



基于负载均衡的 MEH 静态切换机制

刘海鹏^{1,2}

(1. 联想集团, 北京 100085; 2. 北京经贸职业学院, 北京 102488)

摘要: 对欧洲电信标准化协会 (ETSI) 提出的针对移动边缘主机 (MEH) 服务器的切换机制进行了补充和完善, 提出了基于负载均衡的 MEH 静态切换机制。该机制通过将用户设备对应应用实例迁移到与其所属本地高负载服务器相邻的某个低负载服务器, 实现了缓解原控制服务器处瓶颈效应、均衡网络节点间负载、优化信令平均等待时延的目的。分析和仿真结果表明, 提出的静态切换机制能够在均衡网络负载和提升用户满意度方面对 ETSI 规范提供有效的补充和完善。

关键词: 5G; 移动边缘计算; 切换; 负载均衡

中图分类号: TN915

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018183

A static MEH handover scheme based on load balancing

LIU Haipeng^{1,2}

1. Lenovo Group, Beijing 100085, China

2. Beijing Professional Business Institute, Beijing 102488, China

Abstract: Based on the MEH (mobile edge host) handover scheme proposed by ETSI (European Telecommunications Standards Institute), an improved and supplementary static handover scheme was put forward. By transferring the former C-MEH (controlling MEH) to a neighbor server with lighter load, the scheme decreased the average waiting time of a signal message at the former C-MEH site. The negative effect of bottle-neck at the former C-MEH could also be lightened. Analysis and simulation results show that the scheme can provide effective and positive supplementary to the ETSI specification in balancing network load and improving user's satisfaction.

Key words: 5G, mobile edge computing, handover, load balancing

1 引言

伴随着 5G 移动通信技术^[1]大规模商业化的日益临近, 5G 技术所特有的低时延、高带宽、大容量等特性激发起全球电信运营商以及互联网厂商对新兴“杀手级”业务出现的无限憧憬与联想。

与此同时, 正在兴起的车联网 (internet of vehicle)、增强现实 (augmented reality, AR)、物联网 (internet of things, IoT)、智能视频加速 (intelligent video acceleration)、无人机等业务则寄希望于能够得到 5G 网络在诸如端到端时延、网络带宽、存储及计算能力等方面的有力支持, 实现用户体验

和满意度方面质的提升和飞跃。在这种业务与网络相互促进与融合现状的推动下，移动边缘计算（mobile edge computing, MEC, 也被称为多接入边缘计算（multi-access edge computing））^[2-3]技术作为 5G 最重要的代表性业务能力之一，引起了学术界和产业界的高度关注。MEC 通过把网络远端云计算中心侧部分存储和计算能力按需转移到网络边缘侧，使第三方业务厂商能够在更加靠近业务终端的位置向最终用户提供满足超低时延、超高带宽以及超高存储和计算能力的服务。欧洲电信标准化协会（European Telecommunications Standards Institute, ETSI）、第三代合作伙伴计划（3rd Generation Partnership Project, 3GPP）等国际标准化组织正在对 MEC 系统架构及其实现方案进行深入研究、规范和定义。MEC 的系统参考架构如图 1 所示^[4]，通常一个独立的 MEC 系统由若干个移动边缘主机（mobile edge host, MEH）实体及其他软硬件实体协作构成。MEH 实体可被看作移动边缘业务应用（ME application）运行服务器，其软件模块主要包括移动边缘平台（mobile edge platform, MEP）、虚拟基础设施（virtualization infrastructure, VI）以及运行在 MEH 上的多个 ME App 等。移动边缘编排器（mobile edge orchestrator, MEO）是整个 MEC 系统的核心控制管理组件，负责管理和维护当前系统中全部 MEH 及其存储和计算资源，协调各个 ME App 所需承载 MEH 的选择和变换，同时跟踪和触发 ME App 的实例化、运行、终止以及特定场景下 ME App 实例在不同 MEH 之间的迁移等。

2 MEC 系统中的移动性管理

参考现有研究成果^[5]，一种 MEC 系统的典型网络拓扑部署如图 2 所示。整个重叠网络（overlay network）系统分成上下两层，分别是底层的物理通信承载蜂窝网络（LTE（long term evolution）网

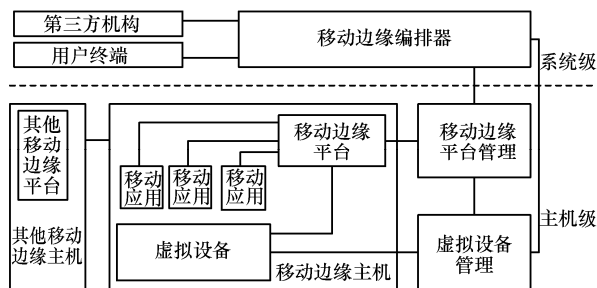


图 1 MEC 系统参考架构

络或者 5G 网络)和上层由 MEC 不同功能实体(如 MEH、MEO 等)构成的 MEC 网络。每个蜂窝小区可以是 LTE 小区或者 5G 小区，一个 MEH 物理上被部署在某一特定蜂窝小区基站处，可能会同时负责管理维护本蜂窝小区和周边多个相邻蜂窝小区内授权用户设备（user equipment, UE）的移动边缘业务。当某一 UE 启动其订购的相应服务后，其所属区域对应 MEH 中会自动实例化一个应用对象，这里称为应用实例（application instance, AI），AI 主要负责保存相应的服务上下文信息，记录当前使用该服务的 UE 信息，同 UE 进行实时的交互，监控和保障服务的正常使用。

