

基于粒子群优化算法的5G网络切片功能迁移机制

陈强^{1,2}, 刘彩霞^{1,2}, 李凌书^{1,2}

(1. 国家数字交换系统工程技术研究中心, 河南 郑州 450002;
2. 移动互联网安全技术国家工程实验室, 北京 100876)

摘要: 在5G的多应用场景中, 数据流量经常出现剧增的情况, 网络切片中虚拟机资源可能无法满足用户正常需求。鉴于此, 提出了一种以负载均衡为目标的网络切片功能迁移机制。该机制基于粒子群优化算法, 将虚拟机模拟成粒子, 每次迁移过程中, 将所有的粒子分成若干个子群, 在群内和群间同时应用粒子群优化算法, 参照历史最优解和当前全局最优解更新粒子位置, 通过选取标记因子较小的粒子实时比较合适度等参数确定最佳目标粒子, 完成迁移过程, 该机制既提高了收敛速度, 又提高了算法精度。通过与其他迁移方法比较, 结果表明, 所提迁移机制具有精度高、收敛快的优点, 并能提升资源的使用效率, 降低了数据中心的能耗, 具有较好的自适应性。

关键词: 功能迁移; 粒子群算法; 网络切片; 5G

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2018066

5G network slicing function migration mechanism based on particle swarm optimization algorithm

CHEN Qiang^{1,2}, LIU Caixia^{1,2}, LI Lingshu^{1,2}

1. National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China
2. National Engineering Laboratory for Mobile Network Security, Beijing 100876, China

Abstract: In multi-application scenarios of 5G, data traffic often increases dramatically. Virtual machine resources in network slicing may not meet the normal needs of users. In view of this, a network slicing function migration mechanism aiming at load balancing was proposed. The mechanism simulates the virtual machine into particles based on particle swarm optimization algorithm. In the process of migration, all particles were divided into several subgroups, and particle swarm optimization algorithm was applied within and among groups. According to the historical optimal solution and the current global optimal solution, the particle location was updated, and the best target particles were determined by selecting the smaller particle size of the particle in real time. The mechanism not only improves the convergence speed, but also improves the accuracy of the algorithm. Compared with other migration methods, the results show that the proposed migration mechanism has the advantages of high accuracy and fast convergence. And it can also improve the efficiency of resource utilization, reduce the energy consumption of data center, and has better adaptability.

Key words: function migration, particle swarm optimization algorithm, network slicing, 5G

收稿日期: 2018-06-06; 修回日期: 2018-07-06

通信作者: 陈强, 1428917764@qq.com

基金项目: 国家高技术研究发展计划基金资助项目(“863”计划)(No.2014AA01A701), 国家自然科学基金资助项目(No.61521003), 科技部支撑计划基金资助项目(No.2014BAH30B01)

Foundation Items: The National High Technology Research and Development Program of China (863 Program) (No.2014AA01A701), The National Natural Science Foundation of China (No.61521003), Ministry of Science and Technology Support Plan (No.2014BAH30B01)

1 引言

随着移动通信技术的快速发展,传统单一的网络模式已经无法满足用户的需求,差异化服务的需求日益提升。移动用户业务的多样性和对时延、速率的高要求,给现有的网络带来了巨大挑战。5G 移动通信技术因为低时延、高可靠和大规模机器通信等特色业务的存在,提出了网络切片的概念,按需求将服务提供给对应的用户^[1]。网络切片作为 5G 的关键技术之一,可以按用户需求灵活地提供一种或多种网络服务,每个切片都可以独立地按照业务场景的需求和话务模型进行网络功能的定制裁剪和相应网络资源的编排管理,提高网络服务的灵活性和资源利用率,增强整体网络的健壮性和可靠性^[2]。但随着用户及各种类型应用数量的增加,会出现各个虚拟机互连接、共同处理同一项任务的情况,服务器虚拟机的承载压力会不断增大,可能出现无法满足网络切片相应功能需求的情况,用户的服务出现中断,直接影响用户的体验质量,也大大降低资源利用率,这与 5G 移动通信技术开始时的设计理念背道而驰。现行的虚拟机迁移技术主要是将多个虚拟机共同迁移到同一服务器上,通过共享服务器的计算、网络等物理资源提高服务器的利用率,但是虚拟机的处理性能必然会受到该服务器上其他虚拟机的影响,当虚拟机无法满足当前的服务需求时,为了保证服务的正常进行,将当前的服务任务迁移到其他虚拟机上是十分必要的,这样可以大大提高系统整体性能。通常不同类型的服务其资源需求具有很大的差别,并且由于处理任务的多样性,应用的负载也呈现出高度的动态性,对物理资源的需求也具有高度的弹性^[3]。而且在实际的场景中,用户所需要的服务请求存在周期性变化的特点,经常随着用户特定的行为习惯或特定的事件出现服务请求的高峰期,这对单一资源评判标准和被动的响应触发机制带来了极大的挑战,使系统无法很好地保证服务质量,大大影响用户的服务体验。所以在 5G 这种数据流量组成十分复杂的情况下,如何更好地满足用户的服务需求,如何在用户数据流量突增的情况下依旧保证正常的服务进行,是目前亟待解决的

