团队科研成果分享 | 面向时延敏感型视频分析的5G RAN与MEC耦合架构

原创 徐冬竹 [imobi Lab](javascript:void(0);)

 2022年10月14日 09:00 北京



我们设计并实现了***5G RAN与移动边缘计算（MEC）耦合架构***，致力于解决5G典型应用响应时延长的问题。研究成果发表于ACM MobiCom 2022。

**一. 研究背景**

5G的兴起带来了许多令人兴奋的潜在应用，例如远程医疗手术、云游戏、移动AR/VR、V2X和视频分析。它们对5G的性能提出了严格要求，包括消耗大量带宽和计算资源、要求超低响应延迟等。尤其在时延敏感型视频分析中，通常需要5G在200毫秒内响应每一图像帧，否则将导致用户体验质量（QoE）下降，甚至在视觉辅助的自动驾驶用例中造成交通事故的发生。

为了满足上述低延迟要求，仅5G空口（NR）的创新是不够的。我们之前的测量研究[1]表明，尽管用户终端和基站之间的数据包级延迟已降低到2毫秒以下，但端到端延迟 (从用户终端到远程云）仍然高达数十甚至数百毫秒（取决于远程云的地理位置）。因此，网络研究者们提出了MEC，并将其视为5G生态的关键组成部分。5G MEC的基本思想是将计算服务器部署于靠近用户终端的位置（例如5G网络边缘侧）而不是在远程云上，以期望缩短传输路径，减少端到端传输延迟。

**二. 发现问题**

虽然5G MEC在概念上是合理的，但其能否在实践中达到上述前景仍未可知。据了解，传统5G MEC研究工作更专注于架构设计、理论或仿真驱动分析以及小型原型系统的实现，常常忽略5G典型应用的实地QoE研究，尤其对时延敏感型应用的性能优化。

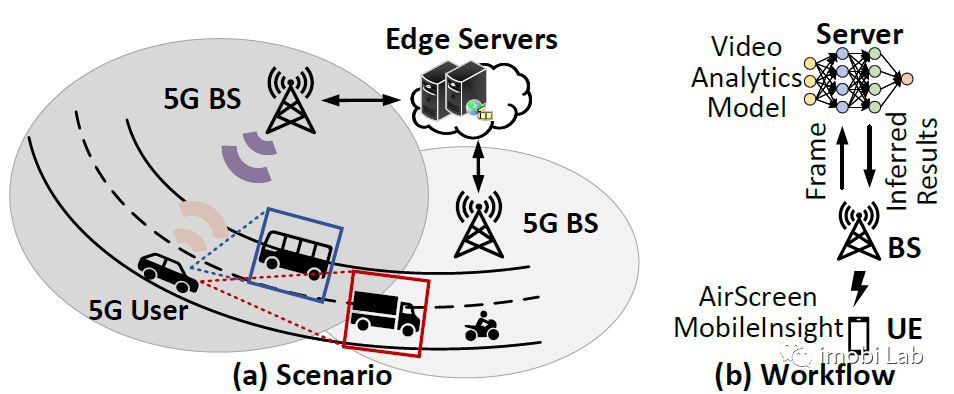


图1：测量场景与时延敏感型视频分析工作流程示意图

通过在运营化5G MEC上开展大规模实地测量，***我们发现当前5G网络与MEC框架之间并不协调，存在“各自为政”且“交互能力弱”的弊端***。这导致时延敏感型视频分析体验了多达300毫秒的响应延迟（仅有44.11%的图像帧可以及时到达边缘服务器）。经过深入分析，我们揭示其***根本原因在于当前5G RAN和MEC之间是松耦合的***。一方面，边缘服务器上的应用程序无法获知5G RAN的无线瞬时动态并及时进行无线资源适配。另一方面，5G RAN不了解上层应用的真实QoE 需求（即，需要多少无线资源而且如何高效授予它们），而只是通过低效率的多轮“请求-授予”操作来分配无线资源。因此，上述两方面因素共同作用导致应用响应延迟过长。

