

5G/后 5G 部署对互联网主干影响的分析与建模

杨 芑^{1,2}, 徐明伟^{1,2}, 陈浩^{1,2}

(1. 清华大学信息科学技术学院, 北京 100084; 2. 北京信息科学与技术国家研究中心, 北京 100084)

摘 要: 针对第五代移动通信技术 (5G) 及后 5G 部署之后对互联网主干产生的影响开展定性和定量研究。首先分析了 5G 所具有的超大带宽、超低时延和海量机器连接的特性对互联网主干在流量、时延、安全等方面带来的挑战, 然后建立了抽象模型用于描述 5G 用户获取内容和服务的特性, 以及 5G、边缘计算、云计算相结合场景下主干网流量的特征。以此为基础, 开展了数值模拟实验, 评价了在不同程度的 5G 部署场景下, 互联网主干网性能和用户体验到的带宽和时延等服务质量指标。研究表明, 5G 的部署会引起互联网主干流量增加, 端到端时延中传播时延所占比例增大, 以及带宽瓶颈由接入网转向主干网等结果。

关键词: 第五代移动通信技术; 互联网主干; 边缘计算; 建模与评价

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019182

Analysis and modeling of Internet backbone traffic with 5G/B5G

YANG Yuan^{1,2}, XU Mingwei^{1,2}, CHEN Hao^{1,2}

1. School of Information Science and Technology, Tsinghua University, Beijing 100084, China

2. Beijing National Research Center for Information Science and Technology, Beijing 100084, China

Abstract: The effects that the 5th generation mobile network (5G) bring to Internet backbone were investigated qualitatively and quantitatively. First, the challenges that the characteristics of 5G, i.e. ultra-high data rate, ultra-low latency, and ultra-large number of connections, introduce to Internet backbone in terms of traffic, latency, and security were analyzed. Second, a model was proposed to capture the characteristics of 5G users and Internet traffic with the coordination of 5G, edge computing, and cloud computing. Then, numerical simulations were used to evaluate the model. The QoS requirements that Internet backbone faces under different extent of 5G deployment were evaluated. According to the study, increment of backbone traffic, increment of the ratio of propagation delay, and movement of bandwidth bottleneck are predicted after 5G/B5G deployment.

Key words: 5G, Internet backbone, edge computing, modelling and evaluation

1 引言

第五代移动通信技术 (5G) 如今已经从实验阶段正式走向了应用阶段。5G 的设计目标和最显著的特征是为用户提供超大带宽、超低时延和海量机器连接的服务。5G 越来越广泛的部署不仅会使个人用户的移动通信体验得到更大的提升, 也会促进一大批新兴应用的诞生, 并推动智能制造、

物联网、智慧医疗等创新领域的发展, 带来产业变革。5G 及后 5G 技术^[1]得到了工业界和学术界的广泛关注。

为了满足 5G 对网络功能虚拟化和灵活部署的需求, 同时提供高带宽、低时延等性能方面的保障, 很多工作针对 5G 通信、系统架构、组网等关键技术展开了研究^[2-4]。然而, 已有研究主要针对接入和边缘网络展开, 例如信道编码、网络功能编排、承

收稿日期: 2019-05-19; 修回日期: 2019-07-01

基金项目: 国家自然科学基金资助项目 (No.61872209, No.61625203, No.61811530336)

Foundation Item: The National Natural Science Foundation of China (No.61872209, No.61625203, No.61811530336)

载网技术等。需要注意的是,移动通信领域中所使用的术语“核心网”,从整个互联网的角度看仍然属于边缘网络,这些边缘网络需要经过互联网主干才能实现全球互联。5G 的广泛部署是否会对互联网主干引入更多具有新特征的流量?互联网主干的容量和现有路由、传输、资源调度等技术能否应对流量和服务质量需求的提高?用户进行远程通信和内容访问时,带宽和时延的瓶颈是否会从接入和边缘网络转移到互联网主干?这些都是 5G 和后 5G 时代互联网主干需要密切关注的问题,然而已有研究对 5G 给互联网主干造成的影响缺乏定性和定量的研究。文献[5]对 FOG (from core to edge) 计算场景下物联网能耗、时延和开销进行了建模分析,体现了“云雾结合”场景的优势,然而针对更一般的 5G 场景,目前还没有看到相关的文献。

为了弥补这一不足,本文对 5G/后 5G 对互联网主干带来的影响进行了分析和建模。首先,对 5G 超大带宽、超低时延、海量连接对互联网主干流量、时延和安全性等方面的影响进行了定性分析。然后建立了一组抽象的模型用于描述 5G 用户的流量特性,通过数值模拟实验验证了模型的合理性,并定量给出了随着 5G 部署程度的增大,互联网主干流量的变化及用户所体验到的带宽、时延情况。为 5G/后 5G 时代互联网主干网络建设和关键技术研究方向提供了一定的参考依据。

2 5G/后 5G 部署对互联网主干的影响

2.1 超大带宽带来的挑战

5G/后 5G 最显著的特征是用户带宽的显著提高,这里的带宽指的是数据速率,即单位时间内传输的数据量。在 4G 技术下,单个用户的数据速率能达到 100 Mbit/s,已经达到了传统的 Wi-Fi 和以太网接入的数量级。5G 技术可以进一步将数据速率提高到 1 Gbit/s,这很可能导致汇入主干网的用户(业务)流量特征发生巨大改变。目前有线网络接入速率也在提高,典型的光纤入户技术可以提供 500 Mbit/s~1 Gbit/s 的最大数据速率,然而 5G 网络不仅支持海量设备的同时接入,而且用户间相互干扰相比 4G 进一步减小,尤其是移动通信场景中容易出现大规模用户聚集在较小地理空间的情况,这更使汇入主干网流量的理论上限大幅提高,可达到数太比特每秒^[6]。一个重要的问题由此产生:互联网主干是否会成为端到端通信数据速率的主

要瓶颈?