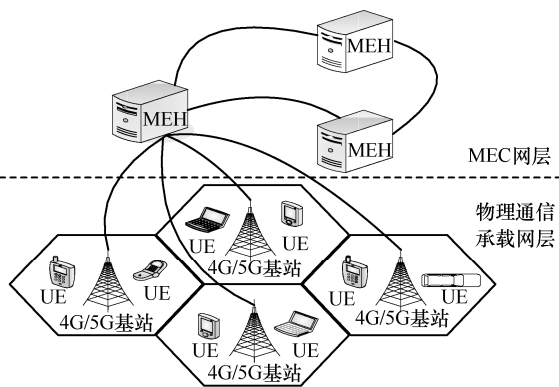


图 2 MEC 系统部署方案示意

在 ETSI 018 规范建议中^[6]，为了保障 UE 上业务连续性和良好的用户业务体验（quality of experience, QoE），当 UE 在底层承载网络相邻蜂窝小区之间移动时，会触发承载网络蜂窝小区基站对 UE 管理权的切换，进一步也有可能触发上层相邻两个 MEH 服务器之间对 UE 上所运行业务应用实例管理权的切换，从而导致 AI 在不同



MEH 之间的迁移。也就是说, UE 的物理移动可能会最终导致特定 AI 从当前源承载 MEH (source MEH, S-MEH) 切换到目标承载 MEH (target MEH, T-MEH), 本文将这种切换方式定义为 MEH 动态切换。同上述场景相对应的是当 UE 在特定承载网络蜂窝小区内处于相对静止状态下 (没有离开当前小区的运动趋势或者动作), 由于当前所属本地 MEH 负载过重导致的服务质量下降和 QoE 异常, 或者其他原因也有可能引起 MEH 管理权的切换, 这里将这种切换方式定义为 MEH 静态切换。

动态切换和静态切换都是将 UE 服务对应的 AI 从 S-MEH 迁移到 T-MEH, 但是两者的触发条件和产生原因是不同的。动态切换发生的最根本原因是 UE 在物理位置上发生了移动和改变, 而静态切换发生的直接原因则是由于当前 MEH 负载过重或者通信链路负载过重等。针对两种切换模式的研究对于保证 MEC 业务连续性和用户业务体验 QoE 都是极其重要的^[4,6], 但是当前针对 MEC 系统中移动性的管理研究^[7-8]以及 MEH 切换技术的研究才刚刚起步, 无论是针对 MEH 动态切换还是 MEH 静态切换的研究成果都很少见, 本文提出一种基于不同 MEH 服务器节点间负载均衡的 MEH 静态切换机制。

3 基于负载均衡的 MEH 静态切换机制

通常情况下 MEC 网络主要由若干 MEH 节点协作构成, 每个 MEH 节点负责维护相应地理范围内全部本地 UE 的 MEC 业务, 这里把特定 UE 当前所属本地 MEH 定义为该 UE 的控制 MEH (controlling MEH, C-MEH), 与 C-MEH 在 MEC 网络拓扑图中相邻的 MEH 被称为参与 MEH (participating MEH, P-MEH)。每个 MEH 节点服务小区的实际管辖范围对应底层一个或者多个蜂窝网络通信小区, 并且 MEH 就部署在某个蜂窝小区的基站侧。为避免混淆, 本文后续在分析 MEH

切换算法时将忽略底层的物理通信承载网络, 只关注上层的 MEH 节点、MEO 节点及 MEC 网络其他相关节点。当特定 UE 在所属本地 MEH 服务小区管辖范围内相对静止 (没有明显离开本 MEH 服务小区的运动趋势或者动作) 时, 有可能发生由于当前 C-MEH 服务器节点负载过重导致的服务质量和用户体验 QoE 下降, 从而触发 MEH 静态切换。

3.1 机制描述

为了方便判定特定 MEH 服务器是否能够被作为待切换目标服务器以及标记与衡量其负载状况, 定义服务器待切换状况参考指数 X 。关于 X 的取值, 首先要考虑该服务器是否能够满足在接管特定 UE 后的业务时延要求, 这可以通过 MEO 定期向指定 UE 及其所属 MEH 发送相关业务时延环回测试指令消息, 并根据返回消息的时延等参数来验证和判定。因为只有当服务器能够满足业务时延要求的前提下, 业务在服务器之间的切换才有实际意义。当服务器满足业务时延要求时, 其对应参考指数 X 的取值主要体现的是服务器当前负载状况, 这可以通过 MEO 定期向相关 MEH 发送指定查询和测量消息, 然后根据接收到反馈消息的内容以及接收到反馈消息的时延等参数进一步遵循特定算法计算得到。进一步定义 X 的有效取值范围 (比如满足条件 $0 < X < 1$), 当 X 取值满足在特定时间段内持续大于特定阈值 (比如 $X > 0.8$), 则认为相应 MEH 处于高负载状况, 需要启动切换算法来进行 UE 所属 C-MEH 服务器的静态切换; 反之, 则不用启动切换算法, 维持当前 C-MEH 不变。这里特定时间段长度和参考阈值的选择要根据具体场景确定, 同时注意避免产生切换的乒乓效应。文中对切换机制的描述重点集中在切换操作的主要触发要素、目标切换节点的选择策略、待切换目标节点队列的更新维护算法等几个主要方面, 关于触发过程中采用的信令交互流程可参照参考文献[6]中关于动态切换信令交互

