

离散时间闸门式多级门限服务的 两级优先级轮询排队系统分析

木文浩,保利勇,丁洪伟,赵一帆

(云南大学信息学院,云南昆明 650500)

摘 要: 本文提出了一种基于闸门式多级门限服务的两级优先级轮询系统. 论文先构建了系统的排队分析模型,然后把模型中各个站点内的等待顾客数合起来设定为系统状态变量,对系统状态变量的概率母函数求二阶偏导后组成迭代方程组,联立求解后得到了站点的平均排队队长、顾客平均等待时间以及服务器查询周期等关键指标的完整数学解析表达式. 最后,对该模型进行了仿真实验分析,实验结果与理论结果相一致. 系统性能分析表明该轮询系统满足了周期性系统服务资源分配过程中业务多样性和弹性服务的发展需求,使得轮询控制策略应用方面更为广泛.

关键词: 轮询; 区分优先级业务; 闸门服务; 平均排队队长; 平均等待时间

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2018)02-0276-05

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.02.003

An Exact Analysis of Discrete Time Two-Level Priority Polling System Based on Multi-times Gated Service Policy

MU Wen-hao, BAO Li-yong, DING Hong-wei, ZHAO Yi-fan

(School of Information Science and Engineering, Yunnan University, Kunming, Yunnan 650500, China)

Abstract: This paper proposed a new system of two-class priority polling based on multi-times gated service policy. A queuing model of the polling system is constructed, in which the number of waiting customers in each site is set together as random variables of the system state. Moreover, iterative equations are set up by taking second-order partial derivatives of the generating function, and then the integral mathematical analytical expressions of the key indexes are derived via simultaneous solution, concerning the mean queue length, the mean waiting time of the key queue and common queues, and the mean polling period time of the server. Finally, computer simulations are conducted on the system and it is found that theoretical and simulation results are identical. The analysis shows that this new polling system satisfies the demand of service diversity and flexible service in the process of resource allocation in a periodic system, and demonstrates that the polling control protocol has wider applications.

Key words: polling; priority services; multi-times gated service; the mean queue length; the mean waiting time

1 引言

轮询在通信网络、交通和工业管理等领域有着广泛应用^[1]. 作为一种周期性的资源动态调度分配方式,轮询系统在系统控制中具有周期性和无冲突调度的特点. 随着高速移动通信网络^[2]、智能交通^[3]、工控网络和物联网等技术的发展,系统服务除了应对终端数目剧增、传送数据量急剧增大、区分业务多样性和实时性保障等变化外,更强调了系统整体服务能力的可迁移、

可伸缩和对不同个体的弹性服务需求^[4].

依据轮询系统的工作机理,轮询系统由相互衔接的顾客到达过程、服务器的查询服务过程和完成服务后的转移过程组成. 从数理建模过程分析,轮询系统的数学模型就由上述三个随机过程构成一个 N 维随机变量的复杂系统. 因为求解系统主要性能指标过程中均包含着对 N 维随机变量的高阶求偏导项,所以轮询系统的建模和求解一直以来都有很大难度和复杂度. 由于基于竞争的服务资源随机调度方式在高负载服务情

况下容易产生拥塞,而轮询方式具有的周期特性所提供的非竞争访问控制机制不仅能提高服务资源在高负载情况下的稳定性,同时也兼具更好的时延特性.因此,对轮询的深入分析和应用推广一直是研究的热点^[5-7].随着网络中业务种类增多和服务多样性的需求,对轮询的研究开始由非优先级到区分优先级业务^[8-10]、由单一服务方式到混合服务方式^[11]、固定站点数目到动态调整站点数目^[12,13]和由单级服务到多级服务的转变^[14].其中,文献[9]提出门限-完全两级优先轮询模型,并求解出该系统平均等待时间理论值,为基于业务优先级的进一步研究奠定了坚实基础.文献[14]中赵东风教授系统深入分析了次门限服务轮询系统特性,改变了一直以来轮询系统静态的服务方式,以动态调整服务次数的方式拓展了轮询系统的灵活性.