一个关键问题^[4]。

频繁的虚拟机迁移会产生大量的迁移流量,从而造成网络运维成本增加,当同一服务器上的虚拟机资源需求无法得到满足时,部分虚拟机就需要被迁移到其他服务器上。一般情况下,要保持服务质量最大,一般的方法是预留一定的资源以防流量突然递增出现资源不足,但这种方法会直接导致系统的资源利用率下降,所以在不采用这种预留资源方法的情况下,满足用户服务质量的最好办法是进行功能迁移以保证相应功能的正常实现。文献[5]提出了一种基于服务质量的虚拟机迁移机制,当虚拟机的服务质量下降并超过某一门限值时会触发虚拟机迁移,将虚拟机迁移到负载低的服务器,以保证其服务质量,该方法也是保持虚拟机性能和提高服务可用性的有效方法,但方法设计比较粗糙。文献[6]提出的机制是用于监控主机资源的使用情况,一旦发现某台主机资源的使用超过了之前设定的阈值或某些用户服务协议(SLA)受到威胁,系统则进行虚拟机迁移,将主机中的某台虚拟机迁移到另一台主机中。这种方法和基于服务质量的虚拟机迁移机制类似。文献[7]提出了一种基于最大资源需求优先的虚拟机迁移算法,即资源需求大的虚拟机首先被迁移。相反,文献[8]则提出了一种最小资源需求的虚拟机优先迁移的方法,认为将资源需求大的虚拟机迁移到其他服务器上,需要有比较多的剩余资源,这可能导致需要开启新的服务器来满足迁移要求,从而导致功耗增多。

本文主要借助负载均衡的思想,根据当前的负载情况将虚拟机的资源进行重新分配,使资源得到合理利用,以实现数据中心整体范围内的负载均衡,获得最佳的资源利用率^[9]。结合粒子群优化算法,尽可能同时满足服务质量和资源利用率最佳,引入标记列表和合适度的概念,使源虚拟机快速搜索到最佳目标虚拟机,缩短收敛时间,减少系统资源开销。

2 虚拟机的功能迁移模型

2.1 功能迁移主要设计的 4 种策略

服务器中虚拟机的虚拟资源特征主要包括

CPU、内存、网络带宽等。而在实际环境中,需要考虑的资源不只是列举的这几种,所以迁移问题变得十分复杂,因此虚拟机中的功能迁移变成一个 NP-Hard 组合优化的问题^[10]。虚拟机功能迁移主要考虑以下 3 个问题:一是哪些虚拟机需要功能迁移;二是哪些虚拟机适合作为目标虚拟机承载源虚拟机的服务功能;三是迁移时机的选择。所以,设计一个改进的迁移模型至关重要,这里首先说明整个模型中的 4 种关键策略。

1) 监控负载策略:主要是对整个主机内负载指标的实时监测,根据监测到的负载信息,计算系统的负载情况。如果有虚拟机的负载情况超过设定阈值,立刻将信息传递给触发策略进行触发。

2) 动态触发策略:该策略主要决定何时虚拟机能够进行迁移,如果满足触发要求则立即进行触发迁移。现在的虚拟机迁移策略一般都设定了阈值,一旦超过阈值则虚拟机立即触发迁移。

3) 计算选择策略:选择合适的目标虚拟机进行迁移,主要是计算负载不均衡度等衡量指标,减少虚拟机迁移带来的系统开销,同时也提高了迁移速度。

4) 迁移策略:主要将源虚拟机的 IP 地址与目标虚拟机的 MAC 地址绑定,同时源虚拟机也向目标虚拟机发送一个 ARP 重定向包进行绑定,完成后将封装好的数据分组直接发送到目标虚拟机,完成整个迁移过程^[11]。迁移完成以后,源服务器中的虚拟机被删除,虚拟机所有的网络通信直接通过目标虚拟机实现,目标虚拟机保持源虚拟机迁移之前的所有状态,接收新来的信息或者任务。

2.2 集中式虚拟机功能迁移模型

集中式虚拟机功能迁移模型主要分为 2 个部分:位于中央管理服务器上的中央控制部分和位于各个主机上的本地迁移部分。其中,中央控制部分主要是从整体上收集各个服务器中虚拟机的资源使用情况,根据相应的策略,实现对本地迁移部分的控制,以达到数据中心整体范围内的负载均衡。本地迁移部分主要负责将源虚拟机中的资源、算法等一起经过中央管理服务器控制迁移

