

5G网络时敏业务的时延特性分析

曹建宇¹, 冯伟², 葛宁²

(1. 桂林电子科技大学计算机与信息安全学院, 广西 桂林 541004;
2. 清华大学电子工程系, 北京 100084)

摘要: 面向工业应用, 5G 接入网需要同时支持时敏与非时敏业务。两类业务共享无线资源和缓冲队列, 常导致时延的复杂耦合与不确定。针对 ON-OFF 时敏业务与非时敏业务共存场景, 基于二阶级联排队模型, 建立包含排队时延的随机时延特性计算方法, 得到时敏业务的时延均值、标准差、越界概率与概率分布。在此基础上, 分析非时敏业务对时敏业务随机时延特性的影响。实验表明, 非时敏业务对 ON-OFF 时敏业务随机时延特性的影响与 ON-OFF 阶段的轮换频率相关; ON-OFF 时敏业务时延的概率分布与伽马分布的右拖尾长度近似相等。

关键词: 5G; 时敏业务; 随机时延特性; 级联排队模型

中图分类号: TN91

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2022.00298

Delay characteristics of time-sensitive services in 5G networks

CAO Jianyu¹, FENG Wei², GE Ning²

1. School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
2. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Abstract: For industrial applications, the time-sensitive services (TSS) and non TSS need to share the radio resources and the same buffer queue in 5G radio access network (RAN). This incurs the complex coupling and uncertainty of the delays. Considering the scenario where ON-OFF TSS and non TSS coexist, the calculation methods for the random delay (including queueing delay) characteristics were established based on the two-stage tandem queueing model, to obtain the mean value, standard deviation, bound violation probability and probability distribution of the delay of TSS. On this basis, the influence of non TSS on the random delay characteristics of TSS was analyzed. It is shown that the influence of non TSS on the random delay characteristics of ON-OFF TSS is related to the rotation frequency of ON-OFF stages. Moreover, the delay probability distribution of ON-OFF TSS has approximately the same right tail length as Gamma distribution.

Key words: 5G, time-sensitive service, random delay characteristic, tandem queueing model

0 引言

网络信息服务向垂直行业不断拓展^[1], 如空天地海一体化^[2-3]、智慧生产、全息通信^[4]等, 将促使移动网络的时敏业务流量日益增加。ITU “Network 2030” 焦点组将时敏业务分为及时 (in-time) 业务

与准时 (on-time) 业务^[5]。及时业务要求端到端时延 T 小于给定阈值, 即 $T \leq T_1$; 准时业务要求端到端时延 T 落在给定时间区间内, 即 $T_1 \leq T \leq T_2$ 。所以, 移动网络的确定性时延服务能力是其满足时敏业务时延需求的重要基础^[6]。为此, 移动边缘计算 (MEC, mobile edge computing) ^[7-8] 通过将远端服务

收稿日期: 2022-05-09; 修回日期: 2022-09-04

通信作者: 冯伟, fengwei@tsinghua.edu.cn

基金项目: 国家重点研发计划 (No.2020YFA0711301); 国家自然科学基金资助项目 (No.61922049, No.61941104)

Foundation Items: The National Key Research and Development Program of China (No.2020YFA0711301), The National Natural Science Foundation of China (No.61922049, No.61941104)

下沉至移动网络边缘可有效减小业务的端到端时延。超可靠低时延通信 (URLLC, ultra-reliable and low-latency communication) 的关键技术^[9-10]与新的多址接入技术^[11]可有效减小空口低层的传输与处理时延。然而, 接入网中无线资源受限与无线环境随机变化使时敏业务在接入网中的时延 (特别是排队等待时延) 随机波动, 给保障确定性时延需求带来挑战。

为应对上述挑战, 在接入网中可采用服务质量 (QoS, quality of service) 优先级调度、无线块 (RB, radio block) 资源预留、载波隔离和物理设备隔离等多个等级的确定性时延保障技术^[12]。这些技术的确定性时延保障能力逐渐提升, 但相应成本也逐级增加。例如, 物理设备隔离为单个行业客户或者重点业务单独建设无线电接入网 (RAN, radio access network) 物理设备形成专网, 但该方案的建设和维护成本较高。软件定义网络和网络功能虚拟化技术不断发展, 使 5G RAN 支持时敏与非时敏业务 (非时敏业务对时延没有特别要求, 或对时延的要求远小于时敏业务, 如下载业务) 共享物理设备和无线资源, 在节约专网建设和维护成本的同时提升无线资源效率^[13]。针对时敏与非时敏业务共享无线资源和缓冲队列的情况, 现有时延控制方法主要关注不同业务共享无线资源时的公平性, 可降低混合业务的整体平均时延, 但难以满足混合业务中时敏业务的确定性时延需求。

时敏业务随机时延特性的计算问题是控制混合业务中时敏业务时延的基础, 但相关计算方法仍然未知^[14]。本文致力于解决此问题, 主要贡献如下。

1) 针对 ON-OFF 时敏业务与非时敏业务共享无线资源和无线链路控制 (RLC, radio link control) 缓冲队列的情况, 建立 ON-OFF 时敏业务在接入网下行链路的随机时延 (含排队等待时间) 特性计算方法, 可计算时延的均值、标准差、越界概率与概率分布。

2) 分析非时敏业务对 ON-OFF 时敏业务随机时延特性的影响。实验表明, 当 ON 与 OFF 阶段的轮换频率不同时, 非时敏业务在 OFF 阶段向 RLC 子层 (RLC 缓冲队列) 交付数据的速率对 ON-OFF 时敏业务的时延均值与标准差的影响程度也不同; ON-OFF 时敏业务时延概率分布的右拖尾长度近似等于伽马分布的右拖尾长度, 即当 ON-OFF 时敏业务的时延均值与方差确定时, 其时延波动的上界可

通过具有相同均值及方差的伽马分布近似给出。

1 相关工作

接入网中多业务共享无线资源和缓冲队列时, 现有排队时延控制方法主要包括主动队列管理 (AQM, active queue management)、限制缓冲队列长度与流量控制。

时延受控 (Codel, controlled delay)^[15-16]是当前被广泛使用的一种 AQM 算法, 根据队列中数据包的逗留时间决定是否丢弃数据包, 既能降低时延, 又能吸收短暂的大流量数据以提升吞吐量。然而, 当多个业务共享缓冲队列时, CoDel 算法不区分业务类型, 只能实现缓冲队列中混合业务时延的整体降低, 此时, 为保障高优先级时敏业务的确定性时延, 需要投入更多无线资源。流队列时延受控 (FQ-Codel, flow queueing CoDel)^[17-18]将流队列与 CoDel 结合, 对链路瓶颈节点的多个缓冲队列进行协同式主动队列管理, 可保障各业务共享资源的公平性, 但在多业务共享无线资源和缓冲队列时不能保障时敏业务的确定性时延需求。

2018 年, Kumar 等^[19]提出动态无线链路控制 (DynRLC, dynamic radio link control) 方法。Imputato 等^[20]将 Linux 内核提供的流量控制 (TC, traffic control) 与字节队列限制 (BQL, byte queue limits)^[21]机制结合提出一种智能积压 (backlog) 管理方法——TC-BQL。DynRLC 与 TC-BQL 分别在分组数据汇聚协议 (PDCP, packet data convergence protocol) 子层内与 PDCP 子层下行入口将不同业务的数据隔离至独立的队列, 但 5G 接入网协议栈在 PDCP 子层对数据做安全性处理后将数据交付到 RLC 子层的缓冲队列。若在 PDCP 子层下行入口或 PDCP 子层识别并重新隔离不同业务数据, 则会增加方法的部署难度。Irazabal 等^[22]基于 5G 接入网协议栈中数据流的映射特征, 提出一种低时延队列管理方法, 其中, 低时延业务数据可以优先向 RLC 子层交付, 但未给出具体的调度方案。之后, Irazabal 等^[23]在 5G-BDP 的基础上提出从服务数据适配协议 (SDAP, service data adaptation protocol) 子层到 RLC 子层的调度算法 e5G-BDP, 使含有短数据包的 QoS 流可以优先向 RLC 缓冲队列交付数据, 避免长数据包的大流量业务长期占用资源而影响短数据包业务的时效性, 保障了各 QoS Flow 使用无线资源的公平性。