部分的建议,这里不再赘述。针对特定 UE 的 MEH 静态切换算法操作步骤描述如图 3 所示。

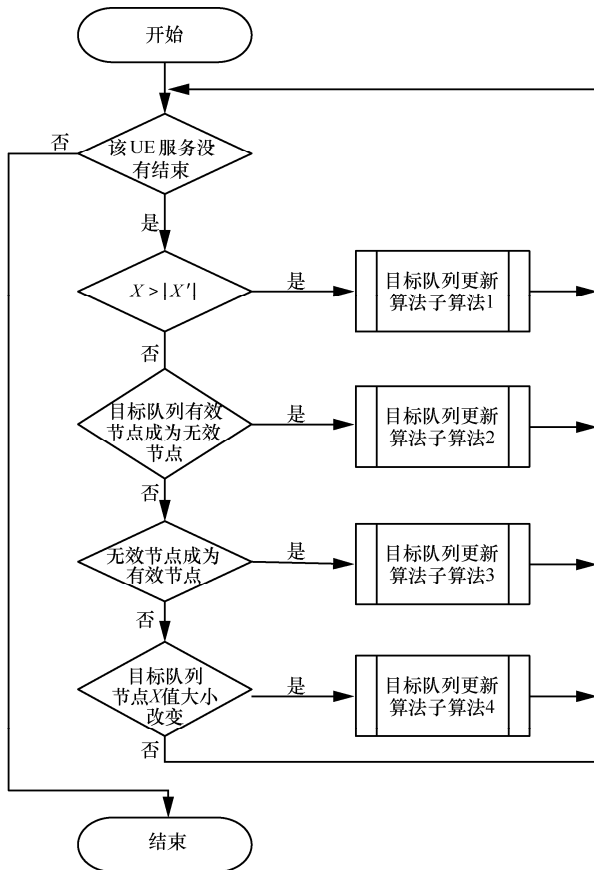


图3 针对特定 UE 的 MEH 静态切换算法操作流程

待切换目标队列更新算法(包括4个子算法)描述如图4所示。

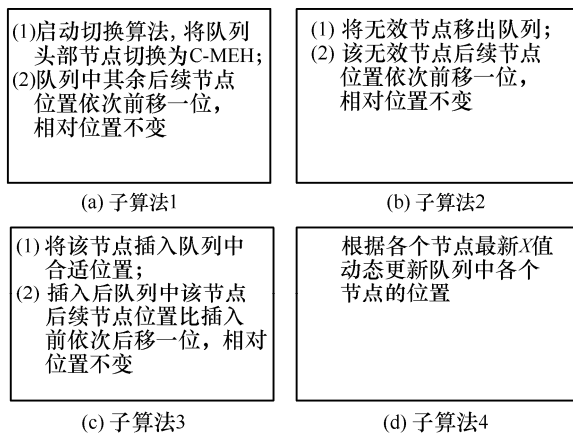


图4 更新待切换目标队列算法

待切换目标队列为当前指定 UE 的 C-MEH 节

点和所有 P-MEH 节点中的有效节点构成的虚拟队列。从队列头至队列尾各个节点按照本身 X 值的大小升序排列,即队列中越是位置靠前(靠近队头)的节点其 X 值越小,负载越低;越是位置靠后(远离队头)的节点其 X 值越大,负载越高。队列头部节点始终为启动切换操作后的首要 T-MEH(队列中负载最低的节点),切换成功完成后,将该头部节点从待切换队列中删除,队列中其他后续节点在队列中的位置均自动前移一位,同时相对位置和顺序保持不变。这样可以保证负载更小的节点有更多的机会被切换成 C-MEH。

队列中有效节点的定义为该节点 X 值不超过阈值的节点,即如果该节点负载状况正常,就可以作为待切换目标节点;反之则为无效节点。由此分析可知,一般来说刚刚完成切换操作的 S-MEH 通常是无效节点(因为该节点 X 通常都是大于阈值的),也不会被插入待切换队列中。总之,待切换队列中的节点都是有效节点,如果队列中某节点 X 值超过了阈值,转换成无效节点,则该节点会实时地从待切换目标队列中被删除,队列中其他节点相对位置和顺序保持不变。反之,如果某无效节点的负载状况有所改善而转换成有效节点,也需要根据该节点的实时 X 值插入队列中相应位置上。另外,队列中所有有效节点在队列中的相对位置和顺序要求根据每个节点最新 X 值实时动态更新。

这里还要补充切换机制要遵循的 2 个基本原则。

- 由于 UE 是相对静止在当前服务小区内,故要求其对应 AI 最远只能由与当前服务小区相邻服务小区的 P-MEH 来接管,这就保证了 UE 始终与当前 C-MEH 距离在 1 跳范围内。距离超出 1 跳范围内的切换机制不在本文讨论范围之内。
- 只要 UE 所在小区的 MEH 服务器节点负载状况恢复正常 (X 小于阈值),就把 UE 所



