

请发明人员自行

复印和保留复本

**发**

**明**

**报**

**告**

**书**

**(1)**

|  |  |  |  |
| --- | --- | --- | --- |
| 提交日： 年 月 日 | 将以本报告书为基础进行专利检索和申请，以下**\***部分为必填项目。 | | |
| **\***发明名称(**中+英**)： 一种车载中继（VMR）辅助的分层联邦学习网络及其切换方案  A VMR-aided Hierarchical Federated Learning Network and Its Handover Scheme  一种分层联邦学习网络下的切换方案  A Handover Scheme in Hierarchical Federated Learning | | | |
| **\***发明概要：(请用100字左右总结该发明)  此方案首先确立了车载中继VMR辅助下的分层联邦学习网络，该网络分为两层：第一层由各个VMR与其服务的用户设备（ UE）组成；第二层由gNB以及与其连接的各个VMR组成。各UE首先进行次迭代本地模型更新，再将结果上传至VMR进行局部聚合（partial aggregation）；各VMR进行次局部聚合后，再将结果上传至基站进行全局聚合（global aggregation）。其次本方案考虑由于VMR或UE的移动性导致的切换问题。gNB根据VMR上报的相关信息进行预测与判断，UE是否需要进行切换。如果需要切换，应尽量保证UE在全局聚合且该模型广播下发后进行，此时各UE，VMR的模型相同，均为全局模型，无需进行额外的模型传输（model transfer），切换代价最小；如不能保证则尽量在VMR完成局部聚合后进行切换，此时VMR与其服务的所有UE模型相同，且能保证UE的训练服务不中断。此外，还可以通过提高发送功率，降低RSRP门限等方法，延长UE与VMR或VMR与gNB的连接时间至局部聚合或全局聚合完成。 | | | |
| **\***Inventor Information Sheet(超过两位请另填Excel附件)  1. 发明人是指对本发明区别于现有技术的部分作出实质性贡献的人员。  2. 发明人员中包括公司外人员，或者本发明是与其他单位联合研发时，请明确注明。   |  |  |  | | --- | --- | --- | |  | 第一发明人 | 第二发明人 | | 英文名（拼音） | Ce ZHENG |  | | 中文名 | 郑策 |  | | 员工号 | 700027697 |  | | 所属单位/部门/联系电话 |  |  | | 居住地 (做出发明时所在的居住地) | Beijing, China |  | | 国籍 | China |  | | 中文通信地址 | 北京市朝阳区太阳宫中路12号冠城大厦701 100028 |  | | 英文通信地址 | Room, 701, Citychamp Building, No. 12 Tai Yang Gong Zhong Lu, Chao Yang District, Beijing 100028, P.R.China |  | | 电子邮箱 |  |  | | 身份证号（中国籍的第一发明人须填写） |  | N/A | | | | |
| **\***技术领域(e.g. 移动通信LTE-A)：  **\***涉及具体项目：□是，项目名称 □否  **\***涉及技术标准：□是，标准名称 □否  发明独特性：□唯一解法 □少数解法中的一种 □多种解法中的一种 | | | **\***研发状态：  □创意阶段  □研发/试作阶段  □产品化阶段，机型 | |
| **\***公开计划：  □无 □有，预定日期  公开方式：  □展示 □销售 □订货 □出厂 □发表论文 □其它 | | **\***现有技术调查：□未调查 □已调查  检索结果：（与发明重点相关的参考文献或专利申请号。空间不够时请添加附件） | | |

以下评价表格为**部门填写**：

|  |  |  |  |
| --- | --- | --- | --- |
| 统括课长 | **\***技术及市场角度综合评价（请领导填写） | 统括部长 | 知产负责人 |
|  | □重点，原因  □一般  □不申请，原因  （e.g.无保护必要，可对外公开；以商业秘密等其他形式保护） |  |  |
| 意见： | | | |

**发**

**明**

**报**

**告**

**书**

(2)

**发明的重点**

专利说明书中的[权利要求范围]（CLAIM）的栏目内容。发明的要点（CLAIM）**至少7项**，其中，第1项发明重点应当包括对最主要创新点的概括，其他项发明重点可以是在第1项基础上附加的创新点的概括，并且请对每一项发明重点中包含的创新点所带来的技术问题及效果进行说明。

