

移动接入网业务加速技术研究

Research on Service Acceleration Technology in Radio Access Network

张 勃,冯 毅(中国联通网络技术研究院,北京 100048)

Zhang Qing, Feng Yi (China Unicom Network Technology Research Institute, Beijing 100048, China)

摘 要:

随着移动游戏产业规模的不断扩大和短视频、直播等移动视频社交用户规模的快速增长,移动互联网业务对移动网络提出了业务加速需求。在简述了移动接入网业务加速技术的分类后,提出了包含调度优先级优化和业务预调度2种实现方式的面向业务加速的资源调度优化方案,与包含域名系统(DNS)缓存和代理确认应答(ACK)2种实现方式的面向业务加速的信令优化方案,最后分析了资源调度优化技术和信令优化技术的典型业务加速应用场景和方案。

Abstract:

With the continuous expansion of the mobile game industry scale and the rapid growth of mobile video social users such as short video and live broadcast, mobile Internet services have made a request of service acceleration for mobile network. Firstly, the classification of service acceleration technology in mobile access networks is briefly described. Secondly, A service accelerated resource scheduling optimization scheme including scheduling priority optimization and service pre-scheduling is proposed. Thirdly, a service optimized signaling optimization technology solution including DNS cache and proxy ACK is proposed. Finally, the typical service acceleration application scenarios and schemes of resource scheduling and optimization technology and signaling optimization technology are analyzed.

Keywords:

Radio access network; Service acceleration; Resource scheduling; Domain name resolution; Network protocol

关键词:

移动接入网;业务加速;资源调度;域名解析;网络协议

doi: 10.12045/j.issn.1007-3043.2019.07.002

中图分类号:TN929.5

文献标识码:A

文章编号:1007-3043(2019)07-0006-05

引用格式:张勃,冯毅. 移动接入网业务加速技术研究[J]. 邮电设计技术,2019(7):6-10.

0 引言

随着智能终端性能的不不断提升和中国游戏产业的快速崛起,中国移动游戏收入在2017年已经突破1 000亿人民币,占据整个游戏产业的54.3%。移动游戏行业也从享受人口红利过渡到提升产品质量和用户体验,而无论是更好的画质、更流畅的动作响应,都对数据的载体——移动网络提出了业务加速需求。另一方面,视频社交开始成为移动视频业务新的亮点

和发展驱动力,中国短视频月活跃用户已达4.6亿,直播视频用户规模则达到了3.98亿,且短视频社交行业用户黏性远高于其他移动互联网业务,无论是短视频要求更快的加载速度还是直播视频要求高分辨率下更少的卡顿次数,都要求移动网络能够实现针对性的业务加速。本文在介绍了移动接入网中业务加速技术的分类后,具体给出了调度优先级优化和业务预调度2种面向业务加速的资源调度优化技术的实现方案,以及主要包含DNS缓存和代理ACK 2种方式的面向业务加速的信令优化技术的实现方案,最后讨论了资源调度优化技术和信令优化技术的应用场景和应

收稿日期:2019-05-04

用方案。

1 业务加速技术的分类

无论是网页内容浏览、视频内容播放还是游戏内容互动,用户的基本诉求之一是内容能够更快地展现在终端上,尤其在业务服务器发生拥塞或网络资源发生拥塞时,仍能够获得最低的延时、最少的卡顿。移动接入网中的业务加速技术通过深度结合接入网的资源和能力,融合部分信息技术,达到对不同类型业务加速的目的。移动接入网中的业务加速技术主要包含资源调度优化技术和信令优化技术2类。

资源调度优化技术是基站根据预设的调度优化策略,为特定用户的特定业务分配无线资源时,采用差异化的手段,使用户的业务在资源调度层面获得特殊的保障,进而直接降低业务延时,提高用户体验。

信令优化技术是指基于TCP协议规则和常见的网络协议流程,打破基站在终端与服务器之间的网络协议数据传输过程中只扮演透明转发的常规角色的固定方式,融合移动网络的特征,优化协议流程,加快数据包的传送,降低业务时延。