在小区 MEH 服务器节点切换回 C-MEH 节点, 即 UE 所在小区本地 MEH 服务器节点拥有随时被切换回 C-MEH 的最高优先级 (其优先级超过待切换目标节点队列中所有节点), 这样可以最大限度地保证 UE 同 AI 之间的信令消息在底层承载网络中传输时延最小。

3.2 工作实例

本节提供一个工作实例, 在图 5 中分别就 MEH 静态切换机制以及待切换目标队列的几种典型工作场景进行描述和阐释。假定 MEC 网络由 5 个 MEH 相邻节点构成, 包括 1 个 C-MEH 和 4 个 P-MEH。初始状态某 UE 工作在 C-MEH 管辖范围内, 且 C-MEH 负载状况正常 ($X < |X^*|$, X^* 为负载参考阈值)。图 5 中每个节点标识含义如下: 分隔符前面数字为节点标识, 分隔符后面数字为该节

点在当前待切换等待队列中的位置标识。比如“3-3”表示节点 3, 其在待切换等待队列中位置为 3, 还有 2 个节点排在其前面, 依次类推。有阴影标识的节点为负载异常节点, 没有阴影标识的节点为负载正常节点。双环的节点为当前 C-MEH 节点, 即 UE 服务对应 AI 当前所在节点, 如图 5 (a) 中的节点 0、图 5 (b) 中的节点 1 等。

图 5 (a) 中, 此时某 UE 位于节点 0 所在小区, 并且节点 0 为当前 C-MEH 节点, 节点 1~4 为相邻的 P-MEH 节点, 5 个节点负载状况均为正常。待切换等待队列按照节点先后顺序更新为 1、2、3、4。无效节点集合为空集 \emptyset 。图 5 (b) 中, 节点 0 负载发生异常, C-MEH 切换到等待队列中队头节点 1。待切换等待队列按照节点先后顺序更新为 2、3、4。无效节点集合为 $\{0\}$ 。图 5 (c) 中, 节点 3 发生负载异常, 将节点 3 移除出等待队列, 待切换等待队