|  |  |
| --- | --- |
| **权利要求** | **技术问题及效果** |
| 1. **一种 车载中继（VMR）辅助的分层联邦学习网络**：   该系统为两层结构：第一层由各个VMR与其服务的用户设备（ UE）组成；第二层由gNB以及与其连接的各个VMR组成。各UE（作为一级节点）首先进行多次本地模型训练更新，再将结果上传至VMR（作为二级节点）进行局部聚合（partial aggregation）；各VMR进行多次局部聚合后，再将结果上传至gNB进行全局聚合（global aggregation）。 | 此可能为本专利的主要创新点（分层联邦学习网络结构已经存在，**但是没有人考虑过用VMR作为二级节点实现局部聚合。此外，二级节点还可以时UAV，RSU等。** |
| 1. 一种 车载中继（VMR）辅助的分层联邦学习网络下的 切换方案：   gNB根据VMR上报的相关信息进行预测与判断，UE是否需要进行切换。如果需要切换，UE尽量保证在全局模型聚合且广播下发后进行切换，此时各UE，VMR的模型相同，无需进行额外的模型传输（model transfer），切换代价最小；如不能保证则尽量在VMR完成局部聚合后进行切换，此时VMR与其服务的所有UE模型相同，且能保证UE的训练服务不中断。此外，还可以通过提高发送功率，降低RSRP门限等方法，延长UE与VMR或VMR与gNB的连接时间至局部聚合或全局聚合完成。 | **此为本专利的主要创新点，该方案适用于所有的分层联邦学习（不止是VMR作为二级节点）** |
| 1. **一种配置在UE端的设备**，主要用于：   接收来自于VMR的局部模型或全局模型；（详见权利要求9）  基于本地数据，进行本地的模型训练与模型更新；上传更新的本地模型至VMR；（详见权利要求6）  对与自身状态相关的信息 进行估计并上报至VMR； （详见权利要求10）  接收来自于VMR或gNB的切换策略，并进行相应的切换。 | 该系统为两层结构：底层由各个VMR与其服务的用户设备（UE）组成；上层由基站（gNB）以及与其连接的各个VMR组成。各UE首先进行次迭代本地模型更新，再将结果上传至VMR进行聚合（aggregation）；各VMR进行次聚合后，再将结果上传至gNB进行全局聚合（global aggregation）。其具体流程为权利要求5--8 |
| 1. **一种配置在VMR端的设备**，主要用于:   接收来自于其服务UE的本地更新模型，并进行VMR的局部模型聚合与更新；将聚合后的局部模型进行下发至UE或者上报至gNB；接收来自于gNB的全局模型，并转发至UE；（详见权利要求7）  接收来自于UE的，并对与自身状态相关的信息 进行估计。上报和；（详见权利要求11）  接收并执行来自gNB的切换策略，同时将其发送至对应的UE。 |
| 1. **一种配置在gNB端的设备**，主要用于：   接收来自于各VMR的局部聚合模型，并进行全局模型的聚合与更新；全局模型的广播下发；（详见权利要求8）  接收来自于VMR的和；  根据和进行全局聚合和局部聚合时间的估计；（详见权利要求12）  根据和进行VMR与gNB间各链路的质量和连接时间的估计，VMR和UE间各链路的质量和连接时间的估计，进而估计出VMR可服务的时间。（详见权利要求13）  做出切换有关的决策，并发送至相应的VMR和UE。 |
| 1. 根据权利要求3，UE端的设备进行本地数据的收集，接收来自于VMR的局部模型或全局模型进行本地模型更新；基于本地数据进行本地模型的迭代计算与更新。其中，每进行次迭代，便将更新后的结果上传至VMR处进行局部聚合。即本地模型的上传是以次迭代为一个周期的，周期时间不是常数。 |  |
| 1. 根据权利要求4，VMR端的设备接收 多个UE上传的本地模型，在接收到所有UE上传的模型后，进行局部的模型聚合与更新。其中，每进行次的局部聚合，便将更新后的局部模型上传至gNB处进行全局聚合。   当VMR的局部聚合的次数不是的整数倍时，VMR将更新后的局部模型下发至UE。当为的整数倍时，VMR接收来自于gNB的全局模型进行局部模型更新，并将此全局模型转发至UE。即局部模型的上传是以次局部聚合为一个周期的，周期时间不是常数。 |  |
| 1. 根据权利要求5，gNB端的设备 接收来自于各VMR的局部模型，用于全局模型的聚合更新，并将更新后的模型下发至各VMR。 |  |
| 1. 根据权利要求3、4和7，UE端的设备接收来自于VMR的模型用于本地模型的更新。   该模型可以是VMR局部聚合后的模型（不是的整数倍时），也可以是VMR转发的全局模型（为的整数倍时）。 |  |
| 1. 根据权利要求3，UE端的设备进行自身状态相关的信息 -- 的估计，其中，包括但不局限于   UE与VMR之间信道的状态信息（如RSRP等）  UE自身的计算能力信息（如CPU占用率等）  本地数据信息（如参与模型训练的样本数量，样本维度等）  电量，位置信息，移动性(如速度，方向，在本地停留时间)等。  并将进行上报至VMR； |  |
| 1. 根据权利要求4，VMR端的设备接收来自于UE的，并对与自身状态相关的信息 进行估计，其中包括但不局限于   UE与VMR之间信道的状态信息（如RSRP等）  VMR与gNB之间的信道状态信息（如RSRP等）  计算能力信息（如CPU占用率等）等  VMR的位置信息，移动性信息（如速度，方向等）  并将和上报至gNB |  |
| 1. 根据权利要求5，gNB端设备对 全局聚合时间和局部聚合时间，包括：   gNB当前时刻到本次全局聚合完成所用的时间  gNB本次全局聚合完成 到 下次全局聚合完成所用的时间  各VMR当前时刻到本次局部聚合完成所用的时间  各VMR本次局部聚合完成 到 下次局部聚合完成所用的时间 |  |
| 1. 根据权利要求5，gNB端设备对 VMR与gNB间链路质量和连接时间，VMR和UE间链路的质量和连接时间 进行估计，进而估计出VMR可服务的时间，具体操作如下:   1）VMR与gNB间链路质量和连接时间，如满足RSRP大于一定阈值的时间  2）VMR和UE间链路的质量和连接时间，如满足RSRP大于一定阈值的时间  同时满足1）和2）的时间即为VMR可服务的时间 |  |
| 1. 根据权利要求2，UE尽量选择在全局模型聚合且广播下发后进行切换，具体操作如下：   1）如果*,*  即VMR和各UE可全程参与完成本次和下次的全局聚合。则在本次全局聚合结束后不进行任何切换  2）如果 ，  即VMR和各UE可参与完成本次全局聚合，但其服务时间内无法完成下一次全局聚合。则在本次全局聚合后切换。此时，各VMR和UE具有相同的全局模型，无需进行模型传输。  3）如果,  即VMR和各UE无法完成本次全局聚合，但VMR服务时间内可以完成本次和下次局部聚合。应采取一定策略保证VMR和各UE在内的服务不中断，如果无法实现，则在本次局部聚合结束后不进行任何切换（详见权利要求15）  4）如果,  即VMR和各UE可以完成本次局部聚合，但VMR服务时间内可以完成下次局部聚合。则在本次局部聚合结束后进行任何切换，此时需要进行模型传输  5）如果  即VMR和各UE无法完成本次局部聚合。立即进行切换，并进行模型传输 |  |
| 1. 根据权利要求2，13和14，针对权利要求14中的 3）。应采取一定策略保证VMR和各UE在内的服务不中断，如提高发送功率，~~分配更多的带宽~~分配更多的传输资源等。如果无法实现，则在本次局部聚合结束后不进行任何切换。 |  |

