



基于业务类型的集中式接入网基站处理资源分配算法

张新苹^{1,2}, 王园园², 田霖², 郝树良¹

(1. 重庆邮电大学, 重庆 400065; 2. 中国科学院计算技术研究所, 北京 100190)

摘要: 现有的资源分配算法研究主要面向数据中心和云计算环境展开, 未能充分考虑基站处理资源的多样性 (CPU、内存、网络带宽、FPGA、DSP) 和业务种类的多样性, 不能直接应用于集中式接入网架构中。针对该问题, 提出了基于业务类型的资源分配算法, 首先采用 Fisher 分割方法, 根据用户业务类型对不同类型计算资源的需求, 对业务进行分类, 然后利用资源分配均衡策略分配基站处理资源。仿真结果表明, 该算法有效地减少了开启物理服务器的个数并提高了物理服务器的资源利用率, 达到了绿色节能的目的。

关键词: 集中式接入网; 基站处理资源分配; 用户业务类型; Fisher 分割方法; 资源分配均衡策略

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018202

Service aware base station processing resource allocation for centralized radio access network

ZHANG Xinping^{1,2}, WANG Yuanyuan², TIAN Lin², HAO Shuliang¹

1. Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Current research on resource allocation algorithms mainly focuses on data centers, and cloud computing environments. These resource allocation algorithms did not consider the diversity of processing resources (CPU, memory, network bandwidth, FPGA, DSP) and the diversity of service types in the centralized base station, resulting in the low processing resource utilization. In order to solve this problem, a service oriented base station processing resource allocation algorithm was proposed. Firstly, the Fisher partition method was used to classify the base station's processing resource requirements according to the user's service request. Then, the resource allocation balancing strategy was used to allocate the base station processing resources. Experimental results show that the algorithm is effective to reduce the number of physical servers and improves the resource utilization of the servers, realizing energy conservation and communications.

Key words: centralized radio access network, base station processing resource allocation, user service type, Fisher partition method, resource allocation balancing strategy

收稿日期: 2017-12-19; 修回日期: 2018-06-08

基金项目: 国家自然科学基金资助项目 (No.61431001)

Foundation Item: The National Natural Science Foundation of China (No.61431001)



1 引言

近年来,蜂窝移动通信系统发展迅猛,网络容量需求呈指数级增长^[1,2]。传统蜂窝移动通信系统采用大规模部署基站的方式提升地区覆盖容量,各基站之间相对独立,基站处理资源按照基站覆盖区域内用户业务峰值容量部署,且资源独占。这种组网架构使得孤立的基站不能处理具有“潮汐效应”的动态网络负载,当负载降低时,基站处理资源的利用率就会明显降低,资源浪费严重。为了解决传统架构的这些问题,业界提出了集中式接入网的概念和架构,例如中国移动的C-RAN (centralized, cooperative, cloud RAN)^[3]和中国科学院计算技术研究所(以下简称中科院计算所)的物理集中、逻辑分布的超级基站架构^[4]。在集中式接入网中,基站处理资源集中部署形成资源池,基于资源水平共享、统计复用,通过资源管控实现处理资源的动态分配^[5-7]。然而集中统一地管理所有处理资源必然给系统管理和资源分配带来新的挑战。如何根据业务实际负载需求,对基站处理资源进行灵活的分配,提高基站处理资源的利用率,是集中式接入网架构的一个重要研究点。

目前,大规模处理资源分配算法的研究主要集中在集中式数据中心、云计算中心等环境下。云计算资源调度通常以虚拟机为资源分配单位^[8-10],将虚拟机调度到一个或多个物理机上,研究的是虚拟机和物理机之间的映射关系。大量的研究将这种问题归结为常见的装箱问题^[11-12],求解目标是用尽可能少的物理资源来满足所有虚拟机的资源需求。参考文献[13-14]在CPU、内存资源约束情况下建立模型,以提高资源利用率为目标,通过遗传算法求解多约束优化问题,获得物理机和虚拟机的优化映射方案。但是采用启发式算法求的只是局部最优解,时间复杂度较高,可扩展性和算法实时性不强。参考文献[15-16]提出了

基于业务量预测的资源分配算法。云服务提供商首先对云用户的业务量进行预测,然后按照所预测的业务量为相应的云用户预先配置虚拟机资源,存在的不足是当突发业务到来时会出现预测业务量不准确的情况,这就会导致预先配置的虚拟机资源过大或过小,从而造成资源利用率较低或不满足业务需求的问题。由于无线接入网所需的资源类型、处理的业务类型等与云计算场景都有较大不同,因此云计算中的资源调度算法虽然具有一定的借鉴意义,但不能直接用于接入网中。在集中式接入网架构下,基站处理资源分配算法主要考虑以下两点:

- 集中式接入网的基站具备类型多样的计算资源,包括CPU、内存、网络带宽、FPGA、DSP等,其中CPU、内存、网络带宽主要完成数据处理和传输;DSP、FPGA作为基带处理的加速器,主要作用是快速地实现各种数字信号处理算法,满足通信系统的实时性要求。
- 通信系统中用户业务的多样性决定了对资源的需求量不同^[17],如对于计算类业务(用户数据处理),它们对计算资源的需求较高,而对于存储类任务(下载音频或视频),它们对存储和带宽资源的需求较高。

针对以上两点,本文提出了基于业务类型的基站处理资源分配算法(station processing resource allocation algorithm, SPRAA),主要思想是在集中式接入网架构下进行基站处理资源分配时,根据用户业务类型对不同类型计算资源的需求,将处理资源需求互补的不同业务组合到一起,按业务组进行处理资源的分配,有助于提高资源利用率。本文以减少物理服务器个数和提高资源利用率为目标对上述问题进行数学建模,通过Fisher业务分类和资源分配均衡策略来完成处理资源的灵活分配,达到提高资源利用率的目的。

2 处理资源分配算法的设计与实现

2.1 算法设计与实现

2.1.1 业务类型分类

在某一个时间窗内到达的业务种类是多样的,不同类型的业务,其数据处理对不同类型处理资源的需求量也会不同^[18-19]。为了降低资源分配的复杂度,首先对到达的业务进行分类,分类标准是业务对基站处理资源(CPU、内存、带宽资源等)的需求大小。为了把业务分成 K 类,同时使得业务分类结果更加精确,避免出现“步骤4根据资源占重比求得业务类型”中业务类型不确定的情况,比如当各类资源占重比为(0.4, 0.1, 0.4, 0.1)时,CPU和带宽的占重比是相同的,确定业务类型时会出现分歧。考虑到业务分类结果的精确性,分类方法采取参考文献[20]的Fisher最优分割法,分类步骤如下。

步骤1 数据标准化

设基站处理资源类型为 P 项,包括CPU、内存、带宽、DSP、FPGA等资源类型。在某个时间窗内到达的业务个数为 N 个,业务集合为 $T=[T_1, T_2, T_3, \dots, T_N]$,对处理资源的需求矩阵为:

$$X_{(P \times N)} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{P1} & x_{P2} & \dots & x_{PN} \end{bmatrix} \quad (1)$$

其中,元素 x_{ij} 表示第 j 个业务对第 i 类计算资源的需求值。为了减小数据的计算量,采用min-max标准化(min-max normalization)对矩阵 X 进行处理得到矩阵 Z ,使得 Z 中每个元素 z_{ij} 大小在 $[0,1]$ 区间内。定义矩阵 Z 中每个 z_{ij} 的计算式如下:

$$z_{ij} = \frac{x_{ij} - \min_{1 \leq j \leq N} \{x_{ij}\}}{\max_{1 \leq j \leq N} \{x_{ij}\} - \min_{1 \leq j \leq N} \{x_{ij}\}} \quad (2)$$

$(i=1, 2, \dots, P; j=1, 2, \dots, N)$

得到矩阵 $Z_{(P \times N)}=[z_{ij}]$ 。

步骤2 计算极差矩阵

在不打乱业务顺序的情况下,把某个周期内到达的 N 个业务进行分割(分类),其所有可能的分割共有 $E = C_{N-1}^1 + C_{N-1}^2 + \dots + C_{N-1}^{N-1} = 2^{N-1}$ 种划分方法。在所有的分割方法中,存在一种分割使得同一类内的所有业务对各类计算资源的需求参数之间的差异最小,而使不同类业务间的资源需求差异最大。在同一类内数据的相似性用极差 d_{ij} 表示,其定义如下:

$$d_{ij} = \begin{cases} \sum_{a=1}^P (\max_{i \leq b \leq j} z_{ab} - \min_{i \leq b \leq j} z_{ab}), & i \leq j \\ 0, & i > j \end{cases} \quad (3)$$

其中, d_{ij} 表示业务 $T=[T_i, T_{i+1}, T_{i+2}, \dots, T_j]$ 对基站处理资源参数的相似性大小, d_{ij} 越小,表示这类业务对所需要的处理资源越相似。由式(3)得到矩阵 D :

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ & d_{22} & \dots & d_{2N} \\ & & \ddots & \\ & & & d_{NN} \end{bmatrix} \quad (4)$$

步骤3 最优 K 分割

用 $b(n, k) (n \geq k)$ 表示 n 个业务分为 k 类的某一种分法,判断某种分法的好坏用分类损失函数 $L[b(n, k)] (n \geq k)$ 来衡量。当 n 和 k 固定时, $L[b(n, k)]$ 越小表示分类越合理。因此要寻找一种分法 $b(n, k)$,使分类损失函数 $L[b(n, k)]$ 达到最小,递推计算式如下:

$$\begin{cases} L[b(n, 2)] = \min_{2 \leq j \leq n} \{d_{1j-1} + d_{jn}\} \\ L[b(n, k)] = \min_{2 \leq j \leq n} \{L[b(j-1, k-1)] + d_{jn}\} \end{cases} \quad (5)$$