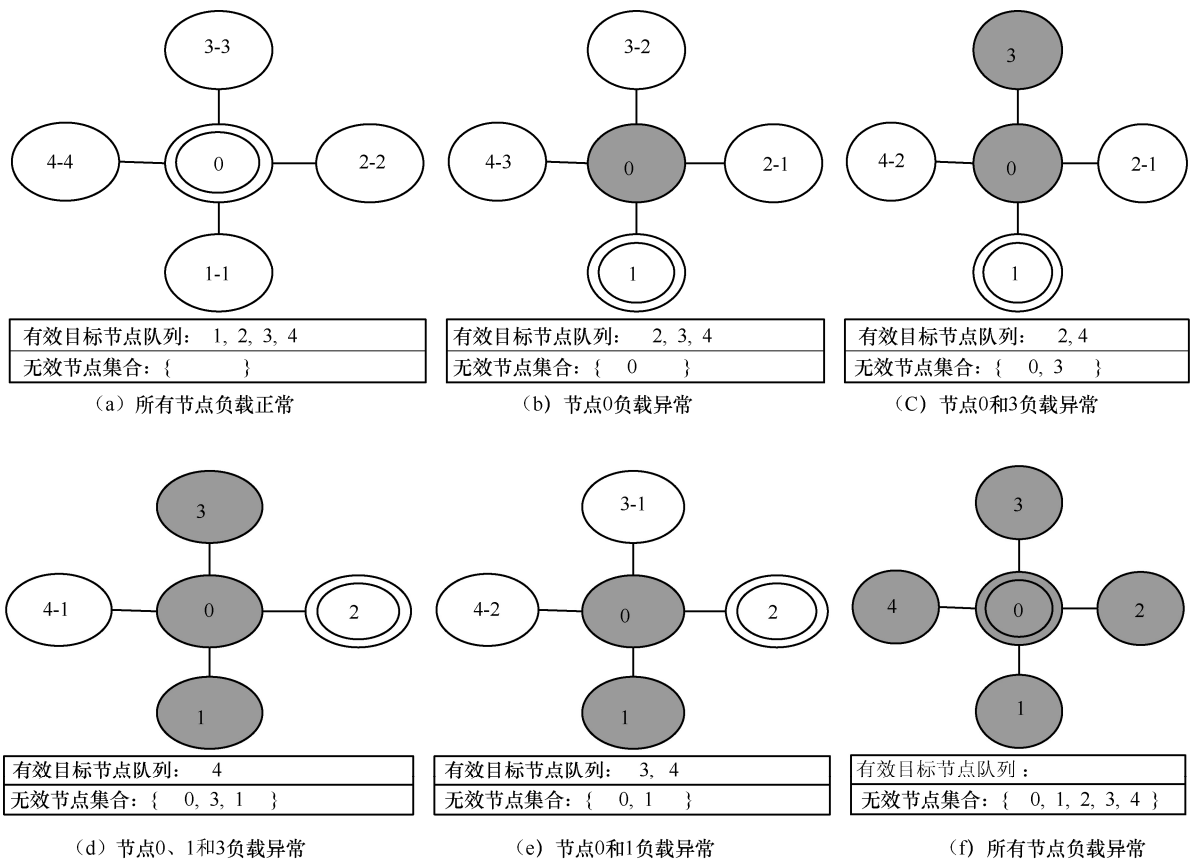


图 5 典型工作场景机制和原理

列按照节点先后顺序更新为 2、4，无效节点集合为 {0, 3}。图 5 (d) 中，节点 1 负载情况发生异常，C-MEH 切换到等待队列中队头节点 2，待切换等待队列按照节点先后顺序更新为 4，无效节点集合为 {0, 1, 3}。图 5 (e) 中，节点 3 负载恢复正常，保持节点 2 作为 C-MEH 不变，将节点 3 插回待切换等待队列相应位置中，待切换等待队列按照节点先后顺序更新为 3、4，无效节点集合相应更新为 {0, 1}。图 5 (f) 中，所有节点负载均发生异常，此时将选定 UE 所属小区对应的本地 MEH 即节点 0 作为 C-MEH，没有下一个待切换目标。

4 数据仿真及分析

基于负载均衡的 MEH 静态切换机制主要作用在上层 MEC 功能节点网络中，如果当前 C-MEH 节点负载过重导致本服务小区下属 UE 服务质量和 QoE 下降，则通过将 C-MEH 功能迁移到相邻的某个负载相对较轻的小区，从而达到平衡 MEH 服务器节点之间负载、减少 UE 用户信令平均等待时延、提升用户体验和 QoE 的目的。下面分别从 MEH 服务器负载均衡性、UE 用户信令

平均等待时延以及该机制的代价 3 个方面进行分析^[9]。为讨论方便且不失一般性，可以假设任意一个 MEH 服务器服务等待队列中来自本小区 UE 的各种信令分组到达率符合泊松分布，每个分组得到 CPU 的处理时间符合负指数分布，只有一个等待处理队列（相当于只有一个服务窗口），且服从“先来先服务”的原则，是一个排队论中 M/M/1 队列模型且服务率保持不变。这里从 UE 所在小区 MEH 服务器节点处的平均服务强度和来自 UE 的信令分组在服务器节点处的平均停留时延两个方面来分析采用静态切换机制前后服务器处负载变化情况，MEC 会话定义相关变量见表 1。

4.1 MEH 服务器负载均衡性分析

4.1.1 UE 所在小区本地 MEH 服务器负载分析

由 M/M/1 队列的性质可知采用静态切换机制前 C-MEH 处平均服务强度为 $\rho = \frac{\lambda}{\mu}$ ，当采用静态切换机制将 C-MEH 功能切换到相邻 MEH 服务器后，原 C-MEH 服务器（已经转换为 P-MEH 服务器）处平均服务强度为 $\rho' = \frac{\lambda'}{\mu}$ 。可

表 1 MEC 会话相关变量定义

变量名称	变量含义
m	UE 所在小区本地 MEH 服务器下属 UE 数目随机变量
k	采用切换机制后从原 C-MEH 迁移出的 UE 数目随机变量
n	采用切换机制后新 C-MEH 所在小区本地下属 UE 数目随机变量
λ	采用切换机制前 C-MEH 处信令分组平均到达率
λ'	采用切换机制后原 C-MEH 处信令分组平均到达率
μ	任意一个 MEH 服务器服务率
ρ	采用切换机制前 C-MEH 处平均服务强度
ρ'	采用切换机制后原 C-MEH 处平均服务强度
d_{UE}	采用切换机制前信令分组在 C-MEH 处平均停留时延
d'_{UE}	采用切换机制后信令分组在原 C-MEH 处平均停留时延
τ_i	每个 UE 在整个会话过程中发出的信令消息总数随机变量，且满足 $i \in (0, m-1)$
Q	切换前在整个 MEC 会话中 C-MEH 到达信令分组总数随机变量
Q'	切换后在整个 MEC 会话中原 C-MEH 到达信令分组总数随机变量
T	单个 MEC 会话总时长随机变量



以假设会话中每个 UE 发出信令消息总数均为 Δn ，得到：

$$\forall j_1, j_2 \in (0, m-1), \tau_{j_1} = \tau_{j_2} = \Delta n \quad (1)$$

由式 (1) 可得切换前在整个会话中 C-MEH 到达的信令分组总数为：

$$Q = \int_0^T \lambda dt = \sum_{i=1}^m \Delta n = m \times \Delta n \quad (2)$$

采用切换机制后 UE 所在小区 MEH 服务器到达的信令分组总数为：

$$Q' = \int_0^T \lambda' dt = \sum_{i=1}^{m-k} \Delta n = (m-k) \times \Delta n \quad (3)$$

定义平均服务强度对比函数：

$$f(m, k) = \frac{\rho}{\rho'} = \frac{\frac{\lambda}{\mu}}{\frac{\lambda'}{\mu}} = \frac{\lambda \times T}{\lambda' \times T} = \frac{Q}{Q'} \quad (4)$$

由式 (2) ~ 式 (4) 可得：

$$f(m, k) = \frac{m \times \Delta n}{(m-k) \times \Delta n} = \frac{m}{m-k} \quad (5)$$

