

文章编号:1007-5321(2018)05-0098-05

DOI:10.13190/j.jbupt.2018-155

# URLLC 业务概率时延约束及资源预留的分析与研究

高月红, 李 岚, 张 欣, 杨鸿文

(北京邮电大学 信息与通信工程学院, 北京 100876)

**摘要:** 分析了第5代移动通信系统(5G)中超可靠低时延通信(URLLC)业务与增强移动带宽(eMBB)业务共存时,资源预留分配策略对URLLC业务的时延保障程度以及对eMBB业务服务质量的影响程度.基于随机网络演算分析理论,推导了满足URLLC业务概率时延约束要求的前提下所需的资源预留比例,即满足URLLC业务时延要求的最小门限值.此外,分析了eMBB业务的时延性能.研究表明,在系统总资源固定的情况下,为URLLC业务预留一定比例的资源后,eMBB业务的时延分布会受到一定程度的影响,其时延分布的恶化程度与资源预留比例密切相关.

**关键词:** 第5代移动通信系统;低时延高可靠;资源预留;概率时延约束

中图分类号: TN929.53

文献标志码: A

## Analysis and Research on the Probabilistic Delay Constraint and Resource Reservation for URLLC Traffic

GAO Yue-hong, LI Lan, ZHANG Xin, YANG Hong-wen

(School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** The delay guarantee of ultra-reliable and low-latency communication (URLLC) traffic under resource reservation policy was analyzed when both URLLC traffic and enhanced mobile broadband (eMBB) traffic co-exist in the fifth generation of mobile communications system (5G) system. The impact on eMBB traffic's quality of service was also analyzed. Based on the stochastic network calculus theory, the needed resource reservation ratio for URLLC traffic with probabilistic delay constraint requirement was investigated. Specifically, a minimum threshold, meeting this delay requirement for URLLC traffic, was derived. In addition, the delay performance of eMBB traffic was also analyzed. The results showed that, when the system resource was fixed, the delay distribution of eMBB traffic would be influenced by the amount of resource reserved for URLLC traffic, and the deteriorate level greatly depended on the reservation ratio.

**Key words:** the fifth generation of mobile communications system; ultra-reliable low latency; resource reservation; probabilistic delay constraint

第5代移动通信系统(5G, the fifth generation of mobile communications system)将渗透到未来生活的各个方面,为未来全球的信息化建设提供坚实的基础<sup>[1-2]</sup>.因此,针对5G的标准化进程以及关键技术

的研究显得尤为重要.从信息交互对象不同的角度,5G系统将涵盖以下3大场景<sup>[1-2]</sup>:增强移动带宽(eMBB, enhanced mobile broadband)、超可靠低时延通信(URLLC, ultra-reliable and low-latency commu-

收稿日期: 2018-07-13

基金项目: 国家科技重大专项项目(2018ZX03001024-006)

作者简介: 高月红(1981—),女,副教授,硕士生导师, E-mail: yhgao@bupt.edu.cn.

nication)和海量机器类通信。这 3 类场景所针对的具体应用场合不同、业务特点不同,对 5G 网络提出的技术需求也不同。当多种业务在 5G 网络中共存时,如何在业务之间进行资源分配和复用,进而满足各业务的技术指标,就成为 5G 网络必须解决的关键问题之一。其中,eMBB 业务与 URLLC 业务预计在 5G 中期(即 2021—2022 年)发展成熟,因此,针对 eMBB 和 URLLC 的复用技术研究是近些年的研究重点。eMBB 业务是未来 5G 网络中的基础业务,占据主导地位,其特点是数据量大,对传输速率的要求较高。相比之下,URLLC 业务的数据包较小,对时延和错误率的要求都很严格。因此,在已有 eMBB 业务的前提下,如何对资源进行合理分配,进而能够保证 URLLC 业务的服务质量要求是一个亟待解决的问题。