如上所述,针对不同业务共享 RLC 缓冲队列的情况,现有的排队时延控制方法主要关注各业务使用无线资源的公平性,可降低混合业务(共享 RLC 缓冲队列的多个业务)的平均时延,但不能有效保障混合业务中高优先级时敏业务的确定性时延需求。控制混合业务中时敏业务的时延需要计算时延的随机特性指标,且获得这些随机时延特性的变化机制,但时延特性的相关计算方法与变化机制仍然未知^[14]。为此,本文将针对时敏与非时敏业务共享无线资源与 RLC 缓冲队列的情况,研究时敏业务随机时延特性的计算方法与变化机制。

2 问题描述

考虑 5G 网络中用户设备(UE, user equipment)同时请求时敏与非时敏业务,且时敏业务按 ON-OFF 模式生成数据(如交互式业务^[24])的场景,5G 网络中下行业务数据的传输过程如图 1 所示,QoS Flow 1 表示 ON-OFF 时敏业务流(目标业务流);QoS Flow 2 与 QoS Flow 3 均表示非时敏业务流(背景业务流)(如无特别说明,本文中的目标业务指 ON-OFF 时敏业务,背景业务指非时敏业务),它们被映射到相同的数据无线承载(DRB, data radio bearer)且共享 RLC 缓冲队列^[25-26]。本文将建立 ON-OFF 时敏业务在接入网下行链路中的随机时延(数据从进入 SDAP 缓冲队列到离开 RLC 缓冲队列所经历的时间)特性计算方法,并分析非时敏业务向 RLC 子层交付数据的速率对 ON-OFF 时敏业务随机时延特性的影响。

3 模型建立

为了计算目标业务的随机时延特性并分析背景业务对目标业务时延特性的影响,建立含有附加到达过程(背景业务到达过程)的离散时间二阶段联排队模型,如图 2 所示,其中,ON、OFF 阶段交替轮换,目标业务在 ON 阶段有数据到达,在 OFF 阶段无数据到达;背景业务在 ON 与 OFF 阶段进入 RLC 缓冲队列的速率不同。连续时间被分隔成长度为 Δt ($0 < \Delta t < \infty$) 的等长时间间隔,时刻 t_i 表示第 i 个时间间隔结束的时刻,即 $t_0 = 0$, $t_{i+1} = t_i + \Delta t$ ($i = 0, 1, 2, \dots$),该模型的构成要素及参数如下。

队列 1: 队列 1 代表目标业务在 SDAP 子层的缓冲区,其可容纳 N_1 ($0 < N_1 < \infty$) 个数据单元。当数据单元到达队列 1 时,若队列 1 饱和,则该数据单元被拒绝进入;否则,该数据单元进入队列 1。

队列 2: 队列 2 代表目标业务与背景业务在 RLC 子层共享的缓冲区,其可容纳 N_2 ($0 < N_2 < \infty$) 个数据单元。当数据单元到达队列 2 时,若队列 2 饱和,则该数据单元被拒绝进入;否则,该数据单元进入队列 2。

队列 1 的到达过程: 队列 1 的数据到达过程包括 ON 与 OFF 两个阶段,ON 阶段有数据到达,OFF 阶段无数据到达,且两阶段交替轮换。假设 ON 与 OFF 阶段的单次持续时间均服从伯努利分布(伯努利分布与伯努利过程分别是指数分布与泊松过程的离散化形式),且数据单元在 ON 阶段按伯努利过程(记为 BAP-1)到达。

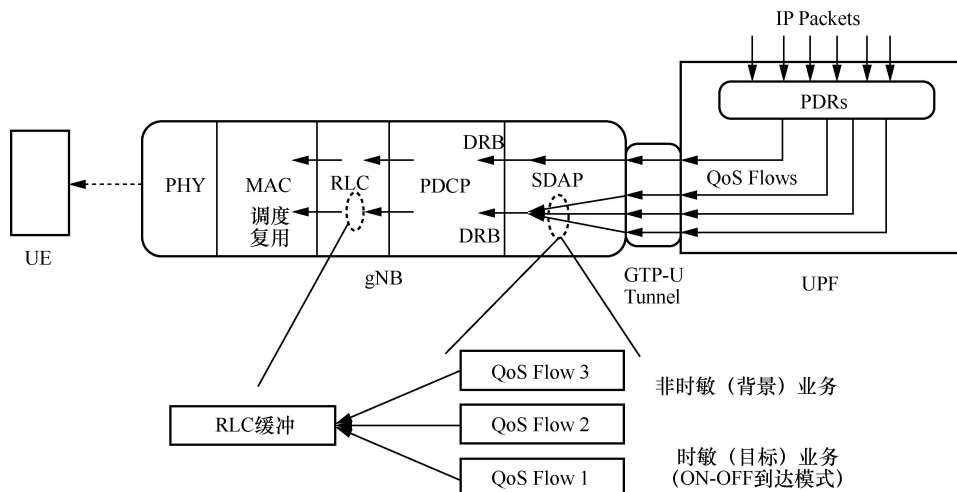


图 1 5G 网络中下行业务数据的传输过程

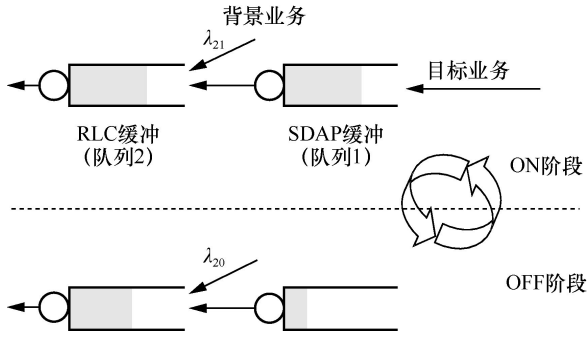


图2 含有附加到达过程的离散时间二阶段级联排队模型

- BAP-1 由参数 p_1 定义, 即 ON 阶段的任意时刻有数据单元到达的概率为 p_1 , 数据单元平均到达率为 $\lambda_1 = p_1/\Delta t$ 。
- ON 阶段单次持续时间所服从的伯努利分布由参数 p_{on} 定义, 在 ON 阶段执行过程中, 每个时间间隔结束时该 ON 阶段以概率 p_{on} 终止; ON 阶段单次平均持续时间为 $\tau_1 = \Delta t/p_{on}$ 。
- OFF 阶段单次持续时间所服从的伯努利分布由参数 p_{off} 定义, 在 OFF 阶段执行过程中, 每个时间间隔结束时该 OFF 阶段以概率 p_{off} 终止; OFF 阶段单次平均持续时间为 $\tau_0 = \Delta t/p_{off}$ 。

队列 2 的附加到达过程: 队列 2 中的数据分为目标业务与背景业务两类, 后者被称为附加数据, 附加数据的到达过程表示背景业务向 RLC 子层交付数据的过程, 也被称为队列 2 的附加到达过程。假设, 在队列 1 数据到达过程的 ON 阶段, 附加数据按伯努利过程(记为 BAP-21)到达队列 2; 在 OFF 阶段, 附加数据也按伯努利过程(记为 BAP-20)到达队列 2。

- BAP-21 由参数 $p_{on,2}$ 定义, 数据单元平均到达率为 $\lambda_{21} = p_{on,2}/\Delta t$ 。
- BAP-20 由参数 $p_{off,2}$ 定义, 数据单元平均到达率为 $\lambda_{20} = p_{off,2}/\Delta t$ 。

