 到 34.0 变化；

情况 13: $\lambda=16.9, \mu_V=1.5, \theta=6.6$ ，且 μ_B 的值从 2.0 到 20.0 变化；

情况 14: $\lambda=16.9, \mu_B=3.4, \theta=6.6$ ，且 μ_V 的值从 0 到 10.0 变化；

情况 15: $\lambda=16.9, \mu_B=3.4, \mu_V=1.5$ ，且 θ 的值从 2.0 到 20.0 变化；

对于上述情况 12-14，弹性数值实验结果分别如图 4.10~4.12。在图 4.10 中，当 λ 增加时， p_{over} 明显减少， p_{under} 显著增加，但 p_{normal} 变化，即先增加后略微减少；图 4.11 表明当 μ_B 增加时， p_{normal} 和 p_{under} 减少，但 p_{over} 极其微小的增加；图 4.12 显示当 μ_V 增加时，弹性概率的变化与图 4.11 中的概率变化相同。但是，对于情况 15，当 θ 增加时， p_{over}, p_{normal} 和 p_{under} 保持基本不变，即 $p_{over}=0.9945, p_{normal}=0.5452e-02, p_{under}=0.3038e-04$ 。因此，系统弹性与控制器的的工作休假率的变化无关。

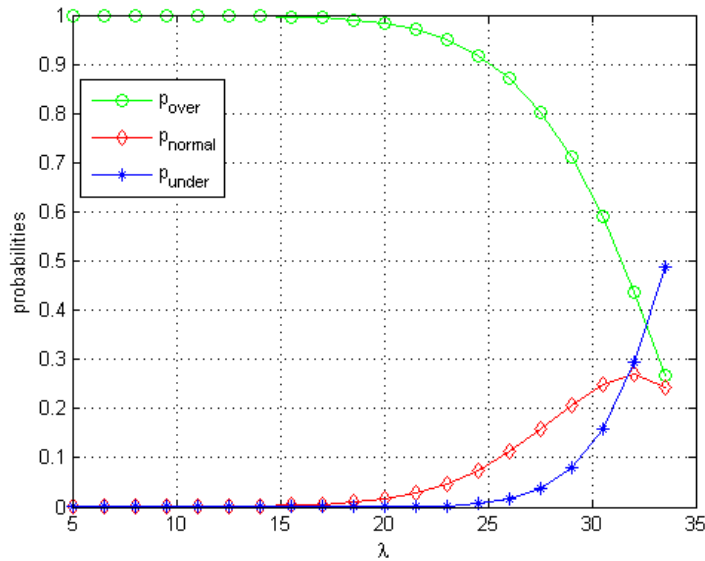


图 4.10 当 $\rho < 1$ 时 p_{over}, p_{normal} 和 p_{under} 随 λ 变化的曲线图

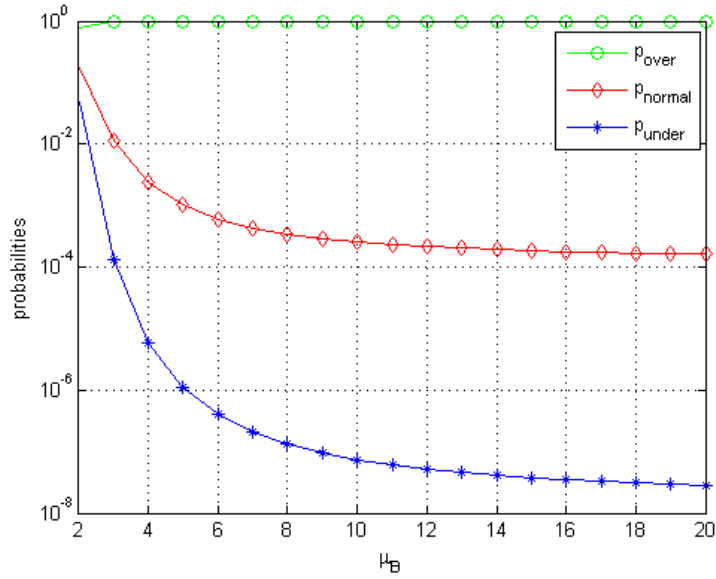


图 4.11 当 $\rho < 1$ 时 p_{over} , p_{normal} 和 p_{under} 随 μ_B 变化的曲线图

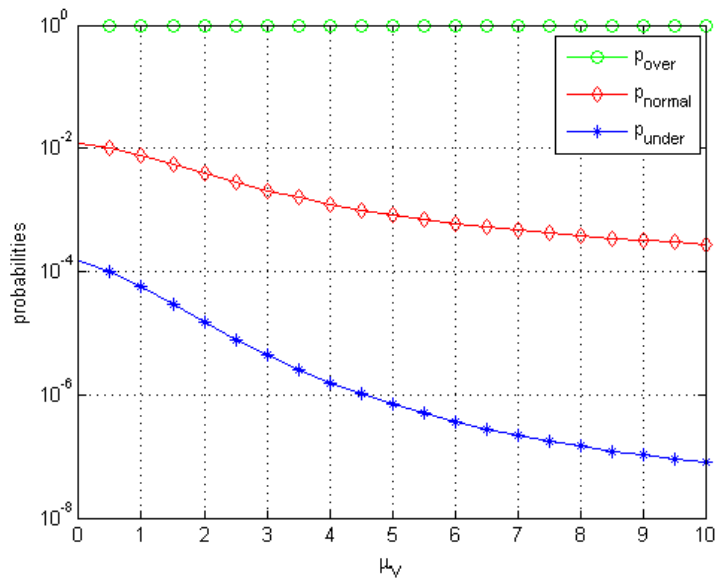


图 4.12 当 $\rho < 1$ 时 p_{over} , p_{normal} 和 p_{under} 随 μ_V 变化的曲线图

4.5.3 系统自适应分析

本节分析系统自适应能力。在 $N/(d, c)-M/M/c/K/SMWV$ 排队系统中，当通信强度 $\rho = \lambda \sigma_0^{-1} < 1$ 时，系统改变半休眠控制器数 d 以适应负载变化(见图 4.2)。但是，当 $\rho \geq 1$ 时，增加控制器在忙期的平均服务率，并设置 $d = 0$ ，使得排队系统变成经典的 $M/M/c/K$ 排队系统。随着平均到达率的减少，当系统中的顾客数小于 c 时，处于空闲状态的控制器开始工作休假，直到半休眠控制器数等于 d 为止。此时，系统再次变为稳定状态。

当平均到达率增加到 $\rho \geq 1$ 时, 系统自适应调节控制器忙期服务率 μ_B , 同时保持系统在稳定状态下 CPR 的最优值 CPR_0 不变。基于 PSO 的最优 CPR 不变自适应调节算法如算法 4.1。

算法 4.1 基于 PSO 的最优 CPR 不变自适应调节算法

输入: 系统平均到达率 λ , 以及精度 $\varepsilon=10^{-7}$ 。

输出: 正常忙期控制器平均服务率 μ_B 。

- (1) 初始化上界 **high** 和下界 **low**, 且 $low < high$;
- (2) 初始化 CPR, 即 $CPR = CPR_0$;
- (3) WHILE TRUE
- (4) FOR 每个粒子 i DO
- (5) $x_i \leftarrow low + rand() \times (high - low)$;
- (6) $v_i \leftarrow low + rand() \times (high - low)$;
- (7) 用式(4.55)评估粒子 i , 令 $pBest_i = x_i$;
- (8) END FOR
- (9) $gBest \leftarrow \min\{pBest_i\}$;
- (10) FOR 每个粒子 i DO
- (11) 更新粒子 i 的速度和位置, 并评估粒子 i ;
- (12) 更新全局最优位置 $gBest$, 且令 $gCPR = CPR(gBest)$;
- (13) END FOR
- (14) IF $abs(gCPR - CPR_0) < \varepsilon$, THEN
- (15) break;
- (16) ELSE
- (17) $mid \leftarrow (high + low) / 2$;
- (18) IF $gCPR \leq CPR_0$
- (19) $temp \leftarrow high$; $high \leftarrow mid$;
- (20) ELSE
- (21) $low \leftarrow high$; $high \leftarrow temp$;
- (22) END IF
- (23) END IF
- (24) END WHILE
- (25) 返回 $gBest$;

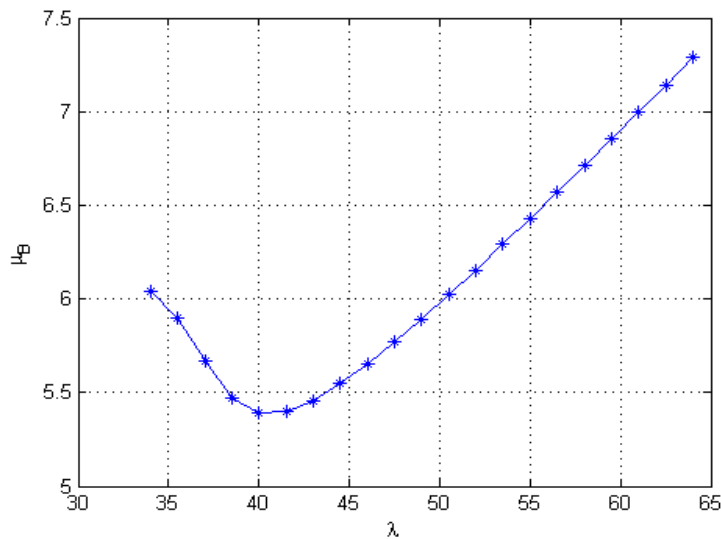


图 4.13 参数 μ_B 与 λ 自适应关系曲线图

在算法 4.1 中, $\text{rand}()$ 表示闭区间 $[0, 1]$ 上的随机函数。算法 4.1 本质上是 PSO 算法的扩展。与 PSO 算法不同之处, 算法 4.1 设置决策变量的上下界, 并通过二分法连续改变上下界, 以缩小决策变量的搜索空间。给定目标函数的最小值 CPR_0 , 算法的终止条件是 CPR 的全局最优值等于给定精度内的最小值 CPR_0 。通过交替减小粒子的值范围来获得粒子的全局最佳位置, 同时 CPR 的最优值从两边交替接近最小值 CPR_0 。因此, 算法 4.1 的时间复杂度为 $O(M \cdot k)$, 其中 M 为种群大小, k 为 while 循环执行次数。在算法 4.1 中, 用 100 个粒子和 0.5 的突变率, 且设初值 $\text{high} = 20, \text{low} = 1.6$, 以及 $\text{CPR}_0 = 53.1134$ 。

根据算法 4.1, 令 $d = QN = 20, K = 20$ 以及 λ 的值从 34.0 到 64.0 变化, 并且忽略 μ_v 和 θ 的值, 能获得 μ_B 和 λ 的关系如图 4.13, 表明随着 λ 的增加, μ_B 减少, 随后明显增加。

由式(4.64), (4.65)和(4.66), 当 λ 从 34.0 到 64.0 变化时, 同时 μ_B 也相应变化, 概率 p_{over} , p_{normal} 和 p_{under} 的变化情况如图 4.14。当 λ 增加时, p_{normal} 和 p_{under} 增加, 但 p_{over} 减少。

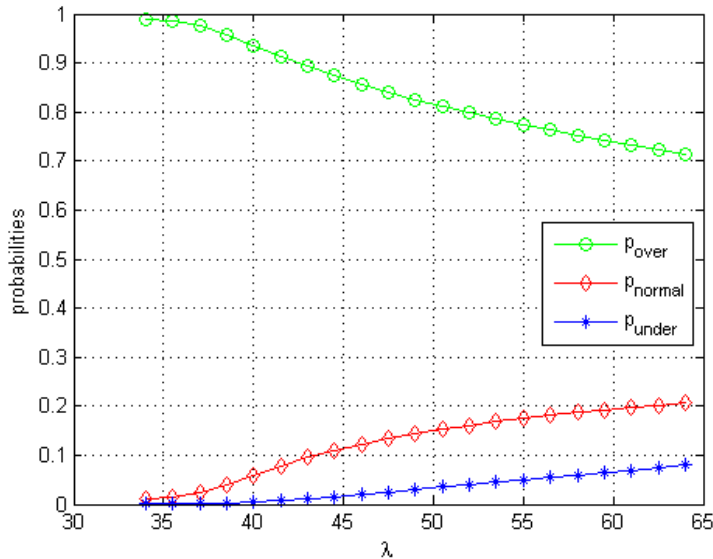


图 4.14 当 λ 变化同时 μ_B 变化时 p_{over} , p_{normal} 和 p_{under} 变化图

4.5.4 比较和分析

在文献[39]中, 计算概率 p_{over} , p_{under} 和 p_{normal} 的模型分别由文献[39]中的式(4)、(5)和(6)给出。系统在稳定条件下, 当 λ 的值从 5.0 到 34.0 变化时, 用文献[39]中的模型计算概率 p_{over} , p_{under} 和 p_{normal} 的变化情况如图 4.15。当平均到率 λ 持续增加时, 在某个时刻总有 $\lambda \geq c\mu_B$, 其中 $\mu_B = 3.4$ 。因此, 系统根据算法 4.1 自适应调节控制器服务率。当到达率 λ 从 34.0 到 64.0 变化, 同时 μ_B 也随之变化时, p_{over} , p_{normal} 和 p_{under} 的变化图如图 4.16。

图 4.15 表明随着 λ 的增加, 概率 p_{normal} 的波动情况比图 4.10 中的波动要大, 且概率值比图 4.10 中显示的值大。类似地, 图 4.14 与图 4.16 比较, 发现图 4.16 中的概率 p_{normal}

也比图 4.14 中的概率大。而且，在图 4.16 中，当 λ 增加时， p_{normal} 增加随后减少，但在图 4.14 中， p_{normal} 持续平稳增加。本章给出的弹性模型符合实际的 CPS 系统情况。

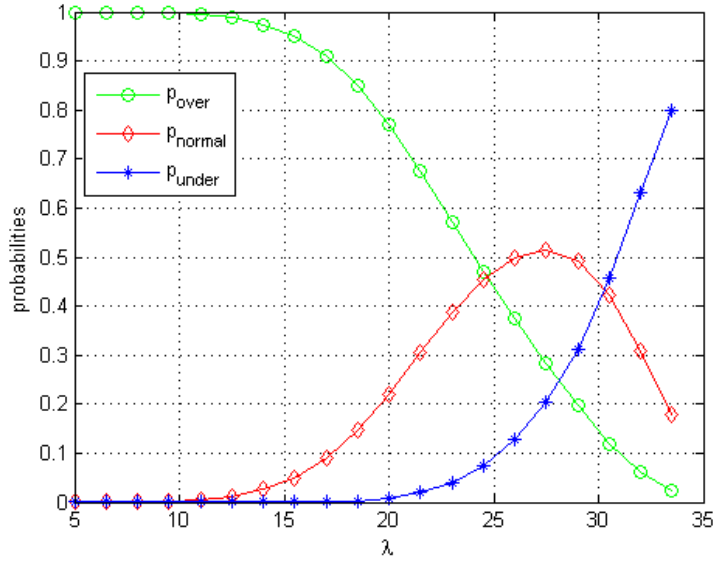


图 4.15 使用[39]中的模型绘制 p_{over} , p_{normal} 和 p_{under} 与 λ 的关系曲线一($\lambda \in [1, 34)$)

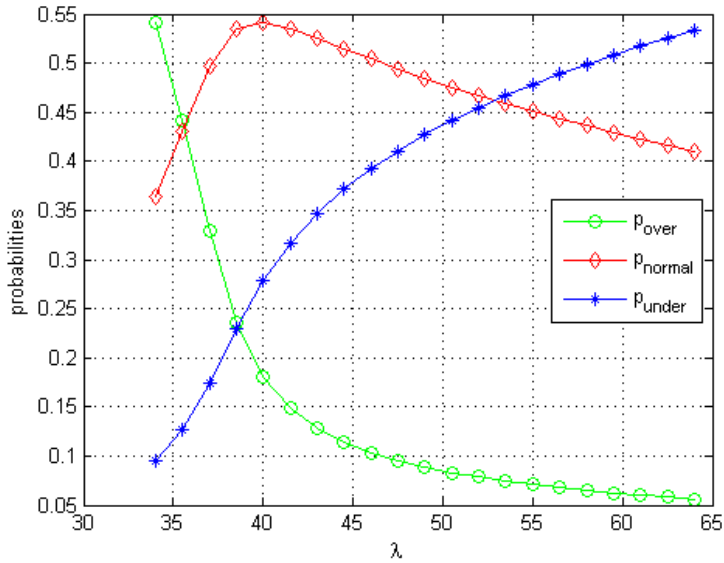


图 4.16 使用[39]中的模型绘制 p_{over} , p_{normal} 和 p_{under} 与 λ 的关系曲线二($\lambda \in [34, 64]$)

用成本函数 F 和标称时延函数 τ_{nom} 比较本文中的 $N/(d,c)-M/M/c/K/AMWV$ 排队模型与第 3 章中的 $N/(d,c)-M/M/c/K/SMWV$ 排队模型。本章中， F 和 τ_{nom} 分别由式(4.54)和(4.52)计算。运用 3.5.3 节中的系统参数值作为这两个模型的输入，并得出比较结果如表 4.4，

表 4.4 两个排队模型的成本函数 F 与标称时延函数 τ_{nom} 的比较结果

Cost/Perform.	$N/(d,c)-M/M/c/K/AMWV$	$N/(d,c)-M/M/c/K/SMWV$
F	234.74mW	237.07mW
τ_{nom}	2.3560e-04s	2.9259e-04s

表明本章中的排队模型的 F 和 τ_{nom} 的值分别比第 3 章中的排队模型的相应值要小。这一结果表明本章中的排队模型有更低的成本和更高的性能。

4.6 高峰期行为

在随机环境下的 M/M/1/K 排队模型中，条件队列长度密度函数 $q_k(j)$ 表示如下：

$$q_k(j) = P\{L=k|J=j\} = \frac{x_{k,j}}{\bar{\pi}_j}, \quad 0 \leq k \leq K, 1 \leq j \leq 3. \quad (4.67)$$

由式 (4.1)，得到系统在稳定条件下的滞弹性值为 $Anelasticity = (1.4172, 5.4617e-03)$ 。

考虑六个队列 A、B、C、D、E 和 F，其到达率向量分别为

$$\begin{aligned} \vec{\lambda}_A &= (73, 1, 1), \vec{\lambda}_B = (37, 37, 1), \vec{\lambda}_C = (25, 25, 25), \\ \vec{\lambda}_D &= (118, 1, 1), \vec{\lambda}_E = (60, 59, 1), \vec{\lambda}_F = (40, 40, 40). \end{aligned}$$

在这六个队列中，前面三个队列的平均到达率为 25，而后面三个队列的平均到达率为 40。对于后面三个队列，平均到达率均大于 $c\mu_B$ ，其中 $c=10, \mu_B=3.4$ 。因此，调用算法 4.1 重新生成控制器忙期的平均服务率 μ_b ，得到 $\mu_b=5.3897$ 。这六个队列的平均服务率向量能分别用式 (4.56) 计算，同时，通信强度 $\tilde{\rho} = (\tilde{\pi}\vec{\lambda})(\tilde{\pi}\vec{\mu})^{-1}$ 也能分别获得。

一个系统处于一种到达率大于等于服务率的环境下的时间称为“高峰期”，否则称为“恢复期”。处在高峰期的队列 A、B、D 和 E 所表现的行为与具有时齐参数的队列（如队列 C 和 F）的行为有本质的不同。由式 (4.1) 和 (4.67)，六个队列在各个位相的滞弹性和标准差如表 4.5，表明这六个队列的条件平均队列长度和相应的标准差都略微减少。大多数高度可变的随机模型的行为仅仅通过均值和二阶矩来描述得到较差的结论^[192]。需要用系统在各个位相的空闲条件概率和队列长度密度的百分位数来讨论队列的耗散特征。在表 4.6 中，通过数值实验表明条件队列长度分布超过 $\alpha\%$ 的情况，其中 α 分别取值为 75、90、95 和 99。

表 4.5 和表 4.6 表明队列 A 和 D 有较好的耗散性，这是因为这两个队列在位相 3 具有较高的空闲条件概率，分别为 0.250 和 0.298。队列 A 和队列 D 的可变性是由于输入连续激增频率很高，并且会相互叠加，导致队列接近 K 的偶然积累。在队列 B 和队列 E 中，在位相 1 的到达率减少，而在位相 2 的到达率增加，但在这种方式下，这两个位相的到达率仍然分别超过了服务率。这两个队列分别表现出比 A 和 D 少得多的变化性，尽管条件标准差的缓慢下降表明叠加积累仍然频繁，但程度要低于队列 A 和 D。队列 C 和队列 F 是两个时齐马尔可夫链，比其余四个队列具有更好的平滑性和耗散性。在表 4.6 表明通信强度 ρ 越小，队列的耗散性越好，顾客丢失的损失概率 $LOSS$ 越小。这符合实际系统的特征。

在表 4.5 和表 4.6 中，从各个位相的条件平均队长、标准差、弹性值 p_{normal} 以及队列长度分布百分位数等指标来看，可以看到，队列 D、E 和 F 的值分别比队列 A、B 和 C 的相应值小，但是，队列 D、E 和 F 在各个位相上的空闲条件概率分别比队列 A、B