到目标服务器上。但由于中央控制部分负责整个数据中心范围内服务器上虚拟机的迁移,一旦中央控制部分失效,则整个数据中心会造成负载失衡,而且集中式功能迁移模型的中央控制部分需要实时控制管理所有主机的资源使用情况,极易造成资源瓶颈^[12]。这种模型的优点是结构简单,便于管理,资源开销比较小。单一地考虑某一种资源时,虚拟机功能迁移模型可以使用这种模型。图 1 为集中式虚拟机功能迁移模型。

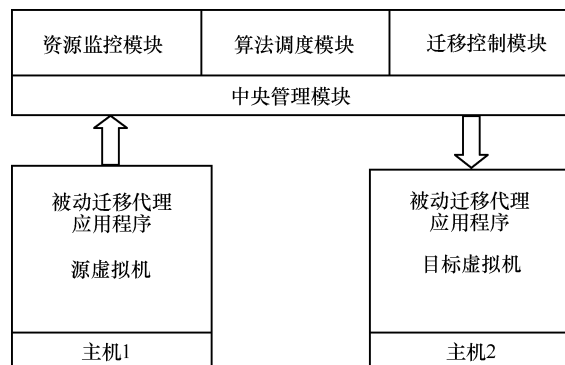


图1 集中式虚拟机功能迁移模型

2.3 分布式虚拟机功能迁移模型

分布式迁移模型特点是每个主机中的虚拟机都有迁移代理可以自行监控其资源的使用情况,一旦超过设定的阈值,系统会发起虚拟机迁移。整个模型省略了中央控制部分,将中央控制部分的功能都移到每个主机中,每个虚拟机可以自行发起迁移,具备迁移的一切条件,包括迁移监控器、迁移控制器、算法池等模块。分布式模型可以增加系统的可靠性和健壮性,克服集中式模型中单点故障造成的负载失衡的缺陷,但是大大增加了资源开销^[13],不便于统一管理,而且一般都使用比较简单的算法,存在比较大的改进空间。图 2 为分布式虚拟机功能迁移模型。

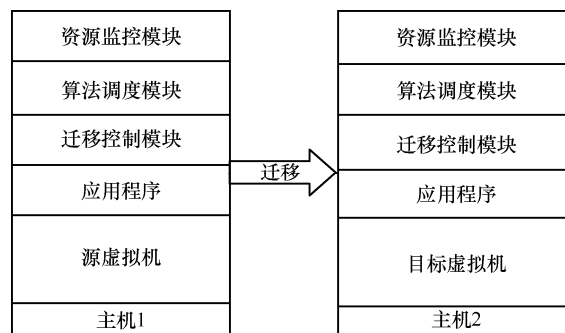


图2 分布式虚拟机功能迁移模型

3 粒子群优化算法

相比目前较为流行的蚁群算法、贪婪算法等,粒子群优化算法由于具有收敛快、顽健性强、计算简单等优点,在求解较为复杂的问题时显示出较好的优越性。由于贪婪算法没有考虑到物理主机和虚拟机的差异性;蚁群算法收敛速度慢、易陷入局部最优,所以两者都不适合作为虚拟机功能迁移机制的算法。

3.1 粒子群算法原理

粒子群算法原理^[14]大致可以描述为:由 N 个没有体积和质量的粒子个体组成的粒子群在维度为 S 的空间中进行搜索,每一个粒子都有速度和位置 2 个参数。算法通过速度更新粒子的位置,而每个粒子的位置代表搜索空间的一个解。用公式表达这些参数:第 i 号粒子当前的位置为 $x_i^{t+1} = x_i^t + v_i^{t+1}$, 粒子当前的速度为 $v_i = (v_{i_1}, v_{i_2}, v_{i_3}, \dots, v_{i_s})$, 粒子的历史最优位置为 $g_i = (g_{i_1}, g_{i_2}, g_{i_3}, \dots, g_{i_s})$ 以及粒子的全局最优位置为 $h_i = (h_{i_1}, h_{i_2}, h_{i_3}, \dots, h_{i_s})$ 。在搜索过程中,粒子每一次的更新过程都必须参照历史最优位置和全局最优位置,如此反复更新直到最后找到最优解。本文中所有的解,都是在空间 S 中进行的,这里不再单独赘述。整个过程用式(1)和式(2)表示。

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (1)$$

$$v_i^{t+1} = v_i^t + \omega_1 \text{rand}1_i^t (g_i^t - x_i^t) + \omega_2 \text{rand}2_i^t (h_i^t - x_i^t) \quad (2)$$

其中, i 为粒子的序号, t 为迭代的次数, ω_1 、 ω_2 为学习因子与惯性系数的比值, ω 主要决定搜索时间和寻优能力, ω 越大则搜索时间越短,但寻优能力越差。 rand_i^t 是 0~1 之间均匀分布的随机数,主要增加粒子个体和群体搜索的随机性。这里可以采用多种群模式,即将所有粒子分为大小相等的 a 个子群,每个子群中有 b 个粒子,先对每个子群中的 b 个粒子进行算法更新,最后在 a 个子群间应用算法进行更新,这样使多个子群中的全部粒子都能向整体历史最优的位置移动,实现在同一时间内快速收敛的目的,从而提高算法的收敛速度。