队列 1 的服务规则: 队列 1 的数据按先进先出(FIFO, first in first out)顺序被交付到队列 2, 假设交付每个数据单元所需的时间服从参数为 θ_1 的伯努利分布, 平均交付速率为 $\mu_1 = \theta_1/\Delta t$ 。

队列 2 的服务规则: 队列 2 的数据按 FIFO 顺序被成功传输到 UE, 假设每个数据单元成功传输到 UE 所用的时间服从参数为 θ 伯努利分布, 平均传输速率为 $\mu = \theta/\Delta t$ 。

4 随机时延特性的计算方法

计算随机时延特性分为 3 步: 建立离散时间马尔可夫链、计算马尔可夫链的平稳分布以及计算 ON-OFF 时敏业务的随机时延特性。

4.1 建立离散时间马尔可夫链

考虑如图 2 所示排队模型在时刻 $t_i (i=0,1,2,\dots)$ 的状态 $\xi(t_i)$, 构造定义在状态空间 $\Psi = \Psi_1 \cup \Psi_0$ 上的同态马尔可夫链 $\Pi = \{\xi(t_i) : i=0,1,2,\dots\}$, 其中,

$$\Psi_1 = \{(n_1, n_2, 1); n_1 \in S_{N_1}, n_2 \in S_{N_2}\} \quad (1)$$

$$\Psi_0 = \{(n_1, n_2, 0); n_1 \in S_{N_1}, n_2 \in S_{N_2}\} \quad (2)$$

其中, $S_{N_1} = \{0,1,\dots,N_1\}$, $S_{N_2} = \{0,1,\dots,N_2\}$, n_1 与 n_2 分别表示队列 1 与队列 2 的长度(数据单元的个数), 1 与 0 分别表示当前队列 1 的数据到达过程处于 ON 与 OFF 阶段。

马尔可夫链 Π 的一步状态转移概率矩阵 M 的结构如下

$$M = \begin{pmatrix} M_0 & M_1 & 0 & 0 & \dots & 0 & 0 \\ M_2 & M_3 & M_4 & 0 & \dots & 0 & 0 \\ 0 & M_5 & M_3 & M_4 & \dots & 0 & 0 \\ 0 & 0 & M_5 & M_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & M_3 & M_4 \\ 0 & 0 & 0 & 0 & \dots & M_5 & M_3' \end{pmatrix} \quad (3)$$

其中, 矩阵块 $M_0, M_1, M_2, M_3, M_4, M_5$ 与 M_3' 分别表示当队列 1 的长度从 0 转移到 0, 从 0 转移到 1, 从 1 转移到 0, 从 n_1 转移到 $n_1 (n_1 \in \{1,2,\dots,N_1-1\})$, 从 n_1 转移到 $(n_1+1) (n_1 \in \{1,2,\dots,N_1-1\})$, 从 n_1 转移到 $(n_1-1) (n_1 \in \{2,3,\dots,N_1\})$, 从 N_1 转移到 N_1 时, 队列 2 的长度及队列 1 的数据到达过程所处阶段(ON 与 OFF 阶段)的转移情况。

4.2 计算马尔可夫链的平稳分布

令 X 表示马尔可夫链 Π 的平稳分布,

$$X = (X_0, X_1, X_2, \dots, X_{N_1}) \quad (4)$$

其中,

$$X_{n_1} = (X_{n_1,0}, X_{n_1,1}, X_{n_1,2}, \dots, X_{n_1,N_2}), \\ X_{n_1,n_2} = (x_{n_1,n_2}^{(1)}, x_{n_1,n_2}^{(0)}), \quad n_1 \in S_{N_1}, \quad n_2 \in S_{N_2} \quad (5)$$

$x_{n_1, n_2}^{(1)}$ 与 $x_{n_1, n_2}^{(0)}$ 分别表示任意时刻队列 1 的数据到达过程处于 ON 阶段与 OFF 阶段, 且队列 1 的长度为 n_1 , 队列 2 的长度为 n_2 的概率。

\mathbf{X} 满足如下关系

$$\begin{cases} \mathbf{X}\mathbf{M} = \mathbf{X} \\ \mathbf{X}\mathbf{e} = \mathbf{1} \end{cases} \quad (6)$$

其中, \mathbf{e} 表示元素均为 1 的列向量, 其维数可根据上下文确定。 \mathbf{X} 可根据 Grassmann、Taksar 与 Heyman 提出的 GTH 算法^[27-28]直接求解方程组(6)获得。当维数较高时, 根据马尔可夫链 Π 的一步状态转移概率矩阵 \mathbf{M} 的结构, 可采用矩阵几何方法^[29-30]计算 \mathbf{X} 。

4.3 计算 ON-OFF 时敏业务的随机时延特性

任意数据单元 U 从进入队列 1 开始到离开队列 2 所需的时间 n (指 n 个时间间隔) 由 3 部分组成, 分别为: 从 U 进入队列 1 之后开始, 排在 U 前面的数据单元全部离开队列 1 所需的时间 (若 U 进入队列 1 之后即排在队首, 则所需时间为 0); 从 U 排在队列 1 的队首开始到其离开队列 1 所需的时间; 从 U 离开队列 1 之后开始, 到其离开队列 2 所需的时间, 其中, 当 U 离开队列 1 时, 若队列 2 未饱和, 则进入, 否则, 被丢弃 (此时不讨论 n)。

命题 1 任意数据单元从进入队列 1 开始到离开队列 2 所需时间 T 不超过 n ($n \geq 1$) 的概率 $P\{T \leq n\}$ 按如下方法计算, 其中, T 与 n 均表示时间间隔的个数。

$$P\{T \leq n\} = \frac{Q_n}{P_{\text{Enter}_{Q2}|\text{Dep}_{Q1}}} \quad (7)$$

其中, $P_{\text{Enter}_{Q2}|\text{Dep}_{Q1}}$ 表示当数据单元离开队列 1 之后进入队列 2 的概率; Q_n 表示从数据单元进入队列 1 开始, 在 n 个时间间隔之内 (不超过 n 个时间间隔), 该数据单元在队列 2 中完成服务并离开队列 2 的概率。

证明 进入队列 1 的任意数据单元在离开队列 1 之后, 只能以一定概率进入队列 2。因此, 当 n 趋于无穷大时, 从某数据单元进入队列 1 开始, 在 n 个时间间隔之内, 该数据单元在队列 2 中完成服务并离开队列 2 的概率 $Q_\infty \leq 1$ 。 $P\{T \leq n\}$ 可通过将 Q_n 归一化后获得, 即

$$P\{T \leq n\} = \frac{Q_n}{Q_\infty} \quad (8)$$

Q_∞ 等价于当排队模型的状态转移处于平稳过程时, 任意数据单元离开队列 1 之后, 可以进入队列 2 的概率, 即 $Q_\infty = P_{\text{Enter}_{Q2}|\text{Dep}_{Q1}}$ 。

命题 2 任意数据单元从进入队列 1 开始到离开队列 2 所需时间的越界概率 W_n 、概率分布 PW_n 、均值 D_{ave} 与标准差 D_{sd} 计算如下

$$W_n = 1 - P\{T \leq n\}, n \geq 1 \quad (9)$$

$$\text{PW}_n = P\{T \leq n\} - P\{T \leq n-1\}, n \geq 1 \quad (10)$$

$$D_{\text{ave}} = \sum_{n=1}^{\infty} n \cdot \text{PW}_n \quad (11)$$

$$D_{\text{sd}} = \sqrt{\sum_{n=1}^{\infty} \text{PW}_n \cdot (n - D_{\text{ave}})^2} \quad (12)$$

证明 当任意数据单元从进入队列 1 开始到离开队列 2 所需时间的累积概率分布已知时, 越界概率、概率分布、均值与标准差均可按定义直接给出。

根据命题 1 与命题 2, 计算随机时延特性的前提是获得变量 Q_n 与 $P_{\text{Enter}_{Q2}|\text{Dep}_{Q1}}$ 的值, 两者的计算方法如下。