针对泛在网络中业务多样性和弹性服务的综合需求^[15],本文通过结合 M 级门限的闸门式调控服务和两级优先轮询系统能够区分业务优先级的特点提出了一种新的两级优先级轮询系统.论文基于嵌入式 Markov 链和概率母函数的分析方法,对系统状态变量求二阶偏导后联立求解方程组,解析得到了平均排队队长、平均查询周期和平均等待时间的理论表达式,并通过计算机仿真实验与理论计算值相互验证.

2 系统模型

2.1 模型定义

该系统是由普通队列和中心队列组成的两层架构的排队模型,包括 N 个普通站点和 1 个中心站点.当服务器对 i 号普通站点服务完成后转移到中心站点进行查询处理,服务完成后再接着对 $i+1$ 号普通站点进行查询处理,服务完后继续转移查询中心站点.按照上述方式,不断循环往复工作下去.在服务过程中对普通站点内的顾客采用 M 次门限的闸门式服务策略进行处理,对中心站点内的顾客采用完全服务策略进行处理.

2.2 系统数学模型构建

根据排队模型的工作机理,设定站点内顾客数、转移时间和服务时间为系统的随机变量,其概率母函数分别为 $A(z)$ 、 $R(z)$ 和 $B(z)$,系统的工作条件以及概率分布的定义与文献[9]相同.

假设 t_n 时刻服务器到达第 i 号普通站点开始服务,此时第 j 号普通站点内等待传输的顾客数设为 $\varepsilon_j(n)$,中心站点内等待传输的顾客数设为 $\varepsilon_h(n)$,此刻系统状态变量设为 $\{\varepsilon_1(n), \varepsilon_2(n), \dots, \varepsilon_N(n), \varepsilon_h(n)\}$,接着在 t_n 时刻中心站点 h 开始接受服务,此时第 j 号普通站点内等待传输的顾客数设为 $\varepsilon_j(n^*)$,中心站点内等待传输的分组数为 $\varepsilon_h(n^*)$,系统状态变量为: $\{\varepsilon_1(n^*), \varepsilon_2(n^*), \dots, \varepsilon_N(n^*), \varepsilon_h(n^*)\}$.随后,在 t_{n+1} 时刻系统开

始为第 $i+1$ 号普通站点进行服务,这时系统的状态变量为: $\{\varepsilon_1(n+1), \varepsilon_2(n+1), \dots, \varepsilon_N(n+1), \varepsilon_h(n+1)\}$.

2.3 系统的状态转移方程和系统状态变量的概率母函数

定义 $u_i(n)$ 为服务器从 i 号站点转向中心站点的查询转换时间; $v_i(n)$ 为服务器对 i 号站点进行服务的时间; $v_i'(n)$ 为服务器对号站点进行第 M 次门限服务的时间; $v_h^*(n)$ 为服务器对中心站点进行服务的时间.根据站点内顾客数的变化规律,列出以下系统的状态转移方程:

$$\begin{cases} \varepsilon_h(n^*) = \mu_h(u_i) + \eta_h(v_i) \\ \varepsilon_i(n^*) = \mu_i(u_i) + \eta_i(v_i') \\ \varepsilon_j(n^*) = \mu_j(u_i) + \eta_j(v_i) + \varepsilon_j(n+1), j \neq i \\ \varepsilon_k(n+1) = \varepsilon_k(n^*) + \eta_k(v_h^*) \\ \varepsilon_h(n+1) = 0 \end{cases}$$

其中, $\mu_h(u_i)$ 为在 $u_i(n)$ 时间内进入中心站点的顾客数; $\eta_h(v_i)$ 为在 $v_i(n)$ 时间内进入中心站点的顾客数; $\mu_j(u_i)$ 为在 $u_i(n)$ 时间内进入第 j 号站点的顾客数; $\eta_j(v_i)$ 为在 $v_i(n)$ 时间内进入第 j 号普通站点的顾客数; $\mu_i(u_i)$ 为在 $u_i(n)$ 时间内进入第 i 号站点的顾客数; $\eta_i(v_i')$ 为在 $v_i'(n)$ 时间内进入第 i 号普通站点的顾客数; $\eta_k(v_h^*)$ 为在 $v_h(n^*)$ 时间内进入第 k 号普通站点的顾客数.

