

# 5G切片架构下具有重传机制的轮询系统研究

杨志军<sup>1,2</sup>,寇倩兰<sup>1</sup>,丁洪伟<sup>1</sup>

(1.云南大学 信息学院,昆明 650500; 2.云南省教育厅教学仪器装备中心,昆明 650223)

**摘要:** 面对海量应用场景及客户需求,传统的轮询系统方案不能满足轮询系统中每个站点所需要的服务质量要求。为降低系统时延,适应更大的网络规模,提出基于5G网络切片的轮询方案。将5G网络切片划分为增强型移动带宽切片、海量机器通信切片和高可靠低时延通信切片,再将这3类切片各自划分出 $N$ 个更低层子切片,构建3个独立的具有重传机制的完全服务轮询模型。信息分组按照先进先出的规则进入各子切片,由通用服务器进行统一发送。在此基础上,采用概率母函数及马尔可夫链建立非理想信道环境下基于5G网络切片架构的轮询系统数学模型,推导系统平均排队队长和平均等待时间的精确表达式,进行信息分组出错概率和重传阈值与轮询系统性能之间的定量关系分析。基于Matlab的仿真实验结果证明了该模型的正确性,其较重传门限服务时延更少,能够为非理想信道环境下基于5G网络切片架构的轮询系统分析提供一种快速评估机制。

**关键词:** 5G网络切片;增强型移动带宽;海量机器通信;高可靠低时延通信;轮询;重传;完全服务

开放科学(资源服务)标志码(OSID):



中文引用格式:杨志军,寇倩兰,丁洪伟.5G切片架构下具有重传机制的轮询系统研究[J].计算机工程,2022,48(10):202-211.

英文引用格式:YANG Z J, KOU Q L, DING H W. Research on polling system with retransmission mechanism in 5G slicing architecture[J]. Computer Engineering, 2022, 48(10): 202-211.

## Research on Polling System with Retransmission Mechanism in 5G Slicing Architecture

YANG Zhijun<sup>1,2</sup>, KOU Qianlan<sup>1</sup>, DING Hongwei<sup>1</sup>

(1.School of Information Science and Engineering, Yunnan University, Kunming 650500, China;

2.Yunnan Province Educational Instruments and Facilities Service Center, Kunming 650223, China)

**[Abstract]** Facing massive application scenarios and customer needs, traditional polling system solutions cannot meet the Quality of Service(QoS) required by each site in the polling system. A polling scheme based on 5G Network Slicing(NS) is proposed to reduce system latency, to enable adaptation to a larger network scale, and meet the needs of different application scenarios. The 5G NSs are divided into enhanced Mobile Broadband(eMBB), massive Machine Type of Communication(mMTC), and ultra-Reliable and Low-Latency Communication(uRLLC) slices, and each of these three types of slices is further divided into  $N$  lower layer subslices to construct three independent exhaustive service polling models with failure retry mechanisms. The message grouping enters each subslice according to the first-in-first-out rule, and each subslice is sent uniformly by the universal server. On this basis, exact expressions for the average queue length and average waiting time of the system are derived, using Markov chain and probability mother function, to establish a quantitative relationship analysis between the error probability and retransmission threshold of message grouping while realizing an acceptable performance for the polling system. The accuracy of the model is proven by simulation experiments. This model has a lower latency than the retransmission-gated service. The analysis model provides a fast evaluation mechanism for the analysis of polling systems based on 5G NS architecture in a non-ideal channel environment.

**[Key words]** 5G Network Slicing(NS); enhanced Mobile Broadband(eMBB); massive Machine Type of Communication(mMTC); ultra-Reliable and Low-Latency Communication(uRLLC); polling; failure retry; exhaustive service

DOI:10.19678/j.issn.1000-3428.0062279

### 0 概述

为满足不同行业需求,5G移动网络必须具备更高

的数据速率和能源效率、更低的网络延迟以及可靠的泛在化连接,5G网络切片(Network Slicing, NS)概念由此被提出。5G网络切片就是将5G网络切出多张逻辑

基金项目:国家自然科学基金(61461054,61461053)。

作者简介:杨志军(1968—),男,研究员、博士,主研方向为通信与信息系统、通信网络;寇倩兰,硕士研究生;丁洪伟,教授、博士。

收稿日期:2021-08-05 修回日期:2021-10-01 E-mail:1141540373@qq.com

独立的专有网络,通过软件定义网络(Software Defined Network, SDN)和网络功能虚拟化(Network Function Virtualization, NFV)技术实现不同网络切片之间计算、存储和网络资源的统一管理和资源切分,进而支持更多业务,最终实现“万物互联”。网络切片技术优势在于运营商基于业务场景的需要,自主定义切片对应的网络特性,包括延迟、速率、连接密度、频谱效率、流量容量、网络效率等,通过网络配置的灵活性保证用户体验<sup>[1]</sup>。文献[2]在5G网络切片概念中对业务进行分类,给予不同优先级的业务不同的资源和不同的服务质量(Quality of Service, QoS),提供了一种较优和有效的方式来应对终端用户的需求和需求的多样性。轮询是一种公平的资源调度方式,5G网络的不同切片可以通过轮询方式实现资源共享。现阶段,针对网络中不同的业务需求,轮询系统出现了预约调度方式,即信息分组按照先进先出的规则进入各个站点排队并等待由服务器统一分配资源对其进行服务。轮询系统一般由一个或多个服务器及 $N$ 个站点组成<sup>[3]</sup>。

由于网络用户量的急剧增加以及用户需求灵活多变,逐渐对网络QoS提出了更高的要求。QoS关键技术指标包括吞吐量、排队队长、时延变化等<sup>[4]</sup>。文献[5]提出一种基于信道感知轮询的MAC协议,通过调整超帧中的轮询周期数来优化信道感知轮询协议的能源效率,以适应动态流量需求和信道波动,显著提高了能源效率,同时保持了较短的时延。文献[6]提出自适应动态轮询MAC协议,通过监测传入业务量变化的系数来切换节点轮询间隔分布,发现该协议流量到达和轮询间隔分布一致时,在时延和能量方面的性能都是最好的。文献[7]提出一种异步的由接收方发起的基于预载波的MAC协议。该协议用于高数据率的浅层水下监测应用,并在不使用任何传输时间表的情况下以节能的方式支持可靠性,解决了数据包碰撞问题,同时提供了良好的能源效率。上述文献通过设计MAC协议降低了系统能源消耗,同时提高了传输效率,但是都没有对传输过程中信息分组的丢失情况进行考虑。

除了设计MAC协议对系统进行优化外,还可以从资源配置角度出发或建立新的模型以实现更好的QoS。文献[8-10]通过调整系统服务资源调度策略并提出新的排队论模型,显著提高了系统吞吐量,降低了排队队长和信息传输时延。针对物联网中优先级业务和实时性的需求,文献[11-12]提出采用两级轮询服务规则,将用户划分为普通用户和优先级用户,同时将服务时间和切换时间并行处理来提高轮询系统利用率和服务效率,保证用户低时延传输。在5G时代,面对海量业务需求,为保证网络性能均衡性和公平性,文献[13]提出一种一队列两服务的轮询控制策略,用于无线传感器网络的媒体接入控制。通过将现有帧结构重新设计以实现

