



一种基于差异化服务的 MEC 应用实例智能迁移机制

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: 基于 ETSI 提出的 MEC 系统移动用户应用实例智能迁移机制, 提出了一种增强的基于差异化服务的应用实例智能迁移机制。根据需要迁移的用户数量和业务数量定义了 4 种状态消息同步模式, 使得运营商可以对具有不同服务性能优先级请求的用户和业务的处理具有更大的灵活性, 为实现差异化服务和计费提供了更多的选择。

关键词: 5G; 移动边缘计算; 切换; 迁移; 差异化服务

中图分类号: TN915

文献标识码: A

doi:10.11959/j.issn.1000-0801.2018242

An intelligent migration mechanism for MEC application instances based on differentiated services

LIU Haipeng^{1,2}

1. Lenovo Group, Beijing 100085, China

2. Beijing Professional Business Institute, Beijing 102488, China

Abstract: Based on the application instance smart relocation scheme proposed by ETSI, a differentiated service based and enhanced scheme was put forward. Four state message synchronization modes were defined according to the number of users to be migrated and the number of services, so that operators could have greater flexibility in handling users and services with different service performance priority requests, providing more choices for differentiated services and billing.

Key words: 5G, mobile edge computing, handover, relocation, differentiated service

1 引言

伴随着 5G^[1]商用大潮日益临近, 网络切片、移动边缘计算 (mobile edge computing, MEC)^[2-3]等 5G 核心技术也越来越引起学术界和产业界的高度关注。MEC 也被称为多接入边缘计算 (multi-access edge computing) 技术, 是 ETSI

(European Telecommunications Standards Institute) 和 3GPP (3rd Generation Partnership Project) 等几个主要国际标准化组织联合华为等几个主要全球电信设备提供商定义的边缘网络计算模式和平台。目的是更好地在网络边缘向终端用户提供超低时延、超高带宽以及超高存储和计算能力的服务。MEC 系统参考架构如图 1^[4]所示, 移动边缘

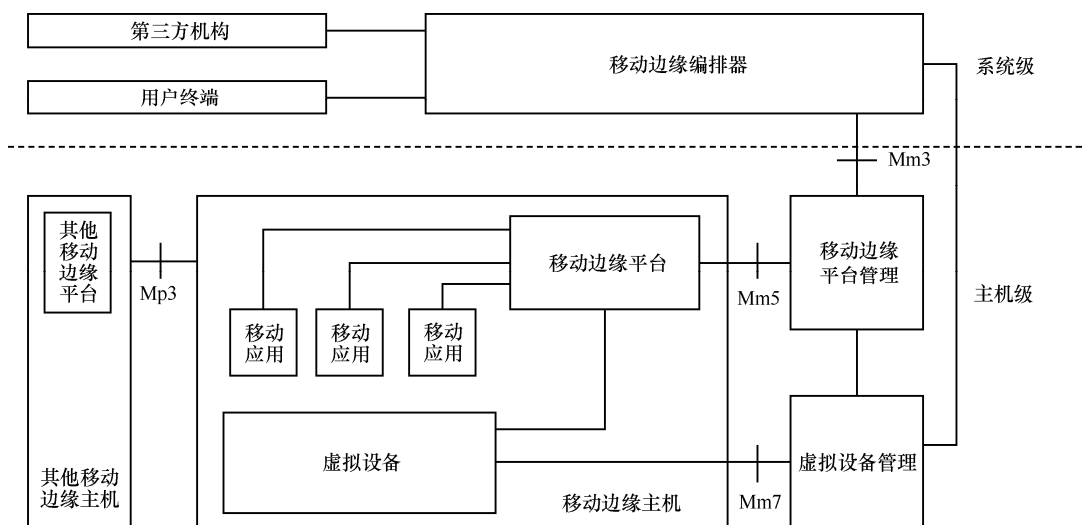


图1 MEC系统参考架构

主机（mobile edge host, MEH）通常由高性能配置服务器承担，负责本MEC小区的业务维护等工作，其上运行着移动边缘平台（mobile edge platform, MEP）、虚拟设备（virtualization infrastructure, VI）、移动边缘业务应用（ME App）等软件模块，是本小区的主要控制节点。多个MEH节点通过Mp3接口之间的协作可以组合成一套完整的MEC系统，其中，移动边缘编排器（mobile edge orchestrator, MEO）是核心控制管理组件，负责管理和维护当前系统中全部MEH及其存储和计算资源，在特定情况下协调各个用户正在使用的业务所对应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。MEO的管理操作主要通过Mm3接口来实现。移动边缘平台管理模块以及虚拟设备管理模块可以对MEH节点中的移动边缘平台（Mm5接口）及虚拟设备（Mm7接口）分别进行管理操作。关于MEC系统中各个模块的功能及其相互之间协作关系的详细描述可参见参考文献[4]中的相关内容。