该轮询系统的状态变化可以用马尔科夫链来描述,在 $\rho_h + N\rho < 1$ ($\rho_h = \lambda_h \beta_h, \rho = \lambda \beta$) 的稳定条件下,服务器在 t_n 时刻服务中心站点时,此刻系统状态变量的概率母函数为:

$$\begin{aligned} G_{ih} &= \lim_{t \rightarrow \infty} E \left[\prod_{j=1}^N z_j^{\varepsilon_j(n^*)} z_h^{\varepsilon_h(n^*)} \right] \\ &= R_i \left[\prod_{j=1}^N A_j(z_j) A_h(z_h) \right] G_i \{z_1, z_2, \dots, \\ &B_i [F_m(A_1(z_1), A_2(z_2), \dots, A_N(z_N), A_h(z_h))] \\ &\prod_{j=1, j \neq i}^N A_j(z_j) A_h(z_h), \dots, z_N, z_h \} \end{aligned} \quad (1)$$

服务器在 t_{n+1} 时刻服务第 $i+1$ 号普通站点时,此刻系统状态变量的概率母函数为:

$$\begin{aligned} G_{i+1} &= \lim_{t \rightarrow \infty} E \left[\prod_{j=1}^N z_j^{\varepsilon_j \langle n+1 \rangle} z_h^{\varepsilon_h \langle n+1 \rangle} \right] \\ &= G_{ih} \left\{ z_1, z_2, \dots, z_N, B_i \left[\prod_{j=1}^N A_j(z_j) F \left(\prod_{j=1}^N A_j(z_j) \right) \right] \right\} \end{aligned} \quad (2)$$

其中,由文献[9,14]可得:

$$\begin{aligned} &F_m(z_1, z_2, \dots, z_N, z_h) \\ &= \begin{cases} A_j \left\{ B_h \left[F_{m-1}(z_1, z_2, \dots, z_N, z_h) \prod_{j=1}^N z_j \cdot z_h \right] \right\}, & m > 1 \\ z_i, & m = 1 \end{cases} \end{aligned}$$

$$F(z_h) = A_h \{B_h [z_h F(z_h) + 1]\}$$

3 系统解析

3.1 平均排队队长

在 t_n 时刻,服务器对第 i 号普通站点进行服务时,平均排队队长的定义与文献[9]相同,由式(1)和式(2)求导计算可得普通站点的平均排队队长为:

$$g_i = \frac{N\lambda\gamma(1-\rho)}{(1-\rho_h - N\rho)(1-\rho^m)}, \quad i=1,2,\dots,N \quad (3)$$

对中心站点来说,同样通过求导计算得到中心站点的平均排队队长为:

$$g_{ih} = \frac{\lambda_h\gamma(1-\rho_h)}{1-\rho_h - N\rho}, \quad i=1,2,\dots,N \quad (4)$$

3.2 平均查询周期

系统的查询周期由对普通站点的服务时间、对中心站点的服务时间和轮询转移时间三部分组成。根据平均排队队长与平均循环周期的关系得该系统的平均循环周期 $E(\theta)$:

$$E(\theta) = \frac{N\gamma}{1-\rho_h - N\rho} \quad (5)$$

3.3 顾客的平均等待时间

顾客的等待时间是指顾客进入站点到其开始被发送的这段时间.由系统普通站点的 M 次门限服务的传输控制机理得到平均等待时间 $\omega_i(n)$ 为:

$$E(\omega_i) = \frac{(1+\rho^m)g_i(i,i)}{2\lambda g_i(i)} - \frac{(1+\rho^m)A''(1)}{2\lambda^2(1+\rho)} + \frac{(1-\rho^{m-1})\lambda B''(1)}{2(1-\rho)} \quad (6)$$

同理,由系统中心站点的完全服务轮询系统的传输控制机理得到平均等待时间 $\omega_h(n)$ 为:

$$E(\omega_h) = \frac{g_{ih}(h,h)}{2\lambda_h g_{ih}(h)} - \frac{A''_h(1)}{2\lambda_h^2(1+\rho_h)} + \frac{\lambda_h B''_h(1)}{2(1-\rho_h)} \quad (7)$$