2 面向业务加速的资源调度优化技术

资源调度优化技术通过直接调整为业务分配无线资源时的方案,采取优先分配资源的方式或采取事先预留资源的方式,实现保障性业务的业务加速。资源调度优化技术包括调度优先级优化和业务预调度2种方式。

2.1 调度优先级优化技术

调度优先级优化技术是由基站在为指定用户的指定业务分配无线网络资源时,调整调度的优先级,从而使需要保障的业务承载优先被调度,在调度层面降低业务调度时延,使业务优先获得无线资源。

在移动通信网络中,资源调度^[1-6]决定了什么用户的哪些业务可以得到何种资源,是接入网侧决定用户业务速率、时延的关键。通常资源调度考虑的要素包括基站目前可用的时域频域资源、无线网络的质量、UE的处理能力、用户的QoS要求等。基站在利用自身的软硬件资源处理资源调度的过程中,不可能对所有用户的所有业务同时分配资源,而是需要按照一定的先后顺序,依次完成资源的分配。资源调度的优先级决定了资源分配的先后顺序,显而易见,被优先调度的用户业务,其等待获得资源的时间更短,业务时延

也会更短;此外,当网络拥塞时,优先被调度的用户更有可能获得更多的无线资源,即更有可能获得更高的无线传输速率或业务所需的无线带宽被保障的可能性更高,对用户来说也就意味着更快地看到业务内容或卡顿更少。

调度优先级优化技术存在2种类型的优化策略,一种是始终提高优先级方式,一种是事件触发型提高优先级方式。

始终提高优先级,可以对指定的高优先级用户、指定的业务类型或指定业务源的业务、指定业务类型或指定业务源业务的指定阶段始终调高优先级。该方式的执行,可以由MEC或能力开放服务器完成用户、业务或业务阶段的识别,再将提高优先级的程度等优化方案下发给基站实现调度优化;也可以仅由MEC或能力开放服务器下发基本优化策略,如对哪些用户提高优先级或对什么类型的业务提高优先级等,而识别和调度优化工作都由基站完成。

事件触发型提高优先级,可以指定事件条件,当满足事件条件或不满足事件条件时,对指定的用户或业务提高调度优先级,最典型的是为指定类型的业务设置最低保障速率或最低保障时延,当识别发现该保障可能达不成时,提高该业务的调度优先级;当识别发现保障可以顺利达成时,撤销该业务的优先调度,恢复为原调度优先级。该方式的执行,通常仅由MEC或能力开放服务器下发基本策略,具体判断是否执行和调度优化都由基站完成。

2.2 业务预调度优化技术

业务预调度技术是根据业务数据包的发送流程和特点,针对必然发生的上行数据包进行预先上行资源预留,同时避免由于不连续接收(DRX)技术^[7-9]对下行数据包的延迟接收而进行持续地预先下行资源预留。业务预调度技术的基础是上行调度请求机制和DRX技术。

在移动通信系统中,当终端不传输上行数据时,基站不会为该终端分配上行资源,这就需要终端在传输上行数据前,通过上行调度请求(SR——Scheduling Request)机制^[10-12]向基站上报自己的需求,以便基站决定是否给终端分配上行资源。终端待传输的上行数据触发终端启动SR机制,终端先利用PUCCH信道或PRACH信道向基站发送调度请求SR,基站基于该请求向终端分配少量的资源并将该授权通知给终端,终端再在授权的资源内上报缓冲区状态报告(BSR

——Buffer Status Report),即告知基站待传输的上行数据量,基站根据调度算法决定是否为用户的该传输需求分配上行资源或分配多少资源,并将授权通知给终端,终端利用分配的资源传输上行数据。

针对如图 1 所示的常见 TCP 业务数据包传输流程,在终端向服务器发送 TCP SYN 消息及服务器向终端发送 TCP SYN ACK 消息后,终端会完成握手机制向服务器发送上行数据包 TCP ACK 消息。