（超出10项请追加）

中英文关键词：

VMR --- Vehicle Mounted Relay 车载中继

gNB --- 5G基站

UE --- User Equipment 用户设备

专利说明书中描述[现有技术]的栏目。请尽可能引用专利公报、学术论文、标准提案等来说明现有技术，例如**概括若干文献中的技术方案**并说明其欠缺点。

**现有技术及其问题点**

|  |  |  |  |
| --- | --- | --- | --- |
| **General background** | 1. **Federated Learning** 2. **V2X** | | |
| **具体内容：** | | | |
| **TS/TR SPEC**  **（if possible）** | **SPEC No.** | | **TR 38.331 TS 22.186** |
| **Chapter No.** | | **6.3.2** |
| **具体内容（直接粘贴-与你技术方案相关的）-<如果标准化，会对标准的哪个章节有影响？>：**   1. **TR 22.874** AMMT   7. Distributed/Federated Learning over 5G system   1. **TS 22.261**   6.42 Mobile base station relays | | | |
| **Meeting/agreements** | **Meeting No.** | | 3GPP TSG-SA WG1 Meeting SA#94-e |
| **Meeting date** | |  |
| **SPEC No.** | |  |
| **具体内容（直接粘贴）：**   1. SA Plenary 94#E Work Item: 5G System Support for AI/ML-based Services (SP-211648, approved):   Objective 1 (WT#1.1):`  c. Enhancements of external parameter provisioning to 5GC (e.g. expected UE activity behaviors, expected UE mobility, etc.) based on Application AI/ML operation  *Objective 3 (WT#1.3):*  *Study whether and how 5GS provides assistance to AF and the UE for the AF and UE to manage the FL operation and model distribution/redistribution (i.e. FL members selection, group performance monitoring, adequate network resources allocation and guarantee.) to facilitate collaborative Application AI/ML based Federated Learning operation between the application clients running on the UEs and the Application Servers.).* | | | |
| **Techical Problem** | | | |
|  | | | |
| **Existing solution for above problem** | | **Other companies’ proposals (one or more)?** | |
| **Company/Tdoc No.** |  |
|  | |  | |
|  | | | |
| **Disadvantage or problem of above existing solution** | | | |
| **解决方案具体说明(可把要点粘贴，并作必要的解释说明)：** | | | |
| **Summary of potential standardized points(你认为可能标准化的点列出来，以标准的形式，like proposal)**  联邦学习的 1）UE的运算能力，数据大小；2）VMR的运算能力；3）UE，VMR，gNB之间的信道状态；4）迭代次数k1和聚合次数k2 共同决定了 **全局模型聚合时间** 和 **VMR局部模型聚合时间**  另一方面，VMR的移动性，决定了**VMR的停留时间**  **以上均属外部参数（external parameters）,这些参数作为新的控制信令信息提供给5G网络，用于VMR的group handover的决策（RAN impact）。** | | | |

**发**

**明**

**报**

**告**

**书**

(3)

1. 请写明为了实现此发明、你认为最佳的实施状态是怎样的（包括发明所要使用的设备、系统及重要部件等全部内容）。
2. 请尽可能详细写明此发明重要部件的结构、运行方式、作用等。
3. **请写出由此发明引申出的、其他相近似的方案的至少两个示例。**
4. **对于所列出的公式请说明公式整体含义及各个参数的具体含义，并给出可选的其他公式示例。**

