

5G 系统通过哪些关键技术和标准优化来提供 URLLC 业务能力

ZH

2024 年 5 月 29 日

1 引言

URLLC (Ultra-Reliable Low-Latency Communications, 超可靠低延迟通信) 作为 5G 网络的三大应用场景之一, 可以在实现空口单向时延低至 0.5ms 1ms 的同时满足可靠性为 99.9999% 的数据业务传输需求。因此, URLLC 被广泛应用于工业自动化、智能制造、车联网、远程手术等场景。下面会针对 URLLC 业务的低时延和高可靠性两方面对其关键技术和优化标准进行分析与梳理。

2 低时延的关键技术与优化标准

2.1 无线接入网方面

为达到降低时延的目的, 5G 新空口 (New Radio, NR) 进行了多项针对性的设计, 主要包括帧结构、调度方式、上行配置授权调度、传输反馈增强、URLLC 与 eMBB 资源复用等方面

- (1) 5G NR 提供更灵活的帧结构。5G NR 在时域上的结构与 4G LTE 相似, 包括帧、子帧和时隙三个概念。在 LTE 中, 子帧长 1ms, 时隙固定长 0.5ms, 由 7 个符号组成。相比于 LTE 采用固定长度的时隙, NR 中引入灵活可配的多种子载波间隔: FR1 频段支持 15kHz、30kHz、60kHz

的子载波间隔配置，而 FR2 频段可以支持 60kHz、120kHz 的子载波间隔配置。子载波间隔的大小决定了 OFDM 符号长度，而每个时隙由 14 个符号组成，所以子载波间隔的大小也就决定了调度时隙的大小，越大的子载波间隔对应越低的单位调度时延，有利于 URLLC 通信业务的低时延保障。

- (2) 5G NR 提供非时隙调度方案。4G 与 5G 网络均支持基于一个时隙（14 个 OFDM 符号）的调度方案，但对于极低时延需求的 URLLC 业务，这种调度粒度仍然有优化的必要。因此，5G 网络引入了基于 mini-slot 的调度方案，将最小调度颗粒从时隙级（14 符号）缩短至符号级（2/4/7 符号），降低业务传输时延，同时，5G NR 提供的灵活可配的调度起始位置可实现在一个时隙内有多次调度机会，从而减少数据的发送等待时延，适用于超短时延小包的业务场景。
- (3) 5G NR 支持配置授权调度。基站负责所有上下行业务的数据调度。当终端存在上行调度需求时，需要向基站申请调度授权，而根据传统的业务调度机制，终端与基站的调度请求及授权交互需要跨越多个时隙，这无疑会带来较大的数据发送等待时延。所以针对 URLLC 业务，基站可以预先为部分终端分配上行传输资源，终端可根据业务需求通过激活相关授权在预分配的资源上直接进行多次上行传输，减少调度时延和开销。5G 网络支持 Type1 和 Type2 两种配置授权激活方案，其中，Type1 方式下，基站发送 RRC 信令配置周期、频域资源、时域偏置、调制编码等参数，终端在接收到 RRC 信令后，根据其中的时域偏置进行授权配置的激活，适用于一些周期和资源等无需频繁变化的特性稳定的业务。Type2 方式下，基站同样发送 RRC 信令进行周期的配置，不过需要由 PDCCH 中的 DCI 消息激活配置授权。
- (4) 传输反馈增强。在 5G 网络中，业务数据以传输块的粒度进行调度传输，在 LDPC 编码的情况下，一个传输块包含了多个编码块，当只有少数编码块出现错误时，基于传输块的 HARQ 机制需要将整个传输块进行重传，这极大浪费了物理层资源，可以将编码块分组形成编码块组，设计基于编码块组的 HARQ 机制进行重传，这样能实现仅重传包含出现错误的编码块的编码块组，降低了重传的资源消耗和传输

时延。此外，相比于 LTE 固定 4ms 间隔的 HARQ-ACK 反馈，NR 支持灵活可配置的 HARQ-ACK 反馈间隔，R16 协议定义了 sub-slot 的 PUCCH HARQ 反馈的时间单元长度，可以实现在一个时隙内完成对多次基于 mini-slot 传输数据的反馈，进一步加强了 URLLC 业务的低时延特性。当初次传输失败时，基于 sub-slot 的 HARQ 反馈可以使得基站更快地接收到重传指示，进而缩短重传的时延，实现在业务时延允许范围内完成重传。

- (5) 5G NR 支持 URLLC 与 eMBB 资源复用。在下行传输方面，资源复用主要包括半静态复用和动态复用两种方式。半静态复用是根据参数配置预留一定比例的频域资源分别供 eMBB 与 URLLC 使用，这种方式能够保证 URLLC 空口资源，但也降低了空口资源的利用率；动态复用是当高优先级业务（URLLC 业务）需求发生时，网络通过向低优先级业务（eMBB 业务）的终端发送相关的 PI（Pre-emption Indicator，抢占指示）信息，将已分配用于低优先级业务调度的空口资源复用于高优先级业务，保障高优先级业务的随到随传。PI 承载在 DCI Format 2_1 格式的信息中，PI 按照小区为单位组包，高层通过 RRC 信令配置每个小区的 PI 在 DCI 负载中的位置，每个 PI 长度固定为 14bit。URLLC 业务通常业务量较小，这种资源抢占通过打孔的方式实现，可以在提供低时延 URLLC 业务传输的同时将对 eMBB 业务产生的影响降到最低。而在上行传输方面，有基于取消指示的和基于功率控制的上行资源复用方案。前者是当业务发生冲突时，低优先级业务的终端发送 UL cancellation 指示取消正在发送或还未发送的业务数据，这种方案可以消除 eMBB 对 URLLC 数据的同频干扰；而后者是在保持低优先级业务原有发送功率的同时增大高优先级业务的发送功率，在保证 URLLC 业务时延可靠性的同时，提高了资源复用效率。