3.2 算法粒子编码

粒子编码是一个将虚拟机简化成粒子并相应编码的过程,以便把问题中的解和标记的粒子位

置进行对应映射。在粒子群优化算法中,合适的编码方式至关重要,不仅可以快速搜索到结果,同时还可以减少搜索解的数量。在应用粒子群优化算法处理虚拟机功能迁移问题时,常见的编码方式有二进制编码和多值编码^[15]。在处理 m 台源虚拟机和 n 台目标虚拟机的迁移问题上,二进制编码的搜索空间为 $2^{m \times n}$,而多值编码的搜索空间为 m^n ,比较得出多值编码的搜索空间更小,并且由于二进制编码需要考虑源虚拟机是否在一个时间间隔内被多次迁移,所以本文选用多值编码对粒子进行编码。对粒子进行多值编码时,第 i 号迁移粒子的当前位置为 $x_i = \{x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_n}\}$, 第 j 号目标粒子的当前位置为 $x_j = \{x_{j_1}, x_{j_2}, x_{j_3}, \dots, x_{j_m}\}$ 。功能迁移向量为 $S_{q \times b} = x_i \times e_j$, 其中, $e_j \in x_j$ 时 e_j 为 1, $e_j \notin x_j$ 时 e_j 为 0。

3.3 算法的具体流程

后面切片的功能迁移机制也是根据该流程演变而来,设置粒子群规模是为了防止粒子跳出区间造成运行时间上的浪费,根据粒子不断移动及速度和位置的更新,从而找到全局最优解。

粒子群算法基本流程如下。

第 1 步:初始化,首先设置一些基本参数:设置粒子总数为 n , 群数为 m , 随机初始化粒子的速度和位置。

第 2 步:在每个子群中,每个粒子都找到历史上最优的位置信息,然后从这些个体历史最优解中找到一个全局最优解,与历史的最优解比较,选出两者中最优的作为历史最优解。

第 3 步:根据式(1)和式(2)进行速度和位置的更新,从而找到新的当前最优解和历史最优解。

第 4 步:子群中出现最优解时停止搜索,然后在群间重复进行第 2 步,直到在整个粒子群中找到最优解,整个过程结束。

4 基于粒子群优化算法的虚拟机功能迁移机制

4.1 功能迁移机制的原理

在介绍功能迁移机制的原理时,由于虚拟资源种类繁多,为了方便讨论,本文只是单纯用 $VR(i)$ (virtual resource) 代替,其代表某一种虚拟资源。事实上,主机的功能迁移不单单迁移一种

资源，是一个多种类型的资源迁移过程。

$$k_i = \frac{VR(i)_{\text{cost}}}{VR(i)_{\text{total}}} \quad (3)$$

式(3)中， k_i 为资源 $VR(i)$ 的利用率， $VR(i)_{\text{cost}}$ 为某种资源服务过程的消耗量， $VR(i)_{\text{total}}$ 为虚拟机中该种资源的资源总量。

$$p_i = \frac{k_i}{\sum_{i=1}^n k_i} \quad (4)$$

$$\sum_{i=1}^n p_i = 1 \quad (5)$$

式(4)和式(5)中， n 为虚拟资源的种类总数，虚拟机的负载均衡值为

$$L = \sum_{i=1}^n k_i p_i \quad (6)$$

式(6)中， p_i 为某种资源利用率占有所有资源利用率总和的比值， L 为主机的负载均衡值。文中根据 L 的大小将整个迁移过程分为4个状态：过载迁移区、负载警告区、负载平衡区和低负载区。整个迁移过程的目的是使负载均衡值最大，这样就达到负载均衡最佳的效果。其中，定义变量 Q 为

$$Q = \frac{VR_{\text{future}}}{VR_{\text{remain}}} \quad (7)$$

Q 是与 L 成正比的变量，可以代表 L 进行性质说明，4个状态的衡量标准为：若 $Q \geq 1$ ，则说明主机现在处于过载迁移区，需要马上进行迁移以保证正常的服务； $Q \geq 0.75$ 则说明主机现在处于负载警告区，需要采取一定的措施，防止继续过载进入过载迁移区；若 $Q \geq 0.5$ 则说明主机现在处于负载平衡区，此时说明主机的资源利用良好，暂时不需要采取措施进行控制，也可以考虑作为目标虚拟机，为其他过载的虚拟机提供资源迁移； $Q \geq 0.25$ 则说明主机现在处于低负载区，应当作为目标虚拟机有效利用其空闲资源，以保证资源得到有效利用。其中， VR_{future} 为虚拟机下一时间预测需要的虚拟资源， VR_{remain} 为虚拟机此时剩余的虚拟资源。