MEC业务是由其下层的4G/5G蜂窝移动通信网络来承载提供的，MEC网络和下层蜂窝移动通信网络构成了一个重叠网络（overlay network），其

部署方案如图2所示。

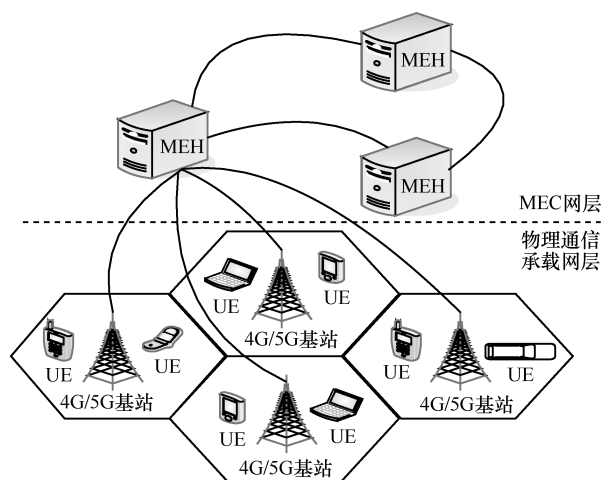


图2 MEC系统部署方案示意图

每个MEH节点对应一个MEC服务小区，同时可能对应着下层物理通信承载网络里的一个或者多个蜂窝基站小区。当某一移动用户设备（user equipment, UE）启动其订购的相应服务后，其所属MEC小区对应MEH中会自动实例化一个应用实例（application instance, AI）对象，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。MEH节点负责跟踪、维护和监管当前MEC小区内所



有 UE 及其相关 AI。如果 UE 只是在 MEC 小区内部移动,并不涉及 MEC 小区之间各个 MEH 节点对 AI 管理权的切换(虽然可能会有底层蜂窝小区基站之间的切换)。但是一旦某个正在接受服务的 UE 从当前所属 MEH 小区移动到其他某个相邻 MEH 小区时,系统为了保障业务连续性,会自动触发相关两个 MEH 之间对该 UE 业务管理权的切换操作,就会引起 AI 及其相关状态信息在 MEH 之间的迁移。ETSI 对 AI 及相关应用状态信息在不同 MEH 之间的迁移场景进行了详细的规范和定义,本文对原规范进行了针对性的完善和补充,主要是增加了系统对差异化服务场景的支持,并且给出了 4 种状态消息同步方法的定义。

2 ETSI 规范中应用实例智能迁移机制

当正在接受 MEC 服务的 UE 从当前所属 MEH 小区移动到其他相邻 MEH 小区时,如果当前业务对应 AI 仍然保持在当前 MEH 处不变,势必会由于 AI 同 UE 的实时物理距离增大而导致业务时延增大,进而导致业务体验和 QoE (quality of experience) 的下降。可见在用户不停机的情况下,实现 AI 从 S-MEH(source MEH)在线迁移到目标 T-MEH(target MEH)是保证用户良好业务体验和 QoE 的必要前提。如果发生业务的迁移,势必会影响到用户的实际感受。迁移速度快,同时用户感受不明显的情况通常

称为高业务连续性;反之,迁移速度慢,用户能够感受到业务受到一定程度影响的情况,则称为低业务连续性。

切换过程中应用实例以及 UE 状态信息迁移^[5]见表 1。

从表 1 中可以看出,ETSI 只对高业务连续性场景进行了分析和探讨,并没有对相对低业务连续性进行讨论。另外,ETSI 规范目前只针对单一 UE 以及单一应用对应 AI 的迁移流程以及消息交互提供了初步建议。在 MEO 处并没有针对切换过程中多个 UE 或者 UE 对应的多个应用进行任何针对性的策略部署以及针对享受不同服务优先级的 UE 和应用区别对待,从而实现 MEC 系统对差异化服务的更好支持。

一套成熟完善的 MEC 系统不仅应该能够为用户提供高优先级、高性价比、高可靠性和连续性^[6]的业务,也应该具备区分用户等级,在特殊场合及场景下提供低优先级、低连续性以及业务保证的服务能力。只有这样平台运营商才能够具有制定更多计费策略和方案、快速回收建设运维成本、最大化收益及 ARPU (average revenue per user) 值的能力^[7]。所以不仅要讨论高业务连续性下 AI 及状态迁移对 MEC 系统和 MEO 的要求,同样也要分析相对低业务连续性下系统的应对策略。可见制定切实可行的分级化、层次化、区别化操作方案对 UE 切换造成的 MEC 业务连续性问题进行处理是十分必要的。

表 1 切换过程中应用实例以及 UE 状态信息迁移

业务移动性	范围	状态	高业务连续性下的 AI 迁移	低业务连续性下的 AI 迁移
MEH 内部移动	任意	任意	不需要	不需要
MEH 之间移动	专用	无状态	AI 迁移	有待研究
		有状态	状态迁移(和/或) AI 迁移	有待研究
		无状态	AI 迁移(特定情况下)	有待研究
	共享	有状态	状态迁移(和/或) AI 迁移(特定情况下)	有待研究

3 基于差异化服务的 MEC 应用实例智能迁移机制

3.1 基于差异化服务的业务切换场景

在第2节 ETSI 规范中只讨论了单个 UE 及其单个应用 (App) 模式下如何进行 AI 在不同 MEH 之间的迁移, 再考虑以下几种服务场景。

(1) 是否有待切换 UE 正在使用的其他业务也需要发生 AI 迁移

某 UE 在自动驾驶的汽车上^[8]同时享受移动视频播放业务, 此时需要发生切换的是针对该 UE 的自动驾驶和移动视频播放两个业务, 甚至可能存在单个 UE 对应的更多个 (大于 2 个) 业务同时需要切换的情况。

(2) 是否有针对同一业务的多个 UE 也在同一时刻进行切换

正在同一个 MEH 小区内享受同一个广播或者多播业务的多个 UE 同时发生切换。比如电影或者其他集会散场后多个用户在时间和空间上同时离开当前 MEH 小区进入相邻小区, 就需要对多个 UE 对应的同一个业务同时进行切换。