和 C 的相应值大。因此，自适应调节排队系统的参数后，队列 D、E 和 F 的平滑性和耗散性分别比队列 A、B 和 C 的相应值要好。表 4.5 和表 4.6 表明在位相 2 的条件平均队列长度越小，弹性值 p_{normal} 越小，系统的平滑性和耗散性越好。因此，弹性是度量系统性能的重要指标之一。

表 4.5 六个队列在各个位相上滞弹性和标准差

Queue	Anelasticity			Standard Deviation		
	1	2	3	1	2	3
A	(10.68, 0.6112)	(9.26, 0.2299)	(8.59, 0.1589)	9.75	9.67	9.57
B	(6.09, 0.7982)	(5.73, 0.1597)	(5.03, 0.0421)	6.30	6.25	6.14
C	(3.82, 0.9039)	(3.79, 0.0869)	(3.74, 0.0092)	4.26	4.25	4.24
D	(8.93, 0.6859)	(7.53, 0.2064)	(6.87, 0.1077)	8.65	8.52	8.38
E	(4.68, 0.8726)	(4.33, 0.1102)	(3.67, 0.0172)	5.00	4.94	4.81
F	(3.19, 0.9360)	(3.16, 0.0599)	(3.14, 0.0041)	3.63	3.62	3.62

表 4.6 六个队列在各个位相上的空闲条件概率和队列长度分布的百分位数

Q	ρ	LOSS	Empty			75%			90%			95%			99%		
			1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
A	0.826	2.57e-03	0.078	0.194	0.250	16	14	13	26	24	23	32	30	29	40	39	38
B	0.815	2.10e-04	0.127	0.163	0.266	9	8	7	15	14	13	19	19	18	29	28	27
C	0.791	1.12e-05	0.204	0.209	0.215	5	5	5	9	9	9	12	12	12	19	19	19
D	0.791	1.28e-03	0.096	0.233	0.298	13	11	10	21	20	19	27	26	25	38	36	35
E	0.768	3.95e-05	0.161	0.207	0.328	7	6	5	11	11	10	15	14	13	23	22	21
F	0.759	2.34e-06	0.236	0.242	0.244	5	5	5	8	8	8	10	10	10	16	16	16

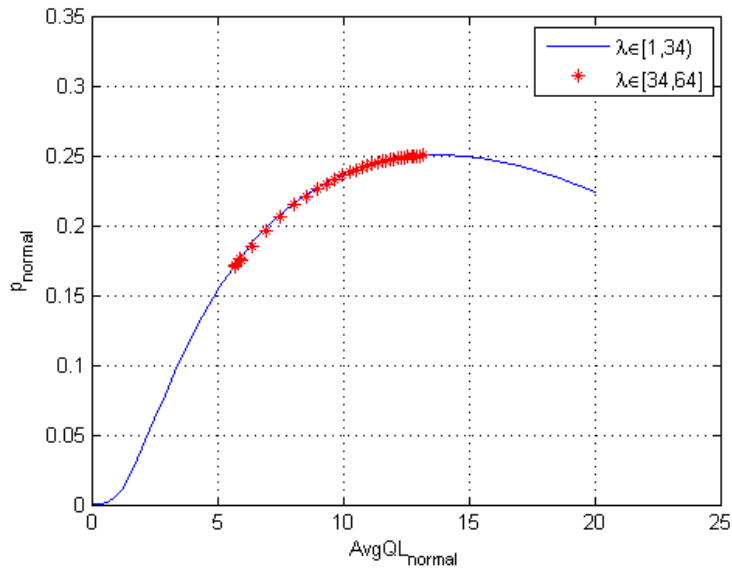


图 4.17 当 $\lambda \in [1, 34) \cup [34, 64]$ 时系统在正常状态下 $AvgQL_{normal}$ 和 p_{normal} 关系曲线

实验表明，当 λ 增加时，队列在位相 2 的条件平均队列长度和弹性值 p_{normal} 不再是线性关系(如图 4.17)。在图 4.17 中，当 $\lambda \in [1, 34)$ ，即 $\rho = \lambda\sigma_0^{-1} < 1$ 时，随着 λ 的增加，

队列在位相 2 的条件平均队列长度增加, 但弹性值 p_{normal} 先增加再稍微减少。然而, 当 $\lambda \in [34, 64]$, 即 $\rho = \lambda \sigma_0^{-1} \geq 1$ 时, 自适应调节排队系统的控制器忙期服务率值为 $\tilde{\mu}_B$ 后, 得到 $\rho = \lambda(c\tilde{\mu}_B)^{-1} < 1$, 此时, 当系统在位相 2 的条件平均队列长度增加时, 弹性值 p_{normal} 仍然增加(见图 4.17 中的“*”), 并且与自适应调节系统参数之前的弹性值基本相同。因此, 当系统平均到达率超过平均服务率时, 自适应调节系统参数后能基本保持系统弹性不变。

4.7 本章小结

考虑一类特殊 CPS 在不确定负载条件下的资源供应问题。在共享总线的 DES 中, 存在与时间相关的资源供应按需特性。滞弹性自然反射了应力和应变之间的滞后特征的按需特性。针对这类特殊的 CPS, 本章提出了滞弹性模型的一种新的、量化以及正式定义, 并求解了这一系统各种参数的最优值, 以配置系统在稳定条件下的系统参数。构建 FSDMC 模型为一种 $N/(d, c)\text{-}M/M/c/K/SMWV$ 排队模型, 最小化系统 CPR, 并获得 3-状态(即过度供应状态、正常状态和不足供应状态)下的各自的服务率。用一种随机环境下的 $M/M/1/K$ 排队模型精确计算 FSDMC 平台的滞弹性值。当排队系统的平均到达率超过平均服务率时, 提出了一种基于 PSO 的自适应调节算法, 动态调节控制器服务率, 同时保持系统在稳定条件下的 CPR 最小值不变。最后, 分析了系统高峰期行为。

第 5 章 具有不确定时延的消息响应时间分析

由于CPS被嵌入到物理设备中，与其他计算机系统相比，它们的寿命更长。这使得有必要在CPS的设计阶段验证它们在整个使用寿命期间的运行可靠和安全。CPS中的控制应用对传感器到执行器的消息响应端到端的时延是敏感的，这些时间必须进行分析 and 最小化。最坏情况下的MRT评估可能不充分，必须通过概率分析来补充。由于图3.1所示的双信道ANCS中的信道B的传输时间相对复杂，因此本章不考虑双信道ANCS的时延问题，仅关注单信道ANCS的MRT分析问题，运用M/PH/1排队系统建模和分析单信道ANCS，协同设计总线参数和控制器服务强度，最优配置总线，计算MRT的概率分布。

5.1 引言

在图 1.6 所示的单信道 ANCS 中，所有控制应用共享从传感器到控制器的前向总线和从控制器到执行器的后向总线以传输消息。对于每个控制应用，由传感器生成的消息由前向总线、控制器、后向总线相继处理，最后进入执行器。在某个时刻，多个消息可能会同时到达共享总线。因此，ANCS 可以被抽象为多队列多服务台串并联混合排队系统。控制应用程序的动态过程符合 M/PH/1 队列的动态行为。M/PH/1 队列是一个拟生灭 QBD 过程，顾客以强度 λ 的泊松过程到达。服务时间具有阶数为 n 且表示为 (β, S) 的相位型(phase-type, PH)分布，其中 μ^{-1} 为服务时间的均值； (β, β_{n+1}) 是该过程的初始概率，

且 $\beta_{n+1}=0$ 。服务时间分布由无穷小生成器 $\begin{pmatrix} S & S^0 \\ 0 & 0 \end{pmatrix}$ 特征化，且 $Se + S^0 = 0$ ，其中 S 是一个

n 阶方阵，其对角元素为负，其他元素都是非负的； S^0 是维数为 n 的非负列向量； e 是一个适当的维列向量，其所有元素都为 1^[192]。本章专注于这一单信道 ANCS 中的 MRT 的概率模型建立。使用具有随机大小批量到达的连续时间 M/PH/1 排队系统建模 ANCS 的每个控制应用为 QBD 过程，以分析该控制应用中消息的不确定响应延迟，并获得早期体系结构探索阶段的 MRT 的 pdf 和 cdf。

在汽车电子系统中，通信协议由定义消息的截止时间和依赖于这一截止时间的开关控制策略；如果消息没有超过截止时间，则使用标称控制器，如果消息超过了截止时间，则消息将被中止。通常假设消息被丢弃并且使用先前的输入，即利用零阶保持算法，因为控制器可用的信息明显延迟^[41]。这种技术可能会使系统不稳定。在设计控制律之前预测当前的延迟是通过控制律来补偿当前延迟的一种可行的方式^[42]。本章使用灰色系统理论^[171]预测ANCS中的延迟，并提出了一种新的灰色模型(1,1)，命名为两参数概率变换的GM(1,1) (PTGM(1,1))，以精确预测每个消息的响应时间。在PTGM(1,1)模型中，需要用获得的MRT的cdf构建预测模型。

为了节省系统实现成本,必须在ANCS中配置共享通信总线。在ET通信协议中,DYN段长度指定为最小时隙(minislot)数。因此,在DYN段,如果在某个特定时隙没有消息发送,则该时隙将具有非常短的长度(等于所谓的最小时隙的长度),否则DYN时隙的长度将等于传输的整个消息所需的最小时隙数^[50]。本章首先使用获得的系统参数值和提出的总线协议配置ANCS中的通信总线,提出了一个约束优化模型来最小化系统成本函数,并设计了一种最优最小时隙算法求解优化模型,得到决策变量minislot的最优值,同时,也得到通信总线的其它参数值,例如每个总线周期的长度和最小时隙数。根据总线配置,可以获得所有消息的响应时间的样本值,并使用这些样本数据来估计每个消息的响应时间的cdf的参数值。就作者最好的知识,本章的技术是首次尝试使用M/PH/1队列模型在考虑最优minislot的情况下计算具有不确定延迟的汽车CPS的MRT的pdf,并使用灰色系统理论预测延迟。

5.2 平台体系结构

在图 3.1 所示的包含五个分布式控制应用的双信道 ANCS 平台体系结构中,假设信道 B 的传输时间忽略不计,仅仅考虑信道 A 的传输时间,即考虑如图 5.1 所示的单信道 ANCS 平台体系结构的 MRT 问题。控制应用被分别映射到不同 PUs 的大量任务,所有 PUs 通过共享总线通信,并运行来自一个或者多个控制应用的不同任务。考虑六个控制应用,分为两类:一类包含控制应用 1 和 2 (如图 3.2); 另一类包含控制应用 3、4、5 和 6 (如图 3.3)。

在如图 5.1 所示的平台体系结构中,通信总线协议采用基于 ET 协议的多层次 FTDMA/FP 总线调度策略 (图 5.2)。上层协议服从总线调度策略,所有控制应用中的控制器 ECU 在一个总线循环内按照固定优先级的次序 $E_1 \rightarrow E_2 \rightarrow E_3 \rightarrow E_4 \rightarrow E_5 \rightarrow E_6$ 依次占用通信总线,其中 $E_i (i = 1, 2, \dots, 6)$ 显示如图 5.2。中间层协议采用 FCFS 规则,运用于每个控制应用中的具有批量到达的消息队列中的消息实例。低层协议如图 3.4 所示,服从

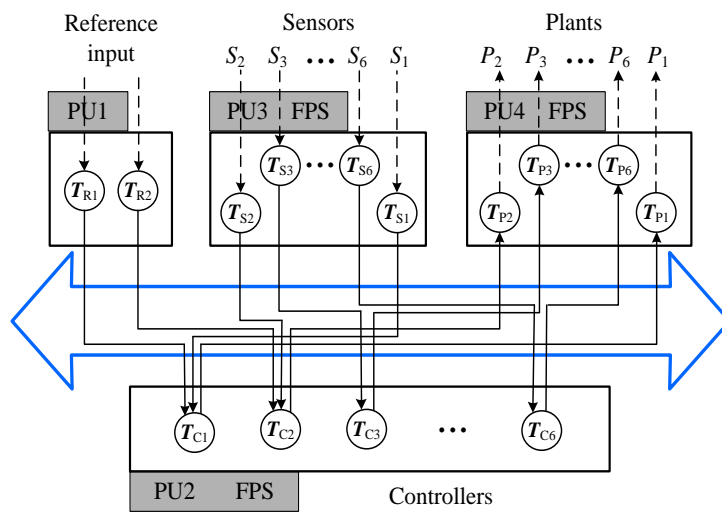


图 5.1 具有六个控制应用的系统体系结构^[11]

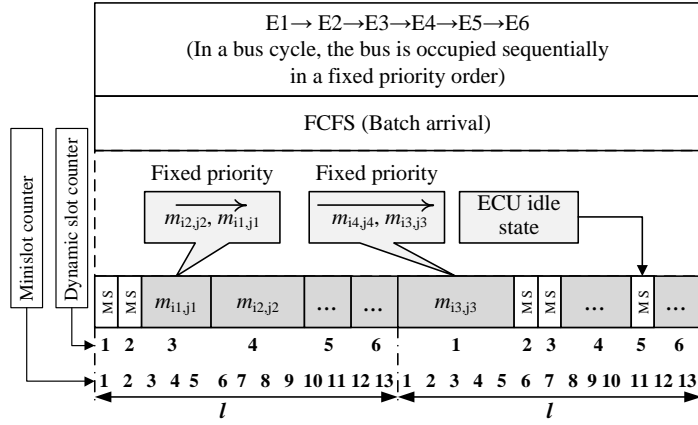


图 5.2 通信总线：多层次 FTDMA/FP 调度器， l 表示 FTDMA 周期长度

ET 协议的分层 FTDMA/FP 总线调度策略。FTDMA 技术的构思主要是为了克服 CAN 的限制，特别支持高性能实时通信^[179]。详见 3.2 节描述。

为了配置总线并获取每个消息的响应时间的 *cdf* 参数值，本章使用 M/PH/1 排队系统对 ANCS 中的控制应用进行建模，并在以部分中分别计算系统的成本函数和 MRT 的 *cdf*。

5.3 排队模型

5.3.1 模型描述

在本节中，ANCS 中的控制应用使用具有随机大小批量到达的连续时间 M/PH/1 排队系统进行建模。在 M/PH/1 队列中，假设 $\beta_{n+1} = 0$ ，这意味着 $\beta e = 1$ 。下文中，假定 e 是其所有元素等于 1 的适当维数的列向量； I 是适当阶数的单位矩阵。而且，服务率和服务时间的拉普拉斯-斯蒂尔切斯变换(Laplace-Stieltjes transform, LST)分别为：

$$\mu^{-1} = -\beta S^{-1} e, \quad f^*(s) = \beta (sI - S)^{-1} S^0. \quad (5.1)$$

在图 5.1 所示的 ANCS 平台体系结构中， c 个控制应用根据图 5.2 所示的通信协议依次共享从传感器到控制器的前向总线和从控制器到执行器的后向总线。分别将前向总线和后向总线抽象为前向节点 T_{fwd} 和后向节点 T_{bwd} 。在平台体系结构中，对于每个控制应用，由传感器的采样数据和参考命令生成相应的子队列， c 个子队列分别由 c 控制应用生成，其中 $c = 6$ 。因此，该平台体系结构被抽象为如图 5.3 所示的多队列多服务台串并联混合排队系统。对于第 i 个控制应用相应的第 i 个子队列，假设来自传感器的消息依据具有速率 $\lambda_i (\lambda_i > 0)$ 的泊松过程到达相应的 ECU _{i} ，也就是说，相继到达之间的时间是

具有均值为 $\frac{1}{\lambda_i}$ 的独立指数随机变量。

在第 i 个子队列中($i=1, 2, \dots, 6$), 向上到达的每一个消息在服务台 \mathfrak{R}_{i1} (即传感器 ECU, 表示为 T_{S_i} 或 E_i 节点) 为空闲时直接进入服务; 否则, 该消息加入队列。当 \mathfrak{R}_{i1} 服务完毕一个消息后, 该消息被传送到服务台 \mathfrak{R}_{i2} (即 FlexRay/CAN 前馈总线, 表示为 T_{fwd} 节点) 接受服务, 依次到达服务台 \mathfrak{R}_{i3} (即控制器 ECU, 表示为 T_{C_i} 节点) 和服务台 \mathfrak{R}_{i4} (即 FlexRay/CAN 反馈总线, 表示为 T_{bwd} 节点) 接受服务, 最后进入相应的执行器 (即执行器 ECU, 表示为 T_{P_i} 节点), 随后, 如果在队列中存在下一个消息, 立即进入服务台 \mathfrak{R}_{i1} 接受服务。因此, 在第 i 个子队列中, 每个到达的消息相继由服务台 $\mathfrak{R}_{i1}, \mathfrak{R}_{i2}, \mathfrak{R}_{i3}$ 和 \mathfrak{R}_{i4} 服务, 最终被执行器 ECU 吸收, 服务时间服从 PH 分布。为方便起见, 称消息为顾客(customer), ECU 或总线为服务台(例如, $\mathfrak{R}_{ij}, i=1, 2, \dots, 6, j=1, 2, 3, 4$)。

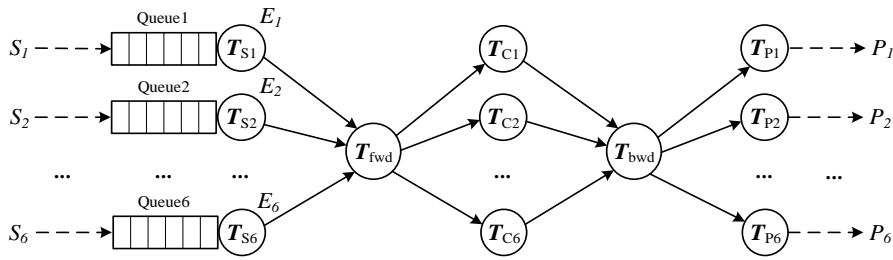


图 5.3 平台体系结构的多队列多服务台串并联混合排队系统

本章给出如下假设:

A1: 在每一个子队列中, 消息按照泊松过程到达服务台 \mathfrak{R}_{i1} (即 E_i)。假设来自用户的消息到达过程也是泊松过程(如控制应用 1 和 2)。独立泊松过程之和也是泊松过程, 因此, 在第 i 个控制应用中, 消息按照速率为 $\lambda_i (\lambda_i > 0)$ 的泊松过程到达。系统中总的消息

到达率为 $\lambda = \sum_{i=1}^c \lambda_i$ 。在图 5.3 中, 为简单起见, 省略了参考命令 T_{R1} 和 T_{R2} 。

A2: 在每一个子队列中, 服务台 \mathfrak{R}_{ij} 的服务时间是一个具有平均服务率 $x_i (x_i > 0)$ 的指数随机变量。一个顾客在 \mathfrak{R}_{ij} 中服务完毕后, 立即进入下一个服务台 $\mathfrak{R}_{i,j+1}$, 直到被吸收为止。一个服务台一次仅仅服务一个顾客。

A3: 控制应用的执行器是高性能处理器, 一旦一个执行器接收控制信号, 将立即驱动被控对象执行相应操作, 执行器的执行时间可以忽略。因此, 在队列模型中, 服务时间分布式是 4 阶位相型分布, 即 $n = 4$, 且执行器的状态为吸收状态, 即 $\beta_{n+1} = 0$ 。

A4: 不同批到达的顾客遵循 FCFS 规则接受服务, 同一批到达的顾客服从固定优先级。

A5: 在每一个控制应用中, M/PH/1 队列中的所有到达间隔时间和服务时间相互独立。

5.3.2 QBD 过程模型

在文献[192]中, M/PH/1 排队系统是一个 QBD 过程, 其状态空间为

$$\Omega = \{0\} \cup \{(i, j) : i \geq 1, 1 \leq j \leq n\},$$

其中状态 $\mathbf{0}$ 表示空队列, 状态 (i, j) 表示系统中有 $i(i \geq 1)$ 个顾客, 且服务过程位于位相 $j(1 \leq j \leq n)$ 。

用状态字典顺序, QBD 过程无穷小生成器表示为如下 Block-Jacobi 矩阵:

$$Q = \begin{pmatrix} -\lambda & \lambda\beta & & & & \\ S^0 & -\lambda I + S & \lambda I & & & \\ & S^0\beta & -\lambda I + S & \lambda I & & \\ & & S^0\beta & -\lambda I + S & \ddots & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

为了分析该 QBD 过程, 需要找到如下矩阵二次方程的最小非负解

$$R^2 S^0 \beta + R(-\lambda I + S) + \lambda I = 0,$$

这里 R 称为率阵。

如果 $\rho = \lambda\mu^{-1} < 1$, 那么上述矩阵方程有最小非负解 $R = \lambda(\lambda I - \lambda e\beta - S)^{-1}$ (见文献[192]中定理 3.2.1)。显然, 容易获得每个控制应用中的顾客平均数为, 即

$$E[L] = (1 - \rho)\beta R(I - R)^{-2} e. \quad (5.2)$$

根据 Little 公式, 顾客在控制应用中的平均逗留时间为

$$E[\omega] = \frac{E[L]}{\lambda}.$$

5.3.3 参考总线的消息队列

应用文献[27]中的实验汽车CAN总线消息集作为本章的测试数据集, 该数据集由6个ECU和69个消息组成, 如表5.1所示。系统中的所有消息 m_i 都以固定的优先级顺序进行索引, 并以毫秒(ms)为单位显示其周期和传输时间。在每个ECU节点处, 总线适配器以固定优先级(ID)顺序排列消息, 并且始终考虑优先级最高的消息用于仲裁。对于每个消息 m_i , T_i 是它的周期, ε_i 是它的传输时间, 而 P_i 是它的固定优先级。对于每次周期激活,