目前已有的研究可以分为 2 个方面:标准化研究和学术研究。标准化研究是以提案的形式给出,重点关注 URLLC/eMBB 帧结构设计<sup>[3-4]</sup>、URLLC 与 eMBB 业务复用机制<sup>[5-6]</sup>、仿真参数配置<sup>[7]</sup>等系统实现相关的内容,部分提案给出了针对错误率的仿真结果<sup>[6-8]</sup>。学术研究以论文的形式给出,通常以特定场景和具体算法实现为落脚点,如干扰协调问题<sup>[9]</sup>、用于增强 URLLC 业务的安全性和可靠性的正交频分复用子载波索引选择技术<sup>[10]</sup>、存在大量、突发性 URLLC 数据包时的业务检测和冲突避免技术<sup>[11-12]</sup>、链路自适应与资源调度的联合算法<sup>[13]</sup>等。已有研究缺少一般性理论建模,未能针对 URLLC 业务的时延特性进行深入探讨。基于此,笔者依托随机网络演算理论,建立用于描述 URLLC 业务和 eMBB 业务复用场景的通用模型,通过分析概率时延分布情况,从理论上推导保证 URLLC 业务时延要求的资源预留边界,为系统设计及性能优化提供理论依据。

笔者研究了资源预留模式下 URLLC 业务和 eMBB 业务的资源分配占比,通过定义时延受限吞吐量将 URLLC 业务的时延和错误率进行联合考虑。利用最小加代数和网络演算分析方法,推导满足一定概率时延约束条件下,系统需要为 URLLC 业务预留资源的下界,同时分析资源预留给 eMBB 业务带来的性能损失。

## 1 系统模型

如图 1 所示的移动通信系统,eMBB 业务和

URLLC 业务以资源预留的方式共享系统资源,其中预留给 URLLC 业务的资源占比为  $\eta$ ,预留给 eMBB 业务的资源占比为  $1 - \eta$ 。eMBB 业务和 URLLC 业务的数据包均依据泊松过程到达系统,到达强度分别为  $\lambda_e$  和  $\lambda_u$ ,数据包的大小分别为  $L_e$  和  $L_u$ 。

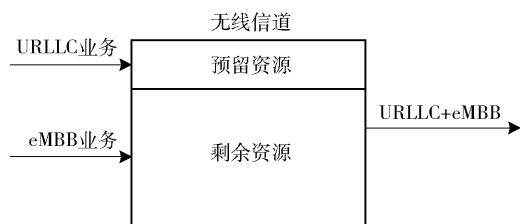


图 1 系统模型

文中所提及的时延定义为从数据包产生到数据包被传输结束的时间差。当数据包经历的时延大于可容忍的最大时延时,该包将被丢弃。被丢弃的数据包,对于接收端而言,即视为一次错误接收。虽然实际系统接收端的错误接收既包含丢包造成的错误,也包含由于信道条件等原因导致的接收失败引起的错误<sup>[14-15]</sup>,但是文中所提及的错误率只考虑丢包引入的接收失败。在此基础上,定义概率时延约束实现对时延和错误率的联合考虑。

**定义 1** 概率时延约束。若业务可容忍的最大错误率为  $\varepsilon$ ,其数据包可容忍的最大时延为  $d$ ,数据包所经历的实际时延为  $D(t)$ ,则概率时延约束定义为

$$P\{D(t) > d\} \leq \varepsilon \quad (1)$$

当 URLLC 业务与 eMBB 业务以资源预留模式复用时,预留资源占比  $\eta$  越大,则对 URLLC 业务的服务质量保证越好,但是对 eMBB 业务的影响也最大。因此,资源预留所需解决的根本问题是:找到最小的资源预留占比  $\eta$ ,在保证 URLLC 业务的概率时延约束的前提下,将对 eMBB 业务的影响降到最小。令  $D_u(t)$ 、 $d_u$  和  $\varepsilon_u$  分别表示 URLLC 数据包的实际时延、可容忍的最大时延和可容忍的最大错误率,则资源预留方案的最优化问题可以归结为

$$\begin{aligned} \min \quad & \eta \\ \text{s. t.} \quad & P\{D_u(t) > d_u\} \leq \varepsilon_u \end{aligned} \quad (2)$$

## 2 随机网络演算基础

在分析业务可以获得的服务质量保证之前,首先需要完成对业务特性以及网络服务能力的数学描述。笔者基于最小加代数和随机网络演算理论所定义的随机到达曲线和随机服务曲线,完成所需

服务质量保证的分析。

网络演算理论是用于分析网络中流问题的一种方法<sup>[16]</sup>,自1991年提出以来,已逐渐演进为2个分支:确定性网络演算<sup>[17]</sup>和随机网络演算<sup>[18]</sup>。本项目研究的是随机性较强的无线网络问题,依托随机网络演算理论进行。