(3) 是否有多个 UE 的多个业务在同一时刻有业务迁移请求

以上两种情况的综合考量, 有时会发生多个 UE 的多个业务在足够短的时间段内发出业务切换和迁移的请求。

综合上述分析, 可以定义以下 4 种状态消息同步模式来满足和支持服务差异化的要求: 单用户单应用 (single UE single App, SUSA) 模式、单用户多应用 (single UE multi App, SUMA) 模式、多用户单应用 (multi UE single App, MUSA) 模式、多用户多应用 (multi UE multi App, MUMA) 模式。系统可以为不同业务订购者提供 4 种模式中的一种, 实现其对不同 QoE、不同服务等级、不同业务感受的要求。如果一个用户为高优先级用户, 那么当他进行切换操作时, 系统会尽最大

努力为其分配 SUMA 模式, 使所有正在使用的应用及服务以最快的速度完成切换。同样地, 如果一个业务为高端业务, 其所有订购使用者在切换过程中会被分配 MUSA 模式, 保证该业务所有用户能够享受最快速的切换服务。SUSA 和 MUMA 则可以根据系统中注册用户的不同优先级有针对性地区分使用。

3.2 基于差异化服务的应用实例及状态智能迁移机制消息交互流程

ETSI 规范中将 AI 及应用状态迁移操作步骤分为有状态和无状态两种情况分别讨论, 详细流程可参见 ETSI 规范^[5]。将图 1、图 2 中所述情况进行合并, 合并后的状态同步消息交互流程如图 3 所示。

对图 3 中主要操作步骤的业务描述如下。

(1) S-MEP 接收到关于发起业务迁移的触发请求 (该请求可能由底层承载网络或者 S-MEH 中相关服务如 S-RNIS 发出)。

(2) 如果 UE 享受的是有状态的服务, 则由 S-MEP 负责记录下来当前 UE 状态信息; 如果是无状态的服务, 忽略此步。

(3) S-MEP 转发该业务迁移请求给 MEO, 请求消息中包含应用标识 (appDid) 和目标底层蜂窝网络节点标识。

(4) MEO 根据请求消息中包含的目标底层蜂窝网络节点选择合适的 T-MEH, 并且通知 T-MEPM 在 T-MEH 中进行 AI 的初始化操作。这里要区分 AI 是某 UE 专用还是同其他 UE 共享使用, 专用时一般都需要重新实例化一个 AI; 共享时则要注意是否已经有相应 AI 被实例化过了, 如果没有, 需要新实例化一个 AI; 否则不需要。

(5) MEO 通过 S-MEPM 发送响应消息给 S-MEP。

(6) 如果是有状态的 AI 转移, 则由 S-MEP 向 T-MEP 发送 AI 状态转移请求, 在请求消息中会包括上述第 (2) 步记录下来的状态信息。注意

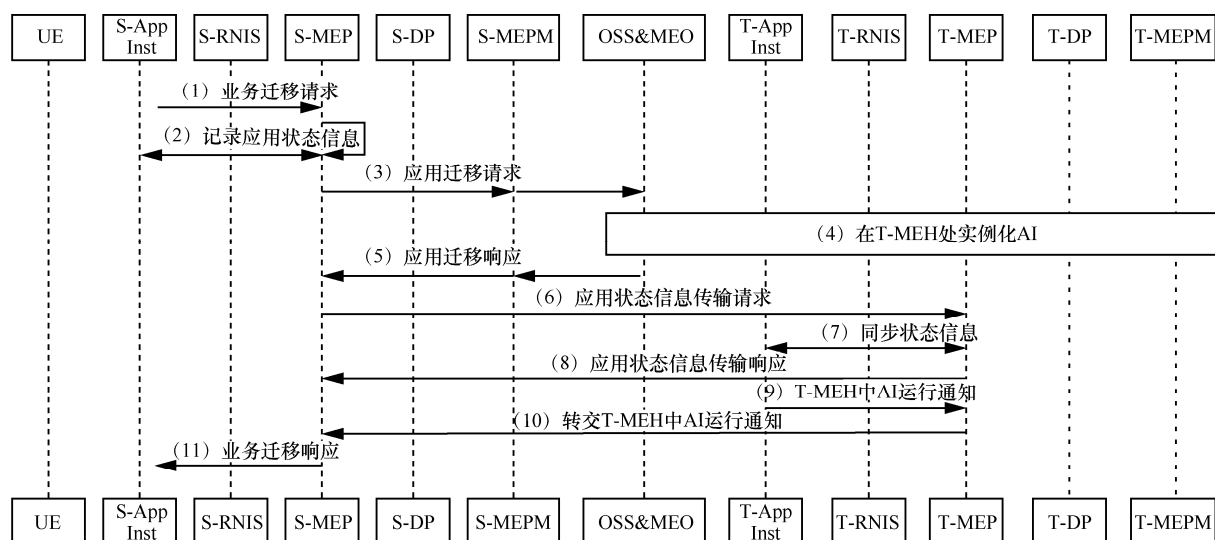


图3 ETSI 规范中 AI 及应用状态迁移消息交互流程示意图

此处 ETSI 并没有对状态信息中需要包含的内容进行深入的定义和划分。

(7) T-MEP 将同步状态信息转发给 T-MEH 中已经被实例化的 AI, 进行状态同步激活。

(8) T-MEP 向 S-MEP 发送 AI 转移完成响应。

(9) AI 在同步完状态信息后向 T-MEP 发送一个正常运行的通知消息。

(10) T-MEP 将该通知消息转发给 S-MEP。

(11) S-MEP 向发出业务迁移触发请求的发送者发出迁移完成响应, 以便其可能触发下一步 ME 移动性的相关操作。

为实现基于差异化的服务体验, 针对图 3 中 ETSI 对 AI 及应用状态迁移消息交互流程定义, 同时结合第 3.1 节中定义的 4 种状态消息同步模式, 本文重点补充了 MEC 系统应该如何采用差异化服务的策略和处理方法的步骤, 重点对图 3 中第 (3) 步~第 (10) 步进行细化、改进及完善。改进后的基于差异化服务的 AI 及应用状态智能迁移机制的信令交互流程如下所示。