消息实例都会到达相应的ECU节点，并由周期性中间件任务排队。将消息 m_i 的第 j 个实例表示为 $m_{i,j}$ 。每个消息 m_i 都与一个唯一的ID(即 P_i)相关联，这也代表了它的优先级。根据固定优先级协议，对于一个消息，其ID号越小，优先级越高。CAN总线速率为500 kb/s，其利用率为60.25%。

表 5.1 汽车 CAN 系统的一个实例^[27]

Msg	ECU	T_i	ε_i	P_i	Msg	ECU	T_i	ε_i	P_i	Msg	ECU	T_i	ε_i	P_i
m ₁	E ₂	10	0.27	1	m ₂₄	E ₃	25	0.23	24	m ₄₇	E ₂	50	0.19	47
m ₂	E ₂	10	0.27	2	m ₂₅	E ₃	25	0.25	25	m ₄₈	E ₄	100	0.27	48
m ₃	E ₃	5	0.19	3	m ₂₆	E ₂	20	0.27	26	m ₄₉	E ₃	100	0.27	49
m ₄	E ₃	10	0.25	4	m ₂₇	E ₅	25	0.27	27	m ₅₀	E ₁	100	0.27	50
m ₅	E ₁	10	0.19	5	m ₂₈	E ₂	20	0.27	28	m ₅₁	E ₃	100	0.13	51
m ₆	E ₆	10	0.27	6	m ₂₉	E ₁	25	0.17	29	m ₅₂	E ₃	100	0.27	52
m ₇	E ₁	100	0.27	7	m ₃₀	E ₂	10	0.21	30	m ₅₃	E ₁	100	0.19	53
m ₈	E ₁	100	0.15	8	m ₃₁	E ₂	20	0.27	31	m ₅₄	E ₃	100	0.13	54
m ₉	E ₁	100	0.17	9	m ₃₂	E ₄	10	0.27	32	m ₅₅	E ₃	100	0.13	55
m ₁₀	E ₅	25	0.27	10	m ₃₃	E ₃	10	0.27	33	m ₅₆	E ₅	100	0.27	56
m ₁₁	E ₁	100	0.15	11	m ₃₄	E ₃	10	0.27	34	m ₅₇	E ₃	100	0.25	57
m ₁₂	E ₁	20	0.19	12	m ₃₅	E ₁	25	0.25	35	m ₅₈	E ₃	100	0.13	58
m ₁₃	E ₂	100	0.19	13	m ₃₆	E ₆	25	0.23	36	m ₅₉	E ₃	100	0.13	59
m ₁₄	E ₁	100	0.19	14	m ₃₇	E ₄	50	0.27	37	m ₆₀	E ₄	100	0.17	60
m ₁₅	E ₅	100	0.27	15	m ₃₈	E ₄	50	0.27	38	m ₆₁	E ₁	100	0.27	61
m ₁₆	E ₁	10	0.23	16	m ₃₉	E ₅	50	0.27	39	m ₆₂	E ₁	100	0.13	62
m ₁₇	E ₂	100	0.25	17	m ₄₀	E ₃	50	0.21	40	m ₆₃	E ₃	100	0.19	63
m ₁₈	E ₂	100	0.27	18	m ₄₁	E ₄	50	0.27	41	m ₆₄	E ₁	100	0.27	64
m ₁₉	E ₂	50	0.25	19	m ₄₂	E ₆	50	0.27	42	m ₆₅	E ₁	100	0.13	65
m ₂₀	E ₃	10	0.27	20	m ₄₃	E ₂	50	0.27	43	m ₆₆	E ₁	100	0.13	66
m ₂₁	E ₆	10	0.27	21	m ₄₄	E ₅	100	0.27	44	m ₆₇	E ₂	50	0.27	67
m ₂₂	E ₆	25	0.27	22	m ₄₅	E ₂	25	0.15	45	m ₆₈	E ₁	100	0.13	68
m ₂₃	E ₃	25	0.23	23	m ₄₆	E ₂	50	0.19	46	m ₆₉	E ₄	100	0.27	69

在图 5.3 中，对于第 i 个控制应用，传感器周期采样的消息实例 $m_{k,j}$ 以速率 $\lambda_k = \frac{1}{T_k}$ 到达相应节点 E_i 。具有不同采样周期的一些消息实例可能在消息周期的公倍数的采样时刻同时到达节点 E_i ，并按照固定优先级顺序在第 i 个子队列中排队。因此，该排队模型是一个随机大小的批量到达的 M/PH/1 排队模型。系统中所有消息周期的最小公倍数称为“超周期”，令 H 表示超周期，那么从表 5.1 可得， $H = 100\text{ms}$ 。

在第 i 个控制应用中，令 $\alpha_{ij} (1 \leq i \leq c, j \geq 1)$ 表示由 j 个顾客组成的任意批次的概率；令 N_i 表示批大小的随机变量，那么 $P\{N_i = j\} = \alpha_{ij}$ 。令 $E[N_i]$ 表示随机变量 N_i 的期望，那么 $E[N_i] = \sum j\alpha_{ij}$ 。假设批量到达依据具有速率 λ_{ia} 的泊松过程发生，并且第 i 个子队列的平均到达速率是 λ_i ，那么 $\lambda_i = \lambda_{ia} E[N_i]$ 。在第 i 个子队列中，假设 $MsgID_i$ 表示消息 ID 集， B_i 表示消息最大到达批数，则 $B_i = \max\left\{\frac{H}{T_k}, k \in MsgID_i\right\}$ ， $\lambda_{ia} = \frac{B_i}{H}$ ，因此，第 i 个子队列的平均到达率由下式计算：

$$\lambda_i = \frac{1}{H} \max\left\{\frac{H}{T_k}, k \in MsgID_i\right\} \sum_{j=1}^{\infty} j\alpha_{ij}.$$

例如，在第 3 个子队列， $\lambda_{3a} = 0.2, j=1, 4, 5, 9, 18$ ，且相应的概率分别为 $\alpha_{3j} = 0.4, 0.1, 0.4, 0.05, 0.05$ ，那么， $E[N_3] = 4.15$ ， $\lambda_3 = 0.83$ 。假设 ζ_i 和 ξ_i 分别表示第 i 个子队列的消息数及消息实例数， φ_{ij} 表示第 j 个消息的实例数，那么， $\zeta_i = \|MsgID_i\|$ ， $\xi_i = \sum_{j \in MsgID_i} \varphi_{ij}$ ，以及 $\varphi_{ij} = \frac{H}{T_j}$ 。假设在一个超周期内的总线环由 ξ 表示，则

$\xi = \max\{\xi_i, 1 \leq i \leq c\}$ 。在一个超周期内，不同子队列的平均到达率、消息数及其消息实例数如表 5.2，第 3 个控制应用的消息队列如图 5.4 所示，其它控制应用的消息队列也能类似给出。

表 5.2 在一个超周期内不同的子队列的消息平均到达率、消息数以及消息实例数

Ctrl. App. ID	1	2	3	4	5	6
Arrival Rates	0.46	0.62	0.83	0.19	0.13	0.30
MsgNo (ζ_i)	18	15	18	7	6	5
InstNo (ξ_i)	46	62	83	19	13	30

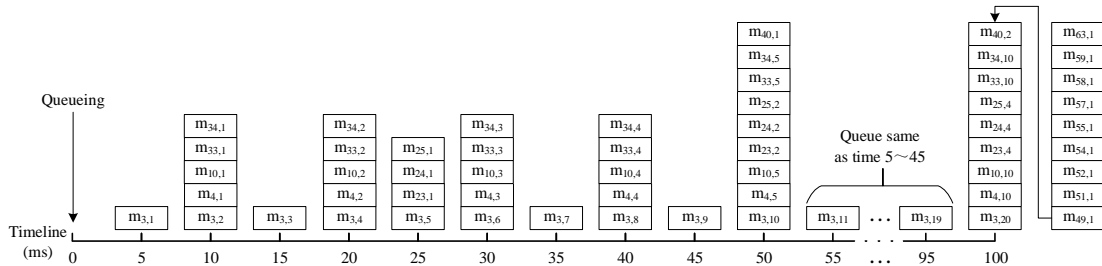


图 5.4 在一个超周期内第 3 个控制应用的消息队列

在排队模型中，顾客接受服务的规则服从 3.1 节中的假设 A4。在第 i 个子队列中，消息 m_k 的消息实例 $m_{k,j}$ 依次接受服务台 $\mathfrak{R}_{i1}, \mathfrak{R}_{i2}, \mathfrak{R}_{i3}$ 以及 \mathfrak{R}_{i4} 的服务。不失一般性，假设每个消息实例 $m_{k,j}$ 在四个服务台的总服务时间是消息 m_k 的传输时间 ε_k 。

5.4 总线配置

5.4.1 最优化模型

本节使用 M/PH/1 队列的平均顾客数来数值评估系统成本，以产生最佳的最小小时隙。首先开发系统在单位时间每个顾客总的期望成本函数。对于每个控制应用，相应的 M/PH/1 队列是一个单服务台多服务位相的排队系统。影响系统成本的因素主要是系统中的顾客数以及顾客驻留系统的时间。定义如下成本参数：

$C_h \equiv$ 驻留在系统中的每个顾客单位时间成本；

$C_d \equiv$ 顾客在系统中延迟时单位时间的惩罚成本。

利用上面的成本参数和 White 等^[193]提出的乘务员服务设备的概念，并考虑 Little 公式，在复杂和非线性成本结构下的单位时间的期望成本函数表示为 $F(\tau)$ ，其中 τ 表示最小小时隙。建立约束优化模型如下：

$$\begin{aligned} \min \quad & F(\tau) = \sum_{i=1}^c \frac{\lambda_i}{\lambda} (C_h E[L_i] + C_d E[\omega_i]) = \sum_{i=1}^c \frac{\lambda_i}{\lambda} \left(C_h + \frac{1}{\lambda_i} C_d \right) E[L_i], \quad (5.3) \\ \text{s. t.} \quad & \begin{cases} \gcd\{\varepsilon_i, 1 \leq i \leq \sigma\} < \tau \leq \frac{H}{\xi c}, \\ \sum_{i=1}^{\delta} \varepsilon_i + \sum_{j=1}^c \tau \left(\left\lfloor \frac{H}{H/(\xi\tau)} \right\rfloor - \xi_j \right) < H, \\ \xi = \max\{\xi_i, 1 \leq i \leq c\}, \\ \sigma = \sum_{i=1}^c \zeta_i, \quad \delta = \sum_{i=1}^c \xi_i, \end{cases} \end{aligned}$$

其中在第 i 个控制应用中的顾客平均数 $E[L_i]$ 由式(5.2)计算， $\gcd\{\varepsilon_i, 1 \leq i \leq \sigma\}$ 表示消息服务时间的最大公约数， σ 和 δ 表示 ANCS 系统中的所有消息数和消息实例数。

本章的目的是确定上述最优化模型(5.3)的决策变量 τ 的最优值。首先讨论每个控制应用的参数估计问题，然后求解最优化问题(5.3)。假设顾客等可能的在每个服务位相接受服务，那么任意控制应用的初始概率可以表示为行向量 $\beta = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)$ 。对于任意控制应用，服务时间 $t(t \geq 0)$ 的 PH 分布的 pdf 表示为

$$f(t) = \beta \exp(St) S^0. \quad (5.4)$$

由 5.3.1 节中的假设 A2, 令 $S = \begin{pmatrix} -x_1 & x_1 & & & \\ & -x_2 & x_2 & & \\ & & -x_3 & x_3 & \\ & & & & -x_4 \end{pmatrix}$. 根据矩阵 S 的结构, 式(5.4)

表示的 pdf 称为双对角 PH 分布密度函数^[196]. 那么, 在式(5.4)中, 用极大似然估计方法仅仅需要估计参数列向量 $\bar{x} = (x_1, x_2, x_3, x_4)^T$ 的值, 其中 T 表示转置。服务时间采样数据来自于表 5.1 中的列 ε_i 。方程(5.4)的对数似然函数表示为

$$\psi(\bar{x}) = \sum_{k \in \text{MsgID}_i} \sum_{j=1}^{\varphi_{ik}} \log f(\hat{t}_{i,k,j}; \bar{x}), \quad (5.5)$$

其中 MsgID_i 表示第 i 个控制应用的消息 ID 集; $\hat{t}_{i,k,j}$ 表示第 k 个消息的第 j 个实例的采样值, 且 $\hat{t}_{i,k,j} = \varepsilon_k$ 对于任意 $j \in [1, \varphi_{ik}]$ 成立, 得到方程(5.5)关于 \bar{x} 的梯度为

$$\nabla \psi(\bar{x}) = \frac{d\psi(\bar{x})}{d\bar{x}} = \sum_{k \in \text{MsgID}_i} \sum_{j=1}^{\varphi_{ik}} \frac{1}{f(\hat{t}_{i,k,j}; \bar{x})} \frac{df(\hat{t}_{i,k,j}; \bar{x})}{d\bar{x}}. \quad (5.6)$$

5.4.2 总线配置算法

总线配置算法包含基于 DFP 方法的参数估计算法和最优 minislot 算法两个算法。DFP 方法是 Davidon^[63]首先提出, 后来又被 Fletcher 和 Powell^[64]改进的算法。在这种方法中, 定义校正矩阵 ΔH_k 为:

$$\Delta H_k = \frac{\bar{p}_k \bar{p}_k^T}{\bar{p}_k^T \bar{q}_k} - \frac{H_k \bar{q}_k \bar{q}_k^T H_k}{\bar{q}_k^T H_k \bar{q}_k},$$

其中 \bar{p}_k 和 \bar{q}_k 分别由算法 5.1 中的第(12)和(13)步定义; H_k 为满足拟牛顿条件的 n 阶对称正定阵, 且初始化为 $H_1 = I$, I 为相应维数的单位矩阵。拟牛顿条件为: $\bar{p}_k = H_{k+1} \bar{q}_k$; DFP 公式由算法 5.1 中的第(14)步给出。

在算法5.1中, 令 $\bar{x}_1 = (1, 1, 1, 1)^T$, $n = 4$, 并且 I 表示适当阶数的单位矩阵。该算法的主要过程来自DFP算法。算法5.1主要计算列向量 $\bar{x} = (x_1, x_2, x_3, x_4)^T$ 的最优值, 以使对数似然函数最大化。为了求解minislot的最优值, 首先从最优化模型(5.3)的约束条件中获得 minislot 的范围, 在该范围内搜索模型(5.3)的最优解。算法5.1的时间复杂度为 $O(n^2)$, 与 DFP 算法的时间量相同。

最优 minislot 算法描述如算法 5.2。在算法 5.2 中, 为了保证所有的消息实例都可以被传输, 选择 c 个控制应用的消息实例的最大数作为一个超级周期中的总线周期数(步骤 3 和 4)。根据最优化模型(5.3)中的约束条件计算决策变量 τ 的下界 a 和上界 b , 以获得最小时隙的范围。在此范围内, 对于每个控制应用, 计算每个总线周期中的最小时隙数和一个超周期中的空闲最小时隙总数, 并调用算法 5.1 来获得 \bar{x}_i 。在步骤 11 中, 假设在第 i 个控制应用中, 一个 ID 号为 0、服务时间为 τ 、实例数为 ν_i 的虚拟(virtual)消息在空闲

算法 5.1 基于 DFP 方法的参数估计算法

输入: $t_{k,j}$.

输出: \bar{x} .

- (1) 初始化向量 \bar{x}_1 以及精度 $\varepsilon = 10^{-6}$;
 - (2) 令 $H_1=I$, $k=1$; 用式(5.6)计算梯度 $\bar{g}_1 = \nabla \psi(\bar{x}_1)$;
 - (3) WHILE $\|\nabla \psi(\bar{x}_k)\| > \varepsilon$ DO
 - (4) $\bar{d}_k = -H_k \bar{g}_k$;
 - (5) 从 \bar{x}_k 出发沿着方向 \bar{d}_k 进行线搜索, 求步长 θ_k , 使它满足 $\psi(\bar{x}_k + \theta_k \bar{d}_k) = \max_{\theta \geq 0} \psi(\bar{x}_k + \theta \bar{d}_k)$;
 - (6) $\bar{x}_{k+1} = \bar{x}_k + \theta_k \bar{d}_k$;
 - (7) IF $k=n$, THEN
 - (8) $\bar{x}_1 = \bar{x}_{k+1}$;
 - (9) $H_1=I$; $k=1$; 用式(5.6)计算 $\bar{g}_1 = \nabla \psi(\bar{x}_1)$;
 - (10) ELSE
 - (11) $\bar{g}_{k+1} = \nabla \psi(\bar{x}_{k+1})$;
 - (12) $\bar{p}_k = \bar{x}_{k+1} - \bar{x}_k$;
 - (13) $\bar{q}_k = \bar{g}_{k+1} - \bar{g}_k$;
 - (14) $H_{k+1} = H_k + \frac{\bar{p}_k \bar{p}_k^T}{\bar{p}_k^T \bar{q}_k} - \frac{H_k \bar{q}_k \bar{q}_k^T H_k}{\bar{q}_k^T H_k \bar{q}_k}$;
 - (15) $k=k+1$;
 - (16) END IF
 - (17) END WHILE
 - (18) $\bar{x} = \bar{x}_k$;
 - (19) 返回 \bar{x} .
-

算法 5.2 最优 minislot 算法

输入: 表 5.1 中的 ECU, T_i , ε_i , 和 P_i , 以及控制应用数 c .

输出: τ_{\min} , $F(\tau_{\min})$.

- (1) 计算第 $i(i=1,2,\dots,c)$ 个控制应用的消息 ID 数, 得到消息 ID 集合 $MsgID_i$, 令 $\zeta_i = \|MsgID_i\|$;
- (2) 计算 $\gcd\{\varepsilon_i, 1 \leq i \leq \sigma\}$ 和超周期 $H = \text{lcm}\{T_i, 1 \leq i \leq \sigma\}$, 其中 $\sigma = \sum_{i=1}^c \zeta_i$, lcm 表示最小公倍数;
- (3) 计算 $\xi_i = \sum_{j \in MsgID_i} \varphi_{ij}$, $\varphi_{ij} = \frac{H}{T_j}$, 以及 $\xi = \max\{\xi_i, 1 \leq i \leq c\}$;
- (4) 计算总线环长度 $\ell = \frac{H}{\xi}$;
- (5) 计算每个控制应用的到达率 $\lambda_i, i=1,2,\dots,c$;
- (6) 根据最优化模型(5.3)的约束条件计算决策变量 τ 的下界 a 和上界 b , 且令 $C_h = C_d = 1$;
- (7) FOR $\tau = a$ TO b STEP h
- (8) FOR $i = 1$ TO c
- (9) 计算每个总线环的 minislot 数, 即 $MSNo = \left\lfloor \frac{\ell}{\tau} \right\rfloor$;
- (10) 计算一个超周期内空闲 minislot 总数, 即 $\nu_i = \xi - \zeta_i$;
- (11) $MsgID_i \leftarrow MsgID_i \cup \{0\}$, 令 $\varphi_0 = \nu_i$, $\varepsilon_0 = \tau$;
- (12) 对每个 $k \in MsgID_i$ 和 $j \in [1, \varphi_{ik}]$, 如果 $\text{mod}(\varepsilon_k, \tau) = 0$, 那么 $\hat{t}_{i,k,j} = \varepsilon_k$; 否则

$$\hat{t}_{i,k,j} = \left(\left\lfloor \frac{\varepsilon_k}{\tau} \right\rfloor + 1 \right) \tau;$$
- (13) 用参数 $\hat{t}_{i,k,j}$ 调用算法 5.1 获得 \bar{x}_i ;
- (14) 用式(5.2)计算 $E[L_i]$;
- (15) END FOR
- (16) 用式(5.3)计算 $F(\tau)$;
- (17) END FOR
- (18) 选择 τ_{\min} 使得 $F(\tau_{\min})$ 最小;
- (19) 返回 τ_{\min} , $F(\tau_{\min})$.

minislot被传递。因此，当计算系统成本时，也考虑了由空闲minislot产生的成本。算法5.2的时间复杂度与控制应用数 c 、决策变量 τ 的下界 a 和上界 b 有关，其时间复杂度为

$$O\left(\frac{c(b-a)n^2}{h}\right), \text{ 其中 } h \text{ 为步长。}$$

5.4.3 实验结果

在实验中，首先从模型(5.3)中得到 $\tau \in [0.01, 0.151)$ ，并令 $a=0.0025$, $b=0.16$ ，以及步长 $h=0.0025$ ；然后，实现算法5.2得到最优化模型(5.3)的最优解为 $\tau_{\min} = 0.01ms$ ，

$F(\tau_{\min}) = 0.2868w$ ，同时获得每一个总线环的长度 ℓ 和minislot数 $MSNo$ ，分别为1.2ms和120个minislot。Minislot和系统成本的关系如图5.5所示。