式(6)和式(7)中的二阶导 $g_i(i,i)$ 和 $g_{ih}(h,h)$ 可由公式(1)和公式(2)求二阶偏导并联立求解得到,其结果为:

$$g_i(i,i) = \frac{N^2\lambda^2(1-\rho)^2}{(1-\rho_h - N\rho)[(1-\rho_h - \rho)(1-\rho^{2m}) + 2\rho\rho_h - 2\rho^{m+1}\rho_h]} \cdot \{ R''(1) + (1-\rho_h)A''(1) + \frac{(1-\rho_h - \rho + 2\rho\rho_h)(1-\rho^{m-1})\gamma A''(1)}{(1-\rho)^2} - \frac{2\rho\rho_h(1-\rho^{m-1})\lambda^2\gamma}{1-\rho} + \frac{1}{(1-\rho_h - N\rho)} \{ \lambda^2\beta_h^2 A''_h(1) + \lambda^2\lambda_h^2 B''_h(1) + (N-1)\lambda^2\gamma\rho(1-\rho_h) + 2\rho_h\lambda^2\gamma(1-\rho_h) + (1-\rho_h)\rho\gamma A''(1) + [(N-1)(1-\rho_h) + 2N\rho(\rho - \rho_h\rho^{m-1} + \rho^{m-1})]\lambda^2\gamma^2 + \frac{N(1-\rho)^2 - (1-\rho)^m(1+\rho_h - \rho)(1+\rho_h - N\rho)\lambda^3\gamma B''(1)}{(1-\rho)^2} \} \} \quad (8)$$

$$g_{ih}(h,h) = \lambda_h^2 R''(1) + \gamma A''_h(1) + \{ 2\lambda_h^2\gamma\beta[1+F_m'(k)] + \lambda_h^2 B''(1)[1+F_m'(h)]^2 + \beta[A''_h(1)(1+F_m'(h)) + \lambda_h^2 F''_m(h,h)] \} g_i(i) + \lambda_h^2\beta^2[1+F_m'(h)]^2 g_i(i,i) \quad (9)$$

4 系统性能分析:

论文采用 MatlabR2008a 平台按照轮询系统的控制机理搭建了系统仿真实验统计分析环境.从结果分析可以看出,本文的系统模型理论计算值与实验值具有较好的一致性,系统性能分析结果如下:

(1)当 $M=1$ 时,论文所得平均排队队长以及平均等待时间公式与文献[9]门限-完全服务系统中的公式相一致,即此时该模型为门限-完全服务系统;当 $M \rightarrow \infty$ 时,文章所得平均排队队长以及平均等待时间公式与文献[16]完全-完全服务系统中的公式相一致,即此时该模型为完全-完全服务系统.

(2)依据图1可分析出,系统平均查询周期随着系统到达率的增大而增大,并且与普通站点的服务次数无关,这与公式(5)表达式所示相一致.

普通站点M级门限不同到达率下系统平均循环周期仿真与数值计算对比

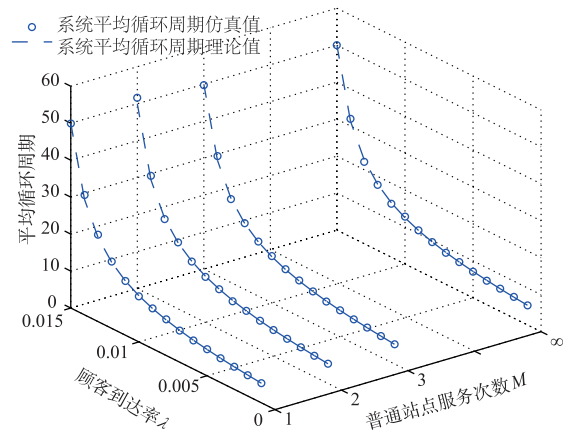


图1 平均循环周期随到达率 λ 变化曲线 ($\beta=\beta_h=10, \gamma=1$)

(3)根据图2所示结果,平均排队队长随顾客到达率增大而增大,且普通站点的平均排队队长随 M 值的增大而减小,这与系统的物理机理和公式(3)相符合,且说明调整 M 的数值实际上是通过调整系统普通站点和中心站点对服务资源的占用率来实现对普通站点的弹性服务.