(1) 同原机制。

(2) 同原机制。

(3) S-MEP 转发该业务迁移请求给 MEO, 请求消息中包含发生切换的 UE 标识、应用标识

(appDId) 和目标底层蜂窝网络节点标识。

(4) MEO 根据请求消息中包含的目标底层蜂窝网络节点选择合适的 T-MEH, 并且通知 T-MEPM 在 T-MEH 中进行 AI 的初始化操作。这里要区分 AI 是某 UE 专用还是同其他 UE 共享使用, 专用时一般都需要重新实例化一个 AI; 共享时则要注意是否已经有相应 AI 被实例化过了, 如果没有, 需要新实例化一个 AI, 否则不需要。

(5) MEO 需要根据第 3.1 节中提供的差异化服务场景进行分析, 然后在给 S-MEPM 的响应消息里要对将要进行的状态信息同步操作模式进行建议和要求, 需要指定是 SUSA、SUMA、MUSA、MUMA 4 种模式中的哪种, 然后 MEO 通过 S-MEPM 发送响应消息给 S-MEP。

(6) S-MEP 会根据第 (5) 步消息中 MEO 的建议按照相应的同步模式和消息格式向 T-MEP 发送同步状态消息传输请求。在请求消息里面将包含请求切换 UE (单个或多个) 和 App (单个或多个) 的状态信息。

(7) T-MEP 将同步状态信息转发给 T-MEH 中已经被实例化的 AI, 进行状态同步激活。这里可能是 T-MEP 向多个 AI 发送状态同步消息, 因为在 SUMA、MUSA 和 MUMA 模式下, T-MEP

需要将第(6)步 S-MEP 发送过来的状态同步消息进行拆分,把属于不同 UE 或者 App 的内容从消息中取出来向对应 AI 转发过去,并且进行状态同步激活操作。

(8) T-MEP 向 S-MEP 发送 AI 转移完成响应。

(9) 所有相关 AI 在状态信息同步完成后向 T-MEP 发送一个正常运行的通知消息。

(10) T-MEP 将该通知消息转发给 S-MEP。

(11) 同原机制。

第(5)步中 MEO 首先会综合考虑如下信息:当前请求切换 UE 和 App 的优先级信息;是否有同一 UE 的其他应用也需要进行切换;是否有针对同一应用的多个 UE 需要切换;是否有必要对多个 UE 和多个 App 同时进行切换。然后 MEO 再为当前切换选择 4 种状态同步消息中的一种,通过响应消息的方式发送给 S-MEP。第(6)步中发起切换的 S-MEP 则根据建议将相关所有 UE 和 App 状态信息打包成一条同步消息一次性发送给 T-MEP。第(7)步中 T-MEP 则将接收到的消息拆包并进行二次分发,把属于不同 UE 或者 App 的内容从消息中取出来向对应 AI 转发过去,并且进行状态同步激活操作。

3.3 4 种状态消息同步模式下消息内容格式定义

根据第 3.1 节中对 4 种状态信息同步模式的定义,对第 3.2 节状态迁移流程中第(6)步包含的应用状态信息传输请求消息(状态同步消息)的内容和格式进行分析。本节给出 4 种模式下状态同步消息内容格式建议。

(1) 单用户单应用(SUSA)

SUSA 模式下一条同步消息需要承载的内容包括一个 UE 以及一个 App 的信息。SUSA 模式下状态同步消息内容格式建议如图 4 所示,其中, AI 信息^[9]主要包含该 AI 的标识信息,如 AI-ID (App instance ID) 等, AI-ID 是 MEPM 在实例化 AI 的时候分配产生的,可以用于标识不同的 AI。特定 UE 信息主要包括当前正在切换 UE 的标识。

特定 App 信息主要包括当前发生切换 UE 正在使用的业务应用标识。

AI信息	特定UE信息	特定App信息
------	--------	---------

图 4 SUSA 同步消息内容格式建议

(2) 单用户多应用(SUMA)

SUMA 模式下一条同步消息需要承载的内容包括一个 UE 以及多个 App 的信息。SUMA 模式下状态同步消息内容格式建议如图 5 所示, AI 信息主要包含该 AI 的标识信息,如 AI-ID 等。特定 UE 信息主要包括当前正在切换 UE 的标识。App1, App2, ..., Appn 信息主要包括当前发生切换 UE 正在使用的所有待切换业务应用标识。

AI信息	特定UE信息	App1 信息	App2 信息	...	Appn 信息
------	--------	------------	------------	-----	------------

图 5 SUMA 同步消息内容格式建议

(3) 多用户单应用(MUSA)

MUSA 模式下一条同步消息需要承载的内容包括多个 UE 的信息以及一个 App 的信息。MUSA 模式下状态同步消息内容格式建议如图 6 所示, AI 信息主要是要包含该 AI 的标识信息,如 AI-ID 等。UE1, UE2, ..., UEn 信息主要包括当前发生切换所有 UE 的标识。特定 App 信息主要包括当前发生切换的所有 UE 同时在使用的同一个业务应用标识。

AI信息	UE1 信息	UE2 信息	...	UE _n 信息	特定App信息
------	-----------	-----------	-----	-----------------------	---------