事实上,随着 5G/后 5G 的部署,还有多个因素会影响互联网主干流量。

1) 5G 的部署将经历一个增量式发展的过程。5G 采用了较高的信号频率,其 FR1 频率能达到 6 GHz,而采用毫米波的 FR2 频率则可达 20~50 GHz。如此高的频率使得 5G 基站的覆盖范围较为有限,导致 5G 的广泛部署需要投入大量的时间和资金成本。不难预测,5G 的部署将从具有一定市场潜力和需求的地区逐渐展开,但在较长的时期内,4G 和 5G 都将共同存在。因此,互联网主干的流量需求不会发生整体上突然的提高,而是从部分节点开始逐步增长。

2) 5G 引入了可灵活部署的边缘计算能力。缓存、计算等网络功能可以从源服务器或云端(数据中心)下沉到距离用户较近的 5G 核心网甚至 5G 接入网,这一方面减小了用户访问网络资源的时延,另一方面也能在一定程度上减少经过互联网主干的流量。例如,边缘网络设备通过缓存流行度较高的内容或预先部署流行的服务,使大量用户不需要访问源服务器或云端就可以快速访问这些内容和服务,该过程中产生的绝大多数流量都被限制在局部网络中。然而,5G 网络功能的下沉对于互联网主干流量减少的效果取决于应用和服务的调度策略及可用资源等诸多因素。边缘网络功能仍然需要从产生内容的源服务器、云端或其他边缘网络功能获取内容和服务,以便本地用户能够快速访问。源服务器、云、边缘相结合的资源调度和内容缓存等机制与云计算场景相比更加复杂,因此相应的主干网流量变化也更加难以预计。

3) 5G 的广泛部署可能会催生新的“杀手”级互联网应用。网络应用的类型和特征在很大程度上决定了互联网流量的特征。移动互联网在 4G 时代诞生的视频直播、高清在线视频、移动支付、短视频等应用使 10 多年前主要由 P2P 流量主导互联网流量这一现象发生了变化。新网络应用的诞生是由技术发展、市场需求等诸多因素决定的,具有难以预测的特点。5G 和后 5G 时代新诞生的网络应用必然也会带来新的互联网流量需求,对互联网路由、传输、资源调度和安全管理带来新的挑战。

值得注意的是,互联网主干链路数据速率在近 10 多年的时间里也得到了飞速发展。当前典型的光纤链路已能够支持 100 Gbit/s 的速率,可以预见,200 Gbit/s、

400 Gbit/s 及 1 Tbit/s 链路将在未来成为现实^[7]。然而,对已有链路和其他网络基础设施进行升级也需要巨大的开销,运营商需要知道主干网如何增长才能满足 5G 和后 5G 移动通信的需求。本文第 3~4 节将通过模型和数值模拟实验的方式展示不同程度的 5G 部署对互联网主干流量的影响。

2.2 超低时延带来的挑战

5G/后 5G 的另一大特点是毫秒级的网络访问时延。相比 4G 网络中用户 30~100 ms 的访问时延^[8],5G 将时延目标定为 1 ms,以实现在大规模实时交互网络应用中用户对时延基本无感知。为了实现这一目标,5G 提出了网络切片、分组核心网络功能的虚拟化下沉等技术。然而,互联网主干的时延受传输距离和光速的限制,具体而言,在忽略处理(如排队)时延的条件下,信号以光速传播 1 ms 的传输距离仅约 300 km。加入处理时延后,互联网主干的典型往返时延能达到数十毫秒甚至两百毫秒以上。在此情况下,5G 和后 5G 超低时延的优势便难以完全体现出来。

为了让部署 5G 和后 5G 的用户能切实体验到访问时延的降低,满足大规模实时交互网络应用的需求,对于时延敏感业务不能按照传统的方法让用户直接访问源服务器或云端,应当尽可能充分地利用边缘访问时间短的优势。例如,将用户需要访问的内容和服务提前推送至位于边缘的计算和存储资源中,从而实现本地快速访问。与 2.1 节讨论的利用边缘计算减少互联网主干流量的情景类似,这同样需要源服务器、云端和边缘相互协同进行调度。这对互联网主干带来的主要挑战在于,主干网应当配合相关机制进行流量调度和路由的优化,例如通过机器学习的方法智能识别时延敏感业务并进行优先调度或选择较优的路由。

2.3 海量设备连接带来的挑战

海量设备连接是 5G 和后 5G 引入的一项新特性。直观上看,这一特性似乎并不会对互联网主干造成特别大的影响。首先,机器之间的通信在网络边缘即可完成。其次,对于需要通过互联网主干进行通信的情况,可以让少数节点充当网关的角色以减少开销,例如文献[9]提出了一种层次化的边缘网络体系结构,先将网络中各物联网(IoT, Internet of things)设备采集的数据汇聚到少数节点,这些节点再与互联网进行通信,从而用较小的开销高效地将海量数据传输到云端。此外,互联网服务提供商

(ISP, Internet service provider)可以在 5G 和后 5G 接入层进行合理的地址划分与分配,在汇聚层和核心层进行流量聚合与整形,使汇入互联网主干的流量容易监测与控制。