步骤 1 计算 Q_n

1) 建立向量 $\mathbf{Y}^{(m)}$ ($m \geq 0$)。

$$\mathbf{Y}^{(m)} = (\mathbf{Y}_0^{(m)}, \mathbf{Y}_1^{(m)}, \mathbf{Y}_2^{(m)}, \dots, \mathbf{Y}_{N_1}^{(m)}) \quad (13)$$

其中,

$$\mathbf{Y}_{n_1}^{(m)} = (\mathbf{Y}_{n_1,0}^{(m)}, \mathbf{Y}_{n_1,1}^{(m)}, \mathbf{Y}_{n_1,2}^{(m)}, \dots, \mathbf{Y}_{n_1,N_2}^{(m)}), n_1 \in \{1, 2, \dots, N_1\} \quad (14)$$

$$\mathbf{Y}_{n_1, n_2}^{(m)} = (y_{n_1, n_2}^{(m,1)}, y_{n_1, n_2}^{(m,0)}), n_1 \in \{1, 2, \dots, N_1\}, n_2 \in S_{N_2} \quad (15)$$

$$\mathbf{Y}_0^{(m)} = (\mathcal{Y}_0^{(m)}, \mathcal{Y}_0^{(m)}) \quad (16)$$

$$\mathcal{Y}_0^{(m)} = (\mathcal{Y}_{0,0}^{(m)}, \mathcal{Y}_{0,1}^{(m)}, \mathcal{Y}_{0,2}^{(m)}, \dots, \mathcal{Y}_{0,N_2}^{(m)}) \quad (17)$$

$$\mathcal{Y}_{0, n_2}^{(m)} = (\mathcal{Y}_{0, n_2}^{(m,1)}, \mathcal{Y}_{0, n_2}^{(m,0)}), n_2 \in S_{N_2} \quad (18)$$

$$\mathcal{Y}_0^{(m)} = (\mathcal{Y}_{0,0}^{(m)}, \mathcal{Y}_{0,1}^{(m)}, \mathcal{Y}_{0,2}^{(m)}, \dots, \mathcal{Y}_{0, N_2}^{(m)}) \quad (19)$$

$$\mathcal{Y}_{0, n_2}^{(m)} = (\mathcal{Y}_{0, n_2}^{(m,1)}, \mathcal{Y}_{0, n_2}^{(m,0)}), n_2 \in S_{N_2} \quad (20)$$

其中, $y_{n_1, n_2}^{(m,1)}$ 表示从目标数据单元进入队列 1 之后, 经过 m 步转移, 队列 1 与队列 2 的长度分别为 n_1 与 n_2 (目标数据单元进入队列 1 与 2 之后, 新进入队列 1 与 2 的数据单元不被计入队列长度), 且队列 1 的到达过程处于 ON 阶段的概率; 与 $y_{n_1, n_2}^{(m,1)}$ 对应, $y_{n_1, n_2}^{(m,0)}$ 表示队列 1 的到达过程处于 OFF 阶段的概率。

$\mathcal{Y}_{0,n_2}^{(m,1)}$ 表示目标数据单元进入队列 1 之后, 经过 m 步转移, 目标数据单元已经进入队列 2, 且队列 2 的长度为 n_2 , 队列 1 的到达过程处于 ON 阶段的概率; 与 $\mathcal{Y}_{0,n_2}^{(m,1)}$ 对应, $\mathcal{Y}_{0,n_2}^{(m,0)}$ 表示队列 1 的到达过程处于 OFF 阶段的概率。 $Y_{0,n_2}^{(m,1)}$ 表示目标数据单元进入队列 1 之后, 经过 m 步转移, 目标数据单元已经被拒绝进入队列 2, 且队列 2 的长度为 n_2 , 队列 1 的到达过程处于 ON 阶段的概率; 与 $Y_{0,n_2}^{(m,1)}$ 对应, $Y_{0,n_2}^{(m,0)}$ 表示队列 1 到达过程处于 OFF 阶段的概率。

2) 计算初始向量 $Y^{(0)}$ 。

$$Y^{(0)} = (Y_0^{(0)}, Z_1, Z_2, \dots, Z_{N_1}), Y_0^{(0)} = (Z_0, \mathbf{0}_{1 \times r_0}) \quad (21)$$

其中,

$$Z = (Z_0, Z_1, Z_2, \dots, Z_{N_1}), Z = \frac{\hat{Z}}{\hat{Z}e}, \hat{Z} = XM \quad (22)$$

M 表示伴随有数据单元进入队列 1 时, 模型状态的一步转移情况, 其结构与 M 的结构一致。

3) 计算 $\mathcal{Y}_{0,0}^{(n)}$ 。

以 $Y^{(0)}$ 为初始向量, 按关系式(23)~式(26)迭代 n 次, 求得 $Y_0^{(n)}$, 进而获得 $\mathcal{Y}_{0,0}^{(n)}$ 。

$$Y_0^{(n)} = Y_0^{(n-1)} \mathcal{M}_0 + Y_1^{(n-1)} \mathcal{M}_2 \quad (23)$$

$$Y_1^{(n)} = Y_0^{(n-1)} \mathcal{M}_1 + Y_1^{(n-1)} \mathcal{M}_3 + Y_2^{(n-1)} \mathcal{M}_5 \quad (24)$$

$$Y_{n_1}^{(n)} = Y_{n_1}^{(n-1)} \mathcal{M}_3 + Y_{n_1+1}^{(n-1)} \mathcal{M}_5, 2 \leq n_1 \leq N_1 - 1 \quad (25)$$

$$Y_{N_1}^{(n)} = Y_{N_1}^{(n-1)} \mathcal{M}'_3 \quad (26)$$

矩阵 $\mathcal{M}_3, \mathcal{M}_5, \mathcal{M}'_3$ 可通过将 M 中的矩阵块 M_3, M_5, M'_3 做相应修改后获得, 即在目标数据单元进入队列 1 之后, 新进入队列 1 的数据单元不被计入队列 1 的长度。 $\mathcal{M}_1 = \mathbf{0}_{2r_1 \times c_1}$, 其中 r_1 与 c_1 分别表示 M_1 的行数与列数。

$$\mathcal{M}_0 = \begin{pmatrix} \mathcal{M}_0^* & \mathbf{0} \\ \mathbf{0} & I_{r_0 \times c_0} \end{pmatrix} \quad (27)$$

其中, $I_{r_0 \times c_0}$ 表示 $r_0 \times c_0$ 维的单位矩阵, r_0 与 c_0 分别表示 M_0 的行数与列数; \mathcal{M}_0^* 可通过将 M 中的矩阵块 M_0 做相应修改后获得, 即新进入队列 1 的数据单元不计入队列 1 的长度。

$$\mathcal{M}_2 = (\mathcal{M}_2^* \quad \mathcal{M}_2^{**}) \quad (28)$$

其中, \mathcal{M}_2^* 的结构与 M_2 一致, 表示当队列 1 长度从 1 减小为 0, 且离开队列 1 的数据单元进入队列 2 时,

队列 2 的长度及队列 1 的数据到达过程所处阶段的转移情况; \mathcal{M}_2^{**} 的结构与 M_2 一致, 表示当队列 1 长度从 1 减小为 0, 且离开队列 1 的数据单元被拒绝进入队列 2 时, 队列 2 长度及队列 1 的数据到达过程所处阶段的转移情况。

4) 根据关系式 $Q_n = \mathcal{Y}_{0,0}^{(n)} \mathbf{e}$ 计算 Q_n 。

步骤 2 计算 $P_{\text{Enter_Q2|Dep_Q1}}$

$$P_{\text{Enter_Q2|Dep_Q1}} = \frac{P_{\text{Dep_Q1,Enter_Q2}}}{P_{\text{Dep_Q1}}} \quad (29)$$