该轮询控制策略,进一步说明了该控制策略在实际网络中的可行性。虽然上述文献从不同角度对轮询系统展开了研究并且在一定程度改善了轮询系统性能,但模型也都是在理想信道基础上建立的,即假设任意信息分组在信道传输过程中每一次都能够成功发送,缺乏对在实际信道环境中因噪声、设备误差以及信道本身对信号的影响导致信息分组在传输过程出错情况的研究。

信道错误和碰撞常常导致无线网络中信息分组传输高概率失败,因此,关于采取重传方案和概率轮询协议减少信息传输延迟,提高信息分组交付率,扩大系统吞吐量的研究具有重大意义<sup>[14]</sup>。文献[15]对一个单服务器多站点的轮询系统进行研究,设置重传等待期,在重传等待期间到达站点的用户停留在该站点的队列中,并在该站点的下一个服务期得到服务。文献[16]研究一种具有动态重传和差异服务机制的时分多址(Time Division Multiple Access, TDMA)时隙分配协议,当数据传输失败时,利用重传时隙提高通信可靠性,通过影子时隙实现有区别服务。文献[17]针对RoF-DAS架构下WLAN差错重传机制与网络接入性能的定量关系进行分析,建立了具有内部数据的门限服务轮询排队模型,利用轮询系统中内部数据的到达和服务过程描述数据差错重传机制。上述研究虽然将信息分组重传纳入了系统模型运行的考虑范围,但针对海量应用场景及客户需求,传统的轮询系统方案已不能满足轮询系统中每个站点所需要的QoS。

为使分析模型更贴合实际网络信道工作环境,本文提出具有差错重传机制的完全服务轮询排队模型,利用5G网络切片架构,针对性地为不同用户提供不同的网络能力,从而满足不同业务场景对网络的需求,以此提高对用户的QoS。该模型以切片为站点,由通用服务器将各个基站中的剩余计算资源进行统一管理和分配。考虑到在传输过程中信道质量对信息分组的影响,设置重传机制以体现非理想信道环境。当信息分组传输出错时则返回源子切片重新排队等待重传,直到正确发送或达到重传阈值仍未发送成功被抛弃。同时,采用概率母函数<sup>[18]</sup>及马尔可夫链分析方法<sup>[19]</sup>建立非理想信道环境下基于5G网络切片架构的轮询系统数学模型,通过对该数学模型的精确解析推导出具有重传机制的完全服务轮询系统平均排队队长和平均等待时间的数学表达式。利用Matlab2019b进行实验仿真,将重传实验的仿真值与理论值对比,同时再与理想信道环境的基础轮询系统对比,验证模型研究的正确性,并进一步研究重传概率和重传阈值对系统性能的影响。此外,将该本文方案与具有重传机制的门限服务轮询策略和无重传机制的两级完全轮询方案进行对比。

## 1 轮询方案及系统工作机制

在5G移动网络中引入网络切片可为每个行业的应用例提供所需的QoS。网络切片是将一个物理网络划分为多个逻辑网络的过程,以支持特定应用场景的按需服务,同时使用同一物理网络。这些逻辑网络被暗指为网络切片<sup>[20]</sup>。如今,5G网络的应用场景大致可以分为增强型移动带宽(enhanced Mobile Broadband, eMBB)、海量机器通信(massive Machine Type of Communication, mMTC)以及高可靠低时延通信(ultra-Reliable and Low-Latency Communication, uRLLC)。这3类网络也称为5G网络下的3类切片。3种应用场景对网络服务的需求是不相同的,下面具体分析这3类5G切片:

1) eMBB。面向4K/8K超高清视频、全息技术、增强现实/虚拟现实等应用, eMBB对网络带宽和速率要求较高。针对多媒体和高性能数据和服务的访问做进一步改善<sup>[21]</sup>。

2) mMTC。由大量的连接设备组成,这些传感器设备是非常密集的,大部分是静止的,交换的数据量很小。这种类型的数据流量对延迟不敏感,由简单的低成本传感器等设备产生,功耗非常低。

3) uRLLC。任务关键性物联网主要应用于无人驾驶、自动工厂、智能电网等领域,主要需求是超低时延和高可靠性。

同时,这3类5G网络切片之间是相互独立的,不会彼此影响,但都是运用QoS来统一管理每个子网络内部不同业务,并且在这3类切片之下,还可以基于同一类子网络继续进行资源划分,形成更低一层子网络,即子切片。分别将eMBB、mMTC、uRLLC这3种场景的切片再划分出 $N$ 个更低层子切片。利用通用服务器将众多基站中的空闲资源统一整理并予以分配,对各场景中不同子切片内的信息分组进行服务,由此满足不同场景对QoS的要求,得到在5G架构下的轮询系统模型,如图1所示。

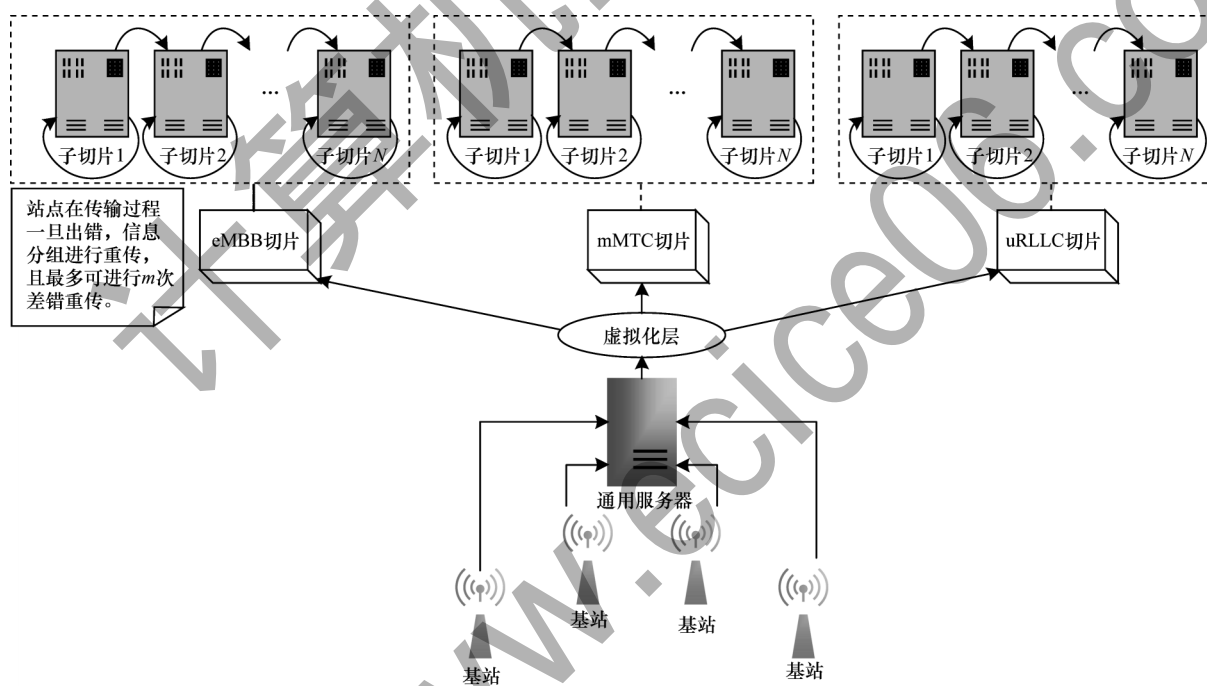


图1 5G网络切片架构下的轮询系统模型

Fig.1 Polling system model in 5G network slicing architecture

面对5G时代背景下部署的大量基站,必定存在某些时间部分基站的计算资源没有用尽,甚至处于空闲状态。因此,本文设立一个通用服务器,将各个基站的空闲运算资源集中起来,利用NFV技术在虚拟化层对这些资源进行统一调配。将eMBB切片、mMTC切片和uRLLC切片各自分别再划分出 $N$ 个更低层子切片作为站点。同时,eMBB、mMTC和uRLLC之间相互独立,即通用服务器可同时分别服务于3个独立的轮询系统,这3个系统之间不会相互影响。定义在