注 参照图纸、图表、流程图等资料时，请编号进行说明、描述；如果有技术报告等材料，请充分利用，加以辅助说明。

发明的具体说明

专利说明书中关于[实施实例]的部分。请按照下列顺序加以详细说明。

Beijing

1. **发明背景**

传统联邦学习网络结构如图1所示

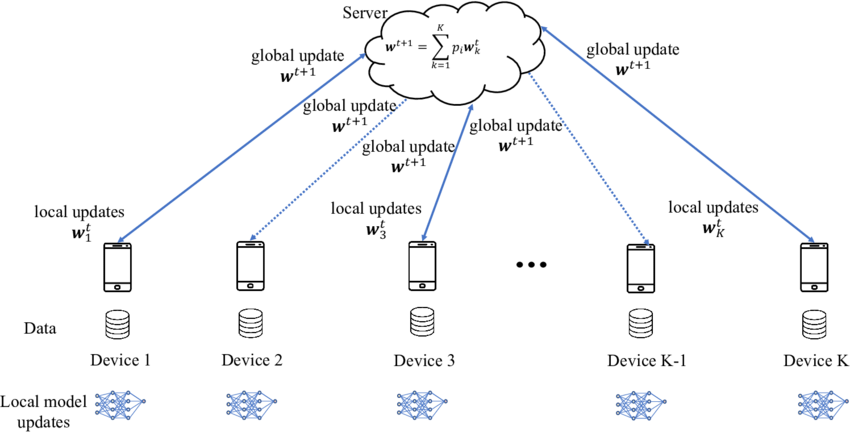


图1. 传统的联邦学习网络结构

各设备通过无线信道接入Server或gNB进行通信，将本地所学习到的模型上传至Server，Server进行聚合后再下发至各设备。具体流程如下：

1. 各设备接入Server，通过Server下行链路传输，获取初始全局学习模型()：
2. 各设备使用存储在本地的数据进行学习，完成一次本地模型更新的迭代:

,

其中表示梯度，表示设备的损失函数，为学习率

1. 各设备通过上行链路，将学习到的本地模型（）或梯度（）上传至Server；
2. Server将收集到的来自各设备的本地模型进行聚合，完成全局模型的更新；

,

其中为来自于各设备的本地模型的权值, 通常设置为 ；其中，表示用户*i*的数据集，表示数据集的大小

1. Server将更新后的全局模型再次下发到各设备，运行 1）--4） 直至模型收敛。

然而，该网络结构有其相应的局限性：

1. 设备通常通过与gNB相连上传至Server，而gNB的覆盖范围有限，对于覆盖范围外的用户或者设备无法提供服务
2. gNB覆盖范围内存在一些通信速率较低的地区，对于该地区的用户，即使在覆盖范围内，服务质量无法保证
3. gNB上下行链路通信资源有限，如频率资源，载波个数等，无法支持过多用户的同时接入。
4. **发明内容**

针对上述问题，我们提出了一种基于车载中继的解决方案，如图2所示。主要内容如下：

1. 提出了一种车载中继（VMR）辅助的分层（双层结构）联邦学习网络结构，
2. 该网络结构下VMR的切换方案以保障服务连续性

图形用户界面

中度可信度描述已自动生成

图2. 车载中继（VMR）辅助的分层联邦学习

注：为了表述方便，第一层的用户我们用UE表示，与gNB直接相连接的UE我们称之为蜂窝UE。

该网络结构由两层组成：第一层由VMR与其服务的用户（UE）组成；第二层由gNB与其连接的VMR以及与gNB直接连接的用户组成。各UE（作为一级节点）首先进行多次本地模型训练更新，再将结果上传至VMR（作为二级节点）进行局部聚合（partial aggregation）；各VMR进行多次局部聚合后，再将结果上传至gNB进行全局聚合（global aggregation）。

该结构优点如下：

* 1. 充分利用了VMR的功能：VMR不只是有单纯作为Relay进行传输的功能，同时也做为第一层结构的聚合者参与计算，完成模型的局部聚合（VMR聚合）。
  2. UE与VMR连接，缩短了通信距离，保障了通信服务质量。
  3. 多个用户与VMR直接通信，VMR再与gNB连接。尽管VMR服务于多个UE，但其与gNB之间的每次通信只上传其局部聚合模型参数。其数据量相当于一个蜂窝UE与gNB通信或者一个UE与VMR通信的数据量。大大降低了gNB的负载。缓解了gNB处通信资源不足的问题。
  4. **车载中继（VMR）辅助的分层联邦学习网络结构**

本小节分别以UE或VMR上传模型和上传梯度为例进行说明,假设有M个VMR，第i个VMR服务的UE的个数为；C个蜂窝UE。

**2.1.1 上传模型参数**

该网络结构下，联邦学习流程如下：

1. gNB首先初始化全局学习模型
2. gNB将全局学习模型下发至蜂窝UE和VMR，VMR，再将该模型下发至UE，此时gNB，所有蜂窝UE，VMR，UE具有相同的学习模型：

其中，

---- VMR#i的模型参数

---- VMR#i服务的第*l*个UE的模型参数

---- VMR#i服务的UE个数

1. 各UE基于本地存储的相关数据，对接收到的全局模型进行本地模型学习更新：

;