其中,

$$P_{\text{Dep_Q1}} = \pi \mathbf{e}, \quad \pi = X\bar{M} \quad (30)$$

$$P_{\text{Dep_Q1,Enter_Q2}} = \hat{\pi} \mathbf{e}, \quad \hat{\pi} = X\tilde{M} \quad (31)$$

\bar{M} 表示伴随队列 1 有数据单元离开时, 模型状态的一步转移情况, 其结构与 M 的结构一致; \tilde{M} 表示伴随队列 1 有数据单元离开且进入队列 2 时, 模型状态的一步转移情况, 其结构与 M 的结构一致。

5 非时敏业务对 ON-OFF 时敏业务随机时延特性的影响

本节首先验证 ON-OFF 时敏业务 (目标业务) 随机时延特性计算方法的正确性; 然后, 针对 ON-OFF 轮换频率不同的时敏业务, 分析非时敏业务 (背景业务) 在 ON 与 OFF 阶段向 RLC 子层交付数据的速率对时敏业务随机时延特性的影响。

5.1 时延随机特性计算方法的正确性

根据命题 1 与命题 2, 时延的概率分布、越界概率、均值与标准差均基于累积分布获得, 因此, 验证时延累积分布计算方法的正确性即可。

采用计算方法与仿真方法分别获得时延累积分布函数, 并进行比较, 所使用参数如下。队列 1 数据到达过程 BAP-1 的参数 $p_1 = 0.5$; 队列 1 向队列 2 交付一个数据单元所需的时间服从参数为 $\theta_1 = 0.67$ 的伯努利分布; 队列 2 向 UE 成功传输一个数据单元所需的时间服从参数为 $\theta = 0.83$ 的伯努利分布; ON 与 OFF 阶段的单次持续时间服从相同的伯努利分布, 其参数为 $p_{\text{on}} = p_{\text{off}} = 0.02$; 队列 2 附加数据单元在 ON 与 OFF 阶段的到达过程 BAP-21 与 BAP-20 的参数为 $p_{\text{on},2} = p_{\text{off},2} = 0.5$ 。队列 1 与队列 2 的容量 $N_1 = N_2 = 10$ 。仿真过程采用序贯法^[31], 其中, 置信水平和相对精度分别设置为 98% 与 0.05。

基于仿真方法与计算方法获得的时延累积分布函数如图 3 所示, 由仿真方法与计算方法获得的时延累积分布函数的曲线相互重合; 另外, 通过计算方法获得的概率值均落在了通过仿真方法获得的置信水平为 98% 的置信区间内, 时延累积概率的计算值与仿真值比较见表 1。以上实验结果验证了随机时延特性计算方法的正确性。

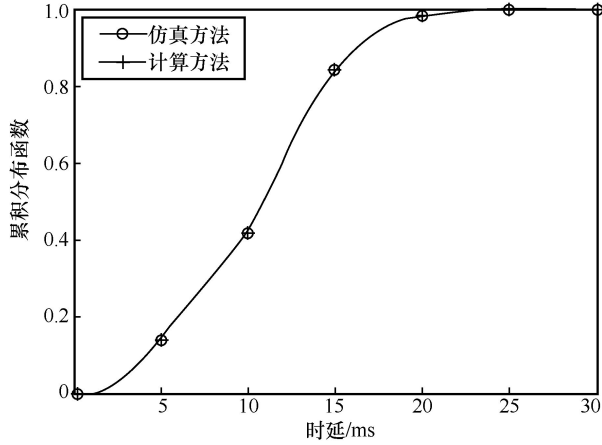


图 3 基于仿真方法与计算方法获得的时延累积分布函数

表 1 时延累积概率的计算值与仿真值比较

时延边界/ms	时延的累积概率		
	计算方法	仿真方法 (置信水平: 98%)	
		概率值	置信区间
5	0.147 9	0.147 8	[0.147 5, 0.148 1]
10	0.429 4	0.429 1	[0.428 6, 0.429 7]
15	0.851 4	0.851 3	[0.851 0, 0.851 7]
20	0.982 1	0.982 1	[0.981 9, 0.982 2]
25	0.998 3	0.998 3	[0.998 2, 0.99 83]
30	0.999 89	0.999 90	[0.999 89, 0.999 90]

5.2 ON-OFF 时敏业务随机时延特性受非时敏业务的影响

数值实验参数见表 2, 其中, λ_1 表示目标业务数据单元到达 SDAP 缓冲队列的平均速率, μ_1 表示 SDAP 缓冲队列向 RLC 子层交付目标业务数据单元的平均速率, μ 表示 RLC 缓冲队列中的数据单元成功传输到 UE 的速率, N_1 与 N_2 分别表示队列 1 与队列 2 的容量, τ_1 与 τ_0 分别表示 ON 与 OFF 阶段的平均持续时间, λ_{21} 与 λ_{20} 分别表示背景业务在 ON 与 OFF 阶段向 RLC 子层交付数据的速率。给定 λ_1 , μ_1 、 μ 、 N_1 与 N_2 , 在 τ_1 与 τ_0 ($\tau_0 = \tau_1$) 取不同值的情况下, 改变 λ_{21} 与 λ_{20} (见表 2), 分析背景业务在 ON 与 OFF 阶段向 RLC 子层交付数据的速率

对目标业务时延的均值、标准差以及概率分布拖尾的影响。

表 2 数值实验参数

参数	取值
$\lambda_1 / (\text{DAUs} \cdot \text{ms}^{-1})$	6
$\mu_1 / (\text{DAUs} \cdot \text{ms}^{-1})$	7
$\mu / (\text{DAUs} \cdot \text{ms}^{-1})$	10
$N_1, N_2 / \text{DAU}$	20
τ_1, τ_0	1 ms, 5 ms, 50 ms, 100 ms, 500 ms, 1 s, 10 s, 20 s, 30 s, 40 s, 50 s, 60 s
$\lambda_{21} / (\text{DAUs} \cdot \text{ms}^{-1})$	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
$\lambda_{20} / (\text{DAUs} \cdot \text{ms}^{-1})$	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

注: $\tau_1 = \tau_0$; $N_1 = N_2$; 离散时间间隔 $\Delta t = 0.05$ ms, DAU 表示数据单元 (data unit)。

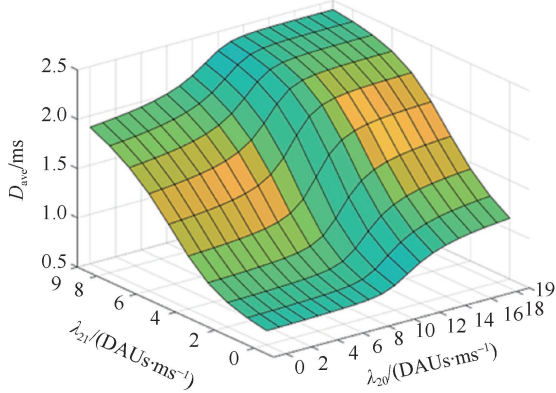
1) 平均时延受非时敏业务的影响

平均时延 D_{ave} 随 λ_{21} 与 λ_{20} 的变化如图 4 所示, 实验结果显示: 当 τ_0 与 τ_1 在 {1 ms, 5 ms, 50 ms, 100 ms} 中取值时, 随着 λ_{21} 或 λ_{20} 从 0 开始增大, 平均时延按“慢→快→慢”的增长率逐渐增长并趋于平稳, 以 $\tau_0 = \tau_1 = 5$ ms 的情况为例, 如图 4(a) 所示。当 τ_0 与 τ_1 取其他值时, λ_{20} 的变化对平均时延无显著影响 (λ_{20} 从 0 增大到 19, 平均时延的增幅小于 0.05 ms); 随着 λ_{21} 从 0 开始增大, 平均时延按“慢→快→慢”的增长率逐渐增长并趋于平稳, 以 $\tau_0 = \tau_1 = 10$ s 的情况为例, 如图 4(b) 所示。