(4)如图3和图4所示,平均等待时间随顾客到达率的增大而增大,且中心站点的平均等待时间随 M 值的增大而变长. M 数值的增大导致中心站点得到的系统服务资源变少,而普通站点所得到的服务资源也就相应增多.

(5)如图5和图6所示,在队长和时延特性方面,中心站点都有着明显优势.随着站点数 N 的增加,各站点

M取不同值时普通站点在不同到达率下平均排队队长仿真与数值计算对比

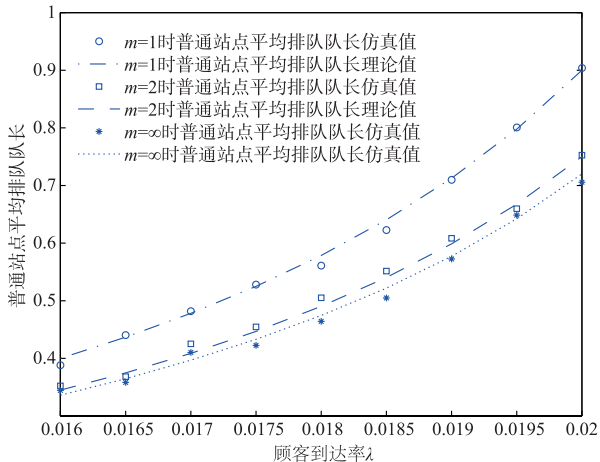


图2 普通站点平均排队队长随到达率λ变化曲线 ($\beta=\beta_n=10, \gamma=3$)

M取不同值时普通站点在不同到达率下平均等待时间仿真与数值计算对比

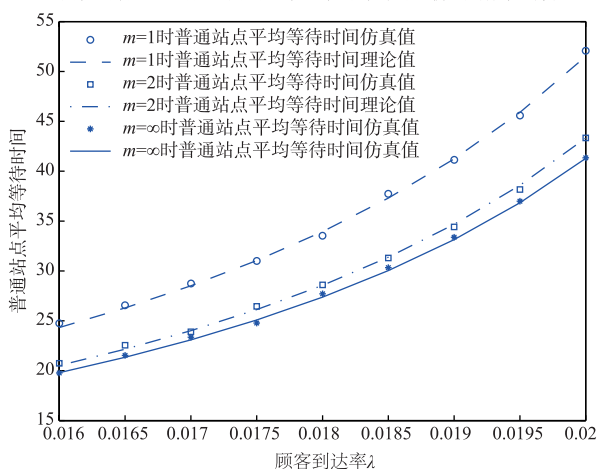


图3 普通站点平均等待时间随到达率λ变化曲线 ($\beta=\beta_n=10, \gamma=3$)

M取不同值时中心站点在不同到达率下平均等待时间仿真与数值计算对比

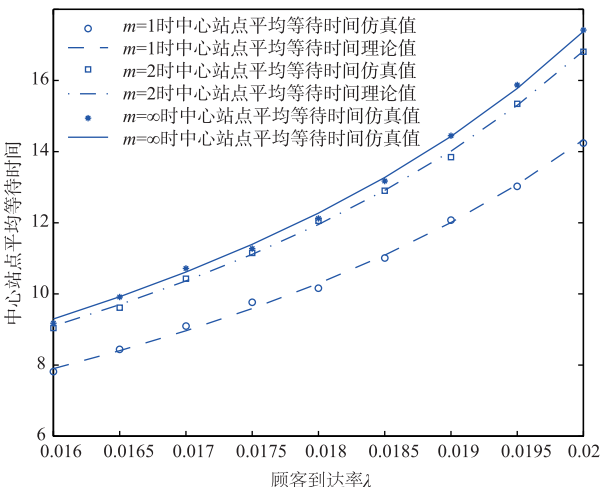


图4 中心站点平均等待时间随到达率λ变化曲线 ($\beta=\beta_n=10, \gamma=3$)

的排队队长和平均等待时间都在增大,这说明系统所

服务的站点数增多将会导致每个站点所分配到的服务资源也随之减小.