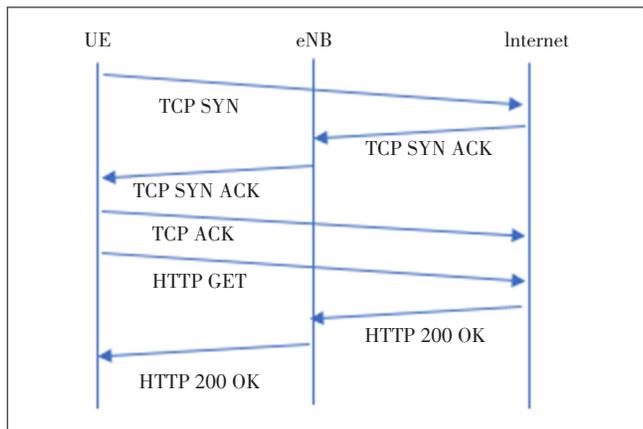


图 1 TCP 业务数据包传输流程

因此,基于该机制,当基站检测到终端和服务器的握手过程时,一方面基站可以优先为终端发送上行资源分配授权的通知,另一方面基站可以为终端将要发送的上行 TCP ACK 消息预留资源,如在收到调度请求后即向终端分配足够的上行资源,使得终端在原本用于传输缓冲区状态报告的时间段内即可完成 TCP ACK 消息的发送,或直接在基站收到服务器向该基站发送的 TCP SYN ACK 消息后,间隔一段时间即直接为该终端分配上行资源用于传输 TCP ACK 消息,从而减少 SR 机制导致的资源分配延时,如图 2 所示。

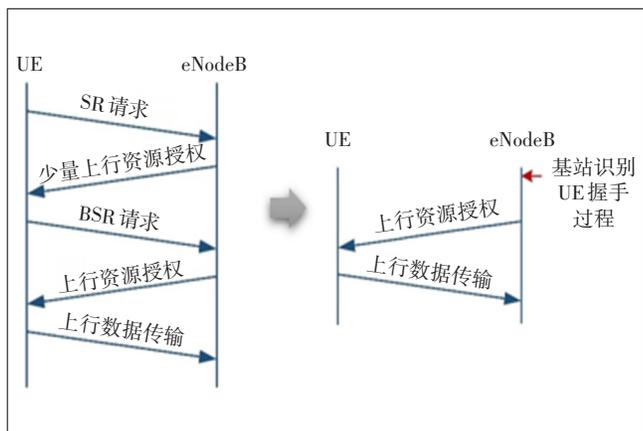


图 2 基于 SR 机制的业务预调度优化数据传输流程

由于移动通信网络中数据业务占据业务的主导地位,而数据包的一个突出特性就是突发性,即在一段时间内传输数据,而在更长的一段时间内暂停数据的传输。DRX 即不连续接收技术,就是利用这种数据突发特点,终端间隔性地监视 PDCCH 信道,并在预定义的没有数据传输的时间段,终端关闭大部分发射与接收系统,从而既提高无线资源利用率,又降低终端的电池消耗^[13-15]。然而,DRX 技术在应用中也存在一定的问题,以图 1 为例,根据目前常用的 DRX 配置,服务器向终端发送的 TCP SYN ACK 消息和 HTTP 200 OK 消息很有可能因为落入 DRX 休眠期内导致不能及时被调度。如果下行包落入第 1 个 DRX 休眠期,将平均延迟 35 ms 发送,如果落入第 2 个 DRX 休眠期,将平均延迟 155 ms 发送。因此,可以针对 TCP 业务的初始阶段,在一段时间内进行持续的下行调度,从而避免 DRX 机制导致的长延时。

业务预调度技术执行的机制,可由 MEC 或能力开放服务器完成指定用户或指定业务的识别,并下发持续调度时长等配置信息,由基站完成上行预调度和下行持续调度时机的识别和选择;也可仅由 MEC 或能力开放服务器下发基本策略,具体识别、判断和执行都由基站完成。