由 Little 公式可知采用静态切换机制前来自 UE 的信令分组在 C-MEH 处的平均停留时延为：

$$d_{\text{UE}} = \frac{1}{\mu - \lambda} \quad (6)$$

当采用静态切换机制将 C-MEH 切换到相邻 MEH 服务器后，来自 UE 的信令分组在原 C-MEH 服务器（该节点角色已转换为 P-MEH）处的平均停留时延为：

$$d'_{\text{UE}} = \frac{1}{\mu - \lambda'} \quad (7)$$

由式 (6)、式 (7) 定义平均停留时延对比函数：

$$g(\rho, m, k) = \frac{m - \rho \times (m-k)}{m \times (1-\rho)} \quad (8)$$

用 MATLAB 仿真工具对式 (5) 和式 (8) 分别分析对比，结果如图 6 和图 7 所示。

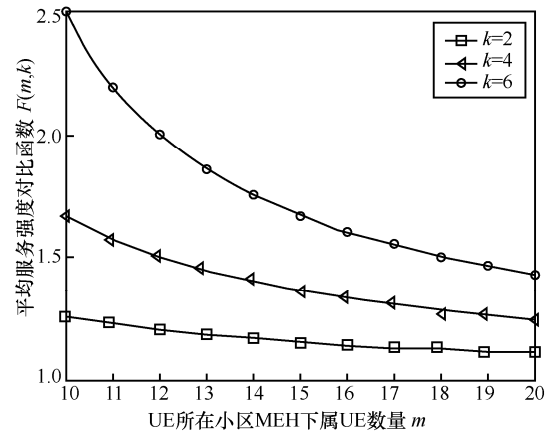


图 6 平均服务强度对比函数

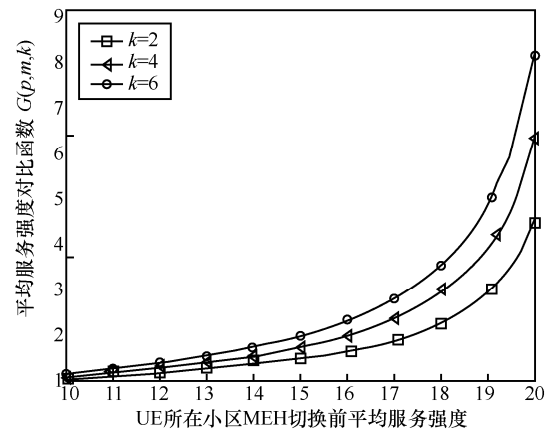


图 7 切换前后平均停留时延对比函数

从图 6 可以看出，当 UE 所在小区本地 MEH 下属 UE 数量固定时， k 值越大， $f(m, k)$ 函数值越大，说明当 UE 数量固定时，迁移出去的 AI 越多，切换后本地 MEH 服务器负载减轻越多。当 k 值固定时，UE 数量越大， $f(m, k)$ 函数值越小，说明相对于较大的 UE 用户数量，只有尽量多地迁移本地 UE 所对应的 AI 到相邻服务器，才能更好地减轻本地 MEH 服务器的节点负载。

从图 7 可以看出，当 k 值固定时，切换前平均服务强度越大（MEH 服务器负载越重）， $g(\rho, m, k)$ 函数值越大，说明切换后信令在服务器处平均等待时延优化效果越好。当切换前平均服务强度固定时， k 值越大， $g(\rho, m, k)$ 函数值越大，说明当 ρ 固定时，迁移出去的 AI 数量越多，对信令平均等待时延的优化效果越好。

4.1.2 切换后新 C-MEH 服务器负载分析

采用 MEH 静态切换机制后,当某 UE 所在小区本地 MEH 服务器负载异常,将 C-MEH 功能迁移到与其相邻且负载相对较轻的节点后,新 C-MEH 节点负载必然会比切换前有所增大。由表 1 中 n 和 k 的定义可知,新 C-MEH 节点在采用切换机制前只是为本地下属 n 个 UE 提供 MEC 服务并且维护相应 AI,采用切换机制后还要接收新迁移过来 k 个 UE 所对应的 AI 的维护管理。由于平均服务强度 $\rho = \frac{\lambda}{\mu}$,其中信令分组到达率 λ 同其下属服务 UE 数量成线性关系而服务率 μ 保持不变,故平均服务强度 ρ 在切换后将会线性增大为切换前的 $\frac{n+k}{n}$ 倍。另外,由于平均停留时延 $d = \frac{1}{\mu - \lambda}$,并且服务率 μ 保持不变,可知切换后该节点处 λ 增大将会导致信令消息在该节点处平均停留时延有所增大。结合对平均服务强度和信令消息平均停留时延的分析可以看出,切换后新 C-MEH 节点负载是增大的,并且增大幅度同迁移过来 UE 的数量正相关。可以认为该节点在采用切换机制后在一定程度上分担了原来只属于被迁移 UE 所在小区的 MEH 服务器(原 C-MEH)负载。

4.1.3 负载均衡性分析

本节分析切换机制前后 MEC 网络各个 MEH 服务器节点之间负载均衡性变化情况。为了讨论方便且不失一般性,可用每个 MEH 服务器所辖 UE 数量随机变量 P 来衡量该服务器处负载情况,下属 UE 数量越多,其负载越大,下属 UE 数量越少,其负载越小。结合参考文献[10,11]中关于判断负载均衡性好坏的算法及方差的物理意义,本文通过随机变量方差的大小来衡量采用切换机制前后系统负载均衡性的好坏,方差越小,说明系统负载均衡性越好,方差越大,说明负载均衡性越差。进一步比较切换前后方差的差值,差值为正且越大说明采用切换对系统负载差异起到的平