当分类数 k 确定时,记分割点为 $(j_2, j_3, j_4, \dots, j_K)$ 。要想找到分割点 $j_k (2 \leq k \leq K)$,即要找到将业务基于处理资源需求参数分为2类、3类... K 类的分割点,依据式(5)求出 $L[b(n, k)] (2 \leq k \leq K)$ 的最小值,最小值对应的 j 即分割点。例如当 $k=2$ 时,求得 $L[b(n, 2)]$ 最小



值, 其对应的 j 即分割点 j_2 ; $k=3$ 时, 求得 $L[b(n,3)]$ 最小值, 其对应的 j 即分割点 j_3 ; 以此类推直到求得分割点 j_k 。

基于以上分析根据业务的处理资源需求参数对业务进行最优二分割: 根据矩阵 D 计算将用户业务分两类的各种分割相应的损失函数 $L[b(j,2)](j=2,3,4,\dots,n)$:

$$\begin{aligned} L[b(2,2)] &= d_{11} + d_{22} \\ L[b(3,2)] &= \min(d_{11} + d_{23}, d_{12} + d_{33}) \\ &\vdots \\ L[b(n,2)] &= \min_{2 \leq j \leq n} (d_{1j-1} + d_{jn}) \end{aligned} \quad (6)$$

找到 $L[b(j,2)](j=2,3,4,\dots,n)$ 的最小值, 则分割点的位置是 j 。由式 (5) 可以求得最优三分割、...一直到 K 分割。在某个周期内, 通过 Fisher 最优分割法, 将到达的业务分成 K 类, 第 $k(1 \leq k \leq K)$ 类中包含的业务个数为 a_k , 业务总数 $N = (a_1 + a_2 + a_3 + \dots + a_K)$ 。本文主要关注集中式接入网基站的 4 类核心资源, 包括 CPU、内存、带宽和 DSP 资源。用 rc_{ij} 、 rm_{ij} 、 rbw_{ij} 、 $rdsp_{ij}$ 分别表示第 $i(1 \leq i \leq K)$ 类中第 j 个业务对 CPU、内存、带宽、DSP 资源的需求, 为了表示方便, 用矩阵 R_i 表示分类后第 i 类中所有业务资源需求参数矩阵的转置:

$$R_1 = \begin{bmatrix} rc_{11} & rm_{11} & rbw_{11} & rdsp_{11} \\ rc_{12} & rm_{12} & rbw_{12} & rdsp_{12} \\ \vdots & \vdots & \vdots & \vdots \\ rc_{1a_1} & rm_{1a_1} & rbw_{1a_1} & rdsp_{1a_1} \end{bmatrix}_{a_1 \times 4} \quad (7)$$

$$R_2 = \begin{bmatrix} rc_{21} & rm_{21} & rbw_{21} & rdsp_{21} \\ rc_{22} & rm_{22} & rbw_{22} & rdsp_{22} \\ \vdots & \vdots & \vdots & \vdots \\ rc_{2a_2} & rm_{2a_2} & rbw_{2a_2} & rdsp_{2a_2} \end{bmatrix}_{a_2 \times 4} \quad (8)$$

$$R_k = \begin{bmatrix} rc_{k1} & rm_{k1} & rbw_{k1} & rdsp_{k1} \\ rc_{k2} & rm_{k2} & rbw_{k2} & rdsp_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ rc_{ka_k} & rm_{ka_k} & rbw_{ka_k} & rdsp_{ka_k} \end{bmatrix}_{a_k \times 4} \quad (9)$$

其中, 矩阵 R_i 表示的是第 i 类业务中 a_i 个业务所需的 4 项基站处理资源参数。 R_i 的第一行参数表示某个业务所需要的各类计算资源参数大小。

步骤 4 确定每类业务的具体业务类型 $type_i$

Fisher 最优 K 分割基于业务处理资源需求参数大小的相似性把业务分成 K 类, 在第 2.1.2 节中需要根据业务的各类资源需求偏重情况来完成资源分配, 所以要确定某类业务的具体业务类型, 包括 CPU 型、内存型、带宽型、DSP 型和平衡型, 具体方法如下。首先将 R_i 第一行业务作为 Fisher 分割算法的分割点, 表明它的资源参数与其他类业务差异很大, 取出各类业务的第一行业务所需资源参数依次放入矩阵 Q , 利用式 (2) 进行归一化, 使得数值大小在 $[0,1]$ 之间。依据式 (10)~式 (13) 求得该业务某类资源占有资源总量的比重如下:

$$p(\text{CPU}) = \frac{rc_{k1}}{rc_{k1} + rm_{k1} + rbw_{k1} + rdsp_{k1}} \quad (10)$$

$$p(\text{内存}) = \frac{rm_{k1}}{rc_{k1} + rm_{k1} + rbw_{k1} + rdsp_{k1}} \quad (11)$$

$$p(\text{带宽}) = \frac{rbw_{k1}}{rc_{k1} + rm_{k1} + rbw_{k1} + rdsp_{k1}} \quad (12)$$

$$p(\text{DSP}) = \frac{rdsp_{k1}}{rc_{k1} + rm_{k1} + rbw_{k1} + rdsp_{k1}} \quad (13)$$