然而,随着海量 IoT 设备的广泛部署,互联网主干的安全性会受到 5G 和后 5G 海量设备连接特性的间接影响。为了节约制造成本,IoT 设备通常更加注重性能、能耗等指标,安全性的优先级较低,这使得这些设备更容易被黑客控制,成为分布式拒绝服务(DDoS, distributed denial of service)攻击、信息窃取等安全攻击的工具^[7]。而 5G 所支持的机设备数量使所受到的安全威胁进一步提升。为此,不论是网络边缘还是互联网主干,都应该加强安全防护机制的建设,除了传统的入侵检测、防火墙等网络功能之外,还应该部署源地址认证和用户身份验证等机制^[10],使人类用户和机器用户的行为均可监测、可控制、可溯源,构建和保护真实、可信、安全的互联网环境。

3 模型建立

为了量化评价 5G 网络部署对互联网主干带来的影响,本节将建立形式化的模型,一方面模拟用户访问内容和服务对网络产生的流量需求,另一方面评价用户所体验到的实际数据速率、时延等服务质量指标。需要指出的是,本节建立的理论建模还比较初级,且由于 5G 尚未广泛部署,模型与真实网络必然存在一定的差距。即便如此,依据当前互联网中已经观察到的规律来建立 5G 场景下的模型仍然具有重要的意义,包括对 5G 和后 5G 互联网流量的趋势进行预测,分析 5G 和后 5G 时代影响互联网流量特征的因素,以及分析互联网面临的挑战和发展趋势等。此外,本节在建立模型的过程中尽可能使模型更加逼近真实网络,并在第 4 节通过数值实验结果(如图 3 和图 4 所示)与真实网络数据(如图 5 所示)进行定性对比来验证所建立模型生成的流量具有一定的可信度。

本节接下来建立的模型和相应公式有多种具体用途。首先,为第 4 节的数值实验提供基础和依据。其次,研究人员可以基于该模型来建立模拟场景,用于测试新开发的路由、拥塞控制、缓存等算法的性能,并与传统算法进行比较,例如模型中采用的算法。此外,模型可以作为进一步研究更加准确的 5G/后 5G 流量模型的基础,起到抛砖引玉的

作用,还能对其他相似场景下建立流量模型提供思路和方法。

3.1 基本网络

互联网主干包含多个自治系统,这里仅考虑单一自治系统的情况。一个自治系统包含多个入网点(PoP, point of presence)。PoP之间通过高速光纤链路相连。形式化地将互联网主干抽象为无向图 $G(V, E)$,其中, V 表示PoP中所有节点的集合, E 表示所有链路的集合。每个PoP连接了一个区域(如一个城市)内多种类型的网络,包括4G/5G移动网、有线接入网、企业网、数据中心网络等,这里将它们统一定义为边缘网络,与互联网主干网络相对。

虽然不同类型的边缘网络内部可以有复杂的结构,但是通过ISP的速率控制等机制,可以认为使用相同接入服务的用户具有相同的数据速率和相似的时延。为了便于描述和抽象,本文模型认为一个PoP直接与其所连边缘网络中的所有用户相连。这里主要考虑移动互联网用户。具体而言,假设5G用户连接到本地PoP的时延在1~10 ms随机取值,边缘网络数据速率设为100 Mbit/s(考虑高峰时段多用户并发的场景);4G用户连接到本地PoP的时延在40~100 ms随机取值,边缘网络数据速率设为10 Mbit/s。

3.2 用户流量

考虑一个时间段 $[T, T + \Delta t]$,在此期间连接到节点 $v \in V$ 的活跃用户数量为 N_v ,其中5G用户所占比例为 μ 。 N_v 受节点 v 所处区域人口和 Δt 的影响,在模型中可以通过调整 Δt 和 N_v 的取值来控制用户的基准流量需求大小(有线和4G接入),再根据建立5G用户访问模型来产生5G用户流量需求大小。

网络中的流量抽象为用户对内容的访问。假设一个普通用户发起内容访问的次数在 $1 \sim H$ 均匀分布,5G用户对内容具有更高的兴趣,假设其访问次数为 $1 \sim \theta H$,其中 $\theta > 1$ 。

每个节点 $v \in V$ 在时间段内可供访问的内容数量为 M_v 。全网总的内容数量为 $M = \sum_{v \in V} M_v$ 。直观上,活跃用户和可访问内容成正相关,本文模型假设 M_v 与 N_v 成正比,即 $M_v = \rho N_v$ 。内容的大小服从Pareto分布^[11],具体而言,产生的内容大小 S 大于任意给定值 s 的概率为

$$P(S > s) = \left(\frac{s}{s_{\min}} \right)^{-k} \quad (1)$$

其中, s_{\min} 为最小的内容大小, k 取值为1~1.5^[11]。

内容被访问的频率服从Zipf分布,且被访问的频率与内容大小无关^[12]。具体而言,将内容按照访问频率由高到低进行排序,排第 i 位的内容被访问的概率为

$$P(i) = \frac{1}{i^\alpha} \left(\sum_{i=1}^M \frac{1}{i^\alpha} \right)^{-1} \quad (2)$$

其中, α 为0~1,根据实际测量的结果取值为0.7^[12]。

3.3 云服务和边缘计算

每个PoP节点连接一个本地的数据中心作为用户可以就近访问的云端或内容分发网络(CDN, content delivery network)节点。通常来说,数据中心的容量是有限的,且不同的内容提供商对内容的分发有各自的策略。为了简单起见,本文模型假设节点 $v \in V$ 所连接的CDN节点可以存储的内容大小为 L_v ,令 $L_v = \tau N_v$, τ 为 L_v 与本地节点用户数量的比例系数。CDN节点以式(2)所描述的概率获取内容直到剩余存储空间耗尽。如果用户不能从本地CDN或本地源服务器获取需要的内容,则访问缓存了该内容的距离最近的CDN节点,如果没有CDN节点缓存了该内容,则访问产生内容的源节点。