随机网络演算的系统模型如图2所示,包括随机到达曲线 $A(t) \sim \langle \alpha, f \rangle$ 、随机服务曲线 $S(t) \sim \langle \beta, g \rangle$ 以及输出数据流 $A^*(t)$ 3个部分。在获得随机到达曲线和随机服务曲线的具体数学描述的基础上,即可实现对输出数据流的特性分析。

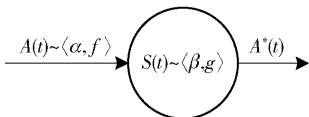


图2 随机网络演算系统模型

随机到达曲线由定义在累积数据量基础上的一个关于时间的函数和一个关于数据量的边界函数构成。

**定义2** 随机到达曲线。令 $A(s, t)$ 表示业务流 $A$ 在时间段 $(s, t]$ 内累积到达系统的数据量,若 $\forall t \geq 0$ 和 $\forall x \geq 0$ ,函数 $\alpha(t)$ 和函数 $f(x)$ 满足

$$P\left\{\sup_{0 \leq s \leq t} \{A(s, t) - \alpha(t-s)\} > x\right\} \leq f(x) \quad (3)$$

则称函数 $\alpha(t)$ 和函数 $f(x)$ 为业务流 $A$ 的随机到达曲线,记为 $A(t) \sim \langle \alpha, f \rangle$ ,其中 $A(t) = A(0, t)$ , $\sup\{\cdot\}$ 表示上确界。

与随机到达曲线的概念类似,随机服务曲线由一组用于描述服务能力的函数对构成,定义如下。

**定义3** 随机服务曲线。令 $A^*(t)$ 表示在时间段 $(0, t]$ 内累积输出服务器 $S$ 的数据量,若 $\forall t \geq 0$ 和 $\forall x \geq 0$ ,函数 $\beta(t)$ 和函数 $g(x)$ 满足

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x) \quad (4)$$

则称函数 $\beta(t)$ 和函数 $g(x)$ 为服务器 $S$ 的随机服务曲线,记为 $S(t) \sim \langle \beta, g \rangle$ ,其中 $A \otimes \beta(t) \equiv \inf_{0 \leq s \leq t} \{A(s) + \beta(t-s)\}$ , $\inf\{\cdot\}$ 为下确界。

在已知业务流的到达特性和服务器的服务特性的前提下,利用最小加代数理论可以获得时延的概率分布的上界。下面直接将相关结论总结在定理1中,详细的推导请参见文献[18]。

**定理1** 时延分布上界。考虑服务器 $S$ 为数据流 $A$ 提供数据传输。若已知数据流 $A$ 的随机到达曲线是 $A(t) \sim \langle \alpha, f \rangle$ ,服务器 $S$ 的随机服务曲线是

$S(t) \sim \langle \beta, g \rangle$ ,则 $\forall t \geq 0$ 和 $\forall x \geq 0$ ,在 $t$ 时刻产生的数据包の時延 $D(t) = \inf\{\tau \geq 0: A(t) \leq A^*(t+\tau)\}$ 的分布受限干

$$P\{D(t) \geq h(\alpha+x, \beta)\} \leq f \otimes g(x) \quad (5)$$

其中 $h(\alpha+x, \beta)$ 表示函数 $\alpha(t)+x$ 与函数 $\beta(t)$ 之间的最大水平距离,可以进一步写为

$$h(\alpha+x, \beta) = \sup_{t \geq 0} \left\{ \inf_{\tau \geq 0} \{ \alpha(t) + x \leq \beta(t+\tau) \} \right\} \quad (6)$$