求上述 4 个参数的最大值 $\max(p(\text{CPU}), p(\text{内存}), p(\text{带宽}), p(\text{DSP}))$, 所对应的资源类型即代表该业务类型 $type_i$ 。当计算 $R_i(1 \leq i \leq K)$ 第一行业务出现两类或三类计算资源所占比重相同的情况时, 则继续计算 $R_i(1 \leq i \leq K)$ 第二行业务, 直到出现 $(p(\text{CPU}), p(\text{内存}), p(\text{带宽}), p(\text{DSP}))$ 4 个值不等时求得业务类型 $type_i$ 。业务类型 $type_i$ 用式 (14) 来表示:

$$type_i = \begin{cases} \text{CPU型}, & i=1 \\ \text{内存型}, & i=2 \\ \text{宽带型}, & i=3 \\ \text{DSP型}, & i=4 \end{cases} \quad (14)$$

当 $(p(\text{CPU}), p(\text{内存}), p(\text{宽带}), p(\text{DSP}))$ 的数值大小都在 $(0.25 \pm \Delta)$ 附近, 将业务类型定义为 $\text{type}_i = \text{平衡型}, i=5$, Δ 的值大小根据对资源的重要程度来确定, 一般取值范围为 0.01~0.1。

基于以上分析可以得出, 在某个周期内业务经 Fisher 最优 K 分割被分成 K 类, 根据式 (10) ~ 式 (13) 确定的业务类型最多是 5 类 (CPU 型、内存型、带宽型、DSP 型和平衡型)。考虑到本文设计的业务类型为 5 类, 而最优 K 分割当 $K=1, 2, 3, 4$ 时, 不能满足将业务分为 5 类的要求, 另外, 虽然 K 值越大, 分类的准确度越高, 但其代价是更高的计算量, 当 $K \geq 6$ 时, K 值每增加 1, 在业务总量为 N 时, 需要增加 N 次加法和 $(N-1)$ 次比较, 综合而言, 本文的 K 取值为 5。

2.1.2 资源分配均衡模型及分析

根据用户业务对基站处理资源 (CPU、内存、带宽、DSP 资源) 需求的差异, 在物理服务器上建立相应的处理实体 (processing entity, PE) 处理用户业务。为了实现业务处理的隔离和安全性, 规定一个处理实体对应一个业务处理。传统的思路会根据处理资源 (以 CPU、内存资源为例) 需求来分配资源, 如图 1(a) 所示, 其中大矩形表示一个物理服务器的 CPU、内存资源的总量, 图 1(a) 的服务器支撑了两个对内存需求较高的用户业务, 分别为 PE1 和 PE2。图 1(a) 中深色区域表示物理服务器所剩的可用处理资源, 由图 1 可见, 此物理服务器还剩余大量的 CPU 资源, 但内存资源已经不足, 难以建立一个新的 PE。如果有第 3 个用户业务请求, 在建立对应的处理实体时需要再开启另一个物理服务器, 这样就造成图 1(a) 的物理服务器 CPU 资源的浪费, 出现了资源分配不均衡的问题。

针对该问题, 需要将分属不同类的处理资源需求互补的 PE 配置在同一个物理服务器, 充分利用 CPU、内存等资源, 尽量减少物理服务器开启的数量, 如图 2(a) 的物理服务器所示。本文以提

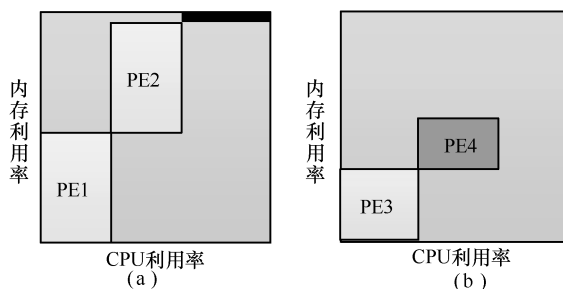


图 1 FFA 下物理服务器的资源利用情况

高资源利用率为目标, 根据 Fisher 分割法将某一周期内的业务根据对不同类计算资源需求不同分类, 考虑到 5 种业务类型, 把分类数 K 设置为 5。定义资源均衡利用率指标体现多类计算资源的综合利用情况, 在资源分配过程中充分考虑了资源总量的约束和分配基站处理资源的顺序, 如图 2 所示, 用资源均衡利用率和各类资源占用率两个指标来分配基站处理资源。相对于经典算法首次适应算法 (first fit algorithm, FFA), FFA 的主要思想是按顺序查找物理服务器, 将 PE 直接部署在满足基站处理资源要求的物理服务器上, FFA 下的物理服务器的资源使用情况如图 1 所示, 本文的算法更有优势。