对于5G网络而言,边缘计算的部署对模型有两方面的作用。一方面,缓存更靠近用户,从而访问时延更短。另一方面,边缘计算与云服务的结合可以视为两级缓存机制。当用户从边缘缓存不能获取所需内容时,如果边缘缓存仍然有剩余存储空间,则由边缘缓存(而不是用户)从最近的CDN节点获取内容。假设节点 $v \in V$ 所连接的边缘网络能缓存的内容大小为 K_v ,令 $K_v = \varphi N_v$ 。与CDN节点不同的是,假设边缘缓存采用被动缓存的方式,总是缓存在时间段 $[T, T + \Delta t]$ 内本地用户访问最频繁(而不是流行度最高)的内容,直到剩余存储空间耗尽。

3.4 数据速率与时延

端用户体验到的数据速率由诸多因素共同决定,位于不同地理位置的用户访问不同位置的内容,其数据速率瓶颈的位置可能也不相同。此外,TCP的拥塞控制算法具有公平性的特征,在传输较长的时间后各用户的数据速率会趋于一致,但是实际网络中用户发起访问的时间和访问内容的大小有很大差别,数据速率难以快速收敛到稳定状态,而是始终处在变化之中。为此,采用简化的模型来

计算用户享有的数据速率。

设边缘网络的最大数据速率为 r_{edge} ，通常由接入技术决定。主干网中的最大数据速率则通过式(3)来估算。

$$r_{\text{backbone}} = \min_{l \in p(s,t)} \omega \frac{C_l}{f_l} \quad (3)$$

其中, $p(s,t)$ 表示从节点 s 到节点 t 的流量所走的路径, 通常采用最短路径; C_l 表示链路 l 的最大数据速率, 即通常所说的带宽; f_l 表示经过链路 l 的流的数量; $\omega > 1$ 是放大因子, 用于消除网络中短流造成计算出的用户数据速率过低的影响。根据文献[13]提供的我国各类应用时长占比, 假设网络视频、游戏、直播、音频使用长流, 所使用时间占比为 31.5%, 而新闻等其他应用使用短流 (不参与带宽竞争), 因此, 取 $\omega = \frac{1}{31.5\%} = 3.2$ 。最终, 用户获得的数据速率为 $\max(r_{\text{edge}}, r_{\text{backbone}})$ 。

用户体验到的访问时延由多个部分组成, 其中接入时延已经在第 3.1 节给出, 传播时延可以由节点间的物理距离和光速计算得到。最后是传输时延和处理 (包括排队) 时延, 由于互联网主干网多采用大容量的核心路由器设备, 传输速率和处理能力都较强, 因此传输时延和处理时延与另外 2 个时延相比可以忽略不计。

4 数值实验分析

本节介绍第 3 节所建立模型的模拟数值实验结果, 侧重于对真实网络拓扑下的数值结果进行直观展示, 并分析导致结果的原因。数值实验程序的实现严格依据第 3 节的模型来编程, 包括流量的产生、内容流行度的分布、云服务和边缘计算的行为、数据速率和时延的计算等。其中 4G 和 5G 用户各自的特点体现在接入时延和带宽的不同、对内容的兴趣不同、是否有边缘计算资源的支持等, 详见第 3.1~3.3 节。

实验使用的拓扑包括 topology-zoo^[14]提供的中国电信主干网 (Chinanet) 和泛欧科研教育数据网 (Geant)。去除没有经纬度数据的节点, Chinanet 包含 38 个节点和 63 条链路, Geant 包含 37 个节点和 58 条链路。虽然 2 个拓扑在规模上很接近, 但在结构上有一定差异。如图 1 和图 2 所示, Chinanet 包含 5 个节点度较大的核心节点, 而 Geant 节点度

的分布较为平均。从后面的实验结果可以看到, 这一差异导致了生成流量的不同。

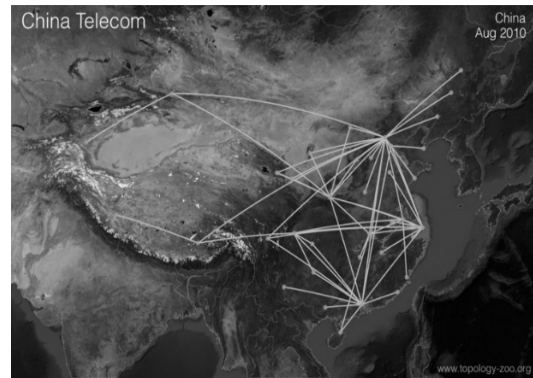


图 1 Chinanet 拓扑

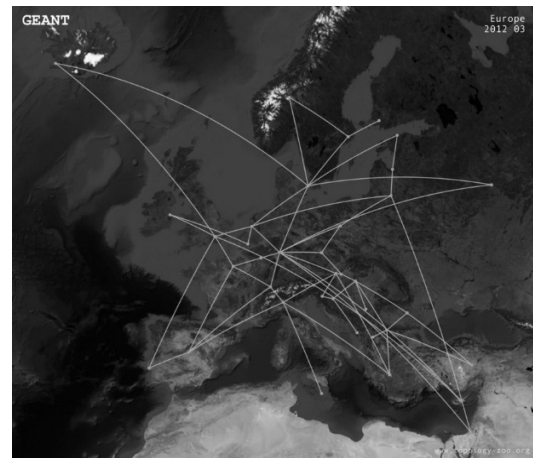


图 2 Geant 拓扑

实验中采用的默认参数设置如下。

节点 v 处的用户数量 N_v 设为该节点的度与其相邻所有节点度之和的 100 倍, 5G 用户占比 μ 为 0.2。4G 用户发起连接数量的最大值 H 设为 10, 5G 用户相比 4G 用户的兴趣倍数 θ 设为 3。时间段长度设为 1 s。上述参数设置可以使单一节点的流数量在数千到数万的量级, 与真实网络的情况^[15]相近。