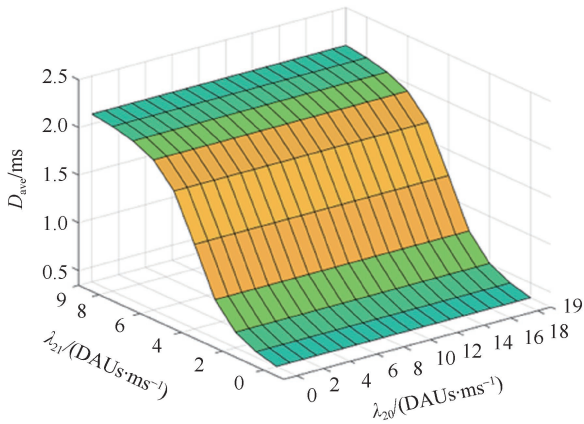
基于上述实验结果可知, 当 ON 与 OFF 阶段的轮换频率较小时 (ON 与 OFF 阶段的持续时间较长), 背景业务在 OFF 阶段向 RLC 子层交付数据的速率对目标业务的平均时延无显著影响; 但随着背景业务在 ON 阶段向 RLC 子层交付数据的速率逐渐增大, 目标业务的平均时延按“慢→快→慢”的增长率逐渐增长并趋于平稳。当 ON 与 OFF 阶段的轮换频率较大时 (ON 与 OFF 阶段的持续时间较短), 随着背景业务在 ON 或 OFF 阶段向 RLC 子层交付数据的速率逐渐增大, 目标业务的平均时延均按“慢→快→慢”的增长率逐渐增长并趋于平稳。

这是因为, 不论 ON 与 OFF 阶段轮换频率的大小, 随着背景业务在 ON 阶段向 RLC 子层交付数据的速率逐渐增大, 当目标业务数据进入 RLC 缓冲队列时, 排在其前面的背景业务数据单元逐渐增多, 所以目标业务的平均时延逐渐增大; 随着 RLC

缓冲队列趋于饱和,排在目标业务数据单元前面的背景业务数据单元的数量趋于稳定,所以,目标业务的平均时延逐渐趋于平稳。当背景业务与目标业务向 RLC 子层交付数据的速率之和接近数据离开 RLC 缓冲队列的速率时,RLC 缓冲队列长度的增幅较明显,所以此时目标业务的平均时延增长较快。



(a) $\tau_0=\tau_1=5\text{ ms}$



(b) $\tau_0=\tau_1=10\text{ s}$

图4 平均时延 D_{ave} 随 λ_{21} 与 λ_{20} 的变化

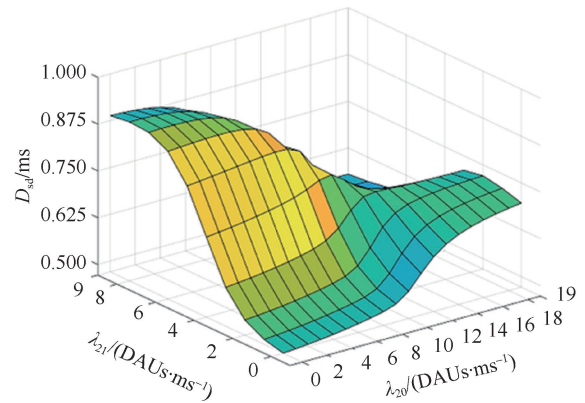
若 ON 与 OFF 阶段切换频率较大,以上描述也可用于解释目标业务平均时延随背景业务在 OFF 阶段向 RLC 子层交付数据速率的变化规律。

若 ON 与 OFF 阶段切换频率较小,当目标业务数据进入 RLC 缓冲队列时,在 OFF 阶段进入 RLC 缓冲队列的背景业务数据仍然留在该队列中的概率较小,所以,背景业务在 OFF 阶段向 RLC 子层交付数据的速率对目标业务的平均时延无显著影响。

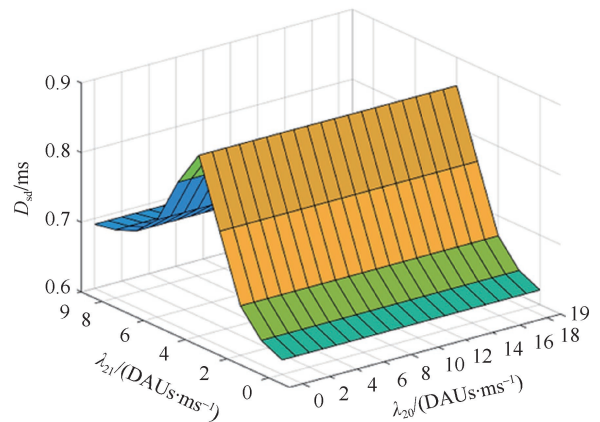
2) 时延标准差受非时敏业务的影响

时延标准差 D_{sd} 随 λ_{21} 与 λ_{20} 的变化如图 5 所示,实验结果显示:当 τ_0 与 τ_1 在 $\{1\text{ ms}, 5\text{ ms}, 50\text{ ms}, 100\text{ ms}\}$ 中取值时,随着 λ_{21} (λ_{20}) 从 0 开始增大,

标准差先增大再减小并趋于平稳, λ_{20} (λ_{21}) 越大拐点越早出现,以 $\tau_0 = \tau_1 = 5\text{ ms}$ 的情况为例,如图 5(a)所示(当 λ_{20} (λ_{21}) 较大时,随着 λ_{21} (λ_{20}) 从 0 开始增大,存在标准差逐渐减小并趋于平稳的情况,此时,减小开始发生的拐点即是起点;当 λ_{20} (λ_{21}) 较小时,随着 λ_{21} (λ_{20}) 从 0 开始增大,存在标准差逐渐增大并趋于平稳的情况,此时,减小开始发生的拐点将不会出现)。当 τ_0 与 τ_1 取其他值时, λ_{20} 的变化对时延标准差无显著影响 (λ_{20} 从 0 增大到 19,时延标准差的增幅小于 0.05 ms);随着 λ_{21} 从 0 开始增大,标准差逐渐增大,在 $\lambda_{21} = 4$ 时达到最大,之后逐渐减小,以 $\tau_0 = \tau_1 = 10\text{ s}$ 的情况为例,如图 5(b)所示。



(a) $\tau_0=\tau_1=5\text{ ms}$



(b) $\tau_0=\tau_1=10\text{ s}$

图5 时延标准差 D_{sd} 随 λ_{21} 与 λ_{20} 的变化

基于上述实验结果可知,当 ON 与 OFF 阶段的切换频率较小时,背景业务在 OFF 阶段向 RLC 子层交付数据的速率对目标业务的时延标准差无显著影响;但随着背景业务在 ON 阶段向 RLC 子层交付数据的速率逐渐增大,目标业务时延标准差先增

大后减小, 其中, 当背景业务与目标业务向 RLC 子层交付数据的速率之和等于数据离开 RLC 缓冲队列的速率时出现拐点。当 ON 与 OFF 阶段的切换频率较大时, 随着背景业务在 ON 或 OFF 阶段向 RLC 子层交付数据的速率逐渐增大, 目标业务时延的标准差先增大后减小, 其中, 拐点在给定区域内出现, 该区域中背景业务与目标业务向 RLC 子层交付数据的速率之和不小于数据离开 RLC 缓冲队列的速率。

这是因为, 不论 ON 与 OFF 阶段切换频率的大小, 随着背景业务在 ON 阶段向 RLC 子层交付数据的速率逐渐增大, 当目标业务数据进入 RLC 缓冲队列时, 该队列长度的波动逐渐增大, 所以目标业务时延标准差逐渐增大; 随着 RLC 缓冲队列趋于饱和, RLC 缓冲队列长度的波动逐渐平稳, 所以目标业务时延标准差随之减小。