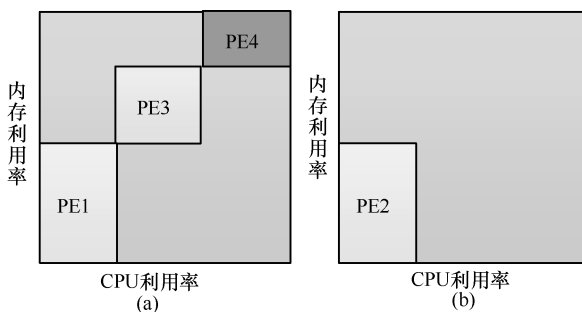


图 2 SPRAA 下物理服务器的资源利用情况

本文综合考虑 CPU、内存、带宽、DSP 资源, 设每个物理服务器的资源量相同, 其中 CPU 总量为 SC , 内存总量为 SM , 带宽总量为 SBW 。DSP 资源是以 DSP 芯片的形式呈现, 用芯片数量 $SD=4$ 来表示 DSP 资源总量。系统中共配置 M 个服务器, 用一个矩阵 S 来描述物理服务器的资源配置情况:

$$S = M \begin{bmatrix} SC & SM & SBW & SD \end{bmatrix} \quad (15)$$



本文通过以下两个指标来衡量业务需求组合的优越性。

(1) 定义单个物理服务器 l 的资源均衡利用率表示为:

$$\mu_l = \alpha u_l^{\text{cpu}} + \beta u_l^{\text{mem}} + \gamma u_l^{\text{bw}} + \chi u_l^{\text{dsp}} \quad (16)$$

其中, α 、 β 、 γ 、 χ 表示资源的重要程度, 且满足 $\alpha + \beta + \gamma + \chi = 1$ 。

假设物理服务器 l 上加载了 L 个处理实体 (PE), 则物理服务器 l 的资源均衡利用率为:

$$\mu_l = \alpha \frac{\sum_{i=1}^L \text{rc}_i}{\text{SC}} + \beta \frac{\sum_{i=1}^L \text{rm}_i}{\text{SM}} + \gamma \frac{\sum_{i=1}^L \text{rbw}_i}{\text{SBW}} + \chi \frac{\sum_{i=1}^L \text{rdsp}_i}{\text{SD}} \quad (17)$$

其中, $\sum_{i=1}^L \text{rc}_i$ 表示 L 个处理实体所需的 CPU 总和, $\sum_{i=1}^L \text{rm}_i$ 表示 L 个处理实体所需的内存资源总和, $\sum_{i=1}^L \text{rbw}_i$ 表示 L 个处理实体所需的带宽资源总和, $\sum_{i=1}^L \text{rdsp}_i$ 表示 L 个处理实体所需的 DSP 资源总和。 μ_l 越大, 资源利用越均衡。当处理实体选择建立的物理服务器时, 通过计算: 若将此处理实体加载到已开启的物理服务器的资源均衡利用率 μ_l 的大小, 将 μ_l 最大值的物理服务器作为拟选定的物理服务器。在 α 、 β 、 γ 、 χ 分别为 0.4、0.2、0.2、0.2 的前提下, 计算 μ_l 可能会出现相等的情况, 所以定义各类计算资源占用情况作为分配资源的另一个指标, 如 (2) 所示。

(2) 物理服务器 l ($1 \leq l \leq M$) 的资源占用情况

用四元组 $u_l = (u_l^{\text{cpu}}, u_l^{\text{mem}}, u_l^{\text{bw}}, u_l^{\text{dsp}})$ 表示, 其中 u_l^{cpu} 、 u_l^{mem} 、 u_l^{bw} 、 u_l^{dsp} 分别表示物理服务器 l 的 CPU、内存、带宽、DSP 资源的利用率。这个四元组在资源均衡利用率相等的情况下, 更加清晰地表示出每类计算资源的占用情况。用这个参数衡量业务需求组合的优越性有一定的指导意义。