4.2 迁移机制的合适度和标记列表

源虚拟机和目标虚拟机的合适度是源虚拟机需求性能和目标虚拟机空闲性能之间的合适程度，是源虚拟机在过载迁移区即将准备迁移时考虑选择合适目标虚拟机进行迁移的最重要指标。目标虚拟机与源虚拟机的合适度越高，说明迁移的效果越好。

$$F_i = \frac{VR(i)_m^{\text{remain}} Q_{\min} - VR(i)_q^{\text{future}} Q_{\max}}{VR(i)_m^{\text{remain}} Q_{\min}} \quad (8)$$

式(8)中，用变量 F_i 表示某一种虚拟资源的源虚拟机和目标虚拟机的合适度， $VR(i)_m^{\text{remain}} Q_{\min}$ 代表资源 $VR(i)$ 下处于低负载区的 Q 最小的目标虚拟机的剩余虚拟资源， $VR(i)_q^{\text{future}} Q_{\max}$ 代表资源 $VR(i)$ 下处于过载负载区的 Q 最大的源虚拟机下一时间预测需要的虚拟资源。 F_i 越大说明两者的合适度越大，目标虚拟机越适合源虚拟机未满足的后续服务，从而可以推出整个主机所有虚拟资源的合适度 F 为

$$F_i = \prod_{i=1}^n \left(\frac{VR(i)_m^{\text{remain}} Q_{\min} - VR(i)_q^{\text{future}} Q_{\max}}{VR(i)_m^{\text{remain}} Q_{\min}} \right) \quad (9)$$

考虑整个主机的合适度时只需要将每一项虚拟资源的合适度相乘即可得到整个主机的合适度 F_i 。式(9)中， $VR(i)_m^{\text{remain}} Q_{\min}$ 代表资源 $VR(i)$ 下 Q 最小的目标虚拟机的剩余虚拟资源， $VR(i)_q^{\text{future}} Q_{\max}$ 代表资源 $VR(i)$ 下 Q 最大的源虚拟机下一时间预测需要的虚拟资源。

源虚拟机和已选择的目标虚拟机都要列入迁移标记列表中，改进的粒子群算法引入了迁移标记列表，将已经迁移的虚拟机和目标虚拟机都进行标记，引入标记因子 $\tau_{mq}(t)$ ，初始时每个虚拟机的标记因子都相同，迁移虚拟机和目标虚拟机每完成一次迁移过程，标记因子就会按照式(10)进行叠加，则 t 时刻每个虚拟机的标记因子为

$$\tau_{mq}(t+1) = (1-\rho)\tau_{mq} + \Delta\tau_{mq}(t) \quad (10)$$

其中， ρ 为一个常数，其大小与负载均衡值 L 和 Q 有关， $\Delta\tau_{mq}(t)$ 为 t 时刻标记因子的增加量。在新的迁移过程开始时，主机自动搜索迁移标记列表，

按照标记因子最小的虚拟机开始搜索, 因为考虑到目标虚拟机在为源虚拟机提供一次资源服务之后, 其所剩资源一般不适合再为其他源虚拟机提供资源服务, 随着用户需求被满足, 目标虚拟机的资源被释放, 标记因子再一次趋于平衡, 迁移过程便又一次回到初始设定。这样避免了重复搜索, 提高了源虚拟机和目标虚拟机的匹配速度, 同时减少了资源消耗, 使改进的粒子群算法能更高效地搜寻。

功能迁移机制主要包含 6 个模块: 参数调度模块、算法模块、决策模块、监测迁移模块、资源管理触发模块、迁移模块。下面具体说明以下几种模块的作用。图 3 为功能迁移模块的组成。

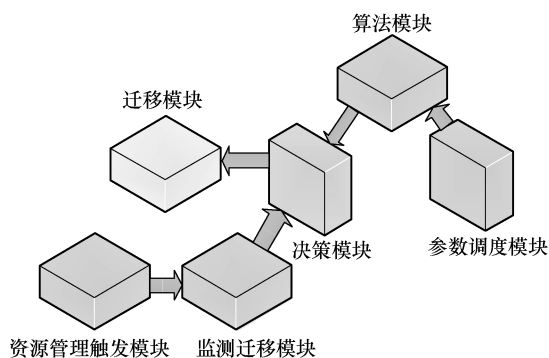


图3 功能迁移模块的组成

1) 参数调度模块: 所有参数都存储在参数调度模块中, 包括各个虚拟机的内存、CPU、硬盘存储、带宽等资源, 直接与数据中心相连, 负责给虚拟机的迁移提供必要的的数据。

2) 算法模块: 在虚拟机迁移过程中, 提供算法必要的公式等, 对虚拟机的迁移提供理论支撑。

3) 决策模块: 利用粒子群优化算法计算迁移虚拟机和目标虚拟机的合适度, 为源虚拟机和目标虚拟机的最佳配对进行决策, 然后在迁移模块进行虚拟机的迁移。

4) 监测迁移模块: 最主要的作用是周期性地获取负载信息后, 通过计算标记因子生成标记列表, 不断更新标记列表, 将虚拟机的标记因子按照从小到大的顺序排列, 等待进入决策模块进行筛选计算源虚拟机和目标虚拟机的合适度。