实际的5G网络切片通信信道中,信息分组在传输过程中出错的概率为 $P_e$ ,重传阈值为 $m$ 。在该模型下,完全服务策略具体是指:服务器在对某一子切片进行服务时,信息分组一旦出错就需要立即返回并加入到源子切片尾部准备重传;若重传了 $m$ 次依旧没有成功发送,则抛弃该信息分组,直到该子切片内所有的信息分组都被服务完毕,转向下一子切片进行服务,依次往后;当第 $N$ 个子切片内的信息分组也被全部发送时,再转向第一个子切片进入下一个轮询周期。



## 2 模型分析

### 2.1 系统工作环境

本文建立非理想信道下基于5G网络切片架构的完全服务轮询排队模型。假设该模型在以下环境下运行:

1) 信息分组进入各子切片并等待发送的过程相互独立且服从同样的概率分布,其分布的概率母函数为 $A(z)$ ,均值为 $\lambda=A'(1)$ ,方差为 $\sigma_\lambda^2=A''(1)+\lambda-\lambda^2$ 。

2) 子切片内的信息分组接受服务器的服务并发送出去所需要的时间相互独立且服从同样的概率分布,其分布的概率母函数为 $B(z)$ ,均值为 $\beta=B'(1)$ ,方差为 $\sigma_\beta^2=B''(1)+\beta-\beta^2$ 。

3) 服务器从一个子切片转换到相邻的下一个子切片进行查询所需要的时间相互独立且服从同样的概率分布,其分布的概率母函数为 $R(z)$ ,均值为 $\gamma=R'(1)$ ,方差为 $\sigma_\gamma^2=R''(1)+\gamma-\gamma^2$ 。

4) 信息分组按照先进入先发送(First Come First Service, FCFS)的原则进入各个子切片内的存储器。

5) 每个子切片存储器容量非常大,信息分组不会溢出。

### 2.2 变量定义

利用马尔可夫链对本文模型进行分析,定义以下变量,其中, $i, j=1, 2, \dots, N$ 。

1) 在服务器从子切片 $i$ 转向子切片 $i+1$ 时间内进入第 $j$ 号子切片内的信息分组数用 $\mu_j(u_i)$ 表示。

2) 在服务器从第 $i$ 号子切片转向 $i+1$ 号子切片所需要的转换查询时间内进入第 $j$ 号子切片内的信息分组数用 $\eta_j(v_i)$ 表示。

若在 $t_n$ 时刻 $i$ 号子切片存储器内的信息分组用 $\xi_i(n)$ 表示,则整个排队系统在此刻的状态可表示为 $[\xi_1(n), \xi_2(n), \dots, \xi_i(n), \dots, \xi_N(n)]$ ,其概率分布为 $P[\xi_i(n)=x_i; i=1, 2, \dots, N]$ 。第 $i$ 号子切片因传输出错需要进行重传的信息分组数为 $\xi_i^*(n)$ ,进行第 $k$ 次重传的信息分组数为 $\xi_{i,k}^*(n)$ ,即 $\xi_i^*(n)=\sum_{k=1}^m \xi_{i,k}^*(n)$ 。由文献[17]推导有:

$$E[\sum_{k=1}^m \xi_{i,k}^*(n)] = E[\xi_i(n)] \frac{p_c(1-p_c^m)}{1-p_c} \quad (1)$$

其中: $S_m = \frac{p_c(1-p_c^m)}{1-p_c}$ 。

### 2.3 概率母函数

系统在 $N\lambda\beta < 1$ 条件下达到稳定状态。系统稳定时的概率分布为:

$$\lim_{n \rightarrow \infty} P[\xi_i(n)=x_i; i=1, 2, \dots, N] = \pi_i(x_1, x_2, \dots, x_i, \dots, x_N) \quad (2)$$

则稳态概率母函数为:

$$G_i(z_1, z_2, \dots, z_i, \dots, z_N) = \sum_{x_1=0}^{\infty} \dots \sum_{x_i=0}^{\infty} \dots \sum_{x_N=0}^{\infty} z_1^{x_1} z_2^{x_2} \dots z_i^{x_i} \dots z_N^{x_N} \pi_i(x_1, x_2, \dots, x_i, \dots, x_N) \quad (3)$$

通过以上分析可知,当服务器在 $t_{n+1}$ 时刻开始对 $i+1$ 号子切片服务时,有:

$$\begin{cases} \xi_j(n+1) = \xi_j(n) + \mu_j(u_i) + \eta_j(v_i), j \neq i \\ \xi_i(n+1) = \mu_i(u_i) + \xi_i^*(n), \text{其他} \end{cases} \quad (4)$$

由此推导得出在 $t_{n+1}$ 时刻系统的稳态概率母函数为:

$$\begin{aligned} G_{i+1}(z_1, z_2, \dots, z_i, \dots, z_N) &= \lim_{n \rightarrow \infty} E \left[ \prod_{j=1}^N z_j^{\xi_j(n+1)} \right] = \lim_{n \rightarrow \infty} E \left[ \prod_{\substack{j=1 \\ j \neq i}}^N z_j^{\xi_j(n) + \mu_j(u_i) + \eta_j(v_i)} \cdot z_i^{\mu_i(u_i) + \xi_i^*(n)} \right] = \\ &= \lim_{n \rightarrow \infty} E \left[ \prod_{\substack{j=1 \\ j \neq i}}^N z_j^{\xi_j(n) + \eta_j(v_i)} \cdot z_i^{\xi_i^*(n)} \right] \cdot E \left[ \prod_{j=1}^N z_j^{\mu_j(u_i)} \right] = \lim_{n \rightarrow \infty} E \left[ \prod_{\substack{j=1 \\ j \neq i}}^N z_j^{\xi_j(n) + \eta_j(v_i)} \cdot z_i^{\sum_{k=1}^m \xi_{i,k}^*(n)} \right] \cdot R_i \left[ \prod_{j=1}^N A_j(z_j) \right] = \\ &= R_i \left[ \prod_{j=1}^N A_j(z_j) \right] \lim_{n \rightarrow \infty} E \left[ \left( \prod_{\substack{j=1 \\ j \neq i}}^N z_j^{\xi_j(n)} \right) \left( \prod_{j=1}^N z_j^{\eta_j(v_i)} \cdot z_i^{\sum_{k=1}^m \xi_{i,k}^*(n)} \right) \right] = \\ &= R_i \left[ \prod_{j=1}^N A_j(z_j) \right] \lim_{n \rightarrow \infty} E \left[ \left( \prod_{\substack{j=1 \\ j \neq i}}^N z_j^{\xi_j(n)} \right) z_i^{S_m} \left( B_i \left( \prod_{\substack{j=1 \\ j \neq i}}^N A_j(z_j) \right) \cdot F_i \left( \prod_{\substack{j=1 \\ j \neq i}}^N A_j(z_j) \right) \right)^{\xi_i(n)} \right] = \\ &= R_i \left[ \prod_{j=1}^N A_j(z_j) \right] [z_1, z_2, \dots, z_{i-1}, z_i^{S_m} B_i \left( \prod_{\substack{j=1 \\ j \neq i}}^N A_j(z_j) \right) \cdot F_i \left( \prod_{\substack{j=1 \\ j \neq i}}^N A_j(z_j) \right)] z_{i+1}, z_{i+2}, \dots, z_N \end{aligned} \quad (5)$$