基于以上分析, 如给内存型业务对应的处理实体 i 分配处理资源时, 依次计算出将内存型的处理实体 (PE) 加载在已开启服务器的资源均衡利用率, 选择资源均衡利用率高的物理服务器作为拟选定的物理服务器。若出现相同的资源均衡利用率的情况, 则利用四元组 $u_l = (u_l^{\text{cpu}}, u_l^{\text{mem}}, u_l^{\text{bw}}, u_l^{\text{dsp}})$ 选择在内存资源占用率低的物理服务器上作为拟选定的物理服务器。将此处理实体 PE i 加载在拟选定的物理服务器 l 的资源总量约束条件为:

$$\text{rc}_i \leq \text{SC}_l^{\text{res}} \quad (18)$$

$$\text{rm}_i \leq \text{SM}_l^{\text{res}} \quad (19)$$

$$\text{rbw}_i \leq \text{SBW}_l^{\text{res}} \quad (20)$$

$$\text{rdsp}_i \leq \text{SD}_l^{\text{res}} \quad (21)$$

其中, rc_i 、 rm_i 、 rbw_i 、 rdsp_i 分别表示处理实体 i 所需要的 CPU、内存、带宽和 DSP 资源。 SC_l^{res} 、 SM_l^{res} 、 $\text{SBW}_l^{\text{res}}$ 、 SD_l^{res} 表示物理服务器 l 剩余的 CPU、内存、带宽和 DSP 资源。

在某个周期内的资源分配完成时, 假设资源池中开启了 M_{phy} 个物理服务器, 则整个资源池的资源利用率 u_{sum} 为:

$$u_{\text{sum}} = \frac{\sum_{l=1}^{M_{\text{phy}}} \mu_l}{M_{\text{phy}}} \quad (22)$$

在某个周期内, 本文根据用户业务对各类计算资源需求参数的不同, 将业务请求分为 $K=5$ 类, 业务类型由式 (10)~式 (14) 来确定。在资源分配过程中尽量将不同业务需求的处理实体分配在同一个物理服务器, 避免物理服务器出现严重的资源分配不均衡现象。在集中式接入网架构下基于用户业务的资源分配步骤如下。

步骤 1 采用 Fisher 最优分割法对用户业务进行分类, 建立业务处理资源需求参数矩阵 R_k ($1 \leq k \leq K$)。

步骤 2 初始化物理服务器 l ($1 \leq l \leq M$) 的各类资源利用率 $u_l = (0, 0, 0, 0)$ 初始时刻每个物理服



源的状态变化（如处理资源的添加或删除等），把这些信息上报给信息维护模块，维持处理资源的实时信息。处理实体管理中心记录了每个物理服务器上加载的 PE 的各类资源参数大小，把这份数据上报给监控设备，为计算整体资源利用率 u_{sum} 提供数据。

3 仿真分析

本文通过在 MATLAB2012a 仿真验证所提算法的性能。综合考虑了物理服务器中的 CPU、内存、网络带宽、DSP 资源，将基站处理实体分为 $K=5$ 类（CPU 型、内存型、带宽型、DSP 型和平衡型），并按上文给的资源分配算法进行处理资源的分配。为了验证算法的有效性，本文与经典算法 FFA 做比较。

在本文的仿真场景中，假设物理服务器的个数能够满足业务需求，在实验中更关注物理服务器开启的个数。假定业务的到达服从泊松过程 ($\lambda=5$)，即平均每秒到达业务数为 5。参考文献[20]，在分类数 K 确定为 5 的情况下，当数据样本数为 50~120 时，Fisher 算法的分割效果达到最佳。为了获得这些样本数，考虑到平均每秒到达业务数为 5，相应的时间区间应以 10~24 s 为宜，以便达到较好的 Fisher 分割效果；同时考虑到业务处理的及时性，处理的时间区间应尽量小，所以本文以 10 s 为一个周期，对 10 s 内到达的业务进行分类和建立处理业务的 PE，在第 10 s 时刻计算总体资源利用率 u_{sum} 。

在建立业务对应的处理实体时，资源开销主要包括两部分：

(1) 业务本身对各类资源的需求，即 $[rc_i, rm_i, rbw_i, rdsp_i]$ ；

(2) 当多个处理实体加载到同一个物理服务器时，这些处理实体对应的数据处理和控制相关进程会在物理服务器中调度执行，这些调度本身也会产生处理资源开销，比如多个处理实体进程

切换时的进程激活，同一个处理实体内部进程间调度，这些都会造成 CPU 资源的开销。

本文主要考虑到业务本身对各类资源的需求，为了表征 (2) 对 CPU 资源的消耗，把 CPU 资源的重要参数设置得大一点。在仿真中，把 CPU 资源的重要参数设置为 0.4，内存、带宽和 DSP 资源统一设置为 0.2。

仿真参数设置见表 1。

表 1 仿真参数设置

参数说明	参数值
物理服务器核数/个	8
物理服务器内存/GB	8
网络带宽/(Mbit·s ⁻¹)	100
DSP 芯片数量/个	4
业务类型（处理实体类型） K	5
某个周期内到达的业务总数	N
开启的服务器个数	M
资源重要程度参数 α 、 β 、 γ 、 χ	0.4、0.2、0.2、0.2
变化量 Δ	0.04