3 面向业务加速的信令优化技术

目前互联网协议主要是基于固定互连网络特征制定的,在移动网络中往往进行直接的移植而未做调整,因此存在协议与移动通信网络特征不适配的问题。信令优化技术主要针对该问题,通过使基站、MEC 或能力开放服务器在面向终端时承担业务服务器的功能,在面向业务服务器时承担终端的功能,提高业务、协议与网络的适配性,从而加速业务数据传输。面向业务加速的信令优化技术包含 DNS 缓存和代理 ACK 2 种。

DNS 缓存是指由基站或 MEC 及能力开放服务器缓存基站或基站组服务的用户中常用的 DNS 域名解析,在识别用户发出 DNS 请求后,直接由基站或 MEC 及能力开放服务器向用户反馈,从而节省了基站向 DNS 服务器索取 DNS 域名解析的过程带来的时延。该过程中,需要基站或 MEC 及能力开放服务器完成请求识别、请求统计、域名的存储、域名更新 4 个方面的工作。具体来说,先对终端发起的 DNS 查询请求进行识别,即先完成请求消息的识别,再识别请求消息中

查询的域名内容;接着基站或 MEC 及能力开放服务器需要对所有服务的终端的所有 DNS 查询请求进行统计分析,以确定请求率较高的域名,从而使得 DNS 域名存储空间有限时优先存储请求率高的域名信息;之后基站或 MEC 及能力开放服务器根据识别的首次发起的 DNS 查询请求代替终端执行向 DNS 服务器和业务服务器的请求操作,获取 CNAME 或 IP 地址后将信息反馈给终端的同时完成域名信息存储,之后再收到终端请求的 DNS 查询时可直接反馈;最后,为存储的域名设置更新周期,在更新周期到期后主动获取一次域名对应的 CNAME 或 IP 地址,以保持存储域名信息的有效性。

代理 ACK 是指在终端与服务器交互 TCP 数据包的过程中,基于终端和基站之间的数据传输通常是保障完成的且通常为顺序传输,因此基站不用等待收到终端发出的 ACK 消息后再转发给业务服务器,而是在向终端转发业务服务器下发的数据后,尚未收到终端向业务服务器发出的 ACK 消息之前,就代替终端发出 ACK 消息,如图 3 所示,从而业务服务器相对提前地向基站发出下一组计划发给终端数据,这样一方面节省基站和终端之间的消息传输时延,另一方面可以在 TCP 慢启动阶段加快窗口的放大过程。考虑到终端对业务数据包的接收与处理存在一定的不确定性,当数据包较大时这一不确定性加剧,如可能因为终端内存分配或 APP 应用层处理等问题,导致终端不一定会对接收到的数据都发出 ACK,因此,代理 ACK 一般只应用于 TCP 慢启动小数据包传输阶段。

4 移动接入网业务加速技术应用方案

资源调度优化技术通过优化传统移动网络资源

调度方案,深度绑定业务与网络,实现业务数据传输的加速。其中,调度优先级优化技术在无线资源有限时为部分业务加速,适用于无线资源利用率较高的热点场景;业务预调度优化技术则更适用于无线资源利用率不高的场景。2 种技术均不限制加速业务的类型。具体来说,调度优先级优化技术可以广泛地为业务提高速率,但往往在网络拥塞时提升性能较为明显。根据优化策略的不同,调度优先级优化技术可以限定在特殊场景中指定高优先级用户,并对高优先级用户执行优先调度,如在园区场景中,为用于安保或灾害预警的终端或模组配置更高的优先级;也可以限定在特殊场景中指定高优先级业务,并对高优先级业务执行优先调度,如在校园教室场景中为来源于学校本地的电子课堂视频、文件业务配置更高的优先级;同时,还可以对指定类型业务的某个阶段执行优先调度,如对于视频点播业务的初始缓冲阶段进行识别并执行优先调度,从而使得用户能够更快地看到第 1 个视频片段。调度优先级优化技术也可以为指定用户的业务或指定业务类型设置最低保障时延或最低保障速率,当该时延或速率不能被满足时,执行优先调度。业务预调度技术对于有频繁包交互的业务能够带来明显的性能提升,如交互类游戏业务,以及常规 HTTP 业务的连接建立初始阶段等。需要注意的是,业务预调度技术由于预调度会过度占用网络资源,并且下行持续调度影响了 DRX 技术进而影响了终端的节电,因此也会导致网络拥塞时系统容量下降和终端耗电增加的问题。