M取2普通站点数不同时系统平均排队队长仿真与数值计算对比

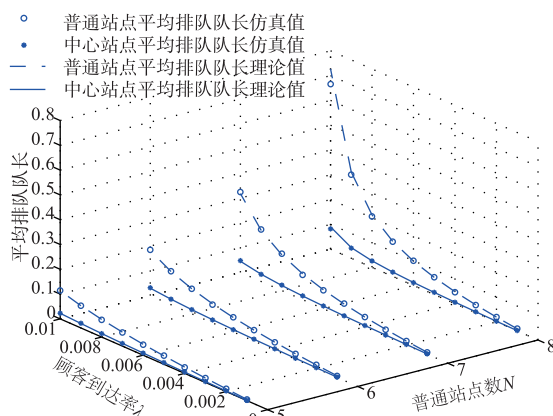


图5 不同站点数下平均排队队长随到达率λ变化曲线 ($\beta=\beta_n=10, \gamma=1$)

M取2普通站点数不同时系统平均等待时间仿真与数值计算对比

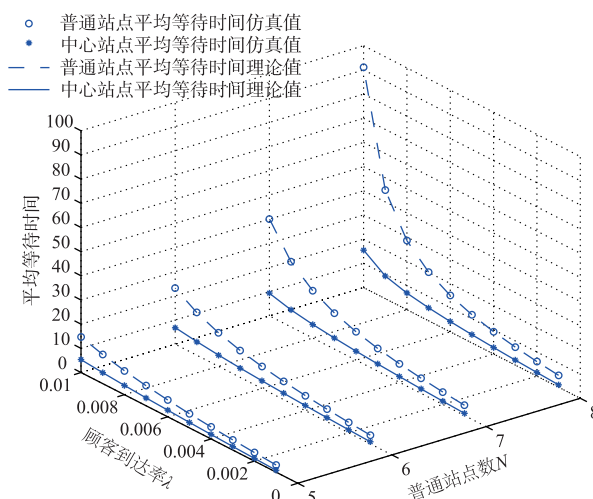


图6 不同站点数下平均等待时间随到达率λ变化曲线 ($\beta=\beta_n=10, \gamma=1$)

5 结论

在现有的两级优先级轮询系统中,对于中心站点和普通站点的服务方式往往是固定不可改变的,系统服务资源分配缺少灵活性.因此,本文根据泛在网络发展需求提出了普通站点采用可调节的M次门限服务和中心站点采用完全服务策略的两级优先级轮询控制策略.系统性能分析表明,在系统服务资源不变的情况下可以通过选取不同的M值实现了对普通站点服务方式的动态控制,从而调整高低优先级站点之间的服务资源分配能力,增强了系统的公平性和灵活性.另外,本文的研究诠释了轮询排队系统中门限-完全到完全-完全服务方式的完整演进过程,进一步完善了轮询系统研究的理论体系.

致谢

我们衷心感谢电子科技大学的查光明教授! 我们永远缅怀云南大学信息学院的赵东风教授! 我们诚挚感谢国家自然科学基金(No. 61072079)为研究提供的大力支持!