图 4 对比了所提算法 SPRAA 和 FFA 的整个资源池物理服务器资源利用率 u_{sum} 随时间变化的情况。可以看出，随着时间的增加，用 FFA 分配资源，资源利用率会起伏不定，主要是 FFA 按顺序查找物理服务器，找到符合条件的就将处理实体加载到物理服务器上，而用所提算法 SPRAA 会考虑到物理服务器的各类资源利用率，将处理资源需求互补的处理实体组合到一起，随着时间增加，到达的业务数增加，能够找到更多处理资源需求互补的业务进行组合，所以资源利用率会逐渐增加。当时间到达一定数量后，SPRAA 的资源利用率会明显高于 FFA，主要是因为 FFA 出现了第 2.1.2 节中图 1 的资源分配不均衡的情况，不能充分利用物理服务器的各类资源；而用 SPRAA 分配资源，可以使得物理服务器上加载的 PE 互补，最大化地利用基站处理资源，SPRAA 优势更明显。在实际中，由于不同类的用户业务对资源需

求参数不可能出现完全互补的情况,使得物理服务器的资源利用率不能达到100%,因此当时间到达一定数量后,物理服务器的资源利用率的增加趋于平缓,可达到的利用率在70%左右。

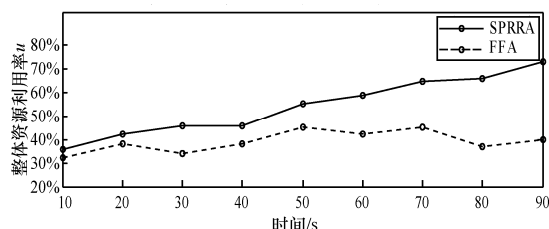


图4 物理服务器资源利用率对比

图5对比了所提算法和FFA的物理服务器开启数目。可以看出,本文提出的算法能减少开启物理服务器的个数,主要因为在资源分配过程中尽量将不同业务需求的处理实体配置在同一个物理服务器上,最大限度地利用CPU、内存、网络带宽和DSP资源。

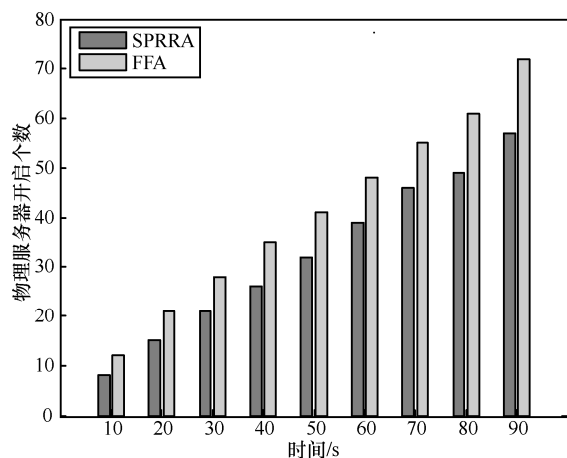


图5 物理服务器开启数目对比

4 结束语

本文在分析了集中式架构下资源分配的研究现状后,考虑到用户业务类型对不同类型计算资源的需求,提出了一种基于业务类型的集中式接入网的基站处理资源分配算法。在资源分配过程中,将不同的处理资源需求互补的业务组合到一起,解决了物理服务器的资源分配不均衡的问题,

仿真表明所提算法减少了开启物理服务器的个数,同时提高了物理服务器的资源利用率。在下一步的研究中,若用户某种业务突然增加,可能会造成PE的迁移,采用的迁移算法仍需要进一步研究。

参考文献:

- [1] LIU L, ZHOU Y Q, TIAN L, et al. Load aware joint CoMP clustering and inter-cell resource scheduling in heterogeneous ultra dense cellular networks[J]. IEEE Transactions on Vehicular Technology, Early Access, 2017, 99(11): 1-14.
- [2] ZHOU Y Q, LIU H, PAN Z, et al. Cooperative multicast with location aware distributed mobile relay selection: performance analysis and optimized design[J]. IEEE Transactions on Vehicular Technology, 2017, 66(3): 8291-8302.
- [3] China Mobile. C-RAN: the road towards green RAN[R]. 2015.
- [4] QIAN M, WANG Y, ZHOU Y, et al. A super base station based centralized network architecture for 5G mobile communication systems [J]. Digital Communications & Networks, 2015, 1(2): 152-159.
- [5] HUANG L, ZHOU Y. Coverage optimization for femtocell clusters using particle swarm optimization[C]//2012 IEEE International Conference on Communications, June 10-15, 2012, Ottawa, ON, Canada. Piscataway: IEEE Press, 2012: 612-615.
- [6] 田霖, 翟国伟, 黄亮, 等. 基于集中式接入网架构的异构无线网络资源管理技术研究[J]. 电信科学, 2013, 29(6): 25-27.
- [7] ZHAI G W, TIAN L, ZHOU Y Q, et al. Research on key technologies of heterogeneous wireless network resource management based on centralized radio access network architecture[J]. Telecommunications Science, 2013, 29(6): 25-27.
- [8] ZHAI G W, TIAN L, ZHOU Y Q, et al. Load diversity based optimal processing resource allocation for super base stations in centralized radio access networks[J]. Science China Information Sciences, 2014, 57(4): 1-12.
- [9] MELL P, GRANCE T. The NIST definition of cloud computing(draft)[R]. 2013.
- [10] 殷波, 张云勇, 房秉毅, 等. 面向成本优化的云计算资源分配方法研究[J]. 电信科学, 2014, 30(11): 22-26, 32.
- [11] YIN B, ZHANG Y Y, FANG B Y, et al. Cloud resource allocation method based on elastic resource adjustment[J]. Telecommunications Science, 2014, 30(11): 22-26, 32.
- [12] 吴清烈, 郭昱, 武忠. 云计算服务与大规模定制模式应用[J]. 电信科学, 2010, 26(9): 74-78.
- [13] WU Q L, GUO Y, WU Z. Study on cloud computing services and mass customization applications[J]. Telecommunications Science, 2010, 26(9): 74-78.
- [14] 徐骁勇, 柯涛, 刘梦娟, 等. 面向云平台的资源分配策略研