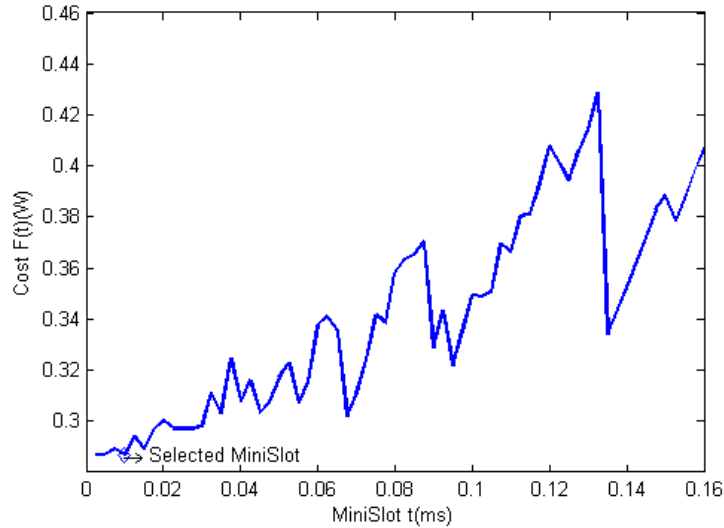


图 5.5 最小时隙与系统成本的关系

实验结果表明，当 $x_1 = x_2 = x_3 = x_4$ 时方程(5.5)中的 $\Psi(\vec{x})$ 最大。不失一般性，令 $S = xS_0$ ，

其中 $S_0 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$ 。那么 $\vec{x} = (x, x, x, x)$ 。对于图5.1中的六个控制应用， x 的值能

得到，分别为 $x = 19.6872, 13.8126, 12.5334, 26.0310, 33.4069$ ，和 19.3936 。因此，方程(5.4)表示的概率密度是一种规范形式的PH分布的密度函数^[196,197]。

在一个超周期内，假设在第 $h(h \leq c)$ 个控制应用中消息实例 m_{ij} 在第 s 个总线环传输，那么第 h 个子队列的 m_{ij} 的等待时间由下式计算：

$$t_{h,i,j,s}^a = (s-1)\ell + \sum_{v=1}^{h-1} \left\lceil \frac{\mathcal{E}_{v,s}}{\tau} \right\rceil \tau - jT_i,$$

其中 $\varepsilon_{v,s}$ 表示在第 v 个控制应用中的第 s 个总线环中的相应消息实例的端到端的服务时间；在第 s 个总线环中，如果存在空闲 minislot，那么 $\varepsilon_{v,s} = \tau$ 。总的端到端的时延表示为

$$t_{h,i,j,s} = t_{h,i,j,s}^q + \varepsilon_{h,s} . \quad (5.7)$$

假设采样时刻由消息实例 m_{ij} 的总的端到端时延和到达时刻组成，即 $k = t_{h,i,j,s} + jT_i$ ，那么，得到系统中的所有消息实例的不确定端到端时延如图 5.6 所示。

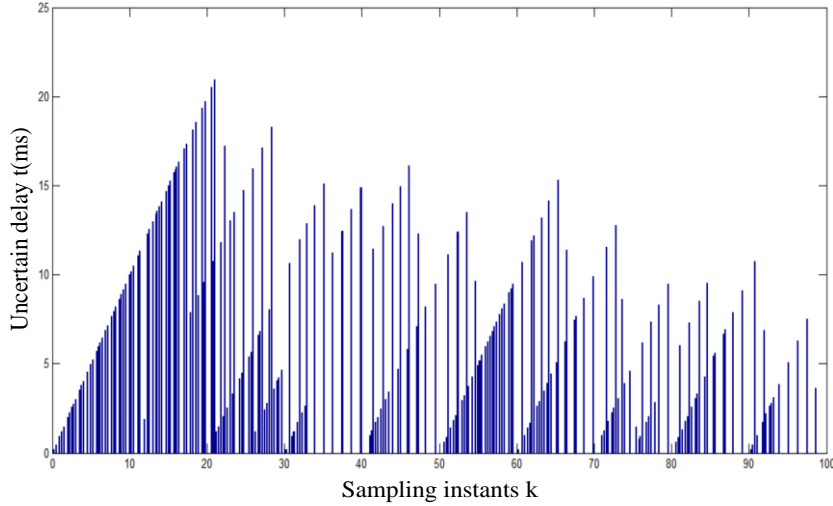


图 5.6 参考系统中的不确定时延

5.5 不确定时延的概率分布

5.5.1 概率密度函数

根据 FCFS 规则，考虑到达顾客在接受服务之前所需的等待时间长度。本章还考虑不确定的 MRT 的长度，包括前面的等待时间和客户的服务时间。用 $W_q(t)$ 和 $W(t)$ 分别表示两种等待时间的平稳分布，用 $W_q^*(s)$ 和 $W^*(s)$ 分别表示它们的 LSTs。根据 Pollaczek-Khintchine 变换公式^[198]，首先获得每个控制应用的 MRT 的 cdf 的 LST，然后，使用 QBD 过程中等待时间分布的计算方法来计算 cdf 的积分形式。通过积分变换，可以得到 cdf 的闭式。

定理 5.1 每个控制应用的 MRT 的 cdf 的 LSTs 表示为

$$W^*(s) = (1 - \rho)\beta \sum_{k=0}^{\infty} \lambda^k [(sI - S)^{-1} e\beta]^k (sI - S)^{-1} S^0 , \quad (5.8)$$

其中 $\text{Re } s \geq 0$ 。

证明. 根据 Pollaczek-Khintchine 变换方程，有

$$W_q^*(s) = \frac{(1 - \rho s)}{s - \lambda + \lambda f^*(s)}. \quad (5.9)$$

式(5.9)表示为

$$W_q^*(s) = (1 - \rho) \sum_{k=0}^{\infty} \left(\frac{\lambda(1 - f^*(s))}{s} \right)^k. \quad (5.10)$$

下列方程成立(见文献[192]第 86 页),

$$f^*(s) = 1 - s\beta(sI - S)^{-1}e. \quad (5.11)$$

将式(5.11)代入(5.10), 得到

$$W_q^*(s) = (1 - \rho) \sum_{k=0}^{\infty} [\lambda\beta(sI - S)^{-1}e]^k, \quad (5.12)$$

由于顾客在系统中的逗留时间等于前面的等待时间与服务时间之和, 根据文献[198]中的定理 2.9.3(d), 有

$$W^*(s) = W_q^*(s)f^*(s). \quad (5.13)$$

将式(5.1)和(5.12)代入式(5.13), 得到方程(5.8)。 □

定理 5.2 每个控制应用的 MRT 的 cdf 由下式计算:

$$W(t) = (1 - \rho)\beta \int_0^t \Phi(u) \exp(-uS) du \cdot \exp(tS)S^0, \quad (5.14)$$

其中 $t \geq 0$, 且

$$\Phi(t) = -V^0S - V^0 \exp[t(-V^0)^{-1}]\lambda e\beta, \quad (5.15)$$

以及

$$V^0 = -(\lambda e\beta + S)^{-1}. \quad (5.16)$$

证明. 由式(5.8), 令 $X = \lambda I$, 并考虑矩阵

$$\Phi^*(s) = \sum_{k=0}^{\infty} X^k [(sI - S)^{-1}e\beta]^k. \quad (5.17)$$

方程(5.17)能写为

$$\Phi^*(s) = I + X\Phi^*(s)(sI - S)^{-1}e\beta. \quad (5.18)$$

定义 $r[C]$ 是由 $m_1 \times m_2$ 阶矩阵 C 的行的直和形成的 $m_1 \times m_2$ 维向量^[192]。如果令

$\phi^*(s) = r[\Phi^*(s)]$, 以及 $v = r[I]$, 那么方程(5.18)表示为

$$\phi^*(s) = v + \phi^*(s)[X^T \otimes (sI - S)^{-1}e\beta], \quad (5.19)$$

其中 X^T 是 X 的转置矩阵, \otimes 表示 Kronecker 乘积算子。

用经典的 Kronecker 积的特征, 式(5.19)变换为

$$\phi^*(s) = v + v(sI \otimes I - I \otimes S - X^T \otimes e\beta)^{-1}(X^T \otimes e\beta). \quad (5.20)$$

令 $\Phi(t)$ 表示 LST $\phi^*(s)$ 的质量函数矩阵, 且令 $\phi(t) = \mathcal{L}[\Phi(t)]$, 此处 $t \geq 0$ 。对变换反演, 式(5.20)导致

$$\begin{aligned} \phi(t) &= v + v \int_0^t \exp[(I \otimes S + X^T \otimes e\beta)u] du (X^T \otimes e\beta) \\ &= v + v(I \otimes S + X^T \otimes e\beta)^{-1} [\exp(I \otimes S \\ &\quad + X^T \otimes e\beta)t - I \otimes I] (X^T \otimes e\beta). \end{aligned} \quad (5.21)$$

令 V^0 是 m 阶方阵, 有

$$\mathcal{L}[V^0] = -v(I \otimes S + X^T \otimes e\beta)^{-1} = v^0,$$

那么 V^0 是下面方程的唯一解

$$V^0 S + X V^0 e\beta = -I. \quad (5.22)$$

将 $X = \lambda I$ 代入式(5.22), 得到方程(5.16).

令 v^2 维向量 $\theta(t)$ 被定义为

$$\theta(t) = v^0 \exp[(I \otimes S + X^T \otimes e\beta)t], t \geq 0, \quad (5.23)$$

且令 $\mathcal{L}[\Theta(t)] = \theta(t)$, 那么由式(5.23), 得知 $\theta(t)$ 满足微分方程

$$\dot{\theta}(t) = \theta(t)(I \otimes S + X^T \otimes e\beta), t \geq 0,$$

其中 $\theta(0) = v^0$ 。这等价于矩阵微分方程

$$\dot{\Theta}(t) = \Theta(t)S + X \Theta(t)e\beta, t \geq 0, \quad (5.24)$$

其中初始条件 $\Theta(0) = V^0$ 。

将 $X = \lambda I$ 代入式(5.24), 得到

$$\Theta(t) = V^0 \exp[t(-V^0)^{-1}]. \quad (5.25)$$

由式(5.21), 矩阵 $\Phi(t)$ 表示为

$$\Phi(t) = I + X V^0 e\beta - X \Theta(t)e\beta. \quad (5.26)$$

由式(5.22)、(5.25)和(5.26), 得到方程(5.15)。

考虑如下矩阵

$$\Psi^*(s) = \Phi^*(s)(sI - S)^{-1}. \quad (5.27)$$

相应的矩阵 $\Psi(t)$ 满足如下矩阵微分方程

$$\dot{\Psi}(t) = \Psi(t)S + \Phi(t), t \geq 0, \quad (5.28)$$

其中 $\Psi(0) = 0$ 。

由式(5.28)，容易得到

$$\Psi(t) = \int_0^t \Phi(u) \exp(-uS) du \cdot \exp(tS). \quad (5.29)$$

根据式(5.8)、(5.27)和(5.28)，得到累积分布函数 $W(t)$ 为

$$W(t) = (1 - \rho)\beta \Psi(t)S^0. \quad (5.30)$$

将(5.29)代入(5.30)，得到方程(5.14)。 \square

引理 5.1 假设 S 是一个非奇异矩阵， t 表示任意实数。如下矩阵方程成立。

$$\textcircled{1} \quad S \exp(tS) = \exp(tS)S, \text{ 以及 } S^{-1} \exp(tS) = \exp(tS)S^{-1}.$$

$$\textcircled{2} \quad \exp(t\lambda e\beta) = I - e\beta + \exp(t\lambda)e\beta.$$

证明. 根据矩阵级数，不难证明上述方程成立。 \square

定理 5.3 每个控制应用的 MRT 的 cdf 和 pdf 分别表示为

$$\begin{aligned} W(t) &= 1 + \beta(S^{-1} - tI - \mu^{-1}tS)\exp(tS)S^0 \\ &\quad + \frac{1}{\lambda}(\beta - \rho\beta + \mu^{-1}\beta S)[\exp(t\lambda) - 1]\exp(tS)S^0, \end{aligned} \quad (5.31)$$

以及

$$\begin{aligned} p(t) &= (1 - \rho)\beta \exp[t(\lambda I + S)]S^0 \\ &\quad + \left[\frac{1}{\lambda} \exp(t\lambda) - \frac{1}{\lambda} - t \right] (\beta + \mu^{-1}\beta S) \exp(tS)SS^0, \end{aligned} \quad (5.32)$$

其中 $t \geq 0$ 。

证明. 将式(5.19)代入式(5.18)，得到

$$\Phi(t) = (\lambda e\beta + S)^{-1}S + \exp[t(\lambda e\beta + S)] - (\lambda e\beta + S)^{-1} \exp[t(\lambda e\beta + S)]S. \quad (5.33)$$

将式(5.23)代入(5.33)，并由引理 5.1，有

$$\begin{aligned} &(1 - \rho)\beta \Phi(t) \exp(-tS) \\ &= \beta \exp(-tS) + (1 - \rho)\beta \exp(t\lambda) + \mu^{-1}\beta S \exp(t\lambda) - \beta - \mu^{-1}\beta S. \end{aligned} \quad (5.34)$$

由式(5.14)和(5.34)，生成式(5.31)。

在式(5.31)中，对 t 求导，得到方程(5.32)。 \square

5.5.2 每个消息的 WCRT

在每个控制应用中，每个消息的响应时间的 cdf 和 pdf 分别使用式(5.31)和(5.32)计算。首先考虑每个消息的响应时间的 pdf 的参数估计问题。在第 h 个控制应用中，对于第 i 个消息，令 $\beta = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ ， $S_i = x_i S_0$ 。每个消息实例的不确定端到端时延的采样数据 $t_{h,i,j,s}$ 来自于方程(5.7)，如图 5.6 所示。假设这些采样数据来自于独立同分布的随机变量。在式(5.32)中，用极大似然估计方法估计式 $S_i = x_i S_0$ 中的唯一参数 x_i 的值。方程(5.32)的对数似然函数为

$$\psi(x_i) = \sum_{j=1}^{\phi_{hi}} \log p(t_{h,i,j,s}; x_i), \quad (5.35)$$

其中 $1 \leq h \leq c$ ， $1 \leq s \leq \xi$ 。

上面方程关于 x_i 求导，得

$$\nabla \psi(x_i) = \frac{d\psi(x_i)}{dx_i} = \sum_{j=1}^{\phi_{hi}} \frac{1}{p(t_{h,i,j,s}; x_i)} \frac{dp(t_{h,i,j,s}; x_i)}{dx_i}. \quad (5.36)$$

因为方程(5.35)是一个复杂的非线性函数，很难直接得出方程(5.36)的闭式。用 Matlab 中的符号计算方法来讨论参数估计问题。Matlab 中的函数 Jacobian() 被用来推导出方程(5.36)的符号表达式。对于每个消息，通过调用算法 5.1，在式(5.35)中的参数 x_i 的值被估计，即对于第 i 个消息，相应的参数 x_i 被生成。第 i 个消息相应的概率密度的参数 x_i 的估计值如表 5.3。

在表 5.3 中，对于第 i 个消息，到达率， $\lambda_i = \frac{1}{T_i}$ 服务率 $\mu_i^{-1} = -\beta S_i^{-1} e$ 。 x_i 是矩阵 S_i 的元素， S_i 是第 i 个消息服务时间的 PH 分布的 pdf 的矩阵参数，且 $S_i = x_i S_0$ 。根据参数 x_i 的估计值，得到第 i 个消息的响应时间的 pdf 的闭式表达式。通过实验，在方程 $p_i(t) = 0$ 中找到一个非负实根，其中 $p_i(t)$ 是第 i 个消息的响应时间的 pdf，该根是第 i 个消息的响应时间的上界，即 WCRT κ_i 。一般来说，随着参数 x_i 的增加，服务率 μ_i 呈现波动性，但呈现上升趋势。每个消息的系统利用率 ρ_i 可以使用 $\rho_i = \frac{\lambda_i}{\mu_i}$ 获得，且 $\rho_i < 1$ 。除第 3 和

第34个消息外，其他所有消息的系统利用率均小于0.5。在所有消息中，88.4%消息的系统利用率均小于0.3。来自于不同和相同控制应用的部分消息的响应时间的pdf曲线图分别如图5.7和图5.8。

表 5.3 第 i 个消息相应参数 x_i 的估计值及其 WCRT

ID	ECU	λ_i	μ_i	x_i	κ_i	ID	ECU	λ_i	μ_i	x_i	κ_i	ID	ECU	λ_i	μ_i	x_i	κ_i
1	E2	0.1	0.6127	1.5316	11.6	24	E3	0.04	0.1662	0.4155	39.9	47	E2	0.02	0.1133	0.2832	54.0
2	E2	0.1	0.4435	1.1087	16.2	25	E3	0.04	0.1547	0.3866	43.1	48	E4	0.01	0.2128	0.5315	26.8
3	E3	0.2	0.3158	0.7895	27.5	26	E2	0.05	0.2985	0.7464	22.3	49	E3	0.01	0.1091	0.2727	51.6
4	E3	0.1	0.2428	0.6070	31.1	27	E5	0.04	0.3271	0.8175	19.9	50	E1	0.01	0.1035	0.2588	54.4
5	E1	0.1	0.4850	1.2126	14.7	28	E2	0.05	0.2534	0.6337	26.5	51	E3	0.01	0.1006	0.2514	56.0
6	E6	0.1	0.3625	0.9062	20.0	29	E1	0.04	0.2918	0.7295	22.3	52	E3	0.01	0.0930	0.2326	60.5
7	E1	0.01	0.8065	2.0187	7.5	30	E2	0.1	0.3039	0.7596	24.3	53	E1	0.01	0.0957	0.2391	58.8
8	E1	0.01	0.4673	1.1706	12.6	31	E2	0.05	0.2017	0.5042	33.7	54	E3	0.01	0.0865	0.2163	65.1
9	E1	0.01	0.3185	0.7973	18.2	32	E4	0.1	0.9681	2.4194	7.3	55	E3	0.01	0.0815	0.2037	69.1
10	E5	0.04	0.5083	1.2708	12.8	33	E3	0.1	0.2079	0.5199	37.2	56	E5	0.01	0.1318	0.3295	42.8
11	E1	0.01	0.2445	0.6114	23.4	34	E3	0.1	0.1962	0.4905	39.8	57	E3	0.01	0.0753	0.1882	74.9
12	E1	0.05	0.3745	0.9363	17.7	35	E1	0.04	0.2435	0.6087	26.8	58	E3	0.01	0.0720	0.1801	78.3
13	E2	0.01	0.4367	1.0906	13.5	36	E6	0.04	0.1788	0.4470	37.0	59	E3	0.01	0.0684	0.1709	82.5
14	E1	0.01	0.1664	0.4157	34.1	37	E4	0.02	0.4938	1.2331	12.5	60	E4	0.01	0.1786	0.4463	31.8
15	E5	0.01	0.3378	0.8456	17.2	38	E4	0.02	0.3478	0.8700	17.5	61	E1	0.01	0.0882	0.2204	63.8
16	E1	0.1	0.3269	0.8173	22.4	39	E5	0.02	0.1953	0.4881	31.0	62	E1	0.01	0.0829	0.2073	67.9
17	E2	0.01	0.2994	0.7488	19.3	40	E3	0.02	0.1116	0.2790	54.8	63	E3	0.01	0.0651	0.1629	86.7
18	E2	0.01	0.2320	0.5807	24.6	41	E4	0.02	0.2639	0.6594	23.0	64	E1	0.01	0.0772	0.1931	72.9
19	E2	0.02	0.2535	0.6333	23.9	42	E6	0.02	0.1028	0.2569	59.7	65	E1	0.01	0.0734	0.1835	76.8
20	E3	0.1	0.2263	0.5658	33.7	43	E2	0.02	0.1458	0.3645	41.7	66	E1	0.01	0.0695	0.1739	81.1
21	E6	0.1	0.3087	0.7719	23.8	44	E5	0.01	0.1721	0.4301	33.0	67	E2	0.02	0.1045	0.2612	58.7
22	E6	0.04	0.1960	0.4899	33.6	45	E2	0.04	0.2065	0.5162	31.8	68	E1	0.01	0.0662	0.1654	85.3
23	E3	0.04	0.1823	0.4558	36.2	46	E2	0.02	0.1218	0.3045	50.1	69	E4	0.01	0.1506	0.3766	37.5

5.5.3 评估拟合精度

用分位数图(Q-Q图)从仿真数据和拟合模型(见文献[27])来图形化比较概率分布。如果被比较的两个分布是类似的，那么Q-Q图上的点将近似于一条直线。从每一个分布取1000个随机采样，那么消息 m_5 和 m_{69} (高优先级和低优先级情况)的Q-Q图分别如图5.9和图5.10所示。这些消息的Q-Q图大致是直线，在 m_{69} 等低优先级消息的情况下具有较高的相对精度。实验的拟合精度也高于文献[27]的拟合精度。