图 6 MUSA 同步消息内容格式建议

(4) 多用户多应用(MUMA)

MUMA 模式下一条同步消息需要承载的内容包括多个 UE 的信息以及多个 App 的信息。MUMA 模式下状态同步消息内容格式建议如图 7 所示, AI 信息主要是要包含该 AI 的标识信息,如 AI-ID 等。UE1, UE2, ..., UEn 信息主要包括当前发生切换所有 UE 的标识。App1, App2, ...,



App n 信息主要包括当前发生切换所有 UE 正在使用的所有待切换业务应用标识。

AI信息	UE1 信息	UE2 信息	...	UE n 信息	App1 信息	App2 信息	...	App n 信息
------	-----------	-----------	-----	--------------	------------	------------	-----	---------------

图7 MUMA 同步消息内容格式建议

需要补充说明的是, 上述 4 种消息内容格式的建议更多体现的是 MEO 在实现差异化服务指导思想的基础上对 AI 及其状态迁移相关联的 S-MEH 和 T-MEH 如何进行应用状态内容的同步进行的策略性指导和建议。有了 4 种模式的内容格式建议, 在发生切换时, 系统可以任意选择其中一种或者几种的组合, 从而为用户提供基于不同感受的业务连续性体验, 方便地实现系统提供差异化服务的最终目标。其中, 消息中每个子域具体包含的内容及其所需数据传输量可根据具体业务来确定。比如 AI 信息中可以包含诸如 AI-ID (可以是字符串 string 类型数据), 也可以再进一步包含 AI 名称 (AI-Name, 可以是字符串 string 类型数据) 或者其他信息详细描述部分内容。UE 信息域以及 App 信息域所含具体内容及数据传输量同样可以根据用户和业务的具体实际需求自行定义, 这里不再赘述。

4 工作实例

本节通过一个工作实例来分析基于差异化服务的 AI 迁移机制下 4 种状态消息转移模式是如何工作的以及每种切换模式具体的工作过程。由于无状态业务不涉及状态信息的同步, 这里不做分析。为能够清晰地表述 4 种切换机制的工作机制和原理并且进行相互之间的对比, 这里仅讨论当用户数 (UE 数量) 和业务数 (App 数量) 均较小时的情况。4 种机制下用户数和业务数较大时的工作原理是类似的, 这里不再赘述。

假设当前 S-MEH 小区有两个 UE, 分别为 UE1 和 UE2。正在享受两种有状态业务, 分别为

App1 和 App2。对应 AI 分别为 AI1、AI2 和 AI3。其中 AI1 为 UE1 和 UE2 共享使用, 对应 App1; AI2 为 UE1 专用, 对应 App2; AI3 为 UE2 专用, 对应 App2。假设 UE1 和 UE2 试图同时从 S-MEH 小区切换到 T-MEH 小区。图 8 列举出了在整个迁移过程中采用不同的迁移模式系统可能存在的全部中间状态。

基于第 3.2 节中状态迁移消息交互流程, 这里分别对基于 4 种不同状态转移模式下的状态迁移过程进行分析和对比。在整个迁移过程中 4 种模式下大部分信令消息交互步骤都是相同的, 只有第 (6) 步和第 (7) 步有所不同, 所以此处重点关注在这两步系统在不同模式下进行状态迁移时的差别。

在 SUSA 模式下, S-MEP 根据 MEO 关于状态同步模式的建议, 向 T-MEP 发送满足 SUSA 内容格式建议的状态同步消息。由于消息中只包含单个 UE 和单个 App 信息, 所以 T-MEP 只需将接收到的消息中属于该 UE 和相应 App 的内容从消息中提取出来向对应 AI 转发过去即可。由于 S-MEP 每次只能转发一个 UE 及其 App 的内容, 所以从初始状态 (图 8(a)) 开始, AI1 的部分内容 (包含 UE1 和 App1 信息) 首先转移到 T-MEP, 如图 8(b)所示。其次是 AI2 的全部内容 (包含 UE1 和 App2 信息) 通过 SUSA 消息转移到 T-MEP, 如图 8(c)所示。再次是 AI1 的部分内容 (包含 UE2 和 App1 信息) 通过 SUSA 消息转移到 T-MEP, 如图 8(d)所示。最后是 AI3 的全部内容 (包含 UE2 和 App2 信息) 通过 SUSA 消息转移到 T-MEP, 如图 8(f)所示。每两个相邻状态之间通过 S-MEH 向 T-MEH 发送符合 SUSA 格式的状态同步消息来实现状态的转换。

在 SUMA 模式下, S-MEP 根据 MEO 关于状态同步模式的建议, 向 T-MEP 发送满足 SUMA 内容格式建议的状态同步消息。由于消息中包含单个 UE 和多个 App 信息, 所以 T-MEP 在接收到