衡作用越好。根据方差的基本性质“方差等于平方的均值减去均值的平方”,可得切换前的方差为:

$$\text{Var}(P) = E(P^2) - [E(P)]^2 = \left(\frac{m^2 + n^2}{2}\right) - \left(\frac{m+n}{2}\right)^2 \quad (9)$$

切换后的方差为:

$$\begin{aligned} \text{Var}(P') &= E[(P')^2] - [E(P')]^2 \\ &= \left[\frac{(m-k)^2 + (n+k)^2}{2}\right] - \left[\frac{(m-k) + (n+k)}{2}\right]^2 \end{aligned} \quad (10)$$

由式(9)和式(10)可得切换前后方差的差值为:

$$\begin{aligned} \text{Var}(P) - \text{Var}(P') &= \left\{ \left(\frac{m^2 + n^2}{2}\right) - \left(\frac{m+n}{2}\right)^2 \right\} - \\ &\left\{ \left[\frac{(m-k)^2 + (n+k)^2}{2}\right] - \left[\frac{(m-k) + (n+k)}{2}\right]^2 \right\} \\ &= k \times [m - (n+k)] \end{aligned} \quad (11)$$

在式(11)中,当 $k=0$ 时,意味着没有发生任何 UE 对应 AI 在 MEH 服务器间的转移,讨论无意义。当 $k>0$ 且满足 $(n+k) < m$ 时,切换前后方差的差值恒大于 0。表明只要满足切换后新 C-MEH 当前下属 UE 数量没有原 C-MEH 下属 UE 数量多,切换后负载均衡性就比切换前更好。注意此时切换后新 C-MEH 当前下属 UE 数量包括本来的 n 个 UE 和迁移过来的 k 个 UE,这里 k 个 UE 并未真正发生物理位置上的改变(相应 UE 所属物理承载网络小区没变),但是这些 UE 在 MEC 网络中对应 AI 所属的 MEH 服务器发生了切换。

如果 k 值保持不变,切换前后负载均衡性的差别主要由 m 和 n 的差值大小决定, m 和 n 差值越大,切换前后负载均衡性差别越大,切换后负载均衡性要比切换前更好,可见切换对平衡系统负载起到了积极作用。另外当不同 MEC 小区下属



UE 数量差别很大时,更容易引起不同 MEH 服务器节点负载之间的巨大差异。在现实网络环境下,各个 MEC 小区当前下属 UE 数量 P 是一个随机变量,因此更证明了本文提出的静态切换方案的必要性。

4.2 等待时延分析

考虑到所有 MEH 服务器硬件配置以及处理能力相当并且用户服务率 μ 保持不变,故主要分析 UE 信令在 C-MEH 服务器处的平均等待处理时延以及信令在下层承载蜂窝网络内的传输时延分别在切换前后的差别即可。定义采用切换机制前从 UE 发送一条信令到达其所在小区 MEH 服务器(原 C-MEH 服务器)的传输时延为 $d_{\text{原-传输}}$,该信令在服务器处平均等待处理时延为

$$d_{\text{原-处理}} = \frac{1}{\mu - \lambda_{\text{原}}};$$

采用切换机制后从 UE 发送一条信令到达新 C-MEH 服务器的传输时延为 $d_{\text{新-传输}}$,该信令在新 C-MEH 服务器处平均等待处理时延为

$$d_{\text{新-处理}} = \frac{1}{\mu - \lambda_{\text{新}}}.$$

$$D_{\text{原}} = 2 \times d_{\text{原-传输}} + \frac{1}{\mu - \lambda_{\text{原}}},$$

$$D_{\text{新}} = 2 \times d_{\text{新-传输}} + \frac{1}{\mu - \lambda_{\text{新}}}.$$

切换前由于信令到达率 $\lambda_{\text{原}}$ 过大导致原 C-MEH 服务器处负载过重,从而引起 $d_{\text{原-处理}}$ 过大和 $D_{\text{原}}$ 过大;切换后新 C-MEH 服务器处信令分组平均到达率会满足 $\lambda_{\text{新}} \leq \lambda_{\text{原}}$,从而 $d_{\text{新-处理}} \leq d_{\text{原-处理}}$ 。因此只要不满足 $d_{\text{新-传输}} \gg d_{\text{原-传输}}$ 即可保证文中切换机制对减少用户总等待时延是有效的。信令分组在下层承载蜂窝网络内传输的时延基本可以认为同通信两端物理距离正相关,因此文中提出的切换机制里把待切换目标 MEH 服务器节点限定在 MEC 网络 1 跳节点距离范围内,这样可以最大限度保证不满足 $d_{\text{新-传输}} \gg d_{\text{原-传输}}$,保证进行切换的目的和意义。