其中,服务器在任意时隙进入任意子切片进行完全服务所需要的时间相互独立且服从相同的概率分布,其概率分布母函数用 $F(z_i)$ 表示,并且有 $F(z_i) = A(B(z_i F(z_i)))^{[22]}$ 。对 $F(z_i)$ 求一阶导和二阶导分别为 $F'(1) = \frac{\rho}{1-\rho}$ 和 $F''(1) = \frac{\beta^2 A''(1)}{(1-\rho)^3} + \frac{\lambda B''(1)}{(1-\rho)^3} + \frac{2\rho^2}{(1-\rho)^2}$ 。

### 2.4 平均排队队长

平均排队队长具体是指任意子切片内等待被服务的信息分组的平均排队长度。定义在 $t_n$ 时刻,第 $i$ 号子

切片开始接受服务时第 $j$ 号子切片内存储的信息分组个数为:

$$g_i(j) = \lim_{z_1, z_2, \dots, z_i, \dots, z_N \rightarrow 1} \frac{\partial G_i(z_1, z_2, \dots, z_i, \dots, z_N)}{\partial z_j}, \quad i, j=1, 2, \dots, N \quad (6)$$

由式(6)计算 $g_{i+1}(j)$ :

$$g_{i+1}(j) = \begin{cases} \gamma\lambda + g_i(j) + \frac{\rho}{1-\rho} g_i(i), i \neq j \\ \gamma\lambda + S_m g_i(i), \text{其他} \end{cases} \quad (7)$$

由式(7)计算  $\sum_{i=1}^N g_{i+1}(j)$ , 并化简得到  $g_i(i)$ , 即平均

排队队长, 如式(8)所示, 并由式(6)计算  $g_i(j)$ , 如式(9)

$$g_i(i) = \frac{(1-\rho)N\gamma\lambda}{1-N\rho-(1-\rho)S_m} \quad (8)$$

$$g_i(j) = \begin{cases} \frac{\gamma\lambda}{1-N\rho-(1-\rho)S_m} \cdot [N-N\rho+i-j-(1-\rho)S_m(i-j)], & N \geq j \geq i \geq 1 \\ \frac{\gamma\lambda}{1-N\rho-(1-\rho)S_m} \cdot \{i-j-N\rho+(1-\rho)S_m[N-(i-j)]\}, & N \geq i > j \geq 1 \end{cases} \quad (9)$$

由式(9)计算  $\sum_{i=1}^N g_i(j)$ , 如式(10)所示:

$$\sum_{i=1}^N g_i(j) = \frac{\gamma\lambda}{1-N\rho-(1-\rho)S_m} \cdot \frac{N(N-1)}{2} [1-2\rho+(1-\rho)S_m]$$

(10)

## 2.5 平均等待时间

平均等待时间具体是指信息分组从进入子切片到其被服务并发送出去所需等待的这段时间。

定义:

$$g_i(j, k) =$$

$$\lim_{z_1, z_2, \dots, z_j, \dots, z_k, \dots, z_N \rightarrow 1} \frac{\partial^2 G_i(z_1, z_2, \dots, z_j, \dots, z_k, \dots, z_N)}{\partial z_j \partial z_k}, \quad i, j, k = 1, 2, \dots, N \quad (11)$$

由式(11)计算得到式(12)~式(15), 由式(12)和式(14)计算  $\sum_{i=1}^N g_{i+1}(j, k)$ , 如式(16)所示, 由式(13)和

式(15)计算  $\sum_{i=1}^N g_{i+1}(j, j)$ , 如式(17)所示, 化简式(16)和式(17), 并联立计算得到式(18):

$$g_{i+1}(j, k) = \lambda^2 R''(1) + \gamma\lambda^2 + \gamma\lambda[g_i(j) + g_i(k)] + g_i(j, k) + \frac{\rho}{1-\rho} \left[ g_i(j, i) + g_i(i, k) + \frac{\rho}{1-\rho} g_i(i, i) \right] + g_i(i) \left[ \frac{\lambda\rho}{1-\rho} + \frac{2\gamma\lambda\rho}{1-\rho} + \frac{2\lambda\rho^2}{(1-\rho)^2} + \frac{\beta\rho^2 A''(1)}{(1-\rho)^3} + \frac{\lambda^2 B''(1)}{(1-\rho)^3} \right], \quad i \neq j \neq k \quad (12)$$

$$g_{i+1}(j, j) = \lambda^2 R''(1) + \gamma A''(1) + 2\gamma\lambda g_i(j) + g_i(j, j) + \frac{\rho}{1-\rho} \left\{ 2g_i(j, i) + \frac{\rho}{1-\rho} g_i(i, i) \right\} + g_i(i) \left[ \frac{2\gamma\lambda\rho}{1-\rho} + \frac{2\lambda\rho^2}{(1-\rho)^2} + \frac{\beta A''(1)(1-2\rho+2\rho^2)}{(1-\rho)^3} + \frac{\lambda^2 B''(1)}{(1-\rho)^3} \right], \quad i \neq j \quad (13)$$

$$g_{i+1}(j, i) = \lambda^2 R''(1) + \gamma\lambda^2 + \gamma\lambda g_i(j) + S_m g_i(j, i) + \frac{S_m \rho}{1-\rho} g_i(i, i) + g_i(i) \left\{ \frac{\gamma\lambda\rho}{1-\rho} + S_m \gamma\lambda + S_m \rho + \frac{S_m \rho^2}{1-\rho} \right\}, \quad i \neq j \quad (14)$$

$$g_{i+1}(i, i) = \lambda^2 R''(1) + \gamma A''(1) + S_m^2 g_i(i, i) + S_m g_i(i) (2\gamma\lambda + S_m - 1) \quad (15)$$