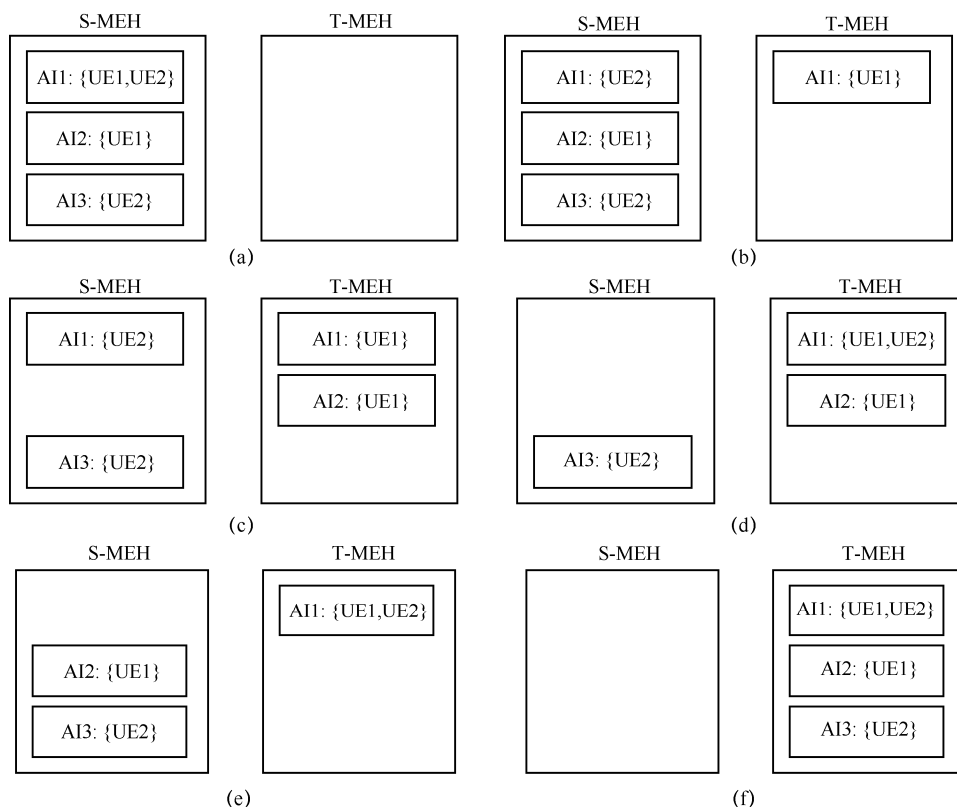


图8 4种模式下AI迁移顺序实时状况分布

消息后需要将接收到的消息中属于该 UE 和相应多个 App 的内容从消息中提取出来分别向对应多个 AI 转发。首先 AI1 的部分内容(包含 UE1 和 App1 信息)以及 AI2 的全部内容(包含 UE1 和 App2 信息)会合并并在一条 SUMA 消息中, 转移到 T-MEP, 如图 8(c)所示。其次是 AI1 的部分内容(包含 UE2 和 App1 信息)以及 AI3 的全部内容(包含 UE2 和 App2 信息)会合并并在一条 SUMA 消息中转移到 T-MEP, 如图 8(f)所示。每两个相邻状态之间通过 S-MEH 向 T-MEH 发送符合 SUMA 格式的状态同步消息来实现状态的转换。

在 MUSA 模式下, S-MEP 根据 MEO 关于状态同步模式的建议, 向 T-MEP 发送满足 MUSA 内容格式建议的状态同步消息。由于消息中包含多个 UE 和单个 App 信息, 所以 T-MEP 在接收到消息后需要将接收到的消息中属于该 App 和相应多个 UE 的内容从消息中提取出来分别向对应多

个 AI 转发。由于 S-MEP 每次转发一个 App 及其相关多个 UE 的内容, 所以从初始状态(图 8(a))开始, 首先 AI1 的全部内容(包含 UE1、UE2 和 App1 信息)会合并并在一条 MUSA 消息中转移到 T-MEP, 如图 8(e)所示。其次是 AI2 的全部内容(包含 UE1 和 App2 信息)以及 AI3 的全部内容(包含 UE2 和 App2 信息)会合并并在一条 MUSA 消息中转移到 T-MEP, 如图 8(f)所示。每两个相邻状态之间通过 S-MEH 向 T-MEH 发送符合 MUSA 格式的状态同步消息来实现状态的转换。

在 MUMA 模式下, S-MEP 根据 MEO 关于状态同步模式的建议, 向 T-MEP 发送满足 MUMA 内容格式建议的状态同步消息。由于消息中包含多个 UE 和多个 App 信息, 所以 T-MEP 在接收到消息后需要将接收到的消息中属于相应多个 App 及其相应多个 UE 的内容从消息中提取出来分别向对应多个 AI 转发。由于 S-MEP 一次性转发多



个 UE 及其相关多个 App 的内容, 所以从初始状态 (图 8(a)) 开始, 只需一条完整的状态同步消息就可以把全部 AI (AI1、AI2、AI3) 的全部内容 (包含 UE1、UE2 和 App1、App2 信息) 合并并在一条 MUMA 消息中转移到 T-MEP, 如图 8(f) 所示。

从上述工作实例可以看出, SUS A 模式下系统经过的中间状态是最多的, 要经过 4 次状态的改变才能达到最后的目标状态; SUM A 模式和 MUS A 模式下系统分别需要经过 2 次状态的改变; MUMA 模式则只需要经过 1 次状态改变即可。

5 效率及适用场景

5.1 效率分析

在第 4 节讨论了当用户数 (UE 数量) 和业务数 (App 数量) 均较小 (均为 2) 时 4 种机制的工作状况。为了具有普适性和代表性, 本节选择用户数和业务数均较大的情况对 4 种机制进行效率的对比和分析。假设当前 S-MEH 小区有 n 个 UE, 分别为 UE1, UE2, ..., UEn。每个 UE 享受的业务数量最多为 m 个 (最少为 0 个), 这 m 个业务分别为 App1, App2, ..., App m 。在同一时刻, 假设 S-MEH 中的所有 AI 全部要求切换到 T-MEH 中, 这里分别从状态同步消息交互数量、数据传输总量、消息交互时延 3 个方面来对比 4 种机制的效率。