并将此时的模型参数上传至其服务VMR，即。 其中，

---- VMR#i服务的第*l*个UE在第次模型更新后的本地模型参数

---- VMR#i服务的第*l*个UE在第次模型更新后的本地模型参数

---- VMR#i服务的第*l*个UE在第次模型更新时的本地梯度

---- VMR#i服务的第*l*个UE在第次模型更新时的本地损失函数

经过次的迭代模型更新后得到本地模型参数，并上传至VMR

1. VMR在接收到其所有服务UE上传的本地模型后，进行一次局部聚合

并将聚合后的模型（此处称之为VMR聚合模型）下发至服务UE。其中，

---- VMR#i第次局部聚合后的局部模型参数

---- VMR#i第次局部聚合后的局部模型参数

---- VMR#i服务的第*l*个用户上传的本地模型对应的权值

, 即VMR每完成一次局部聚合，其服务的UE完成次本地模型更新

1. 重复 3）-- 4）。

当不是的整数倍时，VMR将局部聚合后的模型下发至其服务的各个UE。UE使用接收到的局部聚合模型更新本地模型。

当是的整数倍时，VMR在完成第次局部聚合后，将得到的局部聚合模型上传至gNB。

1. 此外，平行于3）-5），蜂窝UE基于本地数据单独进行本地模型更新。类似于3）

其中，

----第*p*个蜂窝UE在第次模型更新前的本地模型参数

---- 第*p*个蜂窝UE在第次模型更新时的本地梯度

---- 第*p*个蜂窝UE在第次模型更新时的本地损失函数

其进行本地更新次后上传本地模型至gNB。

1. gNB在接收到所有VMR和蜂窝UE上传的局部模型后，进行全局聚合

gNB进行全局模型聚合更新，并将聚合后的模型下发至各VMR，UE和蜂窝UE。其中，

---- gNB在进行第次全局聚合后的全局模型参数

---- 第*m*个VMR上传的局部模型对应的权值

---- 第*c*个VMR上传的本地模型对应的权值

，即gNB每完成一次全局聚合，其连接的VMR完成次局部聚合

=1

1. 并重复2）-- 7）直至模型收敛

详情可参考文献【1】和文献【4】

切换问题：

对于聚合时的权值，即公式（3）和（5）中的权值，许多算法需要与上一次更新的模型作为参考确定权值。如式（3）中，**的确定**

**2.1.2 上传模型梯度**

1. gNB首先初始化全局学习模型
2. gNB将全局学习模型下发至蜂窝UE和VMR，VMR，再将该模型下发至UE，此时gNB，所有蜂窝UE，VMR，UE具有相同的学习模型：

其中，

---- VMR#i的模型参数

---- VMR#i服务的第*l*个UE的模型参数

---- VMR#i服务的UE个数

1. 各UE基于本地存储的相关数据，对接收到的全局模型进行本地模型学习更新：

;

并将此时的模型参数上传至其服务VMR，即。 其中，

---- VMR#i服务的第*l*个UE在第次模型更新后的本地模型参数

---- VMR#i服务的第*l*个UE在第次模型更新后的本地模型参数

---- VMR#i服务的第*l*个UE在第次模型更新时的本地梯度

---- VMR#i服务的第*l*个UE在第次模型更新时的本地损失函数

经过次的迭代更新后得到本地模型参数，并上传**梯度**至VMR：

---- VMR#i服务的第*l*个UE进行次本地模型更新的梯度的和，

注：上传而非

1. VMR在接收到其所有服务UE上传的本地模型后，进行一次局部聚合

并将聚合后的模型（此处称之为VMR聚合模型）下发至服务UE。其中，

---- VMR#i第次局部聚合后的局部模型参数

---- VMR#i第次局部聚合后的局部模型参数

---- VMR#i第次局部聚合时的梯度

---- VMR#i服务的第*l*个用户上传的本地模型（梯度）对应的权值

, 即VMR每完成一次局部聚合，其服务的UE完成次本地模型更新

经过次的局部聚合后得到局部模型参数，并上传**梯度**至VMR：

---- VMR#i进行次局部模型聚合的梯度的和，

注：上传而非

1. 重复 3）-- 4）。

当不是的整数倍时，VMR将局部聚合后的模型下发至其服务的各个UE。UE使用接收到的局部聚合模型更新本地模型。

当是的整数倍时，VMR在完成第次局部聚合后，将得到的局部聚合模型上传至gNB。

1. 此外，平行于3）-5），蜂窝UE基于本地数据单独进行本地模型更新。类似于3）

其中，

----第*c*个蜂窝UE在第次模型更新前的本地模型参数

---- 第*c*个蜂窝UE在第次模型更新时的本地梯度

---- 第*p*个蜂窝UE在第次模型更新时的本地损失函数

其进行本地更新次后得到本地模型参数，并上传**梯度**至gNB:

1. gNB在接收到所有VMR和蜂窝UE上传的局部模型后，进行全局聚合

gNB进行全局模型聚合更新，并将聚合后的模型下发至各VMR，UE和蜂窝UE。其中，

---- gNB在进行第次全局聚合后的全局模型参数

---- 第*i*个VMR上传的局部模型梯度

---- 第*c*个蜂窝UE上传的本地模型梯度

，即gNB每完成一次全局聚合，其连接的VMR完成次局部聚合

=1

为VMR的个数

为蜂窝UE的个数

为第i个VMR的权值

为第j个蜂窝UE的权值

1. 并重复2）-- 7）直至模型收敛
   1. **切换方案**
      1. 现有切换方案及相关问题

我们介绍相关的切换场景，及其带来的问题：

1. 单个或多个UE从VMR#i切换到VMR#j，且切换前VMR#j有服务的UE。如图

此时VMR#j处有局部聚合模型。

现有方案1：

不考虑FL，满足切换条件时直接切换（如接收信号强度 RSRP小于一定门限）。其导致的如下情况：

当 ，即，UE#*l*从VMR#i切换到VMR#j。UE#*l*在完成本轮的次本地模型训练时，即：

a). 对于VMR#i而言，已与UE#*l*断开连接，无法接收到UE#*l*训练好的本地模型。

b).对于VMR#j而言，可以接收UE#*l*上传的本地模型。但是UE#*l*在上传的本地模型，是基于在时，从VMR#i（而不是VMR#j）接收到局部模型训练得到的。VMR#j将会选择丢弃不用。