### 3 URLLC 业务资源预留上界分析

考虑如图1所示的移动通信系统,将其中各部分所对应的随机网络演算描述引入后,如图3所示。

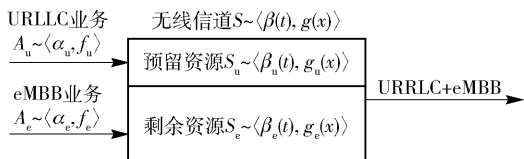


图3 系统模型的随机网络演算描述

设系统可以提供的平均传输速率为 $R$ ,则系统的随机服务曲线可以表示为 $S \sim \langle \beta(t) = Rt, g(x) = 0 \rangle$ 。考虑资源预留后,URLLC业务和eMBB业务可以真正获得的服务能力分别为

$$S_u \sim \langle \beta_u(t) = \eta Rt, g_u(x) = 0 \rangle \quad (7)$$

$$S_e \sim \langle \beta_e(t) = (1-\eta)Rt, g_e(x) = 0 \rangle \quad (8)$$

泊松过程是用于描述业务到达特性的一种典型类型,在通信系统建模中被广泛使用。在泊松过程中,2个数据包到达系统的时间间隔呈负指数分布,时间间隔的均值为 $1/\lambda$ ,即单位时间内到达系统的数据包的平均个数为 $\lambda$ ,称为到达率。URLLC业务和eMBB业务均可以用泊松到达过程来模拟。已知泊松到达过程的随机到达曲线如下<sup>[6]</sup>。

**定义4** 泊松到达过程的随机到达曲线。设业务流的所有数据包的长度 $L$ 均相同,并且数据包依据泊松过程到达系统,到达率为 $\lambda$ ,对于任意的 $r > \lambda L$ ,随机到达曲线可以表示为

$$\alpha(t) = rt$$

$$f(x) = 1 - (1-a) \sum_{i=0}^k \left[ \frac{[a(i-k)]^i}{i!} e^{-a(i-k)} \right] \quad (9)$$

其中: $a = \lambda L/r, k = \lceil x/L \rceil$ 。

用 $A_u \sim \langle \alpha_u, f_u \rangle$ 表示URLLC业务的随机到达曲线, $\alpha_u(t) = r_u t$ ,则依据定理1,可以获得URLLC业务的时延分布上界为

$$P\left\{D_u(t) \geq \frac{x}{\eta R}\right\} \leq f_u(x) \quad (10)$$

若 URLLC 数据包可容忍的最大时延为  $d_u$ 、业务可容忍的最大错误率为  $\varepsilon_u$ , 则可知

$$P\{D_u(t) \geq d_u\} \leq f_u(\eta R d_u) = \varepsilon_u \quad (11)$$

当系统平均服务速率给定时, 保证 URLLC 业务概率时延约束的最小资源预留占比为

$$\eta_{\min} = \frac{f_u^{-1}(\varepsilon_u)}{R d_u} \quad (12)$$

4 eMBB 业务性能损失分析

当系统中所有资源可供 eMBB 业务使用时, 依据定理 1 可知 eMBB 数据包的时延分布上界为

$$P\left\{D_e(t) \geq \frac{x}{R}\right\} \leq f_e(x) \quad (13)$$

即

$$P\{D_e(t) \geq d_e\} \leq f_e(R d_e) \quad (14)$$

假设 eMBB 数据包的长度为  $L_e$ , 以泊松过程到达, 强度为  $\lambda_e$ , 随机到达曲线为  $A_e \sim \langle \alpha_e, f_e \rangle, \alpha_e(t) = r_e t$ .

当系统为 URLLC 业务预留资源时, eMBB 业务的实际可用资源由原来的  $R$  缩减为  $(1 - \eta)R$ , 因此, 其概率时延分布的上界会恶化为

$$P\{D_e(t) \geq d_e\} \leq f_e((1 - \eta)R d_e) \quad (15)$$

5 数值计算结果及分析

参考 5G 相关的仿真结果<sup>[19]</sup>, 设定系统空中接口部分的无线传输平均速率为 20 Mbit/s. 结合 URLLC 业务和 eMBB 业务的特征<sup>[19-20]</sup>, 确定 URLLC 业务的数据包长度为 32 B, 概率时延约束为 (1 ms,  $10^{-5}$ ), eMBB 业务的数据包长度为 0.5 MB, 数据包到达率为 2 pkt/s.