5) 资源管理触发模块: 收集资源使用的数据

和预测数据进行综合计算, 并将所得数据发送给监测迁移模块进行监测。

6) 迁移模块: 执行最后的迁移操作, 在经过决策模块筛选之后, 迁移模块将源虚拟机未能满足服务需求的整个框架全部迁移到目标虚拟机上, 使整个服务不至于中断。

4.3 迁移机制的匹配过程

首先在标记列表中进行标记因子大小排序, 将处于过载负载区和低负载区的虚拟机按照标记因子的大小进行排序, 从标记因子较小的虚拟机中找出 $Q_{\max}$ 和 $Q_{\min}$ 的虚拟机优先进行合适度计算, 其他虚拟机按照类似方法依次进行计算, 合适度最大的虚拟机即为目标虚拟机, 源虚拟机可以进行迁移。

迁移机制匹配过程如下。

第1步: 按照标记列表中标记因子大小排序进行合适度的计算, 利用粒子群优化算法, 参考历史最优合适度和当前全局最优合适度, 确定合适度最大的虚拟机, 即为目标虚拟机。

第2步: 选择合适的目标虚拟机之后将源虚拟机的虚拟应用一并迁移到目标虚拟机上, 后续迁移虚拟机未能满足的功能全部由目标虚拟机提供。

第3步: 迁移过程完成后将标记因子、目标虚拟机的剩余资源等参数进行更新, 以便下次迁移时比较。

第4步: 若仍有迁移虚拟机需要匹配目标虚拟机, 则重复第一步直至匹配成功。

5 实验结果及分析

5.1 实验环境及算法分析

本文采用开源仿真平台 CloudSim 对结论进行验证, 实验的预设变量信息如下。在 CloudSim 中建立一个由 100 台主机组成的数据中心, 这些主机再随机生成 100 台虚拟机, 其中, 主机的主要指标如表 1 所示: CPU 为 8 核, 内存 8 GB, 带宽 100 Mbit/s; 虚拟机主要分 3 种, 各项指标如下: CPU 为 4/2/1 核, 内存 4/2/1 GB, 带宽 50/30/20 Mbit/s, 虚拟机资源总和不超主机资源配置, 粒子的初始量为 100, 粒子群数为 20, 迭代次数为 50, 为了兼顾搜索速度和搜索能力, ω 设为 2, ρ 设为 0.6, τ_{mq} 设置为 0。

表1 虚拟机配置列

虚拟机	CPU/核	内存/GB	带宽/(Mbit·s ⁻¹)
A1	4	2	20
A2	2	4	30
A3	1	2	50

算法分析如下。

第1步：设置各项参数。

第2步：100台虚拟机随机生成100个粒子，将粒子分成20个子群，根据式(1)和式(2)各粒子进行第一次移动，计算每个粒子移动后在当前时刻的变量 Q ，根据4.1节比较当前 Q 处于什么状态，处于过载迁移区的粒子准备进行迁移，处于其他状态的粒子等待下一次移动。

第3步：根据式(8)~式(10)计算标记列表中的标记因子 τ_{mq} 和各个粒子的合适度 F_i 。

第4步：处于过载迁移区的粒子在每个子群中应用标记列表匹配粒子合适度 $F_{imax}^1 F_{imax}^2 \dots F_{imax}^{20}$ 的粒子，再在每个子群之间应用标记列表匹配粒子合适度 F_{imax} 的粒子作为目标粒子进行迁移。

第5步：根据式(10)在每个子群中更新各个粒子的参数及标记列表，下一时间继续进行粒子的移动，并且重复第2步。

第6步：最后在群间重复第2步，直到粒子找到最佳目标粒子为止。

5.2 仿真结果及分析

本文选择用目前在5G网络切片中流行的2种算法——贪婪算法(GA)和蚁群算法(ACO)与改进的粒子群算法(PSO)进行比较，贪婪算法只从单一的指标考虑问题，缺少全局优化，得到的解并非全局最优解，并且贪婪算法只适用于当前最优解与全局最优解无关的问题，但贪婪算法的资源开销比较小，收敛速度快，也受到一些学者的青睐；蚁群算法是目前和粒子群算法同行的比较智能且应用广泛的算法之一，蚁群算法在开销方面可能会随着蚂蚁的增多而增大，且收敛时间也会增加，但可以得到比较精确的最优解；所以，根据目前的理论知识可以在功能迁移方面与改进的粒子群算法进行比较。