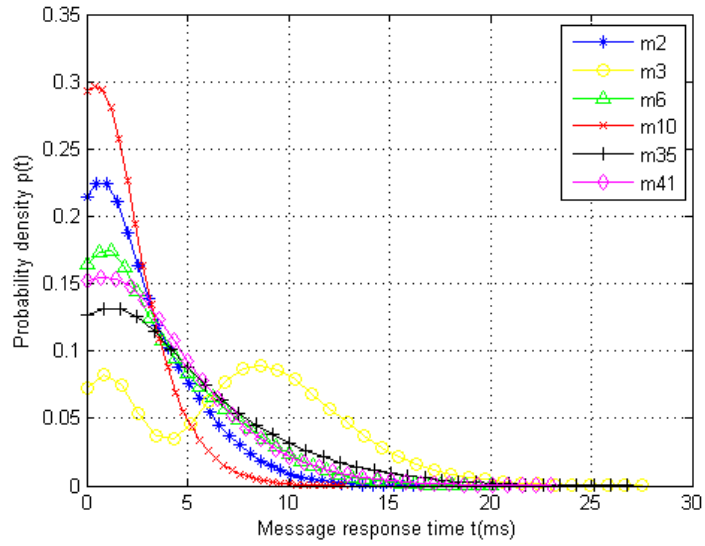


图 5.7 来自于不同控制应用的部分消息的响应时间的 pdf 曲线图

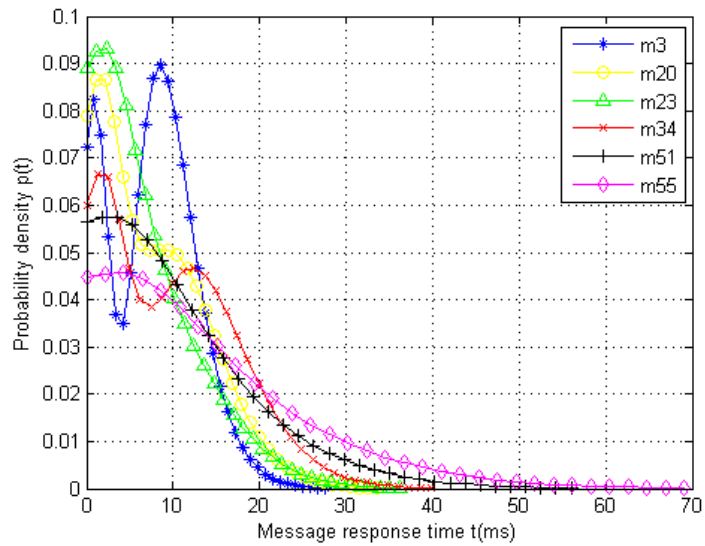


图 5.8 来自于第 3 个控制应用的部分消息的响应时间的 pdf 曲线图

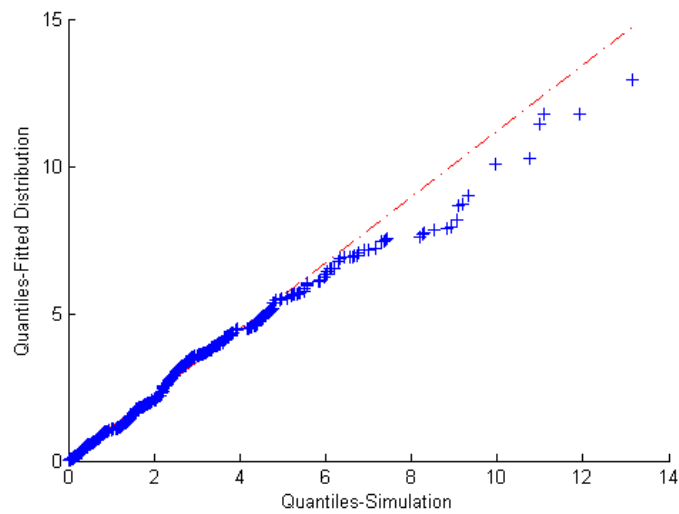


图 5.9 消息 m_5 的仿真和拟合分布采样 Q-Q 图

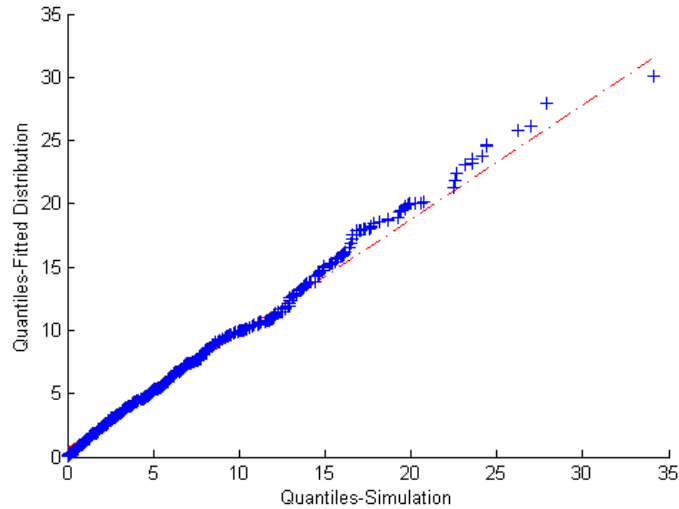


图 5.10 消息 m_{69} 的仿真和拟合分布采样 Q-Q 图

为了量化评估由方程(5.31)近似的MRT的cdf的拟合精度，用几种度量方法评估拟合分布的精度。采用Monte Carlo模拟方法获得10,000次随机采样后MRT的cdf仿真数据。根据表5.3的参数值，计算统计分析指标，诸如根均方误差(root mean squared error, RMSE)，决定系数(R^2)，以及Kolmogorov-Smirnov检验统计量(K-S)，如表5.4。最大RMSE是0.0068， R^2 最小值是0.99939，以及K-S最大值是0.017。与文献[27]中的表5.1比较，对于早期估计，本章的结果更精确。

在本章的工作中，基于排队模型的MRT分析是一种随机分析方法。在文献[27]中，作者比较了[27]中表1所示示例系统的运行时间信息，并指出随机分析比统计分析慢 10^7 倍以上。因此，随机分析显著较慢。在实验中，获得了表5.1中示例系统的运行时间信息。例如，获取消息 m_3 ， m_5 ， m_{12} 和 m_{69} 的pdf的参数值的运行时间值分别为65.17，39.71，27.53和11.75s，其周期分别为5，10，20和100ms。此外，表5.1中所有消息的pdf参数值获得的运行时间值都在区间(10ms,70ms)之间。因此，根据这里的实验结果和[27]中的表8，发现随机分析比统计分析慢大约五个数量级。但是，当完整的消息集可用时，随机分析比统计分析更精确。可以通过比较上面的表5.4和[27]中的表2来得出这个结论。

5.6 预测 MRT

5.6.1 GM(1,1)中的两参数背景值

一阶累积生成运算(AGO)序列为:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\},$$

其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$.

表 5.4 ANCS 中的消息拟合分布统计分析表

ID	RMSE	R ²	K-S	ID	RMSE	R ²	K-S	ID	RMSE	R ²	K-S
1	0.001329	0.999975	0.004683	24	0.003184	0.999863	0.008262	47	0.001621	0.999964	0.006006
2	0.002363	0.999922	0.005917	25	0.002506	0.999916	0.005538	48	0.004086	0.999773	0.010388
3	0.002381	0.999947	0.006896	26	0.001383	0.999973	0.004185	49	0.003639	0.999822	0.010021
4	0.002850	0.999900	0.008480	27	0.003195	0.999856	0.008790	50	0.002926	0.999885	0.007917
5	0.002173	0.999934	0.007809	28	0.001560	0.999966	0.004987	51	0.002196	0.999935	0.006686
6	0.003618	0.999822	0.008879	29	0.003423	0.999836	0.009444	52	0.002469	0.999918	0.009066
7	0.001887	0.999951	0.005760	30	0.006824	0.999388	0.017202	53	0.002816	0.999894	0.009266
8	0.003005	0.999876	0.007352	31	0.003037	0.999875	0.010274	54	0.002488	0.999917	0.006616
9	0.002507	0.999914	0.007729	32	0.002748	0.999892	0.008763	55	0.001458	0.999972	0.005460
10	0.001570	0.999965	0.004782	33	0.002843	0.999907	0.008367	56	0.002525	0.999914	0.008732
11	0.001862	0.999953	0.006612	34	0.003967	0.999826	0.012812	57	0.002594	0.999910	0.009584
12	0.002746	0.999893	0.007708	35	0.004040	0.999773	0.011774	58	0.003771	0.999811	0.012020
13	0.002420	0.999919	0.006488	36	0.002037	0.999944	0.006124	59	0.003071	0.999875	0.009103
14	0.003443	0.999839	0.010404	37	0.001061	0.999984	0.003814	60	0.003554	0.999828	0.008832
15	0.002125	0.999938	0.006530	38	0.002764	0.999893	0.007768	61	0.001524	0.999969	0.006430
16	0.003132	0.999869	0.007855	39	0.001929	0.999949	0.005473	62	0.001854	0.999954	0.004489
17	0.002585	0.999908	0.007566	40	0.004086	0.999775	0.009229	63	0.002602	0.999910	0.006892
18	0.001140	0.999982	0.003719	41	0.002363	0.999922	0.006374	64	0.001936	0.999950	0.005998
19	0.002180	0.999934	0.007505	42	0.003761	0.999810	0.013604	65	0.002043	0.999944	0.006376
20	0.002991	0.999893	0.010112	43	0.002593	0.999908	0.009986	66	0.001643	0.999964	0.006169
21	0.002255	0.999933	0.008959	44	0.002941	0.999882	0.009388	67	0.002081	0.999942	0.006785
22	0.001949	0.999948	0.005840	45	0.002670	0.999902	0.007326	68	0.003184	0.999866	0.008676
23	0.001688	0.999961	0.005010	46	0.002777	0.999895	0.006971	69	0.001670	0.999962	0.005999

背景值数组 $z^{(1)}$ 可以计算为

$$Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\},$$

ANCS 具有不确定时延的灰色系统特征, 用灰色模型预测 MRT。但是, 由于系统时延不确定性, 传统的 GM(1,1)不能直接运用于 ANCS 系统中。在传统的 GM(1,1)模型中, 生成系数通常为 0.5。事实上, 整个系统的生成系数并不总是相同的^[199]。几项研究表明, 应该确定最优的 α 来减少灰色预测的建模误差。Hsu^[173]提出了一种改进的变换 GM(1,1)模型(improved transformed GM(1,1), ITGM (1,1)), 通过遗传算法找到参数 α 的最优值^[200,201]。ITGM(1,1)模型预测方程由下式给出:

$$\hat{x}^{(0)}(k) = \left[\frac{b - ax^{(0)}(1)}{1 + \alpha a} \right] \left(\frac{1 - \alpha a}{1 + \alpha a} \right)^{k-2},$$

其中 $k = 2, 3, \dots, n$.

然而，由于 DES 中的资源约束，单参数 α 背景值 $z^{(1)}(k)$ 的计算方法不适用于 DES 中的预测问题。假定方程(2.10)中的背景值 $z^{(1)}(k)$ 用两个参数 α 和 γ 计算，其中它们的和不总是等于 1，即

$$z^{(1)}(k) = \alpha x^{(1)}(k) + \gamma x^{(1)}(k-1), \quad (5.37)$$

其中 $\alpha, \gamma \in (-\infty, \infty)$ ，且 $k = 2, 3, \dots, n$.

定理5.4 在GM(1,1)中，式(2.11)表示的一阶微分方程，即 $x^{(0)}(k) + az^{(1)}(k) = b$ ，能变换为

$$x^{(0)}(k) = \left[\frac{b - (\alpha + \gamma)ax^{(0)}(1)}{1 + \alpha a} \right] \left(\frac{1 - \gamma a}{1 + \alpha a} \right)^{k-2}, \quad (5.38)$$

其中 a 和 b 分别是发展系数和灰色输入，且 $k = 2, 3, \dots, n$.

证明. 由式(2.9)，有

$$x^{(1)}(k) = x^{(1)}(k-1) + x^{(0)}(k), \quad k = 2, 3, \dots, n. \quad (5.39)$$

由(5.37)和(5.39)，一阶微分方程变换为

$$(1 + \alpha a)x^{(0)}(k) + (\alpha + \gamma)ax^{(1)}(k-1) = b.$$

当 $k = 2$ 时，得到

$$x^{(0)}(2) = \frac{b - (\alpha + \gamma)ax^{(0)}(1)}{1 + \alpha a}.$$

当 $k = 3$ 时，有

$$(1 + \alpha a)x^{(0)}(3) + (\alpha + \gamma)ax^{(1)}(2) = b.$$

那么，得到

$$\begin{aligned} x^{(0)}(3) &= \frac{b - (\alpha + \gamma)ax^{(0)}(1) - (\alpha + \gamma)ax^{(0)}(2)}{1 + \alpha a} \\ &= x^{(0)}(2) \frac{1 - \gamma a}{1 + \alpha a} \\ &= \left[\frac{b - (\alpha + \gamma)ax^{(0)}(1)}{1 + \alpha a} \right] \frac{1 - \gamma a}{1 + \alpha a}. \end{aligned}$$

一般地，能获得

$$x^{(0)}(k) = x^{(0)}(k-1) \frac{1 - \gamma a}{1 + \alpha a}. \quad (5.40)$$

通过式(5.40)的迭代计算，得到方程(5.38)。 □

5.6.2 概率变换 GM(1,1)模型(PTGM(1,1))

提出了一种新的GM(1,1)模型，即具有两参数的概率变换GM(1,1)模型(PTGM(1,1))，以预测ANCS中的MRT。在PTGM(1,1)中，概率值由MRT的cdf计算，背景值 $z^{(1)}(k)$ 用式(5.37)计算。序列 $\hat{x}^{(0)}$ 的预测值由下式定义

$$\hat{x}^{(0)}(k) = \left[\frac{b - \omega(\alpha + \gamma)ax^{(0)}(1)}{1 + \alpha a} \right] \left(\frac{1 - \omega \min\{\alpha, \gamma\}a}{1 + \alpha a} \right)^{k-2}, \quad (5.41)$$

其中 $k = 2, 3, \dots, n$.

在式(5.41)中，令

$$\omega = \begin{cases} 1 - W(x^{(0)}(k-1)), & P\{W(x^{(0)}(j)) < 0.5\} > 0.5, \\ W(x^{(0)}(k-1)), & \text{Otherwise,} \end{cases}$$

其中 $W(x^{(0)}(i))$ ，表示 $x^{(0)}(i)$ 的概率，使用式(5.31)计算， $P\{W(x^{(0)}(j)) < 0.5\}, 1 \leq j \leq k-1$ 表示概率小于0.5的MRT数量的比例。

为了提高预测每个消息响应时间的精确性，优化确定了PTGM(1,1)的两个参数 α 和 γ 。首先定义适应度函数来评估模型中每次迭代的适宜值，使用MAPE来构建适应度函数为

$$MAPE(\alpha, \gamma) = \frac{1}{n-1} \sum_{k=2}^n \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\%, \quad (5.42)$$

其中 $x^{(0)}(k)$ 表示实际值， $\hat{x}^{(0)}(k)$ 表示相应的预测值。使用 DFP 方法最小化 $MAPE(\alpha, \gamma)$ ，得出决策变量 α 和 γ 的最优解。

Martin 和 Witt^[202]提出，预测性能不仅可以在不同的预测技术之间进行比较，而且可以在不同的单位之间进行比较，因此，MAPE 和根均方百分比误差(root mean square percentage error, RMSPE)是最常用的精确度量(见文献[173]中的表 1)。RMSPE 由式(5.43)给出。MAPE 和 RMSPE 的值越低，预测就越精确。

$$RMSPE(\alpha, \gamma) = \sqrt{\frac{1}{n-1} \sum_{k=2}^n \left[\frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right]^2} \times 100\%. \quad (5.43)$$

两参数 PTGM(1,1)的优化算法描述为算法 5.3。

算法 5.3 两参数 PTGM(1,1)的优化算法

输入: 每个消息响应时间的初始序列 $X^{(0)}$.

输出: α , γ , $MAPE(\alpha, \gamma)$, 以及 $RMSPE(\alpha, \gamma)$.

- (1) 由式(2.9)计算 AGO 序列;
- (2) 由式(5.37)计算背景值 $Z^{(1)}$ 关于符号变量 α 和 γ 的符号表达式;
- (3) 由最小二乘方计算参数 a 和 b , 即 $\hat{U} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y$, 其中

$$Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix}, \quad B = \begin{pmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{pmatrix};$$

- (4) 用式(5.41)计算序列 $\hat{X}^{(0)}$ 的预测值;
- (5) 用式(5.42)生成 $MAPE(V)$ 的符号公式, 其中列向量 $V = (\alpha, \gamma)^T$;
- (6) 生成式(5.42)中关于 V 的梯度 $\nabla MAPE(V)$;
- (7) 初始化向量 V 以及精度 $\varepsilon=10^{-5}$;
- (8) 调用算法 5.1 得到决策变量 V 的最优值 \hat{V} ;
- (9) 分别用式(5.42)和(5.43)计算 $MAPE(\hat{V})$ 和 $RMSPE(\hat{V})$ 的最小值;
- (10) 计算 \hat{U} 和序列 $\hat{X}^{(0)}$ 的预测值;
- (11) 返回 \hat{V} , $MAPE(\hat{V})$, 以及 $RMSPE(\hat{V})$.

在算法 5.3 中, 使用 Matlab 中的符号操作方法来生成参数 α 和 γ 的最优值, 分别得到背景值 $Z^{(1)}$ 的符号表达式和 $MAPE(V)$ 的符号公式。符号公式 $MAPE(V)$ 的梯度 $\nabla MAPE(V)$ 是用 Matlab 中的 jacobian() 函数获得的。通过算法 5.1 的实现, 生成了决策变量 V 的最优值。算法 5.3 的时间量与算法 5.1 的时间量相同。

5.6.3 预测性能度量

在评估预测性能时, 由于实验结果表明 GM(1,1)及其改进模型在预测仅包含一个或两个元素的这种特殊的延迟序列时具有较高的精确性, 因此不考虑在一个超时期内只有

一个或两个消息实例的消息的延迟预测问题。在算法 5.3 中，设向量 V 的初始值为 $(0.5, 0.5)^T$ 。对于在一个超周期内包含三个以上消息实例的消息，实现算法 5.3 以得到 PTGM(1,1)模型的最优参数值，如表 5.5 所示。

表 5.5 不同的 GMs(1,1)的最优参数值和性能度量

ID	PTGM (1,1)				ITGM (1,1)			GM (1,1)	
	α	γ	MAPE	RMSPE	α	MAPE	RMSPE	MAPE	RMSPE
1	0.9110	0.0579	10.7793	11.6546	0.5557	11.2592	12.0446	10.9111	16.6843
2	0.6005	-0.2031	18.7455	21.9080	1.3279	19.2495	22.5744	20.5430	23.8058
3	0.4605	0.2319	14.7597	18.3841	0.5845	18.1305	21.8840	18.8721	25.9465
4	0.7146	-0.1339	10.6770	13.5263	0.5323	11.2842	14.4131	12.5637	15.5185
5	0.3588	0.3763	15.5991	24.4904	0.5058	16.4415	20.7637	16.2007	27.2693
6	0.2061	0.8660	40.0277	45.6045	1.5866	65.4939	83.9672	87.0142	127.2786
10	0.5806	0.3986	3.7043	4.7236	0.4912	24.5622	29.2825	25.0412	28.2737
12	-0.2580	1.2511	25.8821	27.1803	0.5079	25.2914	30.0797	26.1496	29.7466
16	-2.9678	3.9124	27.3957	36.6965	0.5261	41.6109	47.5201	40.7243	47.7868
20	0.7768	-0.1137	9.2270	11.3267	0.5300	9.7522	11.7460	10.2999	12.3121
21	0.1671	0.6845	34.4147	46.1168	2.0491	42.8135	53.2240	59.4914	82.3315
22	0.5492	0.4508	38.7443	48.7817	0.9450	51.3058	63.9753	108.6324	142.0622
23	0.5362	0.4014	0.4337	0.5274	0.4959	1.3591	1.5318	1.5590	1.5920
24	0.0384	0.9382	0.3169	0.4530	0.4987	0.4010	0.4486	0.5234	0.5479
25	0.5179	0.4818	0.0186	0.0293	0.5001	0.0209	0.0231	0.2742	0.2866
26	-1.0914	3.2596	20.1295	24.1307	0.4798	23.6006	27.9809	23.7720	28.4443
27	0.0699	0.9066	12.1753	15.1631	0.7937	34.3377	38.4766	34.5011	42.1164
28	-0.8172	2.7321	18.0720	21.2076	0.4860	19.9415	24.1121	20.0489	24.3637
29	0.4922	-0.0591	18.8756	22.1001	0.7958	21.8774	27.7722	41.5503	52.2167
30	0.5489	0.0943	25.7931	45.6623	0.6217	27.0471	47.0546	31.3708	50.4620
31	-0.3446	1.8946	12.9282	15.2651	0.4947	13.3163	16.5628	13.3755	16.5916
32	0.3004	0.6815	20.9190	22.7608	0.0287	21.1864	23.3555	21.4853	24.9075
33	0.4942	0.1605	7.3898	8.3553	0.5410	9.1657	10.6462	9.5682	11.5760
34	0.5168	0.1702	6.8103	7.8237	0.5381	8.5217	9.6392	8.8311	10.2358
35	1.0923	-0.2910	14.3668	16.9134	0.6387	14.4420	16.7770	18.5735	20.2554
36	0.4770	0.2860	37.3835	44.1359	0.6605	78.0861	107.2205	79.3627	101.4740
45	0.4836	0.0186	29.9192	34.3351	0.6603	34.5990	40.6272	38.6430	46.0276

为了性能评估和比较，使用DFP方法生成ITGM(1,1)的最优参数值 α ，并分别获得表5.5中所示的ITGM(1,1)和GM(1,1)的MAPE和RMSPE的值。在表5.5中，对于所有消息，PTGM(1,1)的MAPE值小于ITGM(1,1)和GM(1,1)的MAPE值。较低的MAPE值表示预测值