内容数量与用户数量的比例 ρ 设为 0.1。内容最小值 s_{\min} 取 0.1 Mbit, Pareto 分布参数 k 的默认值约为 1.25。CDN 节点存储容量与本地节点用户数量的比例 τ 设为 0.1 Mbit, 边缘节点存储容量与本地节点用户数量的比例 ϕ 设为 0.01 Mbit。网络中所有的链路容量均设为 100 Gbit/s, 链路的传播时延则由节点间的球面距离除以光速求得。

模拟数值实验一方面需要验证模型的合理性与准确性, 另一方面则希望定量地给出 5G/后 5G

技术部署对互联网主干的影响。下文给出了网络流量、端到端时延、用户数据速率等实验结果,并分析了一些重要参数的变化对结果的影响。

4.1 网络流量

图3和图4分别显示了模型在Chinanet和Geant拓扑中生成的流量矩阵,即各源节点到各目的节点的流量大小。作为对比,图5显示了2004年3月8日Abilene网络中真实测量所得流量矩阵的情况。可以看到,尽管3幅图的流量大小有一定差距,但总体分布趋势十分相似,这表明模型能够在一定程度上反映真实网络的流量特征。虽然生成的Chinanet流量和Geant流量远远大于Abilene流量,但是这是模型参数取值的结果,因为现在距离Abilene流量数据采集的时间已经过去15年,当前真实网络流量大小已经远远大于当时的。从图5中还可以看到,Abilene流量在不同时刻有一定偏差,但总体变化不大,相应地,在图3和图4的结果中引入了一定的随机流量变化。

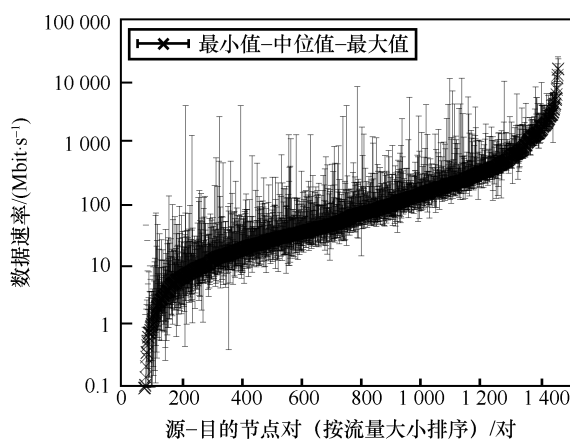


图3 生成的Chinanet流量矩阵

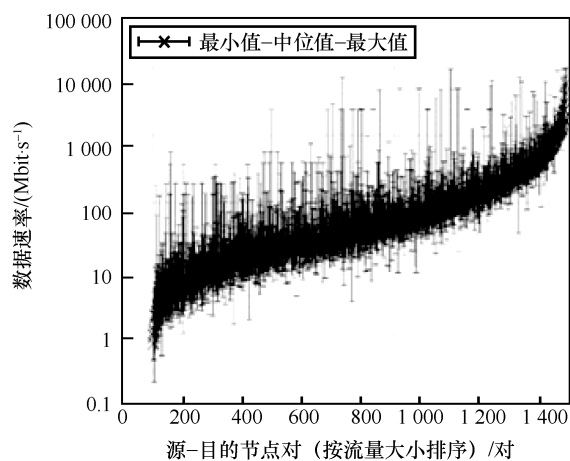


图4 生成的Geant流量矩阵

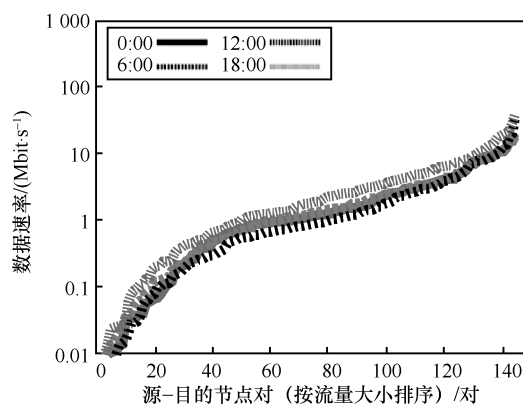


图5 Abilene真实流量矩阵

图6和图7分别显示了Chinanet和Geant拓扑中各链路上的流量大小。可以看到,各链路上的流量跨度很大,最小不到1 Gbit/s,而最大可达20 Gbit/s,这一结果也与当前真实网络的链路流量大体相符。

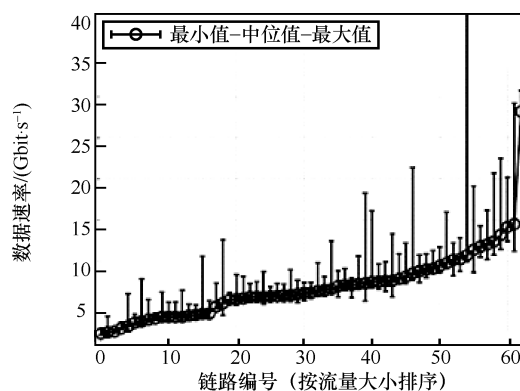


图6 Chinanet链路流量

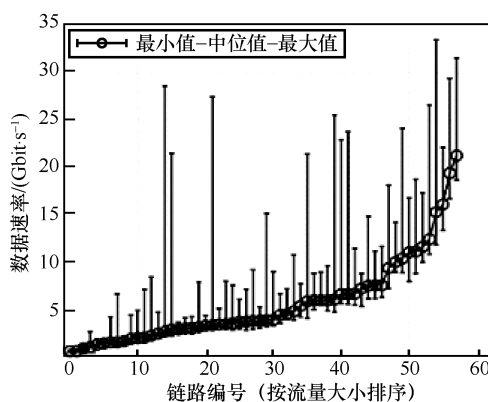


图7 Geant链路流量

图8显示了2个拓扑中生成的节点流量需求。这一结果包括了用户访问本地边缘节点和CDN节点的流量。可以看到,随着5G用户比例的提高,节点流量大致呈线性增长的趋势。当所有用户都使