因此，对于UE#*l*，只有其在第+1次局部聚合过程中完全在VMR#i（或VMR#j）的覆盖范围内，（即第次局部聚合完成模型下发至UE#*l*，到UE#*l*完成次本地模型更新并上传），才可参与VMR#i（或VMR#j）的第+1次局部聚合。

当 时， UE#*l*从VMR#i切换到VMR#j，VMR#i无法接收到UE#*l*在时上传的模型。UE#*l*无法参与本次局部模型聚合，**UE#*l*在VMR#j的模型训练中断。**

**针对此问题，我们可以考虑在局部聚合完成后进行切换。**

注：详情见文献[5]，III-A中的式（13）

现有方案2：

对于UE#*l*，当时，如果UE#*l*从VMR#i切换到VMR#j，UE#*l*可以在完成本轮的次本地模型后上传模型参数或梯度至VMR#j。VMR#j也会将从UE#*l*得到的模型参数用于本轮的局部聚合。但这会导致局部模型的发散（divergence），从而导致系统性能（如模型收敛速度）或者模型准确度的下降。

图示

描述已自动生成

图3 UE从VMR#i切换到VMR#j，且VMR#j有服务UE

1. 单个或多个UE从VMR#i切换到VMR#j，且切换前VMR#j没有服务的UE。

注：该场景可能的发生可由VMR#i的移动导致（如图），也可由UE的移动导致。

**问题1：**如图4，图5所示，此时VMR#j处没有局部聚合模型，只有最初从gNB接收到的全局聚合模型。因此，**VMR#i需要将其局部聚合模型传输给VMR#j。**

**举例说明：**

假设在完成次的局部聚合后，UE从VMR#i切换到VMR#j处。接下来进行次的局部模型聚合，需要次的局部模型参数，如2.1.2**上传模型梯度**时的式（10）。的更新需要，而此时VMR#j处没有。

即使是**上传模型参数**，也需要计算梯度来确定聚合时的权值，详情可见参考文献[6]。

**问题2：**如图5所示，VMR#i服务的用户部分切换到VMR#j。与切换前相比，切换后，VMR#i和VMR#j都缺少了一部分用户参与局部聚合，因此会导致模型发散，准确度降低。**即在局部聚合阶段切换，部分UE的切换也会导致局部模型的发散，降低全局模型的准确性。**

**针上述两个问题，我们可以考虑在完成全局聚合时进行切换。**

图示

描述已自动生成

图4 VMR#i服务的UE全部切换由VMR#j服务，且VMR#j没有服务UE

图形用户界面, 图示

描述已自动生成

图5 VMR#i服务的UE部分切换由VMR#j服务，且VMR#j没有服务UE

* + 1. 新的切换方案及其流程

**针对2.2.1描述的问题，我们提出我们的解决方案：**

**应尽量保证UE在全局聚合且该模型广播下发后进行，此时各UE，VMR的模型相同，均为全局模型，无需进行额外的模型传输（model transfer），切换代价最小；如不能保证则尽量在VMR完成局部聚合后进行切换，此时VMR与其服务的所有UE模型相同，且能保证UE的训练服务不中断。此外，还可以通过提高发送功率，降低RSRP门限等方法，延长UE与VMR或VMR与gNB的连接时间至局部聚合或全局聚合完成。**

我们首先考虑 UE#*l*由 VMR#i 切换到 VMR#j的情况，定义：

和

--- VMR#i对UE#*l*的服务时间，即VMR#i对UE#*l*提供服务的剩余时间的估计值;

--- gNB本轮全局聚合所需的剩余时间，即从当前时刻 到 gNB完成本轮全局聚合的时间；

--- gNB下轮全局聚合所需的时间，即从gNB完成本轮全局聚合起 到 gNB完成下一轮全局聚合的时间；

--- gNB下轮全局聚合所需的剩余时间，即从当前时刻 到 gNB完成下一轮全局聚合的时间，即；

--- VMR本轮局部模型聚合（VMR聚合）所需的剩余时间，即从当前时刻 到 VMR完成本轮VMR聚合的时间；

--- VMR下轮局部模型聚合（VMR聚合）所需的剩余时间, 即从VMR完成本轮VMR聚合起 到 VMR完成下一轮VMR聚合的时间；

---VMR下轮全局聚合所需的剩余时间，即从当前时刻 到 VMR完成下一轮VMR聚合的时间, 即。

其中，对于**，，,** 的估计，由gNB根据各VMR上报的信息(即权利要求中的**，，**和)完成。对于**, ,** 的估计，可由各VMR根据其服务UE上报的信息及其自身信息(即权利要求中的，)完成，也可由gNB完成。