与实际值之间的偏差较小。考虑到表5.5中的结果以及根据[173]中的MAPE标准，PTGM(1,1)模型的所有消息的MAPE值小于50%，20%和10%；这些研究结果分别表明了合理的、良好的和高精度的预测能力。相比之下，在ITGM(1,1)模型中，几个消息，诸如 m_6 、 m_{22} 和 m_{36} ，的MAPE值大于50%，这表明预测结果不精确。GM(1,1)模型的预测结果更不精确。根据表5.5，在PTGM(1,1)模型中，方程 $\alpha + \beta = 1$ 不一定成立。

根据表5.5所示的模型预测结果，本章提出的PTGM(1,1)模型比ITGM(1,1)模型和GM(1,1)模型获得更小的MAPE和RMSPE值。因此，PTGM(1,1)模型是预测ANCS的MRT的合适模型。分别选择三个具有代表性的消息 m_{34} 、 m_3 和 m_6 ，使用PTGM(1,1)模型得到高精度性、良好和合理的预测值。它们的预测值显示在图5.11~5.13中。PTGM(1,1)模型的预测值非常接近MRT的实际值。

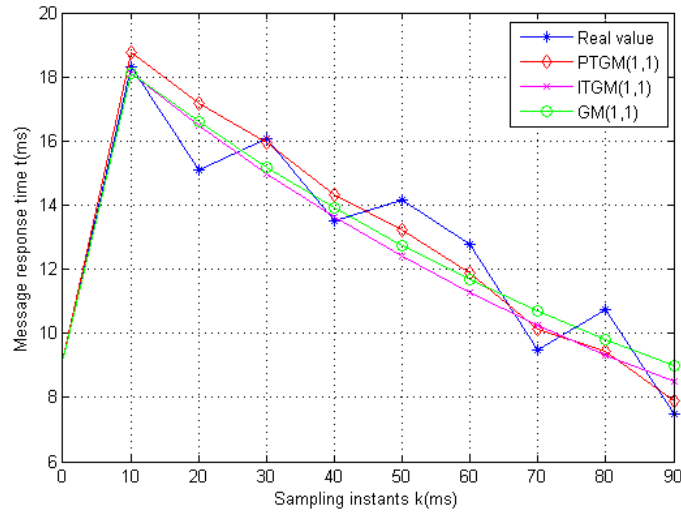


图 5.11 消息 m_{34} 的实际值和预测值比较

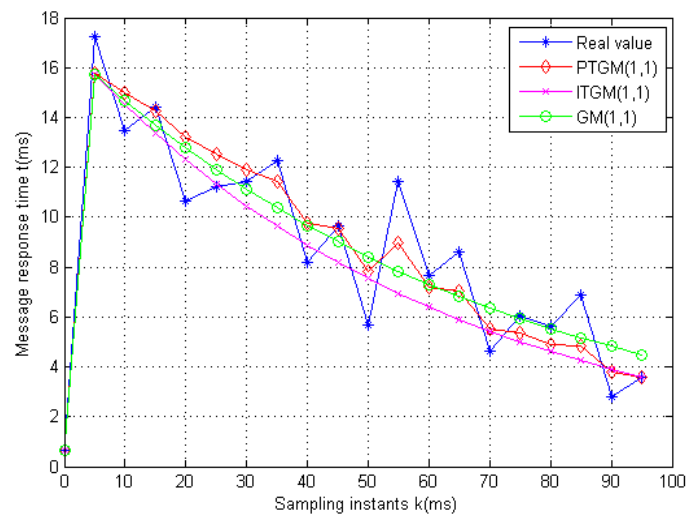


图 5.12 消息 m_3 的实际值和预测值比较

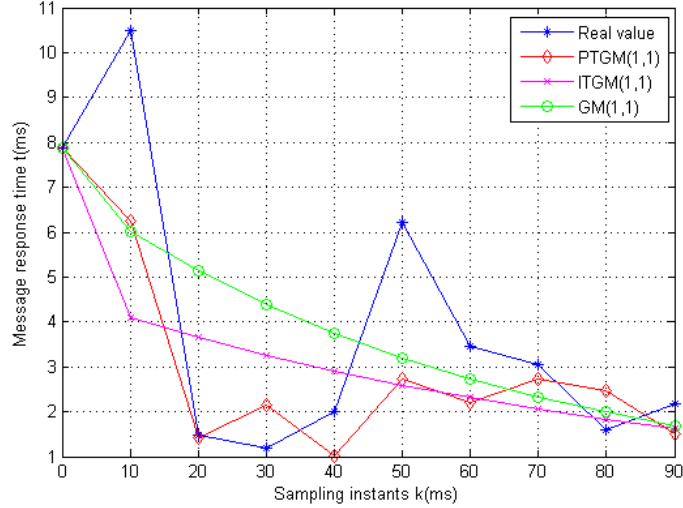


图 5.13 消息 m_6 的实际值和预测值比较

上面三个消息的预测响应时间的 cdfs 显示在图 5.14~5.16 中。预测值越准确，相应的 cdf 就越接近 MRT 的实际值。如图所示，近似的质量不一定取决于消息的优先级，而取决于预测模型。与[27]中预测结果的比较表明，由 PTGM(1,1)模型产生的 m_3 和 m_{34} 预测值的 cdfs 更准确(见文献[27]中的图 13 和 14)。在表 5.5 中，消息 m_6 的 MAPE 值在所有涉及的消息中值最大。因此，由 PTGM(1,1)模型计算的 m_6 预测值在所有涉及的消息中最不准确。然而，其预测值的 cdf 比[27]中的 cdf 更接近实际 cdf (见文献[27]中的图 12)。

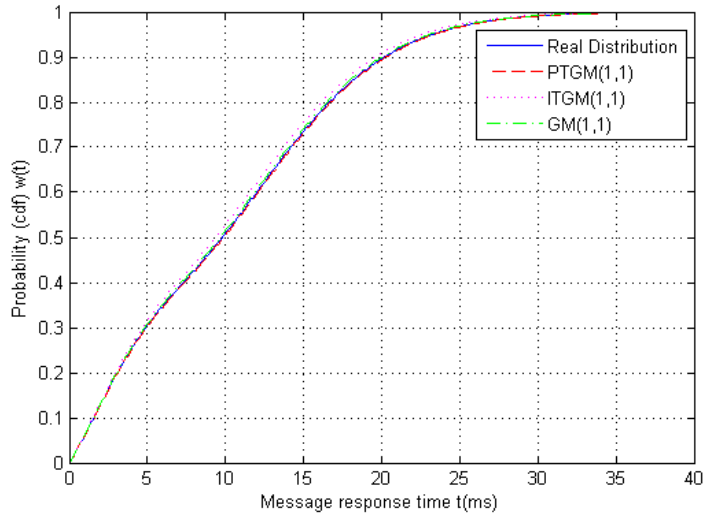
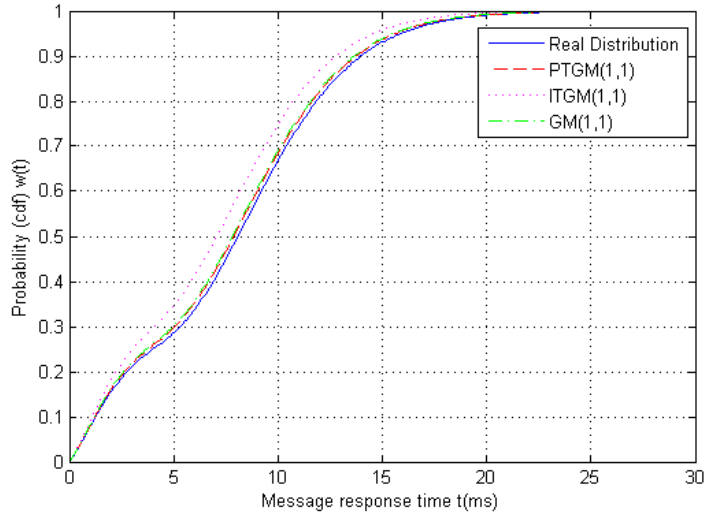
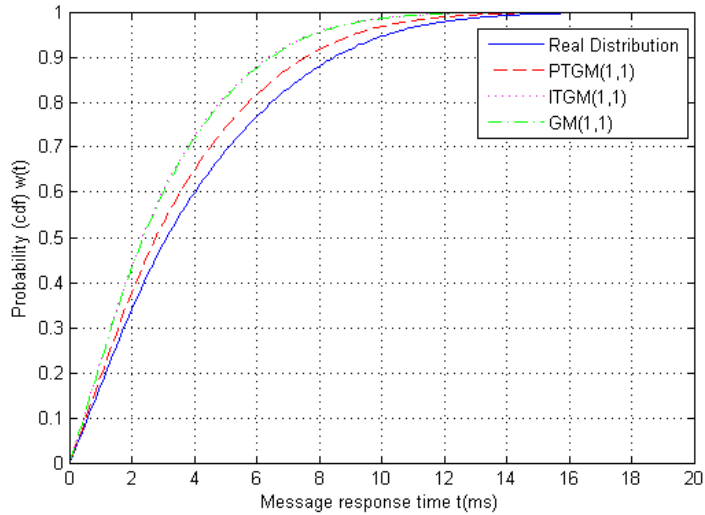


图 5.14 消息 m_{34} 的响应时间的 cdf 的预测

图 5.15 消息 m_3 的响应时间的 cdf 的预测图 5.16 消息 m_6 的响应时间的 cdf 的预测

5.7 本章小结

本章计算了不确定消息响应延迟的概率分布，并预测了ANCS中的延迟。通过分析系统的特征，建立了随机大小批量到达的M/PH/1排队模型，构建了ANCS的成本函数，并获得了MRT的pdf进行早期估计。在此模型和DFP变尺度法的基础上，分别提出了最优最小时隙算法和pdf参数估计算法，分别推导出总线最小时隙的最优值和PH分布的矩阵参数。利用[27]中实验车辆的其中一个CAN总线上的消息作为测试数据。目前的实验结果表明拟合精度高于文献[27]中的精度。本章提出了一个具有两参数的PTGM(1,1)模型来预测每个消息的响应延迟，MAPE和RMSPE用来评估预测的精确性。设计了基于DFP方法的PTGM(1,1)的两参数优化算法，以最小化MAPE并获得每个消息的两个参数的最优值。本章的数值结果表明，PTGM(1,1)的预测结果比现有的ITGM(1,1)模型和GM(1,1)

模型更精确。预测MRT可以用来设计控制器以补偿未来研究中闭环系统的系统延迟。与[27]中预测结果的比较表明，使用PTGM(1,1)模型获得的预测值的cdf比[27]中的更接近实际分布。作者相信这些结果对于汽车CPS早期体系结构探索阶段是非常重要的。使用这些方法，开发人员可以探索新消息的不同优先级，并评估具有不同优先级分配的MRT的不同概率分布。

结 论

1. 主要工作总结

CPS是嵌入物理世界并通过通信网络连接在一起的计算机系统的总称^[1]。CPS通过感知环境、计算所需的动作，以及在有限资源(如有限的通信带宽，有限的计算能力等)的约束下运用动作，直接和大部分时间独立于人类操作员与其物理环境交互，其无所不在的一个原因是，CPS的能力稳步增长，同时其采购和运营成本以更高的速度下降。人们可以确定CPS设计和开发的三个主要技术领域：自动控制、网络控制系统和嵌入式系统。传统上，控制器及其稳定性和性能分析的设计是独立于通信和计算体系结构的设计和分析而完成的，这通常导致非优化和过于保守的设计，进一步加剧了所得到的CPS的验证难度。在这些领域中，存在针对特定问题的解决方案，但挑战在于将这些解决方案组合成一种具有良好理论基础的集成方法——协同设计，以便充分开发利用CPS的全部功能。

传统的NCS研究关注于消息丢失、抖动或在给定网络上的分布式控制环境中出现的变延迟等挑战。但是，在嵌入式平台中，网络本身并不是先验固定的，而是提供了联合设计与优化网络和控制器的可能性。为了强调控制器设计和网络设计的关联，尤其是共享通信媒介的仲裁策略，Annaswamy等设计了一种称为ANCS的特殊CPS体系结构，通过强调消息的仲裁，可以使用关于网络的附加信息来设计控制器。基于这一平台体系结构，本文考虑了网络通信与控制协同设计中的若干问题，研究了以改善系统性能、减少系统成本为目的的多控制器调度、资源约束下的资源动态供应、不确定端到端的MRT等三方面的问题。主要创新点如下：

(1) 提出了一种新的双信道 ANCS 体系结构。在这一 ANCS 中，在轻量级通信情况下，通过构建排队系统，允许 FSDMC 模型中的部分空闲控制器进入半休眠状态，以改善系统性能、减少系统成本。

(2) 构建了一种新的 $N/(d, c)$ -M/M/c/K/SMWV 排队系统。考虑控制器簇的随机调度问题，用这一排队系统量化 FSDMC 模型的各种性能指标，分别构造成本和性能函数。建立一种多目标优化模型，最小化系统的非线性成本函数和标称时延函数，并运用 MOPSO 算法生成 Pareto 前沿和相应的非支配向量集。设计一种最优停止算法和 FSDMC 调度算法分别生成半休眠控制器数的最优值和多目标优化模型决策变量的最优值。这些参数用于配置双信道 ANCS。

(3) 提出了滞弹性系统中的滞弹性计算模型和计算方法，给出了 FSDMC 模型中的滞弹性量化、正式、新的定义，这一量化模型是一个二元组，由正常状态下的平均队列长度和计算资源匹配当前负载的概率组成。用随机环境下的 3-状态 M/M/1/K 排队系统建模 FSDMC 模型，精确计算滞弹性值，衡量系统中计算资源动态供应的自适应缩放能力，以保持 CPR 最小值不变。

(4) 构建了一种新的 $N/(d, c)$ -M/M/c/K/AMWV 排队系统，开发了一种最优化模型，最小化非线性 CPR 函数。PSO 算法用于求解该最优化模型，获得系统平稳条件下的参数最优值。当 $N/(d, c)$ -M/M/c/K/AMWV 排队系统中的平均到达率超过平均服务率时，提出了一种基于 PSO 的最优 CPR 不变自适应算法，动态调整控制器服务率，保持滞弹性值基本不变。

(5) 考虑单信道 ANCS 中不确定 MRT 分析问题，用具有随机大小的批量到达的连续时间 M/PH/1 排队系统建模 ANCS 中的每个控制应用为 QBD 过程，以进行早期体系结构探索。基于 ET 协议的多层 FTDMA/FP 总线调度策略，建立约束优化模型，最小化系统成本函数，获得总线 minslot 的最优值，协同设计总线参数和控制器服务率。用 QBD 过程中等待时间分布的计算方法生成每个控制应用的 MRT 的闭式 pdf。利用获得的 MRT 样本值和最大似然估计方法，提出了一种基于 DFP 方法的参数估计算法，以获得每个消息的 PH 分布矩阵参数的最优值，并生成每个消息相应的 pdf 和 cdf 闭式表达式。

(6) 提出了一种新的灰色预测模型(GM(1,1))，命名为 PTGM(1,1)模型，预测每个消息的响应时间。设计了一种基于 DFP 方法的 PTGM(1,1)两参数优化算法，最小化 MAPE，获得两个参数的最优值。数值结果表明，PTGM(1,1)的预测结果比现有的 ITGM(1,1)模型和 GM(1,1)模型更精确。

这些研究结果能在汽车、列车、航空控制系统、工业自动化系统等 CPS 系统的早期体系结构探索阶段具有重要的应用价值，也能应用到 CPS 系统最优参数配置，达到系统成本最低、性能最优。

2. 未来工作展望

随着对能源、交通和医疗保健中智能和高效工程应用的需求不断增加，CPS 的系统化设计变得越来越重要。该设计的一个重要组成部分是使用嵌入式系统平台实现高级控制算法。传统上，用于实现这些算法的控制设计和硬件/软件技术分别由不相关的研究领域——控制系统和嵌入式系统承担。本文中，考虑 ANCS 中的网络通信与控制协同设计中的若干问题，旨在整体建模整个系统，打破这种孤立的发展，缩小控制理论中保证的稳定性和性能鲁棒性与实现过程中所实现的实际性能之间的差距，从而降低集成、测试和调试成本。然而，本文中获得的結果和直觉对未来挑战和可能的扩展讨论如下：

(1) 双信道 ANCS 的时延问题。本文侧重考虑了单信道 ANCS 中的不确定时延问题，而没有考虑双信道的 ANCS 时延问题。在双信道 ANCS 平台中，消息经过信道 B 转发产生的时延需要进一步度量，基于张量的分析方法可能是一个合适的选择。本文用 M/PH/1 分别建模每一个控制应用，分析每一个消息的响应时间，而没有考虑在总线配置下来自不同控制应用的共享通信总线的消息之间的相互约束，导致 MRT 分析产生偏差。需要寻求合适的排队模型整体建模图 5.3 所示的多队列多服务台串并联混合排队系统，精确分析系统中的消息响应时间；或者采用基于无源性的设计方法分析系统的消息响应时间。考虑 ANCS 中资源受限问题，运用 PJN，通过采用无源采样器和无源保持连接的两种网络使得 c 个无源数字控制器控制 c 个连续时间无源被控对象。这一问题的研究具有挑战性。

(2) FSDMC模型的同步与可靠性问题。在FSDMC模型中,所有控制器构成了具有平等模式的扁平多控制器结构,以维护负载均衡,保持全局一致性。扁平控制方式要求所有控制器都处于同一层次,虽然物理上各个控制器位于不同的控制应用中,但逻辑上所有的控制器均作为全局控制器,掌握全网状态。当网络拓扑变化时,所有控制器将同步进行更新。在扁平控制方式中,虽然每个控制器掌握总线网络全局状态,但只控制单个控制应用中的节点,造成了一定资源的浪费,增加了网络更新时控制器的整体负载。因此,在DES中,在系统成本和性能最优的情况下,如何确保控制器的全局一致性?如何设计信道B的合适的通信协议,在系统性能最优的情况下维护系统负载均衡?此外,当某个控制器发生故障时,如何调度其它控制器完成故障控制器的控制功能,而确保系统成本和性能最优?这一系统可靠性问题是今后进一步研究的课题。

(3) 资源利用率问题。在本文提出的具有N策略、(d, c)多重工作休假的有限容量的M/M/c排队系统中,无论是同步工作休假,还是异步工作休假,都存在资源利用率不高的问题。在稳定条件下,系统处于不足供应状态时,异步多重工作休假模式的系统利用率为 $\rho=0.497$,同步多重工作休假时,系统利用率仅为 $\rho=0.147$ 。在N策略和(d, c)多重工作休假的模式下,当顾客按照泊松速率到达时,系统总的顾客数不断增加,而在顾客数大于c小于N时,这些顾客处于等待服务状态导致时延;当系统中顾客数不断减少到小于c时,在c-d个控制器中总存在一个或者几个控制器处于空闲状态而浪费资源。第5章用M/PH/1排队系统建模ANCS中的控制应用,得到绝大多数消息的系统利用率小于0.5,88.4%消息的系统利用率均小于0.3。因此,在FSDMC模型中,需要寻求新的排队模型建模该系统。

(4) 滞弹性计算。图4.2表明,FSDMC模型中的应变CPR与应力 λ 呈现非线性关系。那么,应变CPR与应力 λ 之间遵循着什么样的方程?出现在这些方程中的参量与应力-应变方程中的有关参量的关系将滞弹性的唯象理论与系统的变化特征联系起来。如何用滞弹性度量系统的变化状态及变化规律?一个热力学系统的内部变量在响应外部变量的变化时,会发生自适应调整的过程,从而随着时间的变化而趋向一种新的平衡状态,这种自调整的过程叫做弛豫(relaxation)。弛豫过程的两个重要参数是弛豫时间和弛豫强度。如何通过滞弹性效应的度量而计算出系统中的控制器服务率的弛豫时间和弛豫强度?这些问题有待进一步研究。

本文用排队系统建模 ANCS,以改善系统性能、降低系统成本为目的,研究了 ANCS 中的多控制器调度问题、资源动态供应问题以及不确定端到端的 MRT 问题。这些问题的研究必将对排队论在 CPS 中的更深入的运用起到推动作用,同时,也将对排队论在 CPS 中进一步的研究起到抛砖引玉的作用。作者相信,将排队论与 CPS 设计技术相结合,必将进一步产生真正对 CPS 的集成设计理论与方法具有指导意义、有影响的研究成果。