$$\sum_{i=1}^N g_{i+1}(j, k) = N\lambda^2 R''(1) + N\gamma\lambda^2 + 2\gamma\lambda \sum_{i=1}^N g_i(j) + \left[ \sum_{i=1}^N g_i(j, k) - g_j(j, k) - g_k(j, k) \right] + 2S_m g_i(j, i) + \frac{2\rho}{1-\rho} \sum_{i=1}^N g_i(j, i) + \frac{\rho^2}{(1-\rho)^2} (N-2) g_i(i, i) + \frac{2S_m \rho}{1-\rho} g_i(i, i) + (N-2) g_i(i) \left[ \frac{\lambda\rho}{1-\rho} + \frac{2\lambda\rho^2}{(1-\rho)^2} + \frac{\beta\rho^2 A''(1)}{(1-\rho)^3} + \frac{\lambda^2 B''(1)}{(1-\rho)^3} \right] + (N-1) \cdot g_i(i) \left[ \frac{2\gamma\lambda\rho}{1-\rho} + 2S_m g_i(i) \left( \gamma\lambda + \rho + \frac{\rho^2}{1-\rho} \right) \right] \quad (16)$$

$$\sum_{i=1}^N g_{i+1}(j, j) = N\lambda^2 R''(1) + N\gamma A''(1) + 2\gamma\lambda \sum_{i=1}^N g_i(j) + \left[ \sum_{i=1}^N g_i(j, j) - g_j(j, j) \right] + S_m^2 g_i(i, i) + \frac{2\rho}{1-\rho} \sum_{i=1}^N g_i(j, i) + \frac{\rho^2}{(1-\rho)^2} (N-1) \cdot g_i(i, i) + (N-1) g_i(i) \left[ \frac{2\gamma\lambda\rho}{1-\rho} + \frac{2\lambda\rho^2}{(1-\rho)^2} + \frac{\beta A''(1)(1-2\rho+2\rho^2)}{(1-\rho)^3} + \frac{\lambda^2 B''(1)}{(1-\rho)^3} \right] + S_m g_i(i) (2\gamma\lambda + S_m - 1) \quad (17)$$

$$g_i(i, i) = \frac{N(1-\rho)^2}{(1-\rho)(1-N\rho)+C} \left\{ \lambda^2 R''(1)(1-S_m+S_m\rho) + \gamma A''(1) \left[ \frac{(\rho^2(N-1)(1-\rho)+D)}{(1-\rho)^2[1-N\rho-(1-\rho)S_m]} + 1 \right] + \frac{(N-1)\gamma\lambda^3 B''(1)(1-S_m)}{(1-\rho)[1-N\rho-(1-\rho)S_m]} + \frac{\gamma\lambda H}{[1-N\rho-(1-\rho)S_m](1-\rho)} \right\} \quad (18)$$

$$\begin{aligned}
C &= S_m^2 N \rho - 2 S_m^2 N \rho^2 + S_m^2 N \rho^3 + S_m^2 \rho + S_m^2 \rho^2 - S_m^2 \rho^3 - S_m^2 + S_m^3 - 3 S_m^3 \rho + 3 S_m^3 \rho^2 - S_m N \rho^2 - 2 S_m \rho^2 - S_m + \\
&\quad 3 S_m \rho - S_m^3 \rho^3 + S_m N \rho^3 \\
D &= S_m N \rho - 3 S_m N \rho^2 + 3 S_m \rho - 3 S_m \rho^2 + 2 S_m N \rho^3 + 2 S_m \rho^3 - S_m \rho^4 - S_m + S_m^2 - 4 S_m^2 \rho + 6 S_m^2 \rho^2 - 4 S_m^2 \rho^3 + S_m^2 \rho^4 \\
H &= N \lambda \rho - \lambda \rho + \lambda \rho^2 - N \gamma \lambda \rho + \gamma \lambda \rho - N \lambda \rho^2 + N \gamma \lambda - \gamma \lambda - S_m N \lambda \rho + S_m \lambda \rho + S_m N \lambda \rho^3 - S_m \lambda \rho^3 - 4 S_m \gamma \lambda \rho + 2 S_m \gamma \lambda \rho^2 - \\
&\quad S_m \rho^2 - S_m N \rho^3 + S_m \rho^3 - S_m^2 N \gamma \lambda + 3 S_m^2 N \gamma \lambda \rho - S_m^2 \gamma \lambda + 3 S_m^2 \gamma \lambda \rho - 3 S_m^2 N \gamma \lambda \rho^2 - 3 S_m^2 \gamma \lambda \rho^2 + S_m^2 N \gamma \lambda \rho^3 + \\
&\quad S_m^2 \gamma \lambda \rho^3 + 2 S_m \gamma \lambda - S_m^2 N \rho - 4 S_m^2 \rho + 2 S_m^2 - S_m^3 + 3 S_m^3 \rho + 2 S_m^2 N \rho^2 + 2 S_m^2 \rho^2 - 3 S_m^3 \rho^2 - S_m^2 N \rho^3 + \\
&\quad S_m^3 \rho^3 + S_m N \rho + S_m \rho - S_m
\end{aligned}$$

在该轮询系统中,服务器按照完全服务策略对各队列进行服务,文献[8]中给出式(19):

$$E(W_E) = \frac{g_i(i, i)}{2\lambda g_i(i)} - \frac{(1-2\rho)A''(1)}{2\lambda^2(1-\rho)} + \frac{\lambda B''(1)}{2(1-\rho)} \quad (19)$$

将式(8)和式(18)代入式(19)即可求得该模型信息分组平均等待时间的精确表达式。

### 3 实验与结果分析

#### 3.1 仿真实验及算法分析

基于上文所建立模型,利用Matlab2019b平台对5G网络切片架构下具有差错重传机制的完全服务轮询排队队列进行仿真模拟。基于式(8)、式(18)、式(19)分别计算出重传完全轮询系统中的平均排队队长和平均等待时间对应的理论值。将理论值与实验仿真值进行对比,再将其与基础轮询模型比较,从而验证理论分析模型的正确性和可靠性。此外,本文还针对重传概率和重传阈值对信息分组平均排队队长和平均等待时间的影响进行分析。最后,将本文模型与其他模型进行对比,对该模型的性能做进一步分析。

实验初始参数设置如表1所示,在此基础上,通过100 000次蒙特卡洛实验,得到本文排队模型性能特点的变化情况,重点针对平均排队队长和平均等待时间进行研究。通过大量实验发现,在一定的信道质量条件下,平均排队队长和平均等待时间均收敛于重传阈值。因此,本文通过经验假设重传阈值为1,并在此条件下展开研究。

表1 排队模型的基础参数设置

Table 1 Base parameters setting of queuing model

参数	重传轮询	基础轮询
到达率 $\lambda$	0.005	0.005
服务时间 $\beta$	1	1
轮询转化时间 $\gamma$	1	1
重传概率 $P_e$	0.01	—
重传阈值 $m$	1	—

本文通过运用排队论建立具有重传机制的完全服务轮询数学模型,采用嵌入式马尔可夫链和概率母函数分析方法推导出系统平均排队队长和平均等待时间的理论表达式。在开始时刻,各个子切片内为空,即无

信息分组需要发送;子切片提出发送请求后,服务器开始对其进行服务;在传输过程中,若有信息分组没有被成功发送出去,则立即返回至源子切片重新排队等待再次发送,直到重传次数大于重传阈值后仍未发送成功,则抛弃该出错数据包;当子切片内所有信息分组全部发送完毕或被抛弃,则以升序转向下一个逻辑上相邻的子切片。本文方案的算法流程如下:

Initialize  $\xi_i(n) = 0, i = 1, 2, \dots, N; s = 0; \text{gk} = \text{zeros}(1, N);$

$P_e = 0.01; m = 1; S_m = \frac{P_e(1-P_e^m)}{1-P_e}; N\lambda\beta < 1;$

while  $s \neq 100\ 000$

for  $k = 1 : N$

$\xi_i(n) = \text{expmnd}\left(\frac{1}{\lambda}, 1, \text{ceil}(\text{gk}(k))\right);$  //子切片内到达的信息

//分组数

while  $\xi_i(n) \neq 0$

$\xi_i(n) = \xi_i(n) \cdot (1 + P_e);$

$v_i(n) = \xi_i(n) \cdot \beta;$  //第i号站点的总服务时间

$\xi_{\text{all}}(n) = \xi_{\text{all}}(n) + \xi_i(n);$  //单个轮询系统中所有子切片的总

//信息分组数

$u_i(n) = \gamma;$  //轮询转换时间

for  $j = 1 : \xi_i(n)$

//单个轮询系统中所有子切片需要的总服务时间

$v_{\text{all}}(n) = v_{\text{all}}(n) + w(j) \cdot (1 + P_e) + (j-1) \cdot \beta;$

end

end

end

$s = s + 1;$

end

上述算法是在文献[12]基础上进行改进的,基于Matlab伪随机数,生成到达率为 $\lambda$ 的序列,该序列服从泊松分布,用于模拟系统各个子切片单位时间内到达的信息分组数。然后,进一步通过10 000次蒙特卡洛实验得到最终的平均排队队长和平均等待时间。由此可知,算法循环次数是一个常数,因此,复杂度与文献[12]方法相同,均为 $O(1)$ 。

#### 3.2 性能分析

实验中,通过改变轮询系统中子切片个数,分析平均排队队长和平均等待时延随子切片个数的变化规律,如图2、图3所示。

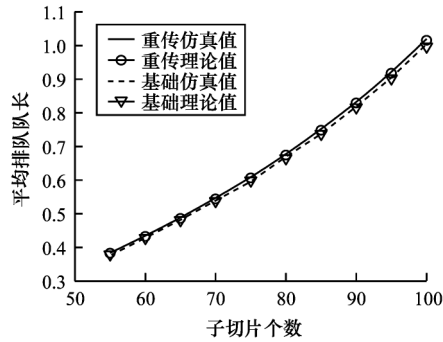


图2 平均排队队长对比分析

Fig.2 Comparison analysis of average queue length

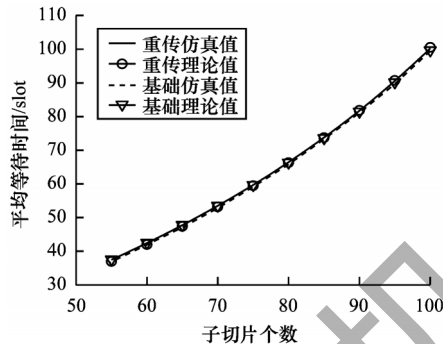


图3 平均等待时间对比分析

Fig.3 Comparison analysis of average waiting time

图2、图3分别表示平均排队队长和平均等待时间随子切片个数变化的情况。从图中可以看出,随着子切片个数不断增加,平均排队队长和平均等待时间也随之增加。重传轮询模型和基础轮询保持相同的变化规律。同时,仿真值和理论值之间的误差较小,且重传轮询队长略大于基础轮询模型,这证明了上述理论分析的正确性。当子切片个数不超过75个时,重传轮询系统的信息分组时延与基本轮询系统几乎相等;当子切片个数超过75个时,重传轮询时延逐渐大于基础轮询时延。这说明在网络规模较小的状况下,重传和不重传情况区别不大,噪声、设备误差等影响信道质量的因素不会对客户等待时间产生明显影响。而面对网络规模较大的状况,因信道质量影响,即使各个客户在信息传递过程中出错,经过重新排队等待再次传输,每次都只增加了微小的延迟,但由于大量客户的逐步累加,也会产生巨大的延迟,因此加入差错重传机制后可以得出更准确的信息分组等待时间。这也再次证明了该模型分析的正确合理性。

基于表1的基础参数设置,将 $P_e$ 改变为0.01、0.03、0.05、0.10、0.15,分析重传概率对该排队模型的影响,结果如图4、图5所示。图4为不同重传概率情况下信息分组平均排队队长随系统吞吐量变化的曲线,图5为基于不同的重传概率信息分组时延随系统吞吐量变化的曲线。吞吐量是指系统在单位时间内处理的信息分组数,用 $T$ 表示,有 $T=N\lambda\beta$ 。以吞吐量为横坐标,评估在相同的时延条件下不同重传概率时的系统吞吐量。

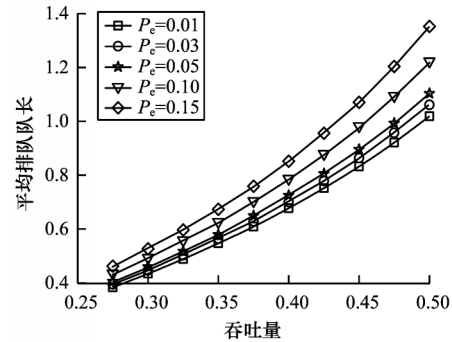


图4 重传概率与队长的关系曲线

Fig.4 Curve of retransmission probability versus queue length

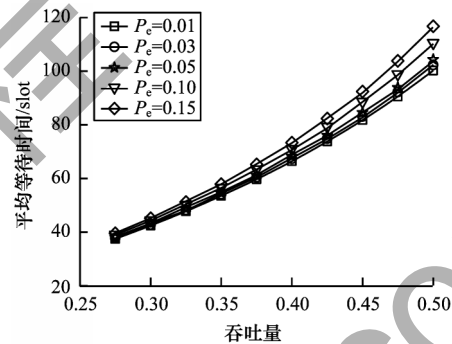


图5 重传概率与时延的关系曲线

Fig.5 Curve of retransmission probability versus delay

从图4、图5中可以看出,平均队长和平均时延与吞吐量呈正相关。在相同的信息分组排队队长和等待时延条件下,更小的重传概率可以使系统实现更大吞吐量。在相同负载情况下,随着重传概率的增加,有更多的信息分组因出错而重新加入到源子切片尾部,等待再次被发送。因此,重传概率越大,平均队长越长,平均时延越高,为保证系统的可靠性,应尽可能降低重传概率,这也是未来针对非理想信道系统研究的方向。

基于表1的基础参数设置,另设置子切片个数 $N=80$ ,改变 $P_e$ 分别为0.01、0.03、0.05、0.10、0.15,重传阈值 $m$ 从1到10以1为步长依次叠加时,通过模拟仿真,得到平均排队队长和平均等待时间随重传阈值的变化情况,如图6、图7所示。