图 4 给出了固定 URLLC 业务数据包到达率  $\lambda_u$  为 100 pkt/s 时, 不同资源预留比例下的概率时延分

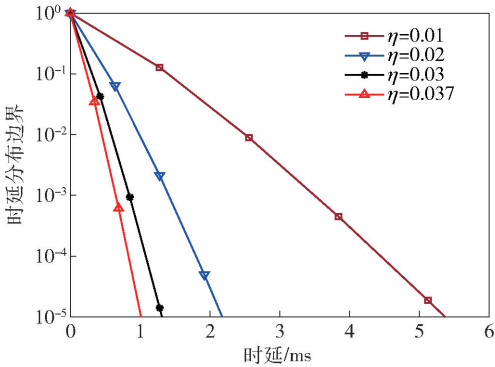


图 4 URLLC 业务的概率时延分布

布结果. 从图 4 中可以看出, 随着资源预留比例的增加, URLLC 业务的概率时延保障越好. 当资源预留比例达到 3.7% 时, URLLC 业务数据包时延超过 1 ms 的概率不超过  $10^{-5}$ , 满足其概率时延约束, 即在此配置下, 系统资源预留比例应不小于 3.7%.

图 5 给出了上述条件下, 不同的资源预留比例对 eMBB 业务的概率时延分布的影响. 从图 5 中可以看出, 随着资源预留比例的增加, eMBB 业务的时延会变大. 但是由于 eMBB 业务数据包本身较大, 当资源预留比例相对较小时, eMBB 业务的时延恶化并不明显. 当 URLLC 业务的到达率增大时, 最小资源预留比例也应当随之增长. 表 1 给出了 URLLC 业务不同到达率配置下, 资源预留比例.

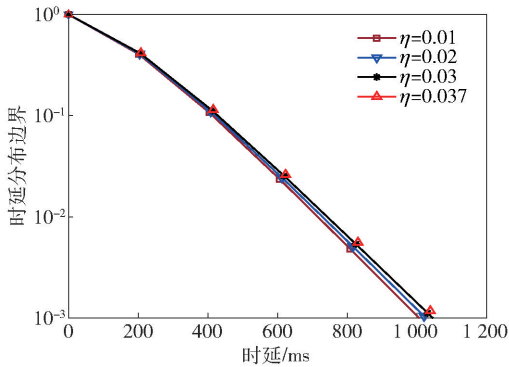


图 5 eMBB 业务的概率时延分布

随着预留给 URLLC 业务的资源量增加, eMBB 业务可使用的资源减少, 当减少量较多时, 其时延性能会明显恶化. 图 6 给出了表 1 中各资源预留比例对应的 eMBB 业务的概率时延分布. 从图 6 中可以看出, 当  $\eta_{\min}$  较小时, 时延的恶化并不明显; 当  $\eta_{\min}$  由 12.7% 增大到 19% 以及 32.5% 时, 概率时延分布曲线显著右移. 例如, 当  $\eta_{\min}$  为 12.7% 时, eMBB 业务数据包时延超过 1 000 ms 的概率不大于  $10^{-2}$ ; 当  $\eta_{\min}$  为 32.5% 时, eMBB 业务数据包时延超过 1 000 ms 的概率接近  $10^{-1}$ .

表 1 资源预留下界

$\lambda_u / (\text{pkt} \cdot \text{s}^{-1})$	$\eta_{\min} / \%$
100	3.7
200	4.25
1 000	6.4
5 000	12.7
10 000	19
20 000	32.5



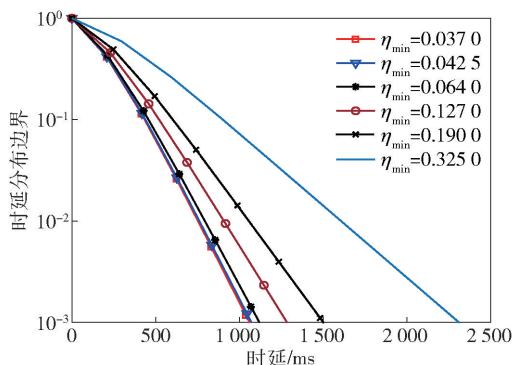


图6 表1对应的eMBB业务的概率时延分布

## 6 结束语

eMBB业务与URLLC业务在系统中共存是未来5G通信系统的典型场景。由于两类业务的生成特性不同、服务质量要求不同,也就对网络的服务特性提出了挑战。资源预留作为一种解决方案,其中所涉及的资源预留比例需要进行合理的配置,才能更好地保障两类业务的服务质量需要。笔者基于随机网络演算分析,构建了eMBB与URLLC业务复用的通用模型,推导了满足URLLC业务概率时延约束的资源预留比例的下界,也分析了资源预留为eMBB业务带来的性能损失。研究表明,当URLLC业务的数据量较多时,资源预留模式下,eMBB业务的时延性能会显著恶化。研究结果为系统设计及资源分配提供了理论依据,资源预留下界的确定,能够在满足URLLC业务服务质量保证需求的前提下,将对eMBB业务性能的损失降至最低。