对于**，，, ，, ,** 的估计，gNB和各VMR可以周期性的进行。也可由某些触发事件触发，如VMR或者UE的突然移动等。

~~VMR根据UE上报的信息及其自身信息(即权利要求中的，)，对~~**~~, ,~~** ~~进行估计。~~

~~gNB需要根据VMR上报的信息(即权利要求中的和)，对。~~

1. 如果,

即VMR#i与UE#*l*连接对其进行服务 可参与完成本轮和下轮的全局聚合，则结束前暂不进行任何切换，如图6所示。

形状

中度可信度描述已自动生成

图6**.** 时，gNB的切换决策（结束前暂不切换）

1. 如果 ，

即VMR#i与UE#*l*连接对其进行服务 可参与完成本轮全局聚合，但其服务时间无法支持下一轮全局聚合的完成，则在本轮全局聚合后切换。gNB将聚合后的全局模型广播下发至VMR *i* 和 VMR *j*，以及其他VMR，蜂窝UE，并通过VMR *i*进一步下发至UE group。此时，UE group 与 VMR *i* 和 VMR *j* 具有相同的模型，切换只需考虑通信链路的建立与释放，无需考虑模型的转移，局部模型的发散等。如图7所示。

形状

中度可信度描述已自动生成

图7. 时，gNB的切换决策（在结束后切换）

1. 如果 ，

即VMR#i与UE#*l*连接对其进行服务 虽然不能完成本轮全局模型聚合，但可以完成本轮和下轮的VMR#i局部模型聚合，则结束前暂不进行任何切换。如图8所示

形状

中度可信度描述已自动生成

图8. 时，gNB的切换决策（结束前暂不切换）

1. 如果 ，

即VMR#i与UE#*l*连接对其进行服务 可参与完成本轮的VMR#i局部模型聚合，但其服务时间无法支持下一轮VMR#i局部模型聚合完成，则在本轮VMR#i局部聚合后切换。VMR#i将聚合后的局部聚合模型下发至其服务的UE group。此时，UE group 与 VMR#i 具有相同的模型。若切换到的VMR#j无服务用户，VMR#i需要将局部模型发送至VMR#j。此切换需要考虑模型的转移。如图9所示。

形状

中度可信度描述已自动生成

图9. 时，gNB的切换决策（结束后切换）

1. 如果 ,

即VMR#i与UE#*l*连接对其进行服务 无法支持本轮VMR#i局部聚合的完成。则直接进行切换。若切换到的VMR#j无服务用户，VMR#i需要将局部模型发送至VMR#j。此切换需要考虑模型的转移。如图10所示。

形状

中度可信度描述已自动生成

图10. 时，gNB的切换决策（直接切换）

接下来，我们考虑UE#*l*由 VMR#i 切换到 gNB的情况，即网络中不存在合适的VMR#j提供服务。情况类似：

1. 如果 , 暂不进行任何切换。如图6所示。
2. 如果 ，在本轮全局聚合后切换，UE#*l*变成蜂窝UE。如图7所示。
3. 如果 ，则暂不进行任何切换。如图8所示。
4. 如果 , 在本轮局部聚合后进行切换。随后各自进行本地迭代模型更新，满足条件后直接上传本地模型进行全局模型聚合。如图9所示。
5. 如果 , 直接切换，各自进行本地模型更新，满足条件后上传模型参与全局模型聚合。如图10所示。

接下来，我们考虑蜂窝UE#*l*由 gNB 切换到 VMR#i的情况，即网络中存在合适的VMR提供服务。此时不存在**，**和的情况。定义

--- gNB对UE#*l*的服务时间，即gNB对UE#*l*提供服务的剩余时间的估计值;

1. 如果 , 暂不进行任何切换。如图11所示

形状

中度可信度描述已自动生成

图11**.** 时，gNB的切换决策（结束前暂不切换）

1. 如果 ，在本轮全局聚合后切换，蜂窝UE#*l*变成Slave UE。如图12所示。

形状

中度可信度描述已自动生成

图12. 时，gNB的切换决策（在结束后切换）

1. 如果 ，直接切换。如图13所示。

形状

中度可信度描述已自动生成

图13. 时，gNB的切换决策（直接切换）

注：此外，为了保障服务连续性，提高系统性能。还可以通过提高发送功率，降低RSRP门限，分配更多的传输资源等方法，延长UE与VMR或VMR与gNB或UE与gNB的连接时间 直至 局部聚合或全局聚合完成。

对于上述切换的决策，均由gNB执行。gNB在做完gNB决策后，可以立即将决策下发致相关VMR和UE，也可以在全局聚合结束后，全局模型广播至各VMR和UE时下发。该决策的下发可通过传统的方式，即Uu链路（Downlink）下发至到相关VMR，再由相关VMR经PC5（Sidelink）下发至相关UE。

注：全局模型是通过广播下发的，各VMR均可接受。但是gNB的切换决策并不上广播的形式，而是至通知需要执行切换的VMR和UE。

**与5G核心网相关的应用实例：**

**NWDAF可以分析车辆移动的时间以及切换的时间从而为AF提供信息以便AF计算最佳的联邦学习时间信息，并将时间信息发送到AMF从而影响UE的移动管理以便实现高效的联邦学习。比方说对，，, ，, , 的估计，，，和作为NWDAF的输入，NWDAF可直接输出，，, ，, , 至AF，或者输出相关的信息至AF，由AF进行计算。**