- 究[J]. 计算机应用, 2013, 33(20): 299-307.
- XU X Y, KE T, LIU M J, et al. Research on resource allocation strategies in cloud computing[J]. Journal of Computer Applications, 2013, 33(20): 299-307.
- [12] 陈小娇, 陈世平, 方芳. 云计算中虚拟机资源分配算法[J]. 计算机应用研究, 2014, 31(9): 2585-2616.
- CHEN X J, CHEN S P, FANG F. Research on virtual machine resource allocation algorithm in cloud computing[J]. Journal of Computer Applications Research, 2014, 31(9): 2585-2616.
- [13] LU J. Improved genetic algorithm-based resource scheduling strategy in cloud computing[C]//2016 International Conference on Smart City and Systems Engineering, Nov 25-26, 2016, Changsha, China. Piscataway: IEEE Press, 2016: 230-234.
- [14] WANG S, GU H, WU G. A new approach to multi-objective virtual machine placement in virtualized data center[C]//IEEE Eighth International Conference on Networking, Architecture and Storage, July 17-19, 2013, Xi'an, China. Piscataway: IEEE Press, 2013: 331-335.
- [15] CHEN N S, FANG X P. A cloud computing resource scheduling scheme based on estimation of distribution algorithm[C]//The 2014 2nd International Conference on Systems and Information, Nov 15-17, 2014, Shanghai, China. Piscataway: IEEE Press, 2014: 304-308.
- [16] GHRIBI C, HADJI M, ZEGHLACHE D. Energy efficient VM scheduling for cloud data centers: exact allocation and migration algorithms[C]//2013 IEEE/ACM 13th International Symposium on Cluster, Cloud, and Grid Computing, May 13-16, 2013, Delft, the Netherlands. Piscataway: IEEE Press, 2013: 671-678.
- [17] IMT-2020(5G)推进组. 5G 概念白皮书[R]. 2015.
- IMT-2020(5G) Propulsion Group. 5G white paper [R]. 2015.
- [18] 翟国伟. 超级基站架构下协议处理资源分配和管理方法的研究[D]. 北京: 中国科学院大学, 2016.
- ZHAI G W. Research on protocol processing resource allocation and management in super BS architecture [D]. Beijing: University of Chinese Academy of Sciences, 2016.
- [19] 王庆扬, 谢沛荣, 熊尚坤, 等. 5G 关键技术与标准综述[J]. 电信科学, 2017, 33(11): 112-122.
- WANG Q Y, XIE P R, XIONG S K, et al. Key technology and standardization progress for 5G[J]. Telecommunications

Science, 2017, 33(11): 112-122.

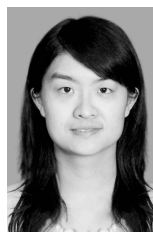
- [20] 武琳琳. 基于 Fisher 最优分割法的聚类分析应用[D]. 郑州: 郑州大学, 2013.

WU L L. Cluster analysis based on the method of Fisher optimal division[D]. Zhengzhou: Zhengzhou University, 2013.

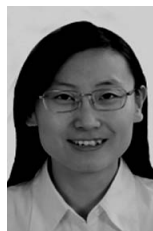
[作者简介]



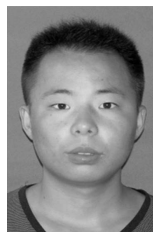
张新苹 (1992-), 女, 重庆邮电大学硕士生, 主要研究方向为绿色无线通信系统、5G。



王园园 (1986-), 女, 博士, 中国科学院计算技术研究所助理研究员, 主要研究方向为绿色无线通信系统、5G、无线接入网络虚拟化等。



田霖 (1980-), 女, 博士, 中国科学院计算技术研究所副研究员, 主要研究方向为绿色无线通信系统与无线资源管理技术。



郝树良 (1991-), 男, 重庆邮电大学硕士生, 主要研究方向为 5G 新型多址接入技术。