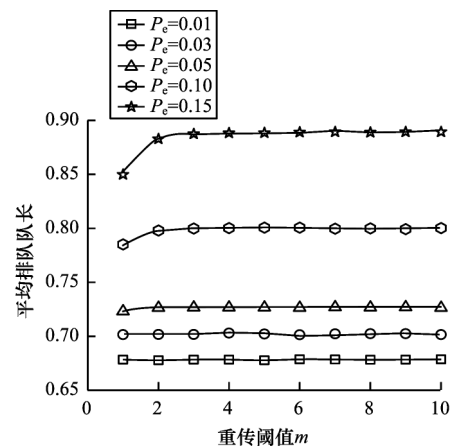


图6 重传阈值与队长的关系曲线

Fig.6 Curve of retransmission threshold versus queue length



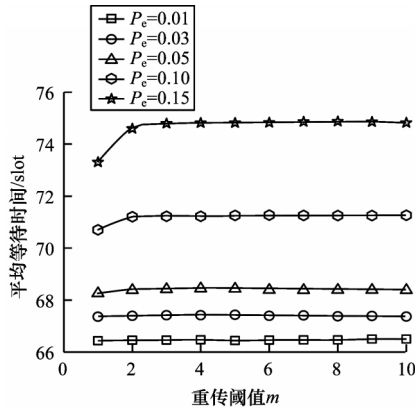


图7 重传阈值与时延的关系曲线

Fig.7 Curve of retransmission threshold versus delay

从图6、图7中可以看出:在相同的负载情况下,当 $P_e$ 为0.01和0.03时,重传阈值并没有对平均队长和平均时延产生明显影响;而当 $P_e$ 为0.05或0.01, $m \leq 2$ 时,信息分组的排队队长和时延都随重传次数的增加而增加, $m > 2$ 时,平均队长和时延分别收敛于一个特定的值,即重传次数对信息分组的平均排队队长和平均时延影响较小;当 $P_e=0.15$ , $m \leq 3$ 时,平均队长和时延都随重传次数呈正比例变化,相比较更小重传概率的情况,平均队长和时延受更大范围的重传次数影响, $m > 3$ 时,平均队长和时延分别收敛于另一个特定的值。同时可以明显看出,对于相同的负载,当重传概率增加时,信息分组的排队队列越长,时延越大。该结论与图4、图5中分析得出的结论一致。由此可见,信道质量影响系统时延,即信息分组在传输过程中出错的概率越高,延迟越长。而当信道质量一定时,系统延迟最终可收敛于重传阈值。

### 3.3 对比分析

将本文方案与文献[17]提出的具有差错重传机制的门限服务轮询方案进行对比分析。基于相同的网络参数条件,对比非理想信道环境中门限和完全2种服务策略的性能,实验结果如图8、图9所示。

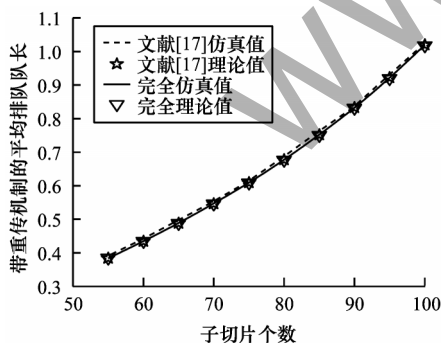


图8 两种服务策略队长比较

Fig.8 Comparison of queue length of two service strategies

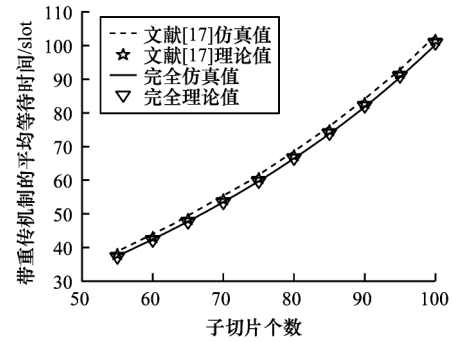


图9 两种服务策略时延比较

Fig.9 Comparison of delay of two service strategies

从图8、图9中可以看出,在非理想信道环境下,重传完全服务的平均队长和平均时延各自的理论值和仿真值一致,而重传门限服务的平均队长和平均时延各自的理论值和仿真值之间存在较小的误差。相比之下,重传完全服务轮询模型更为准确,并且对比重传门限服务,重传完全服务有更小的平均队长和平均时延,在信息处理效率方面,重传完全的性能效果更好。因此,与文献[17]方案相比,本文方案更可靠,同时能够以更快的服务效率满足用户的不同服务需求。

将本文方案与文献[12]提出的两级完全轮询接入方案进行对比分析。文献[12]是在理想环境下进行的,即假设任意信息分组在第一次发送时就发送成功,无需进行重传。其中,文献[12]将子切片分为 $N$ 个普通子切片和1个优先级子切片,服务器在每次服务中,首先对优先级子切片服务再对普通子切片服务。设置文献[12]中的普通子切片和本文方案在相同的网络参数条件下,优先级子切片的到达率为0.05,服务率为2,对比两种方案的平均队长和平均时延,如图10、图11所示。

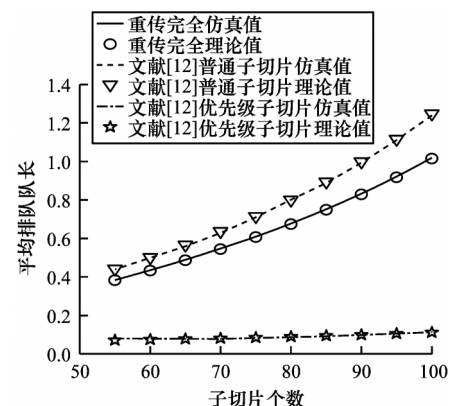


图10 两种方案平均队长的对比分析

Fig.10 Comparative analysis of the average queue length of two schemes



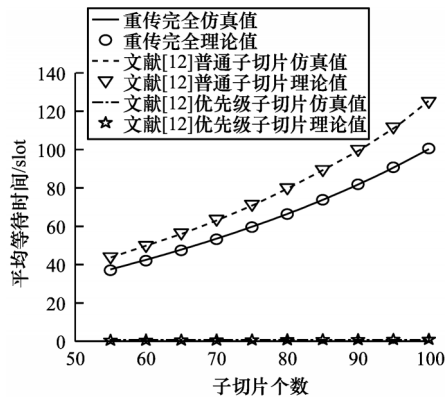


图11 两种方案平均时延的对比分析

Fig.11 Comparative analysis of the average delay of two schemes

从图10、图11中可以看出,两种方案下的理论值和仿真值都保持一致,证明了算法的正确性。文献[12]普通子切片的平均队长和平均时延都明显高于本文方案。虽然文献[12]的优先级子切片的平均队长和平均时延都远低于本文方案,但因为其优先级子切片的信息分组到达率和服务率与本文方案不一致,所以不具有对比性。总体而言,由于文献[12]方案设置必须优先对优先级子切片进行服务,明显降低了普通子切片处理效率,因此即使文献[12]方案不进行重传,本文方案的平均队长和平均时延也明显低于其普通子切片,信息处理效率更快,服务质量更高。