(1) 同步消息交互数量

4 种模式下状态同步消息交互数量对比见表 2。其中, 4 种模式下同步消息交互数量主要是指

第 (6) 步和第 (8) 步所对应的消息条数总和, 不包括其他步骤所对应的指令消息。

根据状态同步消息的定义可知, 无状态业务不涉及状态的转移, 所以表 2 中无状态业务情况下几种状态同步模式不存在消息交互。根据定义, 无论是有状态还是无状态, 专用 AI 的数量都是 mn , 共享 AI 的数量则是业务的数量 m (与用户数量 n 无关)。

AI 专用业务有状态情况下, SUS A 模式第 (6) 步和第 (8) 步分别只传递一个 UE 和一个 App 的状态信息, 所以共需要 $2mn$ 条同步消息。SUM A 模式第 (6) 步和第 (8) 步分别在一条请求消息里包含一个 UE 和 m 个业务的状态信息, 所以共需要 $2n$ 条同步消息。MUS A 模式第 (6) 步和第 (8) 步分别在一条请求消息里包含 n 个 UE 和一个业务的状态信息, 所以共需要 $2m$ 条同步消息。MUMA 模式第 (6) 步和第 (8) 步分别在一条请求消息里包含 n 个 UE 和 m 个业务的状态信息, 所以只需要 2 条同步消息。

AI 共享业务有状态情况下 AI 个数为 m 个, 而专用业务有状态情况下 AI 个数为 mn 个, 4 种模式下需要传递的同步消息数量是相同的。

综合上述分析可以看到 SUS A 模式下同步消息数量最多, 效率相对来说最低。SUM A 模式和 MUS A 模式下同步消息数量分别与 UE 个数 n 以及业务个数 m 正相关, 属于同一数量级, 都比 SUS A 模式下少。MUMA 模式下只需要 2 条同步消息就可以完成状态的同步, 具有最高的效率。

表 2 4 种模式下状态同步消息交互数量对比

状态	范围	AI 个数	SUS A 下状态同步消息交互数量	SUM A 下状态同步消息交互数量	MUS A 下状态同步消息交互数量	MUMA 下状态同步消息交互数量
无状态	专用	mn	无	无	无	无
	共享	m	无	无	无	无
有状态	专用	mn	$2mn$	$2n$	$2m$	2
	共享	m	$2mn$	$2n$	$2m$	2

(2) 数据传输总量

结合第 3.3 节中对 4 种模式消息格式定义以及表 2 中的消息交互总量, 可以假定 AI 信息数据传输长度为 α , 一个 UE 信息数据传输长度为 β , 一个 App 信息数据传输长度为 γ 。得到 4 种模式下状态同步消息数据传输总量, 见表 3。

由于 SUSA 模式下的消息内容只包括一个 AI 信息数据长度、一个 UE 信息数据传输长度和一个 App 信息传输长度, 所以单个 SUSA 消息长度为 $(\alpha+\beta+\gamma)$ 。所需数据传输总量为 $2mn(\alpha+\beta+\gamma)$ 。

SUMA 模式下的消息内容包括一个 AI 信息数据长度、一个 UE 信息数据传输长度和 m 个 App 信息传输长度, 所以单个 SUMA 消息长度为 $(\alpha+\beta+m\gamma)$ 。所需数据传输总量为 $2n(\alpha+\beta+m\gamma)$ 。

MUSA 模式下的消息内容包括一个 AI 信息数据长度、 n 个 UE 信息数据传输长度和一个 App 信息传输长度, 所以单个 MUSA 消息长度为 $(\alpha+n\beta+\gamma)$ 。所需数据传输总量为 $2m(\alpha+n\beta+\gamma)$ 。

MUMA 模式下的消息内容包括一个 AI 信息数据长度、 n 个 UE 信息数据传输长度和 m 个 App 信息传输长度, 所以单个 MUMA 消息长度为 $(\alpha+n\beta+m\gamma)$ 。所需数据传输总量为 $2(\alpha+n\beta+m\gamma)$ 。

当 α 、 β 与 γ 为常量情况下, SUSA、SUMA 与 MUSA 3 种模式下的数据传输总量基本为变量 m 和 n 乘积的函数, 其函数值与 mn 成正相关。数据传输总量基本相当。MUMA 模式下的数据传输总量只与 m 或 n 的值正相关, 当 m 和 n 均很大时

其总量明显要小于其他 3 种模式。可见当用户数很多并且业务数很多的情况下, 如果能够选用 MUMA 模式, 对于减少状态同步过程中的数据传输总量是非常有益的。

(3) 消息交互时延

基于第 3.2 节中对状态同步消息交互流程的定义可知, 4 种模式下信息交互时延的差异主要是由流程中第 (6) ~ (8) 步来决定的。由于 MEH 服务器通常都是配置充足的服务器系统, 所以在发送消息所需时长方面可以对单个消息长度忽略不计, 重点关注所需发送消息的数量。结合表 2、表 3 中的信息, 可以假定发送一条同步消息所需时长为固定值。在第 (6) 步中, SUSA 模式下需发送消息数量最多, 为 m 和 n 乘积的函数, 总共所需发送时长是 4 种模式中最长的。SUMA 模式和 MUSA 模式总时长分别是 n 和 m 的函数, 总共所需发送时长处于中等水平。MUMA 模式下由于发送的消息数量最少, 所以发送总时长是最短的。

在第 (7) 步中, SUMA、MUSA、MUMA 3 种模式均需要 T-MEH 将收到消息中属于不同 UE 或者 App 的内容从消息中取出来向对应 AI 转发过去, 表面看比 SUSA 模式下多了拆分的步骤, 实际上 SUSA 模式下关于每个 AI 的信息在 T-MEH 被转发的总数应该与另外 3 种机制下被转发的总数相等 (因为迁移过来的 AI 是完全相同的)。所以在第 (7) 步中 4 种机制所需总时长应该也是相同的。