**另外AF也可以将本专利所述的联邦学习管理规则发送至PCF并且由PCF将移动管理策略发送至AMF从而控制UE的切换，比如AF将 在全局聚合或局部聚合未完成前，应尽量保持UE与VMR的连接的规则，发送到PCF，PCF进一步指定出提高发送功率或者降低RSRP门限或者传输速率等策略发送至AMF。**

**此外，AF还可以通过NEF获取UE移动切换信息从而控制gNB以及云端的联邦学习应用。**

NWDAF --- Network Data Analysis Function

AF --- Application Function

AMF ---- Access and Mobility Function

PCF --- Policy Control Function

NEF --- Network Exposure Function

-------------------------------------------------------------------------------------------------------

参考《5G核心网---赋能输在时代》这本书，5G核心网SBA（Service-based Architecture基于服务的）架构及其部分网络功能（NF）简介如下

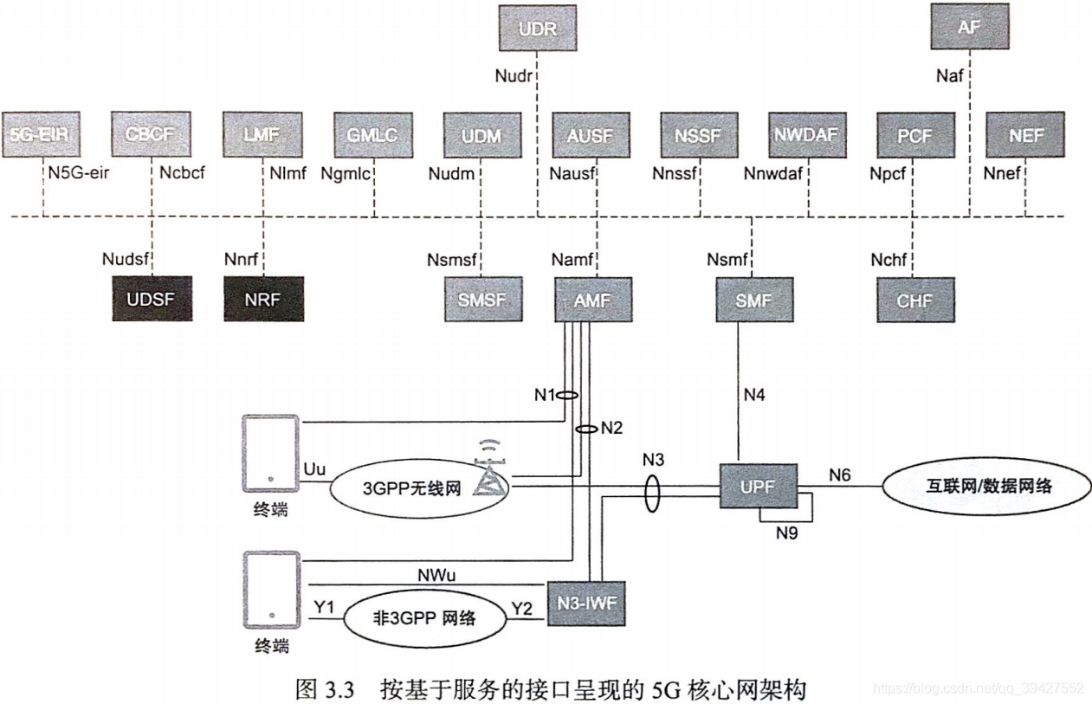


图 14 5G 核心SBA网架构 （该图为基本知识）

**AF: Application Function** ，应用功能。指应用层的各种服务，可以是运营商内部的应用、也可以是第三方的AF（如视频服务器、游戏服务器）。

**NEF：Network Exposure Function** ，网络开放功能。位于5G核心网和外部第三方应用功能体之间，负责管理对外开放网络数据的，所有的外部应用，想要访问5G核心网内部数据，都必须要通过NEF。

**NWDAF：Network Data Analytics Function** 网络数据分析功能。可以收集数据，执行分析并将分析结果提供给其他网络功能，如NEF。

**AMF：Access and Mobility Management Function 接入和移动性管理功能** 负责注册、连接、可达性、移动性及与安全和访问管理和业务授权

**PCF: Policy Control function** 策略控制功能。提供其负责的所有移动性，UE访问选择和PDF会话相关的策略

**参考文献：**

[1] X. Zhang, Y. Liu, J. Liu, A. Argyriou and Y. Han, "D2D-Assisted Federated Learning in Mobile Edge Computing Networks," 2021 IEEE Wireless Communications and Networking Conference (WCNC), 2021, pp. 1-7

[2] 3GPP TR 22.874 – 5G System (5GS); Study on traffic characteristics and performance requirements for AI/ML model transfer, Dec. 2021

[3] 3GPP TR 22.839 – Study on vehicle-mounted relays; Stage 1, Dec. 2021

[4] Liu L, Zhang J, Song S H , et al. Client-Edge-Cloud Hierarchical Federated Learning[J]. 2019.

[5] Feng, Chenyuan, et al. "Mobility-Aware Cluster Federated Learning in Hierarchical Wireless Networks." IEEE Transactions on Wireless Communications (2022).

[6] Wu, Hongda, and Ping Wang. "Fast-convergent federated learning with adaptive weighting." IEEE Transactions on Cognitive Communications and Networking 7.4 (2021): 1078-1088.