用 5G 时, 流量大约增长为 4G 的 3 倍, 这是因为相比 4G 用户, 5G 用户的兴趣倍数 θ 设置为 3。从中位数上看, Chinanet 比 Geant 的流量稍大, 但在最大流量方面, Chinanet 要大得多, 这主要是因为 Chinanet 拓扑中包含节点度较大的核心节点。

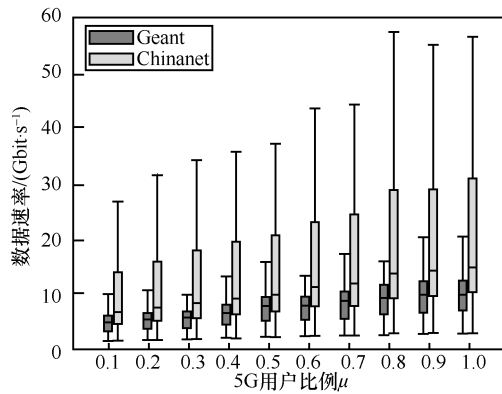


图8 生成的节点流量需求

为了分析 5G 边缘节点存储容量对主干网流量的影响, 模拟实验给出了不同 φ 值下节点流量需求随 5G 用户比例变化的结果。2 个拓扑中的结果分别如图 9 和图 10 所示。

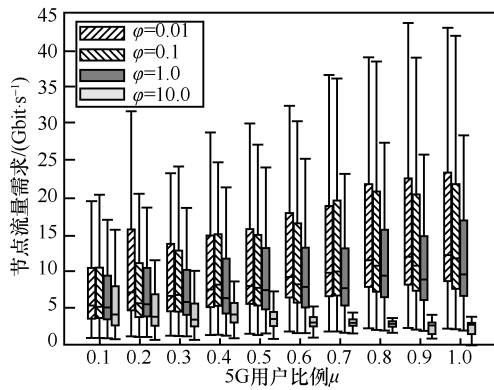


图9 Chinanet 拓扑中节点流量需求随 5G 用户比例变化情况

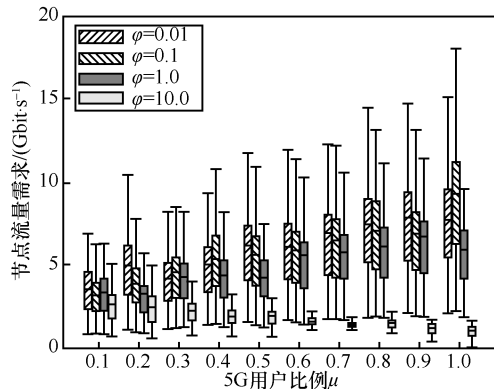


图10 Geant 拓扑中节点流量需求随 5G 用户比例变化情况

注意, 这里的节点流量不包含用户访问本地边缘节点和 CDN 节点的流量, 而仅包含了需要经过主干网链路的流量, 因此当 $\varphi=0.01$ 时的流量比图 8 中的结果更小。当 φ 值较小时, 主干网上的流量随 5G 用户比例的增大而增大, 因为边缘节点存储的内容不能完全覆盖用户访问需求的增加。然而随着 φ 值的增大, 更多的内容可以在边缘节点获取, 因此流量增加的速率逐渐减缓。当 $\varphi=10$ 的时候, 主干网流量随着 5G 用户比例的增大反而减小, 因为此时 5G 所关联的边缘存储和计算资源能够满足 5G 用户访问的所有需求。然而这种情况比较极端, 需要在建设 5G 时部署大量的边缘存储和计算资源, 且需要高效的资源调度算法来使各分散的边缘资源协同工作, 以实现数学模型假设的理想情况。即使在这样的极端假设下, 流量减少的速率也较低, 因此在现实中更加可能出现的情况应该是主干网流量随着 5G 的部署而增加。Geant 拓扑中的实验结果与 Chinanet 拓扑中的结果类似, 但总体上流量更小一些。

上述实验结果表明, 随着 5G 的部署, 互联网主干网将确实面临流量显著增加的挑战。正如第 2.1 节所述, 为了应对这一挑战, 一方面需要在适当扩充主干网容量的同时, 研究更高效的路由寻径、资源调度、传输控制等算法, 以提高主干网资源利用率, 另一方面需要对边缘存储和计算资源的部署和调度方法进行优化, 尽可能使用户需求仅通过访问边缘就得到满足。

4.2 端到端时延

图 11 显示了用户端到端时延的累积概率分布 (又称累积分布函数 (CDF, cumulative distribution function))。可以看到, 2 个拓扑中 5G 用户访问网络的时延大体相同, 在 1~20 ms, 这一数值大约是模型中设定的 5G 接入时延的两倍, 说明在 Chinanet 和 Geant 规模的网络中, 传播时延与 5G 接入时延大体相当, 都在 1~10 ms。此外, 曲线大体呈直线型, 只有首尾的部分略有弯曲, 这是因为弯曲的部分对应于传播时延, 而直线部分则是接入时延取均匀分布的结果。2 个拓扑中 4G 用户访问网络的时延也大体相同, 在 40~110 ms, 这一结果与曲线的形状同样验证了传播时延在 1~10 ms。虽然模型中仅考虑了传播时延而忽略了排队时延和处理时延, 但是上述结果能在一定程度上表明 5G 部署后, 主干网时延会成为远程传输时延的主要组成部分, 这