2.2 核心网方面

- (1) 边缘能力增强。核心网 UPF 网元的部署位置决定了基站到核心网的传输时延，通过 UPF 网元下沉设计，可以从物理上减小基站到 UPF 的传输距离，同时避免了承载网上的传输，减少拥塞的可能性，从而

降低业务传输时延。针对一些特殊场景，还可以考虑将 SMF、AMF 等控制面网元与 UPF 一同部署到边缘云平台，实现控制面信令的实时处理。

- (2) QoS 增强。针对垂直应用，3GPP 定义了新的以时延为主的 GBR 类型，而针对不同的垂直应用，3GPP 定义 82、82、84 等新的 5QI，以支持低时延高可靠通信，与此同时，5G 网络支持 QoS 监控，可以实时监控和测量终端与 PSA UPF 之间数据包的传输时延，保障 URLLC 业务的低时延传输。

3 高可靠的关键技术与优化标准

3.1 无线接入网方面

3GPP 给出的 R15 协议中指出 URLLC 业务的可靠性要达到 99.999%，R16 协议中要求达到 99.9999%。为达到高可靠的要求，NR 在空口和高层协议层进行了针对性设计与增强。提高传输可靠性的思路主要是降低编码率和增加分集增益。

- (1) 下行控制信息增强。R16 协议中引入了 24bit 的 compact DCI，通过减小 URLLC 业务调度所使用的 DCI 的最小比特数，可以降低对物理层资源的需求，并且可以使用较高的 AL（Aggregation Level，聚合等级）来降低编码率，提升传输的可靠性，针对小区中多用户并行业务的情况，这种 DCI 设计可以减少由于缺少 PDCCH 资源导致 PDCCH 被阻塞而无法进行调度的情况，在一定程度上降低了传输时延。
- (2) 调制与编码方案增强。为了提高传输的可靠性，3GPP 定义了支持配置目标 BLER 为 10^{-5} 的 CQI 和 MCS 新表格，通过引入编码效率更低的 $\pi/2$ BPSK 调制方式以及将调制方式的最高等级限制为 64QAM 来增加在恶劣环境和边缘覆盖中的传输成功率，从而达到提高传输可靠性的目的。此外，提高数据单次传输的成功率减少了对重传的需求，间接也降低了业务时延。

- (3) 物理层重复传输。5G 网络支持基于时隙的物理层重传机制，将不同冗余版本的数据在物理层不同时间隙重复传输，让接收端可以获取额外的分集与数据合并增益，从而提高数据传输的可靠性。3GPP 针对 PUSCH 定义了重复传输类型 A 和 B，其中类型 A 是基于时隙的重传，这无疑会增大数据传输时延，而类型 B 是基于 mini-slot 的重传，在提高可靠性的同时兼顾了 PUSCH 的传输时延。对于 PDSCH，R16 提供了多种可配的重复传输次数。
- (4) PDCP (Packet Data Convergence Protocol, 分组数据汇聚协议) 层数据复制。这是一种 5G 网络高层基于载波聚合或双连接提升可靠性的传输方式，在这种方式下，数据包会在 PDCP 层被复制为多份，并分别映射到不同的逻辑信道，对应不同的 RLC 实体，通过不同的无线空口资源进行传输，基于载波聚合的不同的逻辑信道属于相同的 MAC 实体，而基于双连接的属于不同的 MAC 实体，接收端可获取相应的分集增益以提升传输可靠性。R16 中最多可支持四条支路的复制，可通过 CA 复制、DC 复制和 CA 复制 +DC 复制的组合来实现。

3.2 核心网方面

在核心网方面，数据传输过程中的可靠性主要通过建立备份路径来保证。

- (1) 建立备份路径来增强数据传输的可靠性。5G 网络采用基于双 PDU 会话或双核心网隧道的方式进行数据的备份传输，前者是建立两个冗余的 PDU 会话，并采用不相交的用户面路径进行传输，而后者是建立冗余传输，采用两个独立的 N3 隧道部署在 UPF 与 RAN 之间，关联到同一个单独的 PDU 会话，实现可靠性的提升。
- (2) 业务连续性增强。在 PDU 会话锚点更换时，通过与应用功能单元的实时协作，减少业务中断和丢包，保证业务的连续性。
- (3) Multi-TRP 传输机制。R16 中提出的 Multi-TRP 传输机制可以实现从两个或多个 TRP 发送相同的数据，并在接收端进行软合并，提升

解码的成功率。多 TRP 传输的方式包括 SDM、FDM、时隙内 TDM 和时隙间 TDM 四种，同时支持这四种方式的组合及动态转换。