参考文献

- [1] O J Boxma, O Kella, K M Kosiński. Queue lengths and workloads in polling systems[J]. *Operations Research Letters*, 2011, 39(6): 401 – 405.
- [2] MA Al-Maqri, M Othman, BM Ali, ZM Hanapi. Adaptive multi-polling scheduler for QoS support of video transmission in IEEE 802.11e WLANs [J]. *Telecommunication Systems*, 2016, 61(4): 773 – 791
- [3] MY Wang, HW Ding, YF Zhao, ZG Liang, LZ Zhu. The application of the exhaustive polling theory in intelligent traffic system [A]. *Matec Web of Conferences* [C]. Les Ulis, France; EDP Sciences, 2016. 2013 – 2017.
- [4] 李正茂. 通信 4.0 [M]. 北京: 中信出版社, 2016.
- [5] 赵东风, 郑苏民. 周期查询式门限服务排队系统中信息分组的延迟分析 [J]. *通信学报*, 1994, 15(2): 18 – 23.
Zhao Dong-feng, Zheng Su-min. Message waiting time analysis for a polling system with gated service [J]. *Journal of China Institute of Communications*, 1994, 15(2): 18 – 23. (in Chinese)
- [6] 赵东风, 郑苏民. 查询式完全服务排队模型分析 [J]. *电子学报*, 1994, 22(5): 102 – 107.
Zhao Dong-feng, Zheng Su-min. Analysis of a polling model with exhaustive service [J]. *Acta Electronica Sinica*, 1994, 22(5): 102 – 107. (in Chinese)
- [7] JPL Dorsman, SC Borst, OJ Boxma, M Vlasiou. Markovian polling systems with an application to wireless random – access networks [J]. *Performance Evaluation*, 2015, 85(86): 33 – 51.
- [8] MAA Boon, IJBF Adan, OJ Boxma. A polling model with multiple priority levels [J]. *Performance Evaluation*, 2010, 67(6): 468 – 484.
- [9] 杨志军, 赵东风, 丁洪伟, 赵一帆. 两级优先级控制轮询系统研究 [J]. *电子学报*, 2009, 37(7): 1452 – 1456.
YANG Zhi-jun, ZHAO Dong-feng, et al. Research on two class priority based polling system [J]. *Acta Electronica Sinica*, 2009, 37(7): 1452 – 1456. (in Chinese)
- [10] 杨志军, 丁洪伟, 陈传龙. 完全服务和门限服务两级轮询系统 $E(x)$ 特性分析 [J]. *电子学报*, 2014, 42(4): 774 – 778.
YANG Zhi-jun, DING Hong-wei, CHEN Chuan-long. Research on $E(x)$ characteristics of two-class polling system of exhaustive-gated service [J]. *Acta Electronica Sinica*, 2014, 42(4): 774 – 778. (in Chinese)
- [11] M Iftikhar, M Imran, et al. Performance analysis of mixed polling schemes with multiple classes of self-similar traffic input to build comprehensive SLAs [A]. *IEEE Conference on Local Computer Networks Workshops* [C]. Florida: IEEE Press, 2013. 994 – 1000.
- [12] JPL Dorsman, OJ Boxma, RDVD Mei. On two-queue Markovian polling systems with exhaustive service [J]. *Queueing Systems*, 2014, 78(4): 287 – 311.
- [13] 官铮, 杨志军, 何敏, 钱文华. 依托站点状态的两级轮询控制系统时延特性分析 [J]. *自动化学报*, 2016. 42(8): 1207 – 1214.
GUAN Zheng, YANG Zhi-jun, HE Min, QIAN Wen-hua. Study on the delay performance of station dependent two-level polling systems [J]. *Acta Automatica Sinica*, 2016, 42(8): 1207 – 1214. (in Chinese)
- [14] 赵东风, 丁洪伟, 赵一帆. 多级门限服务轮询系统 MAC 离散时间控制协议模型分析 [J]. *电子学报*, 2010, 38(7): 1495 – 1499.
ZHAO Dong-feng, DING Hong-wei, et al. An analytical model of a discrete-time polling MAC protocol for wireless LANs using m-gated services [J]. *Acta Electronica Sinica*, 2010, 38(7): 1495 – 1499. (in Chinese)
- [15] M. Claypool, D. Brown. Inferring user interest [J]. *IEEE Internet Computing*, 2015, 5(6): 32 – 39.
- [16] 梁竹关. 离散时间完全服务两级轮询系统理论研究 [D]. 昆明: 云南大学, 2010.

作者简介



穆文浩 男, 1990 年生于云南怒江. 云南大学硕士研究生. 研究方向为轮询多址接入和随机多址接入理论.
E-mail: 747432952@qq.com



保利勇(通信作者) 男, 1975 年生于云南楚雄. 云南大学信息学院副教授. 2011 年获得博士学位. 研究方向为通信网络 MAC 层多址接入理论、冲突分解和混沌扩频通信等.
E-mail: bly.yx@163.com

丁洪伟 男, 1964 年生于云南景洪. 云南大学信息学院教授, 博士生导师. 研究方向为轮询系统、随机多址通信系统.

赵一帆 男, 1986 年生于云南昆明. 博士研究生. 研究方向为轮询系统、随机多址通信系统、机器人通信等.