与 4G 时的情况相比有很大的不同。

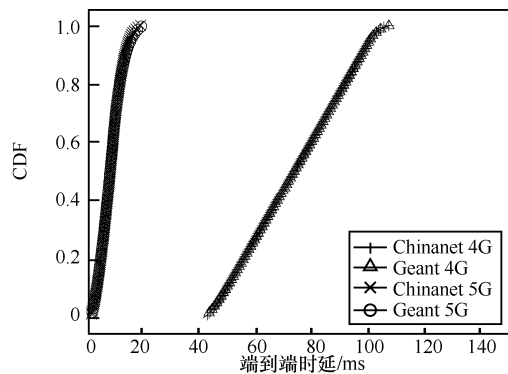


图 11 端到端时延累积概率分布

注意到上述数值实验结果仅包含了传播时延, 因此现实中的端到端时延会更长。为了应对这一挑战, 需要研究更高效的分组处理和路由、传输机制, 以使跨越主干网的访问延时能更加逼近理论极限。同时, 边缘资源的优化利用和调度同样能使用户需求在边缘就得到满足, 起到缩短时延的效果。

4.3 用户数据速率

图 12 和图 13 分别显示了 Chinanet 拓扑中 5G 用户和 4G 用户每条流享有的数据速率的累积概率分布。可以看到, 在数据速率较小值的分布上, 5G 用户与 4G 用户几乎一样。而较大值的分布则呈现了差异。这是因为较小的数据速率主要由主干网的瓶颈链路导致, 而较大值则是用户访问本地缓存的内容时的数据速率, 主要受接入带宽的影响。从图中较为平滑的曲线可以看出, 5G 用户的接入带宽为 100~1000 Mbit/s, 4G 用户为 10~100 Mbit/s, 符合模型中的设定。

同时还可以看到, 当瓶颈在主干网时, 5G 用户占比 μ 值越大, 单流数据速率逐渐减小。具体而言, 当 $\mu=0.2$ 时, 最小数据速率能达到 5 Mbit/s 以上; 当 $\mu=0.5$ 时, 最小数据速率减少了约 1.5 Mbit/s; 当 $\mu=0.8$ 时, 最小数据速率减少了约 1 Mbit/s。这是因为随着 5G 用户比例的提高, 有更多的流加入了带宽的竞争。注意到模拟实验中链路容量均为 100 Gbit/s, 是当前主干网的典型值。这说明随着 5G 的部署, 在不改变现有主干网的容量和路由策略的前提下, 用户远程访问的数据速率会受到较严重的影响。

值得注意的是, 在 5G 用户的结果中几乎没有用户的数据速率分布在 50~100 Mbit/s, 而 4G 用户的数据速率分布更为连续一些。这一现象说明在 5G 场景

下, 网络数据速率的瓶颈主要在主干网, 而 4G 场景下的瓶颈则是主干网和接入网都有可能。

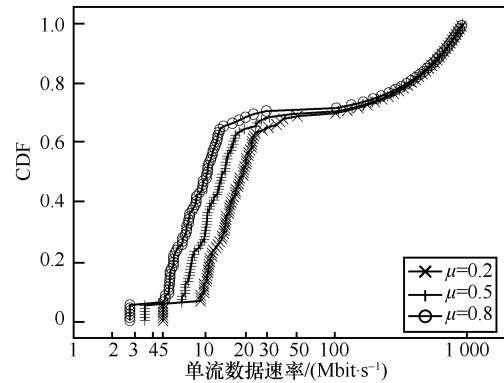


图 12 Chinanet 拓扑 5G 用户单流数据速率

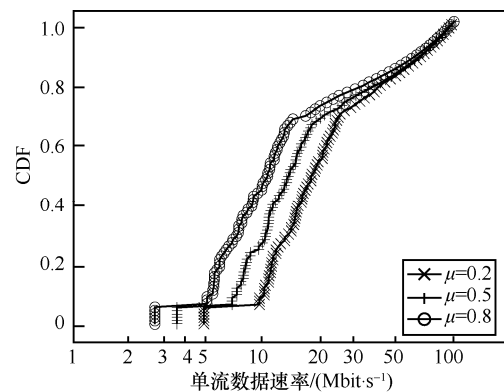


图 13 Chinanet 拓扑 4G 用户单流数据速率

图 14 和图 15 分别显示了 Geant 拓扑中 5G 用户和 4G 用户每条流享有的数据速率的累积概率分布。Geant 拓扑结果大体的趋势与 Chinanet 拓扑中的结果相似, 然而 Geant 拓扑中的最小数据速率略大一些, 这是因为 Geant 拓扑没有节点度特别大的节点, 模型生成的数据流数量少一些, 每条流分得的带宽也就大一些。同样的原因还导致了 5G 用户数据速率的结果在 50~100 Mbit/s 也有一定分布, 而 4G 用户数据速率的结果在 10~30 Mbit/s 不太平滑。