4.3 代价分析

引入 MEH 静态切换机制后整个系统付出的代价可以从 3 个方面进行考虑:首先是 MEO 需要维护用于标记与衡量会话中全部相关 MEH 服务器(包括 C-MEH 和 P-MEH)负载状况的参考指数 X ,这项任务对于通常具备很高配置的 MEO 服务器来说,其代价是可以接受的。其次是系统决定启动切换机制继而触发切换操作直到切换完成后所进行的相关信令交互的代价,相对于 5G 服务控制信令与数据消息交互超大总量来说,此方面也是可以忽略不计的。最后,由于采用切换机制后底层承载通信网络相当于也被切换到了其他服务小区,所以 UE 和新 C-MEH 中相关 AI 进行信令消息交互的物理通信时延相应会有所增大,但是由于本切换机制里限定了上层 MEC 网络中的切换操作仅限制在 1 跳距离(拓扑图中两个相邻 MEH 节点)范围内,所以消息交互时延增大幅度基本可以接受(在第 4.2 节中已经进行过分析)。总之,相对于由于当前 UE 所在小区 MEH 服务器负载状况恶化导致的业务连续性和用户业务体验 QoE 严重下降,上述提到的几方面代价能够换来用户体验和满意度方面的明显提升是值得的。

5 结束语

ETSI 对 MEC 会话中 MEH 服务器的切换操作定义^[6]只提到了当 UE 移动引起实际的底层物理承载网络切换时可能会触发上层 MEC 网络中 MEH 服务器切换,没有定义 UE 不动但是由于 UE 当前小区本地 MEH 负载异常导致的服务质量下降时,系统应该采取怎样的应对策略。本文提出的基于负载均衡的 MEH 静态切换机制对上述应用场景进行了补充定义,该机制能够有效平衡 MEC 会话中各 MEH 节点负载,降低 UE 请求处理信令在服务器处平均等待时延和用户总等待时延,提升服务质量和用户体验,对 ETSI 规范进行了有效补充和完善。后续研究工作将对部署在

NFV 环境中 MEC 网络^[12]的 UE 在同一个网络切片甚至跨不同切片^[13]场景下产生明显运动趋势时,系统如何进行预判并预先启动相应切换策略^[6]进行深入的研究和分析,进一步降低用户对小区业务切换产生影响的感知,使无缝切换更好地服务于 5G 低时延、高带宽、大容量的新业务。

参考文献:

- [1] FOUKAS X, PATOUNAS G, ELMOKASHFI A, et al. Network slicing in 5G: survey and challenges [J]. IEEE Communications Magazine, 2017, 55(5): 94-100.
- [2] TRAN T X, HAJISAMI A, PANDEY P, et al. Collaborative mobile edge computing in 5G networks: new paradigms, scenarios, and challenges[J]. IEEE Communications Magazine, 2017, 55(4): 54-61.
- [3] ETSI. White paper No.11 mobile edge computing a key technology towards 5G[R]. 2015.
- [4] ETSI. Mobile edge computing (MEC); framework and reference architecture: GS MEC 003 V1.1.1[S]. 2016.
- [5] 张建敏, 谢伟良, 杨峰义, 等. 移动边缘计算技术及其本地分流方案[J]. 电信科学, 2016, 32(7): 132-143.
ZHANG J M, XIE W L, YANG F Y, et al. Mobile edge computing and application in traffic offloading[J]. Telecommunications Science, 2016, 32(7): 132-143.
- [6] ETSI. Mobile edge computing (MEC); end to end mobility aspects: GR MEC 018 V1.1.1[S]. 2017.
- [7] 田辉, 范绍帅, 吕昕晨, 等. 面向 5G 需求的移动边缘计算[J]. 北京邮电大学学报, 2017, 40(2): 1-10.
TIAN H, FAN S S, LV X C, et al. Mobile edge computing for 5G requirements[J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(2): 1-10.
- [8] 李子姝, 谢人超, 孙礼, 等. 移动边缘计算综述[J]. 电信科学, 2018, 34(1): 87-101.
LI Z S, XIE R C, SUN L, et al. A survey of mobile edge computing[J]. Telecommunications Science, 2018, 34(1): 87-101.
- [9] 刘海鹏, 廖建新, 朱晓民. PoC 中一种负载均衡与时延优化的 RTP 媒体流转发机制[J]. 通信学报, 2010, 31(8): 105-113.
LIU H P, LIAO J X, ZHU X M. RTP stream transferring scheme with load balancing and delay optimization in PoC[J]. Journal on Communications, 2010, 31(8): 105-113.
- [10] TURGUT D, TURGUT B, DAS S K, et al. Balancing loads in mobile ad hoc networks[C]//The 10th International Conference on Telecommunications, ICT 2003, Feb 23-Mar 1, 2003, Papeete, French Polynesia. Piscataway: IEEE Press, 2003: 490-495.
- [11] 郑相全. 基于负载均衡的无线自组网关键技术与算法研究[D]. 成都: 电子科技大学, 2005: 44-45.
ZHENG X Q. Research on load-balance based key technologies and algorithms for ad hoc networks[D]. Chengdu: University of Electronic Science and Technology of China, 2005: 44-45.
- [12] ETSI. Mobile edge computing (MEC); deployment of mobile edge computing in an NFV environment: GR MEC 017 V1.1.1[S]. 2018.
- [13] 项弘禹, 肖扬文, 张贤, 等. 5G 边缘计算和网络切片技术[J]. 电信科学, 2017, 33(6): 54-63.
XIANG H Y, XIAO Y W, ZHANG X, et al. Edge computing and network slicing technology in 5G[J]. Telecommunications Science, 2017, 33(6): 54-63.

[作者简介]



刘海鹏 (1975-), 男, 博士, 联想集团高级工程师, 北京经贸职业学院副教授, 主要研究方向为网络功能虚拟化 (NFV) 及移动边缘计算。