参 考 文 献

- [1] Voit H. An Arbitrated Networked Control Systems Approach to Cyber-Physical Systems: [dissertation]. Munich: Technical University of Munich, 2013, 1-16
- [2] Lee E A. Cyber physical systems: Design challenges. In: Proc. of 11th IEEE Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC). Orlando, 2008, 363-369
- [3] 李仁发, 谢勇, 李蕊等. 信息物理融合系统若干关键问题综述. 计算机研究与发展, 2012, 49(6): 1149-1161
- [4] 王中杰, 谢璐璐. 信息物理融合系统研究综述. 自动化学报, 2011, 37(10): 1157-1166
- [5] Landau I D, Lozano R, M'Saad M, et al. Adaptive Control: Algorithms, Analysis and Applications(Second Edition). London: Springer-Verlag London Limited, 2011, 1-32
- [6] Wang G, Zhang Q, Yan X. Analysis and Design of Singular Markovian Jump Systems. Cham: Springer International Publishing Switzerland, 2015, 229-249
- [7] Kim K -D, Kumar P R. Cyber-Physical Systems: A Perspective at the Centennial. Proceedings of the IEEE, 2012, 100(5): 1287-1308
- [8] Hespanha J, Naghshtabrizi P, Xu Y. A survey of recent results in networked control systems. Proceedings of the IEEE, 2007, 95(1): 138-162
- [9] Baillieul J, Antsaklis P. Control and communication challenges in networked real-time systems. Proceedings of the IEEE, 2007, 95(1): 9-28
- [10] Mindell D A. Digital Apollo - human and machine in spaceflight. Cambridge: MIT Press, 2008, 1-42
- [11] Annaswamy A M, Soudbakhsh D, Schneider R, et al. Arbitrated Network Control Systems: A Co-Design of Control and Platform for Cyber-Physical Systems. In: Proc. of the Workshop on Control of Cyber-Physical System. Baltimore, 2013, 339-356
- [12] Annaswamy A M, Chakraborty S, Soudbakhsh D, et al. The arbitrated networked control systems approach to designing cyber-physical systems. IFAC Proceedings Volumes, 2012, 45(26): 174-179
- [13] Voit H, Schneider R, Goswami D, et al. Optimizing Hierarchical Schedules for Improved Control Performance. In: Proc. of 5th International Symposium on Industrial Embedded Systems (SIES). Trento, 2010, 9-17

- [14] Voit H, Annaswamy A M. Adaptive Control of a Networked Control System with Hierarchical Scheduling. In: Proc. of the American Control Conference (ACC). San Francisco, 2011, 4189-4194
- [15] Voit H, Annaswamy A M, Schneider R, et al. Adaptive Switching Controllers for Systems with Hybrid Communication Protocols. In: Proc. of the American Control Conference (ACC). Montreal, 2012, 4921-4926
- [16] Voit H, Annaswamy A M, Schneider R, et al. Adaptive Switching Controllers for Tracking with Hybrid Communication Protocols. In: Proc. of 51st IEEE Conference on Decision and Control (CDC). Maui, 2012, 4121-4126
- [17] Chakraborty S, Künzli S, Thiele L. A general framework for analysing system properties in platform-based embedded system designs. In: Proc. of Design, Automation and Test in Europe Conference and Exhibition (DATE). Munich, 2003, 190-195
- [18] Soudbakhsh D, Phan L T X, Annaswamy A M, et al. Co-design of arbitrated network control systems with overrun strategies. *IEEE Trans. Control Netw. Syst.*, 2018, 5(1): 128-141
- [19] Lian F, Moyne J, Tilbury D. Network design consideration for distributed control systems. *IEEE Transaction on Control Systems Technology*, 2002, 10(2): 297-306
- [20] 徐锬. 资源约束实时控制系统反馈调度的研究: [硕士学位论文]. 杭州: 浙江工业大学信息工程学院, 2008, 1-6
- [21] Rajkumar R, Lee I, Sha L, et al. Cyber-physical systems: the next computing revolution. In: Proc. of 47th ACM/IEEE Design Automation Conference (DAC). Anaheim, 2010, 731-736
- [22] Zeng H, Natale M D, Ghosal A, et al. Schedule Optimization of Time-Triggered Systems Communicating Over the FlexRay Static Segment. *IEEE Trans. Ind. Informat.*, 2011, 7(1): 1-17
- [23] Dvorak J, Hanzalek Z. Using Two Independent Channels with Gateway for FlexRay Static Segment Scheduling. *IEEE Trans. Ind. Informat.*, 2018, 12(5): 1887-1895
- [24] Hu M, Luo J, Wang Y, et al. Holistic Scheduling of Real-Time Applications in Time-Triggered In-Vehicle Networks. *IEEE Trans. Ind. Informat.*, 2014, 10(3): 1817-1828
- [25] Kang M, Park K, Jeong M -K. Frame Packing for Minimizing the Bandwidth Consumption of the FlexRay Static Segment. *IEEE Trans. Ind. Electron.*, 2013, 60(9): 4001-4008

- [26] Goswami D, Schneider R, Masrur A, et al. Challenges in Automotive Cyber-physical Systems Design. In: Proc. of IEEE International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation (IC-SAMOS). Samos, 2012, 346-354
- [27] Zeng H, Natale M D, Giusto P, et al. Using Statistical Methods to Compute the Probability Distribution of Message Response Time in Controller Area Network. IEEE Trans. Ind. Informat., 2010, 6(4): 678-691
- [28] Gao S, Luo H, Chen D, et al. A Cross-Domain Recommendation Model for Cyber-Physical Systems. IEEE Trans. Emerg. Topics Comput., 2013, 1(2): 384-393
- [29] Chakraborty S, Faruque M A A, Chang W, et al. Automotive Cyber-Physical Systems: A Tutorial Introduction. IEEE Design & Test, 2016, 33(4): 92-108
- [30] Schmidt D C, White J, Gill C D. Elastic Infrastructure to Support Computing Clouds for Large-scale Cyber-Physical Systems. In: Proc. of 17th IEEE Int. Symp. Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC). Reno, 2014, 56-63
- [31] Balasubramanian J. Resource-aware deployment, configuration, and adaptation for fault-tolerant distributed real-time embedded systems: [dissertation]. Nashville: Institute for Software Integrated Systems, Vanderbilt University, 2009, 1-21
- [32] Ai W, Li K, Lan S, et al. On Elasticity Measurement in Cloud Computing. Scientific Programming, 2016, 2016: 1-13
- [33] Dustdar S, Guo Y, Satzger B, et al. Principles of elastic processes. IEEE Internet Computing, 2011, 15(5): 66-71
- [34] Nikam S, Ingle R. Resource Provisioning Algorithms for Service Composition in Cyber Physical Systems. In: Proc. of 2014 Int. Conf. Advances in Computing, Communications and Informatics (ICACCI). New Delhi, 2014, 2797-2802
- [35] Moldovan D, H. -L. Truong. A Platform for Run-time Health Verification of Elastic Cyber-physical Systems. In: Proc. of 24th IEEE Int. Symp. Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). London, 2016, 379-384
- [36] Moldovan D, Copil G, Dustdar S. Elastic systems: Towards cyber-physical ecosystems of people, processes, and things. Computer Standards & Interfaces, 2018, 57: 76-82
- [37] Candra M Z C, Truong H -L, Dustdar S. Modeling Elasticity Trade-Offs in Adaptive Mixed Systems. In: Proc. of 22nd IEEE Int. Workshop Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). Hammamet, 2013, 21-26

- [38] Shih C -S, Wang Y -H, Yang C -M, et al. Elastic Computation Middleware for Interactive Wearable Devices in Cyber-Physical Systems. In: Proc. of 3rd IEEE Int. Conf. Cyber-Physical Systems, Networks, and Applications (CPSNA). Hong Kong, 2015, 1-6.
- [39] Li K. Quantitative Modeling and Analytical Calculation of Elasticity in Cloud Computing. IEEE Trans. Cloud Comput., In press, 2017.
- [40] Charette R N. This car runs on code. IEEE Spectrum, 2009, 46(46): 1-6
- [41] Soudbakhsh D, Phan L T, Sokolsky O, et al. Co-design of Control and Platform with Dropped Signals. In: Proc. of 4th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPs). Philadelphia, 2013, 129-140.
- [42] Ge Y, Chen Q, Jiang M, et al. SCHMM-based modeling and prediction of random delays in networked control systems. Journal of the Franklin Institute, 2014, 351(5): 2430-2453
- [43] Joseph M, Pandya P. Finding response times in a real-time system, The Computer Journal, 1986, 29(5): 390-395
- [44] Audsley N C, Burns A, Richardson M, et al. Applying new scheduling theory to static priority pre-emptive scheduling. Software Engineering Journal, 1993, 8(5): 284-292
- [45] Audsley N C, Burns A, Davis R I, et al. Fixed priority preemptive scheduling: an historic perspective. Real-Time Syst., 1995, 8(2-3): 173-198
- [46] Tindell K, Hansson H, Wellings A. Analysing real-time communications: controller area network (can). In: Proc. of 15th IEEE Real-Time Systems Symposium (RTSS). San Juan, 1994, 259-263
- [47] Mubeen S, Maki-Turja J, Sjodin M. Extending worst-case response-time analysis for mixed messages in controller area network with priority and fifo queues. IEEE Access, 2014, 2(1): 365-380
- [48] Mubeen S, Maki-Turja J, Sjodin M. Response time analysis with offsets for mixed messages in can supporting transmission abort requests. In: Proc. of 19th IEEE Int. Conf. Emerging Technology and Factory Automation (ETFA). Barcelona, 2014, 291-294
- [49] Mubeen S, Maki-Turja J, Sjodin M. Integrating mixed transmission and practical limitations with the worst-case response-time analysis for controller area network. Journal of Systems and Software, 2015, 99: 66-84
- [50] Pop T, Pop P, Eles P, et al. Timing analysis of the flexray communication protocol.

- Real-Time Syst., 2008, 39(1): 205-235
- [51] Ouedraogo L, Kumar R. Computation of the precise worst-case response time of exray dynamic messages. *IEEE Trans. Autom. Sci. and Eng.*, 2014, 11(2): 537-548
- [52] Tanasa B, Bordoloi U D, Eles P, et al. Probabilistic timing analysis for the dynamic segment of flexray. In: *Proc. of 25th Euromicro Conf. Real-Time Syst. (ECRTS)*. Paris, 2013, 135-144
- [53] Zeng H, Natale M D, Giusto P, et al. Stochastic analysis of can-based real-time automotive systems. *IEEE Trans. Ind. Informat.*, 2009, 5(4): 388-401
- [54] Xie G, Zeng G, Kurachi R, et al. Gateway modeling and response time analysis on can clusters of automobiles. In: *Proc. of 17th IEEE Int. Conf. High Performance Computing and Communications (HPCC)*. New York, 2015, 1147-1153
- [55] Xie G, Zeng G, Li Z, et al. Adaptive dynamic scheduling on multi-functional mixed-criticality automotive cyber-physical systems. *IEEE Trans. Vehicular Tech.*, 2017, 66(8): 6676-6692
- [56] Dong Z, Ma Y, Peng W. Stability analysis of networked control system. *International Journal of Applied Mathematics and Statistics*, 2013, 43(13): 255-263
- [57] Cervin A, Alriksson P. Optimal on-line scheduling of multiple control tasks: A case study. In: *Proc. of 18th Euromicro Conference on Real-Time Systems (ECRTS)*. Dresden, 2006, 141-150.
- [58] Schneider R, Goswami D, Zafar S, et al. Constraint-driven synthesis and tool-support for FlexRay-Based automotive control systems. In: *Proc. of 9th IEEE/ACM International Conference on Hardware/Software-Codesign and System Synthesis (CODES+ISSS)*. Taipei, 2011, 139-148.
- [59] Baronti F, Petri E, Saponara S, et al. Design and Verification of Hardware Building Blocks for High-Speed and Fault-Tolerant In-Vehicle Networks. *IEEE Trans. Ind. Electron.*, 2011, 58(3): 792-801
- [60] FlexRay Consortium. FlexRay Protocol Specification (Version 3.0.1). <http://www.flexray.com>, 2010
- [61] Xie Y, Zeng G, Chen Y, et al. Worst case response time analysis for messages in controller area network with gateway. *IEICE Trans. Information and Systems*, 2013, 96-D(7): 1467-1477.
- [62] Kennedy J, Eberhart R. Particle swarm optimization. In: *Proc. of IEEE Int. Conf. Neural Networks*. Perth, 1995, 1942-1948

- [63] Davidon W C. Variable metric method for minimization, *Siam J. Optimization*, 1991, 1(1): 1-17
- [64] Fletcher R, Powell M J D. A rapidly convergent descent method for minimization. *Computer Journal*, 1963, 6(2): 163-168
- [65] Lewis C D. *Industrial and business forecasting methods: A practical guide to exponential smoothing and curvetting*. London: Butterworths Scientific, 1982, 22-51
- [66] Porter J, Hemingway G, Nine H, et al. *The ESMoL Language and Tools for High-Confidence Distributed Control Systems Design. Part I : Design Language, Modeling Framework, and Analysis*. Institute for Software Integrated Systems, Vanderbilt University, Technical Report, ISIS-10-109, 2010
- [67] Henzinger T, Horowitz B, Kirsch C. Giotto: a time-triggered language for embedded programming. *Proceedings of the IEEE*, 2003, 91(1): 84-99
- [68] Zou J, Matic S, Lee E A, et al. Execution strategies for PTIDES, a programming model for distributed embedded systems. In: *Proc. of the 15th IEEE Real-Time and Embedded Technology and Application Symposium (RTAS)*. San Francisco, 2009, 77-86
- [69] Samii S, Cervin A, Eles P, et al. Integrated scheduling and synthesis of control applications on distributed embedded systems. In: *Proc. of Design, Automation and Test in Europe Conference and Exhibition (DATE)*. Nice, 2009, 57-62
- [70] Lincoln B, Cervin A. Jitterbug: A Tool for Analysis of Real-Time Control Performance. In: *Proc. of the 41st IEEE Conference on Decision and Control (CDC)*. Las Vegas, 2002, 1319-1324
- [71] Naghshtabrizi P. *Delay impulsive systems: A framework for modeling networked control systems: [dissertation]*. Santa Barbara: University of California, 2007, 12-44
- [72] Zhang W, Branicky M S, Phillips S M. Stability of networked control systems. *IEEE Control Systems*, 2001, 21(1): 84-99
- [73] Naghshtabrizi P, Hespanha J P. Analysis of distributed control systems with shared communication and computation resource. In: *Proc. of the American Control Conference (ACC)*. Saint Louis, 2009, 3384-3389
- [74] Tabuada P. Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Transactions on Automatic Control*, 2007, 52(9): 1680-1685
- [75] Zhao X, Zhang L, Shi P, et al. Stability and stabilization of switched linear systems with mode-dependent average dwell time. *IEEE Transactions on Automatic Control*, 2012,

57(7): 1809-1815

- [76] Mason O, Shorten R. On common quadratic Lyapunov functions for stable discrete-time LTI systems. *IMA Journal of Applied Mathematics*, 2004, 69(3): 271-283
- [77] Morelli M, Di Natale M. Control and Scheduling Co-design for a Simulated Quadcopter Robot: A Model-Driven Approach. *Lecture Notes in Computer Science*, 2014, 8810: 49-61
- [78] Glass M, Teich J, Zhang L. A co-simulation approach for system-level analysis of embedded control systems. In: *Proc. of International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation (IC-SAMOS)*. Samos, 2012, 355-362
- [79] Kauer M, Soudbakhsh D, Goswami D, et al. Fault-tolerant control synthesis and verification of distributed embedded systems. In: *Proc. of 17th Design, Automation and Test in Europe (DATE)*. Dresden, 2014, 1-6
- [80] Saifullah A, Wu C, Tiwari P B, et al. Near optimal rate selection for wireless control systems, In: *Proc. of 18th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*. Beijing, 2012, 231-240
- [81] Chen D, Nixon M, Han S, et al. WirelessHART and IEEE 802.15.4e. In: *Proc. of IEEE International Conference on Industrial Technology (ICIT)*. Busan, 2014, 760-765
- [82] Maasoumy M, Zhu Q, Li C, et al. Co-design of control algorithm and embedded platform for building HVAC systems. In: *Proc. of 4th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*. Philadelphia, 2013, 61-70
- [83] Li M, Li P. Crowdsourcing in Cyber-Physical Systems: Stochastic Optimization with Strong Stability. *IEEE Trans. Emerg. Topics Comput.*, 2013, 1(2): 218-231
- [84] Xia F, Sun Y. Control-scheduling codesign: A perspective on integrating control and computing. *Dynamics of Continuous, Discrete and Impulsive Systems - Series B*, 2006, 13(S1): 1352-1358
- [85] Xu H, Sahoo A, Jagannathan S. Stochastic adaptive event-triggered control and network scheduling protocol co-design for distributed networked systems. *IET Control Theory and Applications*, 2014, 8(18): 2253-2265
- [86] Peng C, Han Q -L. Output-based event-triggered H_∞ control for sampled-data control systems with nonuniform sampling. In: *Proc. of American Control Conf. (ACC)*. Washington, 2013, 1727-1732

- [87] Donkers M C F, Heemels W P M H. Output-Based Event-Triggered Control With Guaranteed L_∞ -Gain and Improved and Decentralized Event-Triggering. *IEEE Trans. Autom. Control*, 2012, 57(6): 1362-1376
- [88] Heemels W P M H, Donkers M C F, Teel A R. Periodic event-triggered control based on state feedback. In: *Proc. of 50th IEEE Conf. Decision and Control and European Control Conf. (CDC-ECC)*. Orlando, 2011, 2571-2576
- [89] Guan Y, Han Q -L, Peng C. Decentralized event-triggered control for sampled-data systems with asynchronous sampling. In: *Proc. of American Control Conf. (ACC)*. Washington, 2013, 6565-6570
- [90] Al-Areqi S, Gorges D, Liu S. Event-based networked control and scheduling codesign with guaranteed performance. *Automatica*, 2015, 57(4): 128-134
- [91] Al-Areqi S, Gorges D, Liu S. Event-Based Control and Scheduling Codesign: Stochastic and Robust Approaches. *IEEE Trans. Autom. Control*, 2015, 6(5): 1291-1303
- [92] Majdoub T, Nours S L, Pasquier O, et al. Performance evaluation of an automotive distributed architecture based on a high speed power line communication protocol using a transaction level modeling approach. *Journal of Real-Time Image Processing*, 2014, 9(1): 281-295
- [93] Postoyan R, Tabuada P, Neši D, et al. A Framework for the Event-Triggered Stabilization of Nonlinear Systems. *IEEE Trans. Autom. Control*, 2015, 60(4): 982-996
- [94] Cao X, Cheng P, Chen J, et al. An Online Optimization Approach for Control and Communication Codesign in Networked Cyber-Physical Systems. *IEEE Trans. Ind. Informat.*, 2013, 9(1): 439-450
- [95] Gunes V, Peter S, Givargis T, et al. A Survey on Concepts, Applications, and Challenges in Cyber-Physical Systems. *KSII Trans. Internet and Information Systems*, 2014, 8(12): 4242-4268
- [96] Ravindran B, Li P. DPR, LPR: Proactive Resource Allocation Algorithms for Asynchronous Real-Time Distributed Systems. *IEEE Trans. Comput.*, 2004, 53(2): 201-216
- [97] Marti P, Lin C, Brandt S A, et al. Draco: Efficient Resource Management for Resource-Constrained Control Tasks. *IEEE Trans. Comput.*, 2009, 58(1): 90-105
- [98] Tan L, Zhu Z, Ge F, et al. Utility Maximization Resource Allocation in Wireless Networks: Methods and Algorithms. *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2015, 45(7): 1018-1034

- [99] Samii S, Eles P, Peng Z, et al. Dynamic Scheduling and Control-Quality Optimization of Self-Triggered Control Applications. In: Proc. of 31st IEEE Real-Time Systems Symposium (RTSS). San Diego, 2010, 95-104
- [100] Lozoya C, Martí P, Velasco M, et al. Resource and performance trade-offs in real-time embedded control systems. *Real-Time Syst.*, 2013, 49(3): 267-307
- [101] Chantem T, Hu X S, Lemmon M D. Generalized Elastic Scheduling for Real-Time Tasks. *IEEE Trans. Comput.*, 2009, 58(4): 480-495
- [102] Xie G, Chen Y, Liu Y, et al. Resource Consumption Cost Minimization of Reliable Parallel Applications on Heterogeneous Embedded Systems. *IEEE Trans. Ind. Informat.*, 2017, 13(4): 1629-1640
- [103] Chang W, Chakraborty S. Resource-aware Automotive Control Systems Design: A Cyber-Physical Systems Approach. *Foundations and Trends in Electronic Design Automation*, 2016, 10(4): 249-369
- [104] Garcia-Valls M. A proposal for cost-effective server usage in CPS in the presence of dynamic client requests. In: Proc. of 19th IEEE Int. Symp. Real-Time Distributed Computing (ISORC). York, 2016, 19-26
- [105] Huang J, Li R, An J, et al. Energy-Efficient Resource Utilization for Heterogeneous Embedded Computing Systems. *IEEE Trans. Comput.*, 2017, 66(9): 1518-1531
- [106] Zeng X, Liu Z, Hui Q. Energy Equipartition Stabilization and Cascading Resilience Optimization for Geospatially Distributed Cyber-Physical Network Systems. *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2015, 45(1): 25-43
- [107] Khaitan S K, McCalley J D. Design Techniques and Applications of Cyber Physical Systems: A Survey. *IEEE Systems Journal*, 2015, 9(2): 350-365
- [108] Simmon E, Kim K -S, Subrahmanian E, et al. A vision of cyber-physical cloud computing for smart networked systems. NIST Interagency/Internal Report, NISTIR-7951, 2013
- [109] Gai K, Qiu M, Zhao H, et al. Resource Management in Sustainable Cyber-Physical Systems Using Heterogeneous Cloud Computing. *IEEE Trans. Sustain. Comput.*, 2018, 3(2): 60-72
- [110] Fu T Z J, Ding J, Ma R T B, et al. DRS: Auto-Scaling for Real-Time Stream Analytics. *IEEE/ACM Trans. Networking*, 2017, 25(6): 3338-3352
- [111] Chaâri R, Ellouze F, Koubâ A, et al. Cyber-physical systems clouds: A survey. *Computer Networks*, 2016, 108: 260-278