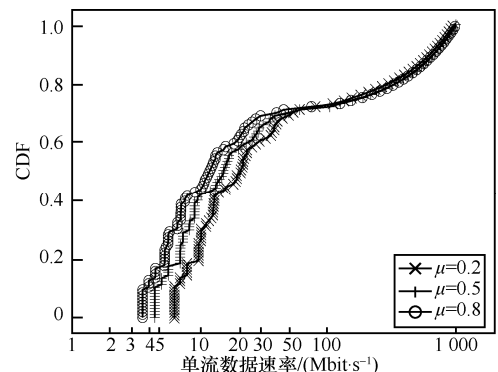


图 14 Geant 拓扑 5G 用户单流数据速率

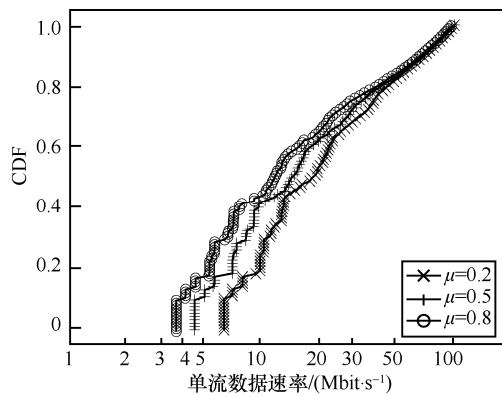


图15 Geant 拓扑 4G 用户单流数据速率

上述实验结果表明,网络数据速率瓶颈会随着 5G 的部署从边缘网逐渐向主干网链路移动。为了应对这一挑战,同样需要从主干网扩容和高效利用、边缘网优化调度等几个方面同时着手。

5 结束语

移动通信技术已经进入了 5G 时代,对于互联网来说既是机遇也是挑战。本文对 5G 及未来的后 5G 部署对互联网主干带来的挑战进行了分析,并建立了抽象模型用于描述 5G 用户的流量特征。模拟数值实验结果表明,即使在引入海量边缘计算和存储资源的情况下,5G 的部署也会引起互联网主干流量的显著增加;移动用户远程访问的时延将大幅减少,但传播时延所占比例将大幅增加。此外,移动用户远程访问的带宽瓶颈位置有由接入网转向主干网的趋势。这些结果为未来网络路由、传输、资源调度和管理、安全等方面的研究提供了一定参考依据。

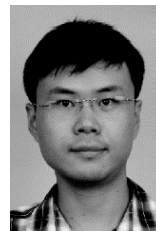
参考文献:

- [1] 张平. B5G: 泛在融合信息网络[J]. 中兴通讯技术, 2019, 25(1):55-62.
ZHANG P. B5G: ubiquitous fusion information networks[J]. ZTE Technology Journal, 2019, 25(1):55-62.
- [2] 张平, 陶运铮, 张治. 5G 若干关键技术评述[J]. 通信学报, 2016, 37(7): 15-29.
ZHANG P, TAO Y Z, ZHANG Z. Survey of several key technologies for 5G[J]. Journal on Communications, 2016, 37(7):15-29.
- [3] PARVEZ I, RAHMATI A, GUVENC I, et al. A survey on low latency towards 5G: RAN, core network and caching solutions[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4):3098-3130.
- [4] 李继蕊, 李小勇, 高云全, 等. 5G 网络下移动云计算节能措施研究[J]. 计算机学报, 2017, 40(7): 1491-1516.
LI J R, LI X Y, GAO Y Q, et al. Energy saving research on mobile

cloud computing in 5G[J]. Chinese Journal of Computers, 2017, 40(7):1491-1516.

- [5] SARKAR S, CHATTERJEE S, MISRA S. Assessment of the suitability of fog computing in the context of Internet of things[J]. IEEE Transaction on Cloud Computing, 2018, 6(1):46-59.
- [6] 赵春华. 5G 承载网的架构演进及带宽分析[J]. 电信科学, 2019(2): 79-83.
ZHAO C H. Architecture evolution and bandwidth analysis of 5G bearer networks[J]. Telecommunications Science, 2019(2):79-83.
- [7] HUSTON G. Another 10 years later[EB]. The Asia Pacific Network Information Centre, (2018-06)[2019-05-19].
- [8] 周紫阳, 魏松杰. 移动蜂窝网络中被动模式时延测量方法[J]. 计算机应用与软件, 2017, 34(11): 147-151.
ZHOU Z Y, WEI S J. A delay measurement for passive mode in mobile cellular networks[J]. Computer Applications and Software, 2017, 34(11): 147-151.
- [9] HU C, BAO W, WANG D, et al. sTube+: an IoT communication sharing architecture for smart after-sales maintenance in buildings[J]. ACM Transactions on Sensor Networks, 2018, 14(3-4):1-10.
- [10] 吴建平, 李丹, 毕军, 等. ADN: 地址驱动的网络体系结构[J]. 计算机学报, 2016, 39(6): 1081-1091.
WU J P, LI D, BI J, et al. ADN: address driven Internet architecture[J]. Chinese Journal of Computers, 2016, 39(6):1081-1091.
- [11] SHI W, COLLINS E, KARAMCHETI V. Modeling object characteristics of dynamic web content[J]. Journal of Parallel and Distributed Computing, 2003, 63(10):963-980.
- [12] BRESLAU L, CAO P, FAN L, et al. Web caching and zipf-like distributions: Evidence and implications[C]// IEEE International Conference on Communications. IEEE, 1999:126-134.
- [13] 中国互联网中心. 《中国互联网络发展状况统计报告》[R]. 北京: 中国互联网中心, 2019.
CNNIC. Statistical report on internet development in china[R]. Beijing: China Internet Network Information Center, 2019.
- [14] KNIGHT S, NGUYEN H X, FALKNER N, et al. The Internet topology zoo[J]. IEEE Journal on Selected Areas in Communications, 2011, 29(9): 1765-1775.
- [15] CAIDA. Trace statistics for CAIDA passive OC48 and OC192 traces[Z]. [S.L.: s.n.], 2019.

[作者简介]



杨芄(1984—), 男, 四川乐至人, 博士, 清华大学助理研究员, 主要研究方向为互联网体系结构、下一代互联网、路由与传输协议。

徐明伟(1971—), 男, 辽宁朝阳人, 博士, 清华大学教授, 主要研究方向为网络体系结构、下一代互联网、网络空间安全。

陈浩(1995—), 男, 北京人, 清华大学博士生, 主要研究方向为下一代互联网、网络功能虚拟化。