若 ON 与 OFF 阶段切换频率较大, 以上描述也可用于解释目标业务时延标准差随背景业务在 OFF 阶段向 RLC 子层交付数据速率的变化规律。

若 ON 与 OFF 阶段切换频率较小, 当目标业务数据进入 RLC 缓冲队列时, 在 OFF 阶段进入 RLC 缓冲队列的背景业务数据仍然留在该队列中的概率较小, 所以, 背景业务在 OFF 阶段向 RLC 子层交付数据的速率对目标业务时延标准差无显著影响。

3) 时延概率分布的拖尾受非时敏业务的影响

将目标业务时延的概率分布与具有相同均值及方差的伽马分布及正态分布相比较, 实验结果显示: 目标业务时延概率分布的右拖尾长度近似等于伽马分布的右拖尾长度, 即当目标业务时延的均值与方差确定时, 其时延波动的上界可通过具有相同均值及方差的伽马分布来近似给出。下面以 $\tau_0 = \tau_1 = 10\text{ s}$ 与 $\tau_0 = \tau_1 = 5\text{ ms}$ 两种情况为例列举实验结果。

当 $\tau_0 = \tau_1 = 10\text{ s}$ 时, λ_{20} 的变化对时延概率分布无显著影响。以 $\lambda_{20} = 10$, λ_{21} 分别取 4 与 6 为例, 如图 6 所示, 当 $\lambda_{21} + \lambda_1 \leq \mu$, 以 $\lambda_{21} = 4$ 为例, 目标业务时延概率分布的右拖尾长度与伽马分布的右拖尾长度基本一致; 当 $\lambda_{21} + \lambda_1 > \mu$ 时, 以 $\lambda_{20} = 6$ 为例, 伽马分布的右拖尾长度略小。当伽马分布的右拖尾长度略小时, 用其计算目标业务的时延上界时, 所得时延上界会略小, 但误差并不显著。

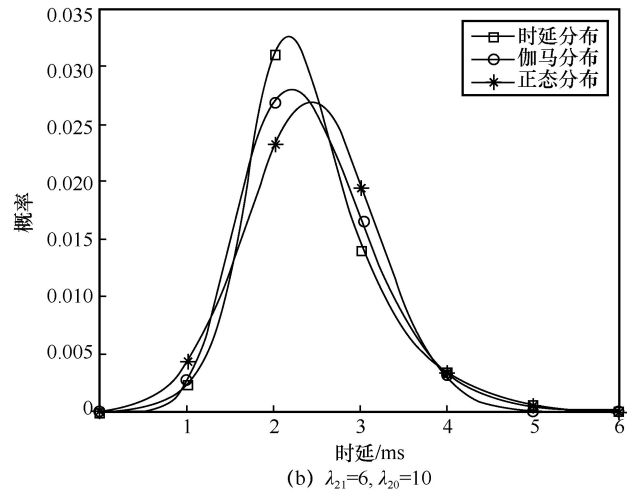
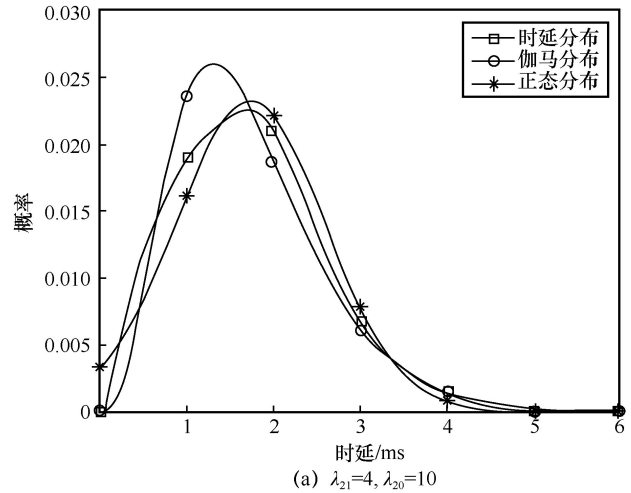


图 6 时延概率分布与伽马分布及正态分布的拖尾比较 ($\tau_0 = \tau_1 = 10\text{ s}$)

当 $\tau_0 = \tau_1 = 5\text{ ms}$ 时, 时延概率分布与伽马分布及正态分布的拖尾比较如图 7 所示。目标业务时延概率分布的右拖尾长度与伽马分布的右拖尾长度或基本一致 (如图 7(a)), 或后者略大 (如图 7(b)、图 7(c)), 或后者略小 (如图 7(d))。当伽马分布的右拖尾长度略大时, 用其计算目标业务的时延上界时, 所得时延上界会略大, 这样可为时延保障策略的制定提供较强的约束。

6 结束语

在 5G 接入网中, 当时敏与非时敏业务共享无线资源和 RLC 缓冲队列时, 基于 QoS 优先级调度的接入网资源共享方案仍缺乏有效的排队时延控制方法, 以致无法满足高优先级时敏业务的确定性时延需求。本文针对接入网中 ON-OFF 时敏业务与非时敏业务共享数据无线承载和 RLC 缓冲区的情况, 建立 ON-OFF 时敏业务在接入网下行链路的随机时延特性计算方法, 可计算时延的均值、标准差、

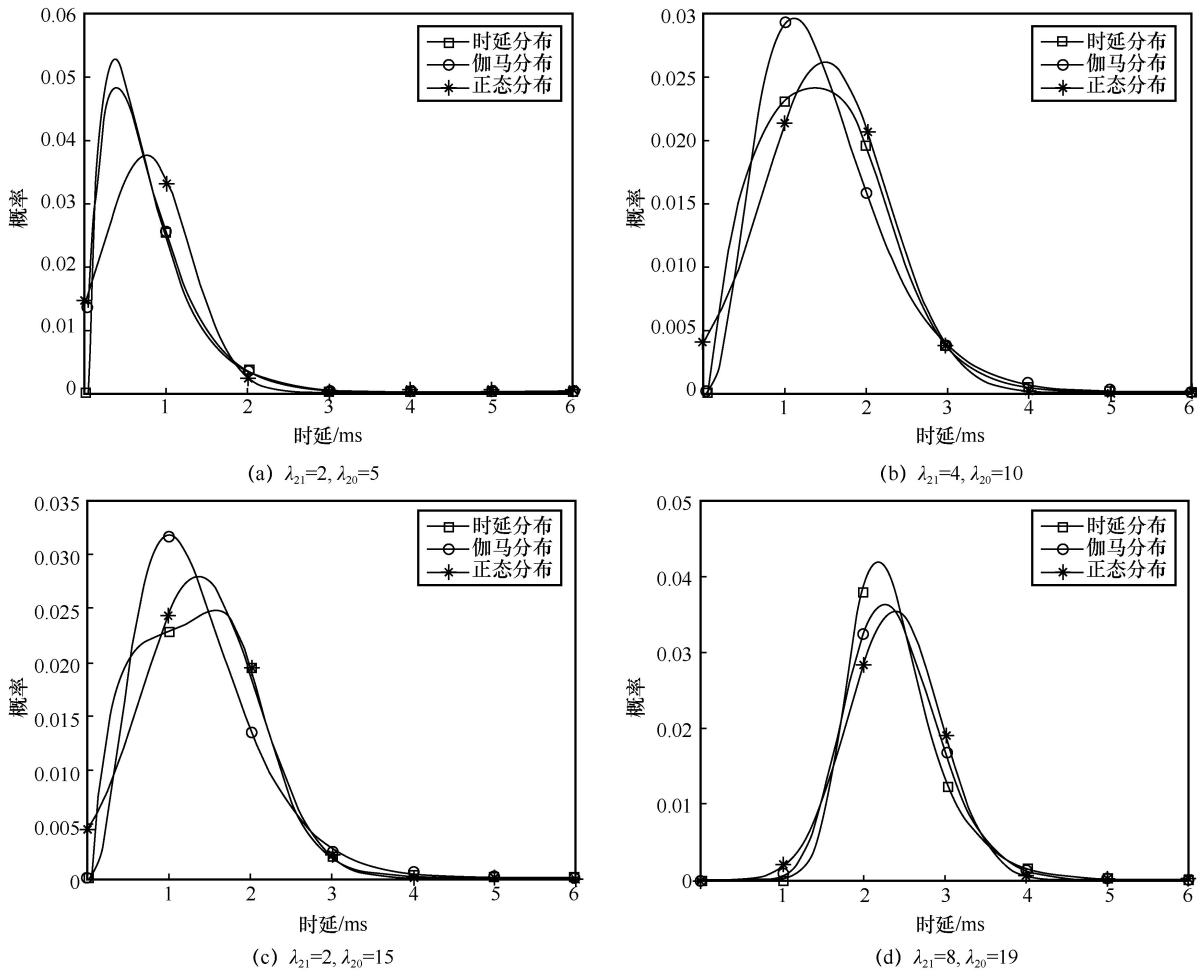


图7 时延概率分布与伽马分布及正态分布的拖尾比较 ($\tau_0 = \tau_1 = 5 \text{ ms}$)