信令优化技术通过深入 Internet 网络协议,结合移动网络特性,优化移动网络业务访问流程,实现业务数据传输的加速。其中,DNS 缓存技术通过缓存 DNS

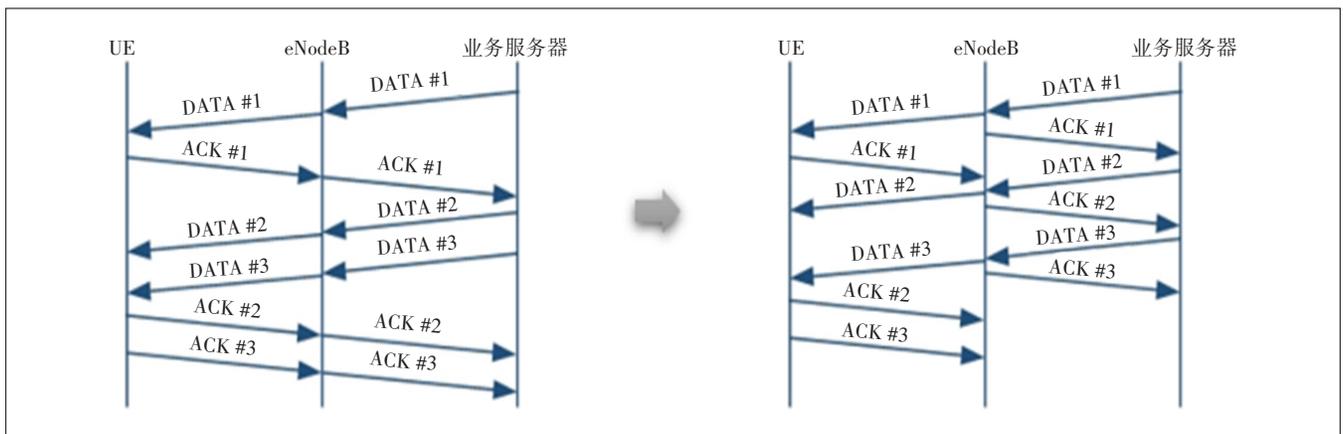


图3 代理ACK数据传输流程

与 CNAME 或 IP 地址信息对的方式,直接减少终端获取 CNAME 或 IP 地址信息的信息交互次数,从而达到快速响应终端请求的目的,因此,一方面,DNS 缓存技术更适于域名与 IP 地址对应关系相对稳定的业务,虽然 DNS 缓存技术可以通过域名更新机制适应域名信息的变化,且该更新机制由基站或 MEC 及能力开放服务器发起,不增加终端的请求延时,但若域名信息更新较快,仍有可能造成缓存的域名信息更新不及时而使得用户获取业务地址错误的问题;另一方面,DNS 缓存技术适于业务访问请求有一定集中度的场景,因为基站或 MEC 及能力开放服务器的缓存能力有限,且只有当请求具有一定集中度时,才能提高缓存的域名信息命中用户请求的概率,从而充分降低用户 DNS 请求响应的时延。代理 ACK 技术由基站充当终端角色,向服务器发出 ACK 消息,从而达到缩短 ACK 消息传输时延的目的,因此,一方面代理 ACK 技术只适用于 TCP 业务;另一方面,该技术不适合拥塞场景和频繁切换场景,因为在拥塞场景中,终端未必能够及时、完整地收到基站转发的服务器发送的业务数据,使用该技术基站由于提前反馈了 ACK 消息将会过早地收到服务器下发的新的需要转发给用户的数据,此时若上一次基站转发给用户的数据尚未完成传输,就需要基站缓存新下发的数据,依此发展将可能导致基站需要缓存过多的数据,对于频繁切换场景,则可能出现基站代替终端向服务器发出的 ACK 消息与终端实际发出的 ACK 消息在切换过程中不能完全适应与匹配的问题,从而导致业务数据丢包或重复收包的问题。