- [112] Molina E, Jacob E. Software-defined networking in cyber-physical systems: A survey. *Computers and Electrical Engineering*, 2018, 66(2): 407-419
- [113] Lien S -Y. Resource-Optimal Heterogeneous Machine-to-Machine Communications in Defined Networking Cyber-Physical Systems. *Wireless Pers. Commun.*, 2015, 84(3): 2215-2239
- [114] Pradhan S, Otte W, Dubey A, et al. Towards a Self-adaptive Deployment and Configuration Infrastructure for Cyber-Physical Systems. Institute for Software Integrated Systems, Vanderbilt University, Nashville, USA, Technical Report, ISIS-14-102, 2014
- [115] Rieger C G, Gertman D I, McQueen M A. Resilient Control Systems: Next Generation Design Research. In: *Proc. of 2nd Conf. Human System Interactions (HSI)*. Catania, 2009, 632-636
- [116] Lakshmanany K, de Niz D, Rajkumary R, et al. Resource Allocation in Distributed Mixed-Criticality Cyber-Physical Systems. In: *Proc. of 30th IEEE Int. Conf. Distributed Computing Systems (ICDCS)*. Genova, 2010, 169-178
- [117] Woo H, Yi J, Browne J C, et al. Design and Development Methodology for Resilient Cyber-Physical Systems. In: *Proc. of 28th Int. Conf. Distributed Computing Systems Workshops (ICDCS)*. Beijing, 2008, 525-528
- [118] Pereira A, Rodrigues N, Barbosa J, et al. Trust and Risk Management Towards Resilient Large-scale Cyber-Physical Systems. In: *Proc. of 2013 IEEE Int. Symp. Industrial Electronics (ISIE)*. Taipei, 2013, 1-6
- [119] Davis R I, Burns A, Bril R J, et al. Controller area network (can) schedulability analysis: Refuted, revisited and revised. *Real-Time Syst.*, 2007, 35(3): 239-272
- [120] Davis R I, Navet N. Controller area network (can) schedulability analysis for messages with arbitrary deadlines in fifo and work-conserving queues. In: *Proc. of 9th IEEE Int. Workshop on Factory Comm. Syst. (WFCS)*. Lemgo, 2012, 33-42
- [121] Davis R, Kollmann S, Pollex V, et al. Schedulability analysis for controller area network (can) with fifo queues priority queues and gateways. *Real-Time Syst.*, 2013, 49(1): 73-116
- [122] Schmidt K W, Alkan B, Schmidt E G, et al. Controller area network with priority queues and fifo queues: improved schedulability analysis and message set extension. *Int. J. Vehicle Design*, 2016, 71(1-4): 335-357
- [123] Xie G, Zeng G, Kurachi R, et al. WCRT analysis of can messages in gateway-integrated in-vehicle networks. *IEEE Trans. Vehicular Tech.*, 2017, 66(11): 9623-9637

- [124] Yomsi P M, Bertrand D, Navet N, et al. Controller area network (can): Response time analysis with offsets. In: Proc. of 9th IEEE Int. Workshop on Factory Comm. Syst. (WFCS). Lemgo, 2012, 43-52
- [125] Szakaly A. Response time analysis with offsets for can: [Master's thesis]. Goteborg: Department of Computer Eng., Chalmers Univ. Tech., 2003, 1-31
- [126] Mubeen S, Maki-Turja J, Sjodin M. Worst-case response-time analysis for mixed messages with offsets in controller area network. In: Proc. of 17th IEEE Int. Conf. Emerging Technology and Factory Automation (ETFA). Krakow, 2012, 1-10
- [127] Chen Y, Kurachi R, Takada H, et al. Schedulability comparison for can message with offset: Priority queue versus fifo queue. In: Proc. of 19th Int. Conf. on Real-Time and Network Syst. (RTNS). Nantes, 2011, 181-192
- [128] Mubeen S, Maki-Turja J, Sjodin M. Extending offset-based response-time analysis for mixed messages in controller area network. In: Proc. of 18th IEEE Int. Conf. Emerging Technology and Factory Automation (ETFA). Cagliari, 2013, 1-10
- [129] Mubeen S, Maki-Turja J, Sjodin M. Response time analysis for mixed messages in can supporting transmission abort requests. In: Proc. of 7th IEEE Int. Symp. on Industrial Embedded Syst. (SIES). Karlsruhe, 2012, 291-294
- [130] Lehoczky J P. Real-time queueing theory. In: Proc. of 17th IEEE Real-Time Systems Symposium (RTSS). Washington, 1996, 186-195
- [131] Kim J, Shin K G. Execution time analysis of communicating tasks in distributed systems. *IEEE Trans. Comput.*, 1996, 45(5): 572-579
- [132] Navet N, Song Y Q, Simonot F. Worst-case deadline failure probability in real-time applications distributed over can (controller area network). *J. Syst. Architecture*, 2000, 46(7): 607-617
- [133] Nolte T, Hansson H, Norstrom C. Probabilistic worst-case response-time analysis for the controller area network. In: Proc. of 9th IEEE Real-Time and Embedded Tech. Appl. Symp. (RTAS). Toronto, 2003, 200-207
- [134] Broster I, Burns A, Rodriguez-Navas G. Probabilistic analysis of can with faults. In: Proc. of 23rd IEEE Real-Time Systems Symposium (RTSS). Austin, 2002, 269-278
- [135] Manolache S, Eles P, Peng Z. Schedulability analysis of applications with stochastic task execution times. *ACM Trans. Embed. Comput. Syst.*, 2004, 3(4): 706-735
- [136] Diaz J L, Garcia D F, Kim K, et al. Stochastic analysis of periodic real-time systems. In: Proc. of 23rd IEEE Real-Time Systems Symposium (RTSS). Austin, 2002, 289-300

- [137] Lopez J M, Diaz J L, Entrialgo J, et al. Stochastic analysis of real-time systems under preemptive priority-driven scheduling. *Real-Time Syst.*, 2008, 40(2): 180-207
- [138] Mary G I, Alex Z C, Jenkins L. Response time analysis of messages in controller area network: A review. *J. Comput. Netw. and Comm.*, 2013, 2013: 1-11
- [139] Kumar P, Goswami D, Chakraborty S, et al. A hybrid approach to cyber-physical systems verification. In: *Proc. of 49th Annual Design Automation Conference (DAC)*. San Francisco, 2012, 688-696
- [140] Yu M, Wang L, Chu T, et al. Stabilization of networked control systems with data packet dropout and network delays via switching system approach. In: *Proc. of 43rd IEEE Conference on Decision and Control (CDC)*. Nassau, 2004, 3539-3544
- [141] 包玥. FlexRay车载网络的调度与控制协同设计: [硕士学位论文]. 武汉: 武汉理工大学信息工程学院, 2014, 9-20
- [142] 华长春, 楚艳钢, 关新平. 基于主动丢包的无线网络控制系统反馈控制. 见: 中国自动化学会控制理论专业委员会第三十届中国控制会议论文集B卷. 中国烟台, 2011, 837-840
- [143] 陈晓杜. 基于时间片划分的MIMO网络控制系统建模与研究: [硕士学位论文]. 南京: 南京理工大学自动化学院, 2012, 57-68
- [144] Hamdaoui M, Ramanathan P. A dynamic priority assignment technique for streams with (m, k) -firm deadlines. *IEEE Transactions on Computers*, 1995, 44(12): 1443-1451
- [145] Poggi E, Song Y, Koubaa A, et al. Matrix-DBP for (m, k) -firm real-time guarantee. In: *Proc. of Real-Time and Embedded System*. Paris, 2003, 457-482
- [146] Quan G, Hu X. Enhanced fixed-priority scheduling with (m, k) -firm guarantee. In: *Proc. of 21st IEEE Real-Time Systems Symposium (RTSS)*. Orlando, 2000, 79-88
- [147] Ramanathan P. Overload management in real-time control applications using (m, k) -firm guarantee. *IEEE Transactions on Parallel and Distributed Systems*, 1999, 10(6): 549-559
- [148] Sztipanovits J, Koutsoukos X, Karsai G, et al. Toward a Science of Cyber- Physical System Integration. *Proceedings of the IEEE*, 2012, 100(1): 29-44
- [149] Haddad W M, Chellaboina V S. *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton: Princeton University Press, 2008, 135-206
- [150] Hirche S, Matiakis T, and Buss M. A distributed controller approach for delay-independent stability of networked control systems. *Automatica*, 2009, 45(8): 1828-1836

- [151] Stramigioli S, Secchi C, van der Schaft A J, et al. Sampled data systems passivity and discrete port-Hamiltonian systems. *IEEE Transactions on Robotics*, 2005, 21(4): 574-587
- [152] Koutsoukos X, Kottenstette N, Hall J, et al. A passivity approach for model-based compositional design of networked control systems. *ACM Transactions on Embedded Computing Systems*, 2013, 11(4): 1-31
- [153] Arcak M. Passivity as a design tool for group coordination. *IEEE Transactions on Automatic Control*, 2007, 52(8): 1380-1390
- [154] Bai H, Arcak M, Wen J T. Rigid body attitude coordination without inertial frame information. *Automatica*, 2008, 44(12): 3170-3175
- [155] Ihle I -A F, Arcak M, Fossen T I. Passivity-based designs for synchronized path-following. *Automatica*, 2007, 43(9): 1508-1518
- [156] Leblanc H, Eyisi E, Kottenstette N, et al. A passivity-based approach to deployment in multi-agent networks, In: *Proc. of 7th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*. Funchal, 2010, 53-62
- [157] Kottenstette N, Antsaklis P J. Stable digital control networks for continuous passive plants subject to delays and data dropouts. In: *Proc. of 46th IEEE Conference on Decision and Control (CDC)*. New Orleans, 2007, 4433-4440
- [158] Fettweis A. Wave digital filters: Theory and practice. *Proceedings of the IEEE*, 1986, 74(2): 270-327
- [159] Hill D J, Moylan P J. Stability results for nonlinear feedback systems. *Automatica*, 1977, 13(4): 377-382
- [160] Kottenstette N, Antsaklis P J. Control of multiple networked passive plants with delays and data dropouts. In: *Proc. of the American Control Conference (ACC)*. Seattle, 2008, 3126-3132
- [161] Kottenstette N, Hall J, Koutsoukos X, et al. Digital control of multiple discrete passive plants over networks. *International Journal Systems, Control and Communications*, 2011, 3(2): 194-228
- [162] Niemeyer G, Slotine J -J E. Telemanipulation with time delays. *International Journal of Robotics Research*, 2004, 23(9): 873-890
- [163] Kottenstette N, Chopra N. L_2^m -stable digital-control networks for multiple continuous passive plants. In: *Proc. of 1st IFAC Workshop on Estimation and Control of Networked Systems (NecSys)*. Venice, 2009, 120-125

- [164] Kottenstette N, Koutsoukos X, Hall J, et al. Passivity-based design of wireless networked control systems for robustness to time-varying delays. In: Proc. of Real-Time Systems Symposium (RTSS). Barcelona, 2008, 15-24
- [165] Kottenstette N, Karsai G, Sztipanovits J. A passivity-based framework for resilient cyber physical systems. In: Proc. of 2nd International Symposium on Resilient Control Systems (ISRCS). Idaho Falls, 2009, 43-50
- [166] 田乃硕, 岳德权. 拟生灭过程与矩阵几何解. 北京: 科学出版社, 2002, 16-82
- [167] 潘峰, 李位星, 高琪. 粒子群优化算法与多目标优化. 北京: 北京理工大学出版社, 2013, 137-192
- [168] Coello C A C, Lamont G B, Van Veldhuisen D A. Evolutionary algorithms for solving multi-objective problems (Second Edition). Alphen aan den Rijn: Kluwer Academic Publishers, 2007, 1-52
- [169] Engelbrecht A. 计算群体智能基础. 谭营译. 北京: 清华大学出版社, 2009, 27-84
- [170] Carlos A C C, Gregorio T P, Maximino S L. Handling Multiple Objectives With Particle Swarm Optimization. IEEE Trans. Evolut. Comput., 2004, 8(3): 256-279
- [171] Deng J L. Introduction to grey system theory. The Journal of Grey System, 1989, 1(1): 1-24
- [172] Zhao H, Zhao H, Guo S. Using gm(1,1) optimized by mfo with rolling mechanism to forecast the electricity consumption of inner Mongolia. Applied Sciences, 2016, 6(1): 1-18
- [173] Hsu L C. Using improved grey forecasting models to forecast the output of optoelectronics industry. Expert Systems with Applications, 2011, 38(11): 13879-13885
- [174] Lawler G F. Introduction to Stochastic Processes (Second Edition). Boca Raton: CRC Press, Taylor & Francis Group, LLC, 2006, 73-84
- [175] 曹万科. CAN协议车载网络若干关键理论研究: [博士学位论文]. 沈阳: 东北大学机械工程与自动化学院, 2008, 15-42
- [176] 关学峰. CAN总线网络的实时性能分析、改进和测试: [硕士学位论文]. 成都: 西华大学机械工程与自动化学院, 2008, 1-5
- [177] Fu Y, Bi J, Wu J, et al. A Dormant Multi-Controller Model for Software Defined Networking. China communications, 2014, 11(3): 45-55
- [178] ONF White Paper. Software-Defined Networking: The New Norm for Networks. <https://www.opennetworking.org/component/content/article/105-module-content/related-documents/840-white-paper>, 2012

- [179] Cena G, Valenzano A. On the Properties of the Flexible Time Division Multiple Access Technique. *IEEE Trans. Ind. Informat.*, 2006, 2(2): 86-94
- [180] Yadin M, Naor P. Queueing system with a removable service station. *J. Oper. Res. Soc.*, 1963, 14(4): 393-405
- [181] Heyman D P. Optimal Operating Policies for M/G/1 Queueing System. *Operations Research*, 1968, 16(2): 362-382
- [182] Chang Y C, Pearn W L. Optimal management for infinite capacity N-policy M/G/1 queue with a removable service station. *Int. J. Syst. Sci.*, 2011, 42(7): 1075-1083
- [183] Zhang Z G, Tian N. Analysis on queueing systems with synchronous vacations of partial servers. *Perform. Evaluation*, 2003, 52(4): 269-282
- [184] Xu X, Zhang Z G. Analysis of multiple-server queue with a single vacation (e, d)-policy. *Perform. Evaluation*, 2006, 63(8): 825-838
- [185] Ke J -C, Lin C -H, Yang J -Y, et al. Optimal (d, c) vacation policy for a finite buffer M/M/c queue with unreliable servers and repairs. *Appl. Math. Model.*, 2009, 33(10): 3949-3962
- [186] Servi L D, Finn S G. M/M/1 queues with working vacations (M/M/1/WV). *Perform. Evaluation*, 2002, 50(1): 41-52
- [187] Zhang Z -J, Xu X. Analysis for the M/M/1 queue with multiple working vacations and N-policy. *Int. J. Inform. Management Sci.*, 2008, 19(3): 495-506
- [188] Lin C -H, Ke J -C. Multi-server system with single working vacation. *Appl. Math. Model.*, 2009, 33(7): 2967-2977
- [189] Yang D -Y, Wu C -H. Cost-minimization Analysis of a Working Vacation Queue with N-policy and Server Breakdowns. *Comput. Ind. Eng.*, 2015, 82: 151-158
- [190] Jain M, Upadhyaya S. Synchronous working vacation policy for finite-buffer multiserver queueing system. *Appl. Math. Comput.*, 2011, 217(24): 9916-9932
- [191] Lehoczky J P. Real-Time Queueing Network Theory. In: *Proc. of 18th IEEE Real-Time Systems Symposium*. San Francisco, 1997, 58-67
- [192] Neuts M F. *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Baltimore: Johns Hopkins University Press, 1981, 41-132
- [193] White J A, Schmidt J W, Bennett G K. *Analysis of Queueing Systems*. New York: Academic Press, Inc., Harcourt Brace Jovanovich Publishers, 1975, 1-102
- [194] Park I, Sunwoo M. FlexRay Network Parameter Optimization Method for Automotive Applications. *IEEE Trans. Ind. Electron.*, 2011, 58(4): 1449-1459

- [195] Guo P. A Fast Newton-Shamanskii Iteration for a Matrix Equation Arising from M/G/1-Type Markov Chains. *Mathematical Problems in Engineering*, 2017, 2017: 1-8
- [196] Okamura H, Dohi T, Trivedi K S. A refined em algorithm for ph distributions. *Perform. Evaluation*, 2011, 68(10): 938-954
- [197] Okamura H, Dohi T. PH fitting algorithm and its application to reliability engineering. *Journal of the Operations Research Society of Japan*, 2016, 59(1): 72-109
- [198] Allen A O. *Probability, Statistics, and Queueing Theory With Computer Science Applications (Second Edition)*. San Diego: Academic Press, INC, 1990, 9-85
- [199] Wen J C, Huang K H, Wen K L. The study of alpha in gm(1,1) model, *Journal of the Chinese Institute of Engineering*, 2000, 23(5): 583-589
- [200] Hung K C, Chien C Y, Wu K J, et al. Optimal alpha level setting in gm(1,1) model based on genetic algorithm. *Journal of Grey System*, 2009, 12(1): 23-31
- [201] Wang C H, Hsu L C. Using genetic algorithms grey theory to forecast high technology industrial output. *Applied Mathematics and Computation*, 2008, 195(1): 256-263
- [202] Martin C A, Witt S F. Forecasting tourism demand: A comparison of the accuracy of several quantitative methods. *International Journal of Forecasting*, 1989, 5(1): 7-19

致 谢

本文是在我的导师李仁发教授的指导下完成的。李老师对本人的研究工作给予了热情支持，从论文的选题、写作、初稿修改直至最后定稿，都倾注了李老师大量的精力和心血。李老师看待问题的思路及研究视野始终站在国际最前端，在学业上给予我的精心指导，使我少走了许多弯路，是我顺利完成论文的研究以及撰写工作的坚强支柱，他渊博的学识、孜孜不倦奋发向上的精神让我深感敬佩，他严谨的治学态度、实事求是、一丝不苟、勤勤恳恳的工作作风对我起到了表率作用。在此，谨向我的导师表示衷心的感谢！

在这里我要感谢嵌入式与网络计算湖南省重点实验室在我求学生涯中给予我的良好的研究条件。感谢国家“千人计划”特聘教授李克勤教授，感谢他对我论文的撰写和投稿给予的悉心指导，衷心感谢他对我诸多鼓励和帮助！感谢实验室的安吉尧、刘彦、陈巍巍、付彬、谢国琪老师，感谢他们给予了我诸多的帮助！我还要感谢杨帆、刘尧、刘丽、高楠、陈伟宏、吴武飞、黄晶、李万里、周佳、白洋、黄一智等师兄弟，感谢他们一直以来的陪伴、关心和支持！我还要感谢湖南大学信息科学与工程学院的各位老师，感谢他们在我学习期间的辛勤工作和默默付出，为我营造了一个良好的学习环境。

特别感谢我的妻子许波女士，在我四年的学习和研究工作中，在精神上给予我热情的鼓励，在生活上对我细致入微地照顾，也给予了我奋发向上的动力和源泉。

我在此向所有支持和帮助我的师长、同学、亲友以及朋友致以最诚挚的谢意！

最后，感谢所有的评审专家及答辩老师百忙之中对本文的审阅和指导。

附录 A 攻读学位期间所发表的学术论文

- [1] Hongfang Gong, Renfa Li, Jiyao An, Weiwei Chen, and Keqin Li. Scheduling algorithms of flat semidormant multicontrollers for a cyber-physical system. *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1665-1680, Aug. 2017
- [2] Hongfang Gong, Renfa Li, Jiyao An, Yang Bai, and Keqin Li. Quantitative Modeling and Analytical Calculation of Anelasticity for a Cyber-Physical System. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, 2018
- [3] Hongfang Gong, Renfa Li, Yang Bai, Jiyao An, and Keqin Li. Message Response Time Analysis for Automotive Cyber-Physical Systems with Uncertain Delay: An M/PH/1 Queue Approach. *Performance Evaluation*, vol. 125, pp. 21-47, Sept. 2018

附录 B 攻读学位期间所参与的科研课题

- [1] 国家自然科学基金项目，61173036，以汽车为例的 CPS 若干问题研究，2012/01-2015/12，已结题，参与
- [2] 国家自然科学基金项目，61370097，一类 CPS 中基于模糊理论的系统建模方法研究，2014/01-2017/12，已结题，参与
- [3] 国家自然科学基金项目，61370095，异构并行环境下的 MapReduce 资源调度模型与方法研究，2014/01-2017/12，已结题，参与
- [4] 国家重点研发计划项目，2016YFB0200405，体系结构感知的程序优化方法研究，2016/07-2018/06，在研，参与
- [5] 国家自然科学基金项目，61672217，新一代汽车嵌入式系统功能安全的建模与算法研究，2017/01-2020/12，在研，参与
- [6] 国家自然科学基金项目，61702172，基于AUTOSAR 新平台标准的汽车CPS自适应安全调度，2018/01-2020/12，在研，参与