未来5G移动通信最重要的技术指标是时延，低时延也成为5G未来三大场景之一，所以

收敛时间是需要考虑的问题之一，但收敛时间只是说明某一算法收敛到本身最优解的时间，并非收敛到全局最优解的时间，当然这个解也可能是全局最优解，通过图4可以看出，随着虚拟机数量的增加，采用贪婪算法机制的切片完成迁移过程的时间最短；改进的粒子群算法和蚁群算法在刚开始虚拟机数量比较少的情況下，收敛时间都很短，和贪婪算法不相上下，但随着虚拟机数量的增加，蚁群算法的收敛时间明显大于改进的粒子群算法。所以应用到现实中，在收敛时间这方面，改进的粒子群算法性能优于蚁群算法。图4为3种机制下切片中虚拟迁移时间。

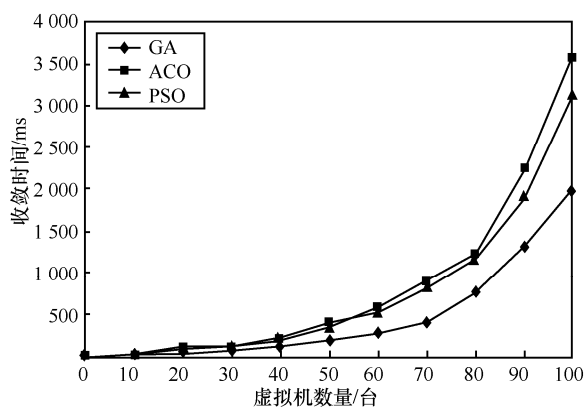


图4 3种机制下切片中虚拟机迁移时间

5G移动通信将进入一个流量爆炸增长的时代，所以虚拟资源短缺将是面临的主要挑战之一，在切片的功能迁移阶段考虑系统开销也是必要的，也正是因为资源不足才导致虚拟机的功能迁移。本文中列举了CPU、带宽、内存等虚拟资源，在仿真时主要通过主机的剩余CPU资源比较3种算法在虚拟机功能迁移上的优劣，其余资源同理。通过图5的仿真结果可知，应用贪婪算法的迁移机制所消耗的CPU资源总体上最少，其他2种机制所消耗的主机CPU资源都比较大，开始时应用粒子群算法的主机资源消耗要大于应用蚁群算法的主机资源消耗，但是随着虚拟机数量的增加，蚁群算法的消耗量大于改进的粒子群算法的消耗量。拓展到实际中，虚拟机数量会十分巨大，所以，在资源消耗量这方面，改进的粒子群算法性能明显优于蚁群算法。图5为3种机制下切片主机剩余CPU。

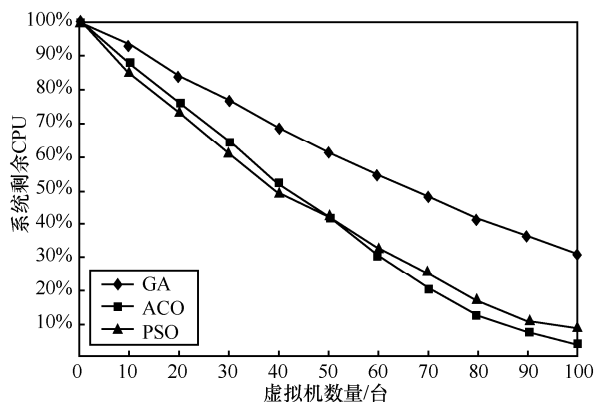


图5 3种机制下切片主机剩余CPU

迁移机制最主要的考察指标就是应用某种机制后是否可以达到最优解或者所得到的解占最优解的百分比,这直接影响源虚拟机是否可以迁移到最优的目标虚拟机上,也影响到后续的用户需求能否得到满足。5G白皮书中明确指出,首要任务是可以最大化满足用户的通信需求,所以网络切片的功能迁移也是在不影响用户需求的前提下进行的,如果用户当前的通信活动因为虚拟机的迁移而中断或者后续的需求得不到满足都将使虚拟机的功能迁移变得毫无意义。通过图6的仿真结果可知,应用贪婪算法的机制随着虚拟机数量的增加,所得解的最优程度越来越差,几乎不能满足5G的基本要求;应用改进的粒子群算法求得的解的最佳度明显高于蚁群算法,可以满足当前5G对用户需求定义,所以在解所能达到最优解百分比这方面,改进的粒子群算法性能明显优于蚁群算法。图6为3种机制下所达到全局最优解的百分比。