5 结束语

本文从移动游戏和视频社交的高速发展对移动网络提出了业务加速需求出发,介绍了移动接入网中业务加速技术的分类,提出了面向业务加速的资源调度优化技术方案,包括调度优先级优化和业务预调度 2 种实现方式,并提出了面向业务加速的信令优化技术方案,包括 DNS 缓存和代理 ACK 2 种实现方式,最后分析了资源调度优化技术和信令优化技术的典型业务加速应用场景和应用方案。

参考文献:

[1] SADIQ B, MADAN R, SAMPATH A. Downlink scheduling for multi-class traffic in LTE[J]. *Eurasip Journal on Wireless Communications & Networking*, 2009, 2009(1): 510-617.

- [2] KWAN R, LEUNG C, ZHANG J. Proportional Fair Multiuser Scheduling in LTE[J]. *IEEE Signal Processing Letters*, 2009, 16(6): 461-464.
- [3] LUO H, CI S, WU D, et al. Quality-driven cross-layer optimized video delivery over LTE[J]. *IEEE Communications Magazine*, 2010, 48(2): 102-109.
- [4] LEI H, FAN C, ZHANG X, et al. QoS Aware Packet Scheduling Algorithm for OFDMA Systems[C]// *IEEE, Vehicular Technology Conference*. IEEE, 2007: 1877-1881.
- [5] PIRO G, GRIECO L A, BOGGIA G, et al. Two-Level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks[J]. *IEEE Transactions on Multimedia*, 2011, 13(5): 1052-1065.
- [6] PIRO G, GRIECO L A, BOGGIA G, et al. A two-level scheduling algorithm for QoS support in the downlink of LTE cellular networks[C]// *Wireless Conference*. IEEE, 2010: 246-253.
- [7] STEA G, VIRDIS A. A comprehensive simulation analysis of LTE Discontinuous Reception (DRX)[J]. *Computer Networks*, 2014, 73(38): 22-40.
- [8] ZHANG Y F, GAO S T, TIAN H, et al. Delay analysis of DRX in LTE-advanced considering carrier aggregation[J]. *The Journal of China Universities of Posts and Telecommunications*, 2011, 18(6): 1-7.
- [9] 李小文,徐志文,张峥. TD-LTE 系统 DRX 机制的分析与设计[J]. *电子技术应用*, 2012, 38(3): 44-46.
- [10] 孙延洲. LTE 系统中资源调度机制的研究与实现[J]. *数字通信世界*, 2010(5): 6-10.
- [11] 万翠,陈发堂,姚胜. LTE 系统中 MAC 层 BSR 的研究与设计[J]. *广东通信技术*, 2013(1): 33-37.
- [12] 张德民,蒋忠均,虞勤. Improvement of BSR and SR delivery in LTE systems[J]. *光通信研究*, 2012(4): 56-59.
- [13] FOWLER S. Study on power saving based on radio frame in LTE wireless communication system using DRX [C]// *GLOBECOM Workshops*. IEEE, 2011: 1062-1066.
- [14] KOC A T, JHA S C, VANNITHAMBY R, et al. Device Power Saving and Latency Optimization in LTE-A Networks Through DRX Configuration[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(5): 2614-2625.
- [15] ZHANG Y, JIN S F, YUE W Y. An enhanced energy saving strategy for an active DRX in LTE wireless networks[J]. *International Journal of Innovative Computing Information & Control*, 2013, 9(11): 4373-4387.

作者简介:

张勃,工程师,硕士,主要从事移动通信新技术研究和 5G 行业应用新业务孵化等工作;冯毅,高级工程师,硕士,主要从事移动通信系统网络规划及新技术研究和 5G 垂直行业业务探索工作。