在第 (8) 步中, SUMA 模式下 T-MEP 向 S-MEP

表 3 4 种模式下状态同步消息数据传输总量对比

状态	范围	AI 个数	SUSA 下状态同步消息数据传输总量	SUMA 下状态同步消息数据传输总量	MUSA 下状态同步消息数据传输总量	MUMA 下状态同步消息数据传输总量
无状态	专用	mn	0	0	0	0
	共享	m	0	0	0	0
有状态	专用	mn	$2mn(\alpha+\beta+\gamma)$	$2n(\alpha+\beta+m\gamma)$	$2m(\alpha+n\beta+\gamma)$	$2(\alpha+n\beta+m\gamma)$
	共享	m	$2mn(\alpha+\beta+\gamma)$	$2n(\alpha+\beta+m\gamma)$	$2m(\alpha+n\beta+\gamma)$	$2(\alpha+n\beta+m\gamma)$



发送 mn 条 AI 转移完成响应消息。SUMA 模式下, T-MEP 向 S-MEP 发送 n 条 AI 转移完成响应消息。MUSA 模式下 T-MEP 向 S-MEP 发送 m 条 AI 转移完成响应消息。MUMA 模式下 T-MEP 向 S-MEP 只需发送 1 条 AI 转移完成响应消息。所以, SUSA 模式下发送响应的总时长最长, SUMA 和 MUSA 两种模式发送响应总时长介于中间, MUMA 模式下发送响应的总时长最短。

综上所述, 在总的消息交互时延方面, SUSA 模式所需总时延最长, SUMA 和 MUSA 模式所需总时延介于中间, MUMA 模式所需总时延最短。

5.2 适用场景

进一步分析 4 种状态消息同步模式分别适合的业务场景, 见表 4。

如果没有针对特定 UE 和相关服务的优先级方面的要求, 则适合采用 SUSA 模式。当 UE 享受的所有服务都拥有高优先级, 在切换时则使用 SUMA 模式来进行状态迁移。当 UE 同时享有多个服务且每个服务拥有不同的服务优先级时, 在切换状态下, 系统会根据不同服务优先级的高低区分对待。将高优先级的相关服务使用 SUMA 模式来进行状态迁移, 其他低优先级服务则使用 SUSA 模式来进行状态迁移。

如果某特定业务属于高优先级业务, 只要在使用该业务的 UE 全都是高优先级用户, 当发生切换时, 该业务所对应的用户将全部享受 MUSA 模式来进行状态迁移。即一旦发生正在使用该业务的某个用户的切换事件, MEO 会监

控是否同时存在其他使用该业务的用户也在进行切换, 如果有, 则将对所有正在切换的 UE 状态进行消息合并, 采用 MUSA 方式进行状态迁移。

当普通用户享受单个高优先级业务时, 建议使用 MUSA 模式进行切换操作, 这样可以保证该高优先级业务不区分用户的等级对所有用户一致对待, 使所有用户尽可能有相同的业务感受和 QoE 体验。当高优先级用户享受普通业务时, 建议采用 SUMA 模式进行切换操作, 这样可以保证该用户的高标准服务等级, 不会进一步区分其选择的业务等级高低, 从而保证该用户对所有业务的满意度和业务体验。

MUMA 模式则适用于所有类型的 UE 及 App 组合。这需要根据平台运营商结合套餐和资费的实际情况来灵活掌握。

6 结束语

本文基于差异化服务的 AI 智能迁移机制对 ETSI 原有的机制进行了改进与增强, 通过合并可能同时需要发生切换的其他 UE 或者 App 请求, 为 AI 迁移定义了 4 种可选的同步模式供系统选择, 从而最大限度地实现了特定情况下同步消息数量的减少和切换效率的提升, 最终实现了系统对差异化服务的支持。增强的机制使得运营商可以对具有不同服务性能优先级请求的 UE 和 App 的处理具有更大的灵活性, 对整个 MEH 切换过程中 QoE 提升提供了更多的操作选择, 对 ETSI 规

表 4 4 种状态迁移消息模式适应场景对比

模式	高优先级用户/ 所有业务	高优先级业务/ 所有用户	相同高优先级用户组/ 相同高优先级业务组	普通用户/ 普通业务	普通用户/ 高优先级业务	普通业务/ 高优先级用户
SUSA				√		
SUMA	√					√
MUSA		√			√	
MUMA	√	√	√	√	√	√

范进行了有力补充和完善。下一步研究将针对如果 UE 对应的多个 App 属于不同的移动网络切片^[10], 当 UE 发生在线切换时系统将如何应对。

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] ETSI. Mobile edge computing (MEC); end to end mobility aspects: GR MEC 018 V1.1.1[S]. 2017.
- [6] 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.
- [7] 何晓明. EV-DO 系统支持差异化服务的 QoS 应用研究[J]. 电信科学, 2009, 25(6): 90-92.
HE X M. The QoS application survey of differentiated service based on EV-DO system[J]. Telecommunications Science, 2009, 25(6): 90-92.
- [8] ETSI. Mobile edge computing (MEC); service scenarios: GS MEC-IEG 004 V1.1.1[S]. 2015.
- [9] ETSI. Mobile edge computing (MEC); mobile edge management; part 2: application lifecycle, rules and requirements management: GR MEC 010-2 V1.1.1[S]. 2017.
- [10] 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。