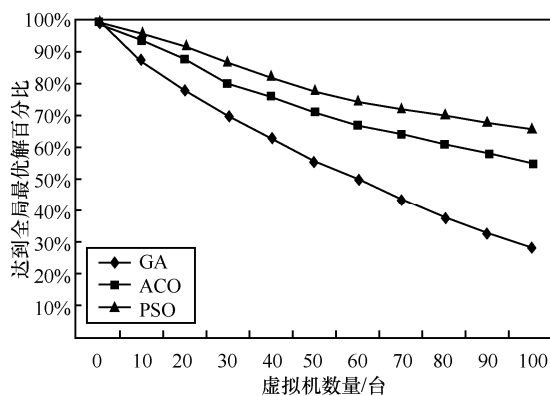


图6 3种机制下所达到全局最优解的百分比

综合观察各个仿真结果来看,本文中的改进粒子群算法在整体上的性能明显优于其他2种算

法,可以满足网络切片虚拟机功能迁移的各项要求。

6 结束语

本文的主要贡献在于研究了流量激增及需求不断变化下的虚拟机功能迁移机制,旨在时变环境下达到较好的服务质量和资源利用率,降低系统功耗。由于服务器内的资源使用具有竞争性和动态性等特点,这导致在虚拟机的迁移问题上存在提高服务质量或资源利用率的权衡^[16]。通过分析现有虚拟机迁移算法存在的不足,本文提出了一种改进粒子群算法,并且结合5G网络切片的大背景,完全可以应用到切片的功能迁移中;在充分考虑到5G未来面临的诸多挑战和现有粒子群算法的一些不足,提出了新的标记因子和标记列表的概念,同时比较源虚拟机和目标虚拟机的合适度,使收敛时间和解的最优程度都得到了提高,并且也大大降低了系统开销。理论分析和实验结果表明,本文中提出的改进粒子群算法具有很好的实用性,对于未来5G网络切片的虚拟机功能迁移具有很好的理论支撑。

参考文献:

- [1] WANG Q, CHEN J, LIAO G Q. Live migration algorithm of virtual machine for reduce energy consumption[J]. Journal of East China University of Science and Technology, 2018(1): 58-61.
- [2] REN C, MA R T. Network slicing: building customizable 5G network[J]. Journal of East China University of Science and Technology, 2018(1): 26-30.
- [3] SU X, GONG J J, ZENG J. Wireless resources allocation for 5G network slicing[J]. Electronic Engineering & Product World, 2017, 24(4): 30-32.
- [4] WEI H, ZHANG Z, FAN B. Network slice access selection scheme in 5G[C]//IEEE Information Technology, Networking, Electronic and Automation Control Conference. 2017: 352-356.
- [5] CHEN W, ZHANG C. Research on virtual machine dynamic migration strategy based on load trend prediction[J]. Journal of Suihua University, 2016(12): 145-149.
- [6] ZHAO C, YAN L, CUI Y H, et al. Dynamic adjusting threshold algorithm for virtual machine migration[J]. Journal of Computer Applications. 2017.
- [7] LIU J J, CHEN G L, HU C X. Virtual machine migration scheduling strategy based on load characteristic[J]. Computer Engineering, 2011, 37(17): 276-278.
- [8] PENG J Y. Research on the energy-efficient based datacenter resource consolidation mechanism[D]. Chengdu: University of Electronic Science and Technology of China, 2013.

- [9] ZHAO S, JI X S, CHENG G Z, et al. Research on dynamic migration of critical virtual machines for multi-tenancy[J]. Chinese Journal of Network and Information Security, 2017, 3(8): 8-17.
- [10] SHI Y, ZHANG H G, WU F S. A method of constructing the model of trusted virtual machine migration[J]. Journal of Computer Research and Development, 2017, 54(10): 2284-2295.
- [11] TIWARI P K, JOSHI S. A review on load balancing of virtual machine resources in cloud computing[C]//International Conference on Information and Communication Technology for Intelligent Systems. 2016: 369-378.
- [12] LIU L, YU H F, ZHENG S P. Comprehensive stable VM migration techniques for managing time-varied workload[J]. Application Research of Computers, 2016, 33(2): 534-539.
- [13] WANG H, CHEN S, AI M, et al. Mobility driven network slicing: an enabler of on demand mobility management for 5G[J]. The Journal of China Universities of Posts and Telecommunications, 2017, 24(4): 16-26.
- [14] YANG J, ZHANG H J, ZHAO S N, et al. Virtual machine deployment strategy based on particle swarm optimization algorithm[J]. Journal of Computer Applications, 2016, 36(1): 117-121.
- [15] AMENDOLA D, CORDESCI N, BACCARELLI E. Bandwidth management VMs live migration in wireless fog computing for 5G networks[C]//IEEE International Conference on Cloud Networking. 2016.
- [16] ROST P, BERBERANA I, MAEDER A, et al. Benefits and challenges of virtualization in 5G radio access networks[J]. IEEE Communications Magazine, 2015, 53(12): 75-82.

[作者简介]



陈强（1993-），男，辽宁本溪人，国家数字交换系统工程技术研究中心硕士生，主要研究方向为新一代移动通信。

刘彩霞（1974-），女，山东烟台人，国家数字交换系统工程技术研究中心副教授，主要研究方向为移动通信网络、新型网络体系结构。

李凌书（1992-），男，湖北恩施人，国家数字交换系统工程技术研究中心博士生，主要研究方向为新型网络体系结构、网络空间安全。