#### 4 结束语

在5G时代,网络规模和业务量的增加导致信道拥堵,信息处理效率低下。因此,本文引入5G网络切片架构对海量业务信息进行分类,进而针对性地提供不同的资源给予更好的服务。建立具有差错重传机制的完全服务轮询排队模型,将eMBB、mMTC和uRLLC切片各自划分出 $N$ 个子切片,分别在各层切片采用该排队模型进行服务。同时,基于马尔可夫链和概率母函数分析方法,获得具有差错重传机制的完全服务轮询系统排队状态概率母函数,以及信息分组平均排队队长和平均等待时间的精确表达式。实验结果表明,重传模型的仿真实验数据值与理论计算值结果一致,且重传的平均队长和分组平均时延略大于基本轮询服务,验证了本文方案的正确性。与无重传机制的两级完全服务轮询系统的普通队列相比,本文方案网络延迟更低且信息处理效率更高。本文采用固定的信息分组出错概率,下一步将针对动态的差错概率研究信道质量变化对系统性能的影响。同时,也将划分多优先级以满足优先级业务的实时性需求。

#### 参考文献

- [1] 李红伟,赵一荣,李金艳,等. 基于能力开放的5G网络切片管理研究[J]. 电子技术应用,2020,46(1):1-5,11.
- [2] ALQAHTANI S A, ALTAMRAH A S. Supporting QoS requirements provisions on 5G network slices using an efficient priority-based polling technique [J]. Wireless Networks, 2019, 25(7): 3825-3838.
- [3] DE HAAN R, AL HANBALI A, BOUCHERIE R J, et al. Transient analysis for exponential time-limited polling models under the preemptive repeat random policy [J]. Advances in Applied Probability, 2020, 52(1): 32-60.
- [4] YANG Z J, MAO L, GAN J H, et al. Performance analysis and prediction of double-server polling system based on BP neural network [J]. Chinese Journal of Electronics, 2020, 29(6): 1046-1053.
- [5] LIN C H, LIN K C J, CHEN W T. Channel-aware polling-based MAC protocol for body area networks: design and analysis [J]. IEEE Sensors Journal, 2017, 17(9): 2936-2948.
- [6] SIDDIQUI S, GHANI S, KHAN A A. ADP-MAC: an adaptive and dynamic polling-based MAC protocol for wireless sensor networks [J]. IEEE Sensors Journal, 2018, 18(2): 860-874.
- [7] ROY A, SARMA N. RPCP-MAC: receiver preamble with channel polling MAC protocol for underwater wireless sensor networks [J]. International Journal of Communication Systems, 2020, 33(9): 1-5.
- [8] 罗启鹏,保利勇,丁洪伟,等. 离散时间并行优化门限服务轮询系统排队特性解析[J]. 电子学报,2019,47(9):1937-1942.
- [9] LUO Q P, BAO L Y, DING H W, et al. Performance analysis of discrete time polling queuing system based on parallel optimization gated service [J]. Acta Electronica Sinica, 2019, 47(9): 1937-1942. (in Chinese)
- [10] 孔维东,王永斌,刘宏波. 基于排队论模型的轮询协议数据链系统时延分析[J]. 火力与指挥控制,2017,42(3):100-103.
- [11] KONG W D, WANG Y B, LIU H B. Research on time delay of polling protocol data link based on queuing theory model [J]. Fire Control & Command Control, 2017, 42(3): 100-103. (in Chinese)
- [12] ISMAIEL B, ABOLHASAN M, NI W, et al. Analysis of effective capacity and throughput of polling-based device-to-device networks [J]. IEEE Transactions on Vehicular Technology, 2018, 67(9): 8656-8666.
- [13] 杨志军,苏杨,丁洪伟. 完全服务和非对称门限服务两级轮询系统特性分析[J]. 自动化学报,2018,44(12):2228-2237.
- [14] YANG Z J, SU Y, DING H W. Analysis of two-level polling system characteristics of exhaustive service and asymmetrically gated service [J]. Acta Automatica Sinica, 2018, 44(12): 2228-2237. (in Chinese)
- [15] 杨志军,毛磊,丁洪伟,等. 连续时间两级完全轮询接入MAC协议分析[J]. 计算机工程与应用,2022,58(9):136-143.
- [16] YANG Z J, MAO L, DING H W. Analysis of continuous time two-level exhaustive polling access MAC protocol [J]. Computer Engineering and Applications, 2022, 58(9): 136-143. (in Chinese)
- [17] 杨志军,刘征,丁洪伟,等. 基于轮询控制的无线传感器网络MAC协议性能分析[J]. 云南大学学报(自然科学版), 2020, 42(1): 1-5, 11. (in Chinese)

- 2020,42(3):437-443.
- YANG Z J, LIU Z, DING H W, et al. Performance analysis of wireless sensor network MAC protocol based on polling control[J]. Journal of Yunnan University (Natural Sciences Edition), 2020,42(3):437-443. (in Chinese)
- [14] HUANG T P, CHEN H M, CUI L, et al. An effective discriminator for differentiating the root causes of packet transmission failures in indoor WSNs[J]. Transactions on Emerging Telecommunications Technologies, 2017,28(7):1-5.
- [15] KIM B, KIM J. Analysis of the waiting time distribution for polling systems with retrials and glue periods[J]. Annals of Operations Research, 2019,277(2):197-212.
- [16] YANG Z J, MAO L, YAN B, et al. Performance analysis and prediction of asymmetric two-level priority polling system based on BP neural network[J]. Applied Soft Computing, 2021,99:106880.
- [17] 官铮,熊磊,贾徕,等. RoF-DAS架构下具有差错重传的预约式 WLAN MAC 协议研究[J]. 通信学报,2020,41(3):102-111.
- GUAN Z, XIONG L, JIA Y, et al. Research on scheduled WLAN MAC protocol with failure retries on RoF-DAS architecture[J]. Journal on Communications, 2020,41(3):102-111. (in Chinese)
- [18] OMMEREN J K, AL HANBALI A, BOUCHERIE R J. Analysis of polling models with a self-ruling server[J]. Queueing Systems, 2020,94(1/2):77-107.
- [19] 杨志军,刘征,丁洪伟. 连续时间完全服务与门限服务两级轮询系统性能研究[J]. 计算机应用,2019,39(7):2019-2023.
- YANG Z J, LIU Z, DING H W. Research of continuous time two-level polling system performance of exhaustive service and gated service [J]. Journal of Computer Applications, 2019,39(7):2019-2023. (in Chinese)
- [20] ZHANG H J, LIU N, CHU X L, et al. Network slicing based 5G and future mobile networks: mobility, resource management, and challenges[J]. IEEE Communications Magazine, 2017,55(8):138-145.
- [21] MOHSEN N, HASSAN K S. C-RAN simulator: a tool for evaluating 5G cloud-based networks system-level performance [C]//Proceedings of the 11th International Conference on Wireless and Mobile Computing, Networking and Communications. Washington D. C., USA:IEEE Press, 2015:302-309.
- [22] 孙洋洋,杨志军. 依据站点状态的完全服务轮询控制系统[J]. 云南师范大学学报(自然科学版),2018,38(4):31-35.
- SUN Y Y, YANG Z J. Full service polling control system based on site status [J]. Journal of Yunnan Normal University (Natural Sciences Edition), 2018,38(4):31-35. (in Chinese)

编辑 金胡考