**三. 关键技术**

我们提出***Tutti——一种新型的软件定义的5G RAN和MEC耦合架构，以满足时延敏感型应用的需求***。

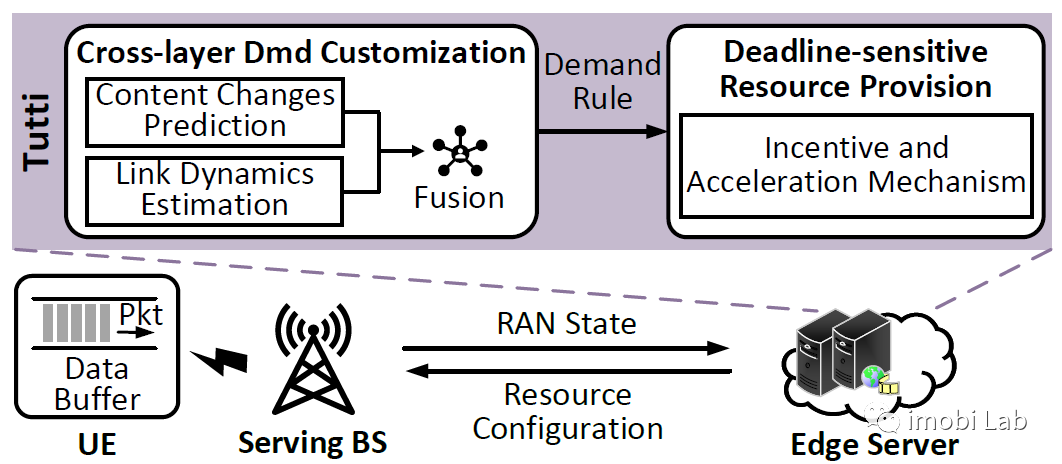


图2：Tutti系统概述图

为了实现这种耦合能力，Tutti首先深度理解5G RAN的网络动态性和上层应用的内容变化，精准定制应用的服务需求。然后，Tutti通过一系列实时的5G RAN与边缘服务器“微”交互操作，自适应调整无线资源授予。例如，当边缘服务器知道图像帧即将到达临界期限时，它会立即敦促基站优先满足其资源授予。更重要的是，当前5G运营商正采用网络切片技术来划分专用的存储、传输、计算资源切片以迎合5G应用的性能。Tutti的设计思想符合这一发展趋势。其跨层融合和轻量级设计方法为网络切片实例化提供了一个可行参考。