越界概率以及概率分布，并分析共享数据无线承载的非时敏业务对 ON-OFF 时敏业务随机时延特性的影响。相关计算方法与分析结果可用于建立 ON-OFF 时敏业务的排队时延控制方法，使基于 QoS 优先级调度的资源共享方案能够有效保障混合业务中 ON-OFF 时敏业务的确定性时延需求。

参考文献:

[1] 宋光敏, 王群青, 黄占兵, 等. 赋能 5G 垂直行业: 基于端边协同保障时延确定性技术研究“5G 行业应用”[EB]. 2022. SONG G, WANG Q, HUANG Z, et al. Empowering 5G vertical industries: research on the edge-end collaboration framework for deterministic network “5G industry applications”[EB]. 2022.

[2] WANG Y M, FENG W, WANG J, et al. Hybrid satellite-UAV-terrestrial networks for 6G ubiquitous coverage: a maritime communications perspective[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(11): 3475-3490.

[3] 冯伟, 唐睿, 葛宁. 星地协同智能海洋通信网络发展展望[J]. 电信科学, 2020, 36(10): 1-11. FENG W, TANG R, GE N. Perspectives on coordinated satel-

lite-terrestrial intelligent maritime communication network[J]. Telecommunications Science, 2020, 36(10): 1-11.

[4] 中国移动研究院. 6G 全息通信业务发展趋势白皮书[R]. 2022. China Mobile Research Institute. White paper on development trend of 6G holographic communication service[R]. 2022.

[5] ITU FG-NET-2030. Network 2030 a blueprint of technology, applications and market drivers towards the year 2030 and beyond[R]. 2019.

[6] HUANG Y, WANG S, HUANG T, et al. Cycle-based time-sensitive and deterministic networks: architecture, challenges, and open issues[J]. IEEE Communications Magazine, 2022, 60(6): 81-87.

[7] WEI P, GUO K, LI Y, et al. Reinforcement learning-empowered mobile edge computing for 6G edge intelligence[J]. IEEE Access, 10: 65156-65192.

[8] LIU C X, FENG W, TAO X M, et al. MEC-empowered non-terrestrial network for 6G wide-area time-sensitive Internet of Things[J]. Engineering, 2022, 8: 96-107.

[9] ZHANG X, WANG J Q, POOR H V. Statistical delay and error-rate bounded QoS provisioning over mmWave cell-free M-MIMO and FBC-HARQ-IR based 6G wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(8): 1661-1677.

[10] ZHANG X, WANG J Q, POOR H V. Statistical delay and error-rate bounded QoS control for URLLC in the non-asymptotic re-

- gime[C]//Proceedings of 2022 IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2022: 2112-2117.
- [11] QI T, FENG W, CHEN Y F, et al. Double QoS guarantee for NO-MA-enabled massive MTC networks[J]. IEEE Internet of Things Journal, 2022, PP(99): 1.
- [12] 5G 确定性网络产业联盟. 5G 确定性网络架构产业白皮书[R]. 2021. 5G Deterministic Networking Alliance. 5G deterministic networking architecture industry white paper[R]. 2021.
- [13] 中国通信学会. 5G 专网前沿报告[R]. 2022. China Institute of Communication. The frontier report for 5G private network[R]. 2022.
- [14] 徐文伟. 后香农时代, 数学决定未来发展的边界[R]. 中国工业与应用数学学会数学促进企业创新发展论, 2020. XU W W. In the post Shannon era, mathematics determines the boundary of future development[R]. China Society for Industrial and Applied Mathematics, Forum on Mathematics Promoting Enterprise Innovation and Development, 2020.
- [15] NICHOLS K, JACOBSON V, MCGREGOR A, et al. Controlled delay active queue management: IETF RFC 8289[S]. 2018.
- [16] JUNG S, KIM J, KIM J H. Intelligent active queue management for stabilized QoS guarantees in 5G mobile networks[J]. IEEE Systems Journal, 2021, 15(3): 4293-4302.
- [17] HOEILAND-JOERGENSEN T, MCKENNEY P, TAHT D, et al. The flow queue CoDel packet scheduler and active queue management algorithm: IETF RFC 8290[S]. 2018.
- [18] Best routers for 2021[EB]. 2020.
- [19] KUMAR R, FRANCINI A, PANWAR S, et al. Dynamic control of RLC buffer size for latency minimization in mobile RAN[C]//Proceedings of 2018 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2018: 1-6.
- [20] IMPUTATO P, PATRICIELLO N, AVALLONE S, et al. Smart backlog management to fight buffer bloat in 3GPP protocol stacks[C]//Proceedings of 2019 16th IEEE Annual Consumer Communications & Networking Conference. Piscataway: IEEE Press, 2019: 1-6.
- [21] BQL: byte queue limits[EB]. 2011.
- [22] IRAZABAL M, LOPEZ-AGUILERA E, DEMIRKOL I, et al. Dynamic buffer sizing and pacing as enablers of 5G low-latency services[J]. IEEE Transactions on Mobile Computing, 2022, 21(3): 926-939.
- [23] IRAZABAL M, LOPEZ-AGUILERA E, DEMIRKOL I, et al. Preventing RLC buffer sojourn delays in 5G[J]. IEEE Access, 2021(9): 39466-39488.
- [24] 3GPP. Technical specification group radio access network; continuous connectivity for packet data users; 1.28 Mcps TDD:3GPP TR 25.929 V12.0.0[S]. 2014.
- [25] ETSI. 5G; NR; radio resource control (RRC) protocol specification: ETSI TS 138 331 V15.5.1[S]. 2019.
- [26] ETSI. 5G; study on scenarios and requirements for next generation access technologies: ETSI TR 138 913 V15.0.0[S]. 2018.
- [27] GRASSMANN W K, TAKSAR M I, HEYMAN D P. Regenerative analysis and steady state distributions for Markov chains[J]. Operations Research, 1985, 33(5): 1107-1116.
- [28] CAO J Y, XIE W X. Optimization of a condition-based duration-varying preventive maintenance policy for the stockless production system based on queueing model[J]. Journal of Industrial & Management Optimization, 2017, 13(5): 1-11.
- [29] CAO J Y, FENG W, GE N, et al. Delay characterization of mobile-edge computing for 6G time-sensitive services[J]. IEEE Internet of Things Journal, 2021, 8(5): 3758-3773.
- [30] ALFA A S. Applied discrete-time queues[M]. New York: Springer-Verlag, 2016.
- [31] ZHANG H. Analytical solution to a discrete-time model for dynamic learning and decision making[J]. Management Science, 2022, 68(8): 5924-5957.

[作者简介]



曹建宇 (1987-), 男, 博士, 桂林电子科技大学副研究员, 主要研究方向为智能无线通信、移动确定性网络、排队理论等。



冯伟 (1982-), 男, 博士, 清华大学研究员, 主要研究方向为广域通信、机器通信等。



葛宁 (1971-), 男, 博士, 清华大学研究员, 主要研究方向为通信片上系统、短距离无线通信、宽带无线网络等。