## 参考文献:

- [1] 朱红梅, 林奕琳, 刘洁. 5G URLLC 标准、关键技术及网络架构的研究[J]. 移动通信, 2017 (17): 28-33.  
Zhu Hongmei, Lin Yilin, Liu Jie. Research on URLLC standard, key technique and network architecture for 5G [J]. Mobile Communications, 2017(17): 28-33.
- [2] 温萍萍. 5G 网络中支持 URLLC 业务的数据复制传输技术[J]. 无线互联科技, 2018(10): 20-22.  
Wen Pingping. Data duplication technique supporting URLLC service in 5G network [J]. Wireless Internet Technology, 2018(10): 20-22.
- [3] 3GPP R1-167127. Frame structure considerations for URLLC[R]. [S. l.]: Intel Corporation, 2016.
- [4] 3GPP R1-1608957. URLLC and eMBB frame structure and multiplexing[R]. [S. l.]: ZTE Microelectronics, 2016.
- [5] 3GPP R1-1611222. DL URLLC multiplexing considerations[R]. [S. l.]: Huawei HiSilicon, 2016.

- [6] 3GPP R1-1704215. On pre-emption indication for DL multiplexing of URLLC and eMBB[R]. [S. l.]: Huawei HiSilicon, 2017.
- [7] 3GPP R1-166398. URLLC system level simulation assumptions[R]. [S. l.]: Qualcomm Incorporated, 2016.
- [8] 3GPP R1-1609057. Evaluation results of eMBB with HARQ retransmission in case of multiplexing eMBB and URLLC[R]. [S. l.]: Samsung, 2016.
- [9] Abreu R, Berardinelli G, Jacobsen T, et al. A blind retransmission scheme for ultra-reliable and low latency communications[C]//2018 IEEE 87<sup>th</sup> Vehicular Technology Conference. Porto: IEEE, 2018: 1-5.
- [10] Singh B, Tirkkonen O, Li Z, et al. Interference coordination in ultra-reliable and low latency communication networks[C]//2018 European Conference on Networks and Communications. Ljubljana, Slovenia: IEEE, 2018: 251-255.
- [11] Jehad M, Hamamreh, Ertugrul Basar, et al. OFDM-subcarrier index selection for enhancing security and reliability of 5G URLLC services[J]. IEEE Access, 2017(5): 25863-25875.
- [12] Jacobsen Thomas, Abreu Renato, Berardinelli Gilberto, et al. System level analysis of uplink grant-free transmission for URLLC [C]//2017 IEEE Globecom Workshops. Singapore: IEEE, 2017: 1-6.
- [13] Pocovi G, Pedersen K I, Mogensen P. Joint link adaptation and scheduling for 5G ultra-reliable low-latency communications[J]. IEEE Access, 2018(6): 28912-28922.
- [14] Zhang Yan, Jiang Yuming. Performance of data transmission over a Gaussian channel with dispersion[C]//2012 International Symposium on Wireless Communication Systems. Paris, France: IEEE, 2012: 721-725.
- [15] Fu Songwei, Zhang Yan, Ceriotti Matteo, et al. Modeling packet loss rate of IEEE 802.15.4 links in diverse environmental conditions [C]//2018 IEEE Wireless Communications and Networking Conference. Barcelona, Spain: IEEE, 2018: 1-6.
- [16] Cruz R L. A calculus for network delay: part I: network elements in isolation and part II: network analysis [J]. IEEE Transactions on Information Theory, 1991, 37(1): 114-141.
- [17] Boudec J-Y Le, Thiran P. Network calculus: a theory of deterministic queueing systems for the internet[M]. Berlin Heidelberg: Springer, 2001: 1-274.
- [18] Jiang Y, Liu Y. Stochastic network calculus[M]. London: Springer-Verlag, 2008: 1-207.
- [19] 3GPP R1-1611222. DL URLLC multiplexing considerations[R]. [S. l.]: Huawei HiSilicon, 2016.
- [20] 3GPP R1-167127. Frame structure considerations for URLLC[R]. [S. l.]: Intel Corporation, 2016.