**四. 部署与验证**

***我们在开源的软件定义平台上部署了Tutti原型系统***。

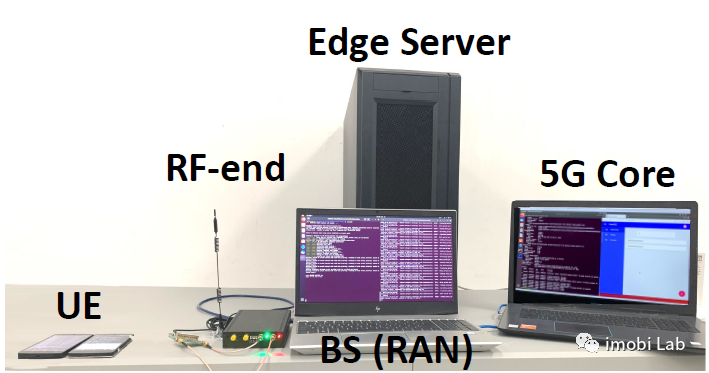


图3：Tutti系统原型测试床

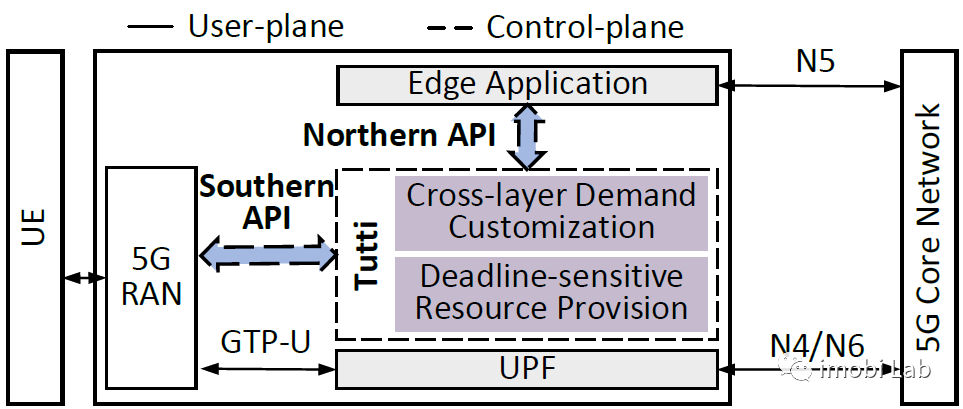


图4：标准协议兼容的交互接口设计示意图

实验结果表明，与现有松耦合的5G MEC架构（如srsEdge）相比，***Tutti将响应延迟降低了61.69%***。此外，Tutti实现手段极其轻量——仅增加约2毫秒信令交互成本以及2.69%的CPU/GPU负载压力。这不仅有利于运营商或云服务商的商用部署，还能促进“端-边-云”协传、协算。

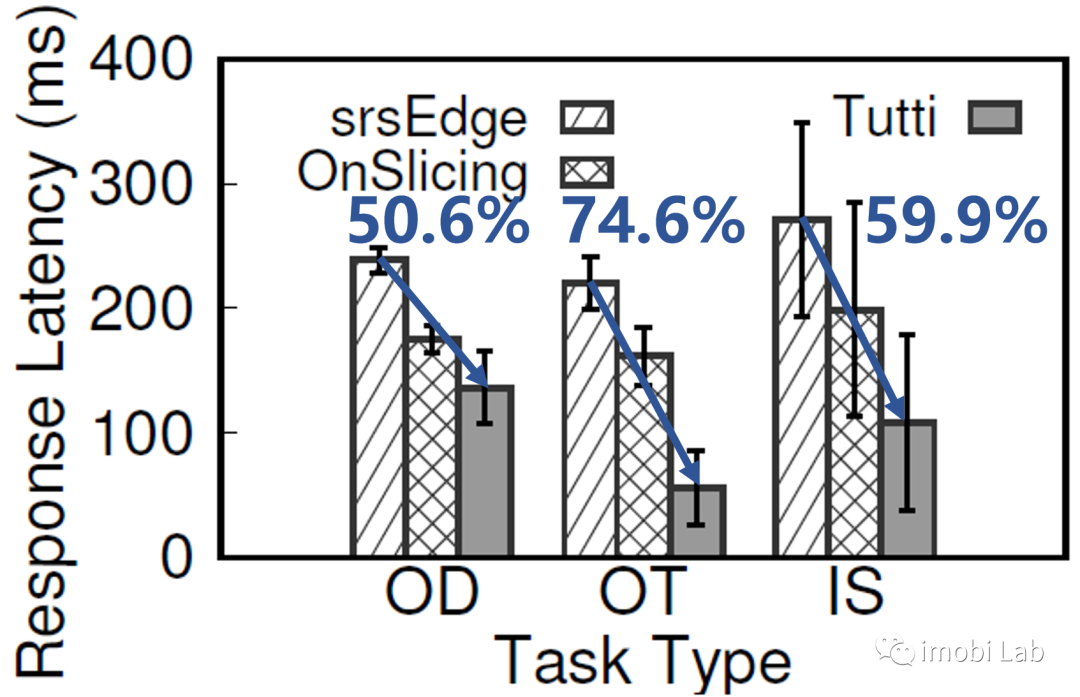


图5：图像帧响应延迟降低结果图

上述工作被计算机网络领域顶尖国际会议ACM MobiCom 2022收录。更多细节请参见原文（论文链接请点击下方“阅读原文”）：

Dongzhu Xu, Anfu Zhou, Guixian Wang, Huanhuan Zhang, Xiangyu Li, Jialiang Pei, Huadong Ma. 2022. Tutti: Coupling 5G RAN and Mobile Edge Computing for Latency-critical Video Analytics. In The 28th Annual International Conference On Mobile Computing And Networking (ACM MobiCom' 22), October 17–21, 2022, Sydney, NSW, Australia. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3495243.3560538