

doi:10.14132/j.cnki.1673-5439.2019.05.011

基于粒子群优化的数据中心负载均衡机制

宋文文,王 珺,杜 晔,徐京明,李成星

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

摘要:针对数据中心网络中传统多径路由方法易造成大象流(Elephant flow)冲突,进而导致网络链路拥塞等问题,文中提出了一种基于粒子群优化的动态负载均衡机制(Dynamic load balancing mechanism based on particle swarm optimization,DLB-PSO)。该机制结合SDN具有全局网络拓扑信息可视化等技术优势,基于粒子群优化算法,根据当前网络链路资源状态等信息,以最小化最大链路利用率为优化目标,为大象流计算出最佳传输路径。实验结果表明,相比于传统的等价多径路由(Equal-cost multi-path routing,ECMP)、全局优先匹配(Global first fit,GFF)及面向SDN的LABERIO机制,文中提出的机制能够更加有效地改善网络吞吐量,降低流的时延抖动。

关键词:数据中心;软件定义网络;负载均衡;粒子群优化

中图分类号:TP393.09 **文献标志码:**A **文章编号:**1673-5439(2019)05-0081-08

Load balancing technology for data center based on particle swarm optimization

SONG Wenwen, WANG Jun, DU Ye, XU Jingming, LI Chengxing

(College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: With the rapid development of big data and cloud computing, existing IP multipath protocols cause substantial bandwidth losses due to long-term collisions in data center networks. To avoid the elephant flow transmission collisions caused by traditional multipath routing methods in the data center networks, a dynamic load balancing mechanism based on particle swarm optimization (DLB-PSO) algorithm is proposed. Aiming to minimize maximum link utilization, the integer linear programming of traffic scheduling in data center network is used and the global optimal solution by the particle swarm optimization (PSO) algorithm. The simulation results show that, compared with equal-cost multi-path routing (ECMP), global first fit (GFF) and software defined network (SDN) oriented LABERIO mechanisms, the proposed mechanism can improve the network throughput and decrease delay jittering of flow in the staggered communication patterns.

Keywords: data center; software defined network (SDN); load balancing; particle swarm optimization (PSO)

近年来,随着云计算和大数据的快速发展,海量的数据备份和迁移等业务需求造成数据中心内部数据流量急剧增加,而传统的数据中心网络由于拓扑

结构简单,容易造成核心链路拥塞。为保证高效的数据传输,提出了很多胖树拓扑(Fat-tree)^[1]、BCube^[2]等具有“富连接”特性的多根树网络拓扑结

收稿日期:2019-04-23;修回日期:2019-06-07 本刊网址: <http://nyzr.njupt.edu.cn>

基金项目:国家自然科学基金(61571233)资助项目

作者简介:宋文文,男,硕士研究生;王珺(通讯作者),女,博士,副教授, wang_jun@njupt.edu.cn

引用本文:宋文文,王珺,杜晔,等.基于粒子群优化的数据中心负载均衡机制[J].南京邮电大学学报(自然科学版),2019,39(5):81-88.

构。为利用 Fat-tree 等拓扑结构提供的多条冗余链路,ECMP^[3]作为一种传统的静态哈希路由机制,根据不同数据流的哈希值,可以将数据流均匀地映射到多条等价路径上。但文献[4]研究表明,当前数据中心网络内部流量呈现重尾分布,即由数据备份和文件传输等业务产生的大象流(Elephant)数量占总流量数目的9%,但这些大象流所传输的字节数占据了总字节数的90%;相反,由web搜索和网页浏览等应用产生的小鼠流(Mice)虽然数目较多,但传输的总字节数在数据中心网络内部占比非常少。而传统的ECMP作为一种静态路由属性,在流量调度过程中并未考虑当前网络状态,可能会将拥有相同IP地址的多条大象流分配到相同的路径上,由于大象流占用链路带宽较多,容易造成部分链路拥塞,其他链路空闲,最终导致数据中心网络链路负载不均衡。

相比于传统网络,SDN^[5]作为一种新型网络结构,利用软件可编程的思想,将控制层与数据层分离;此外,SDN拥有网络全局信息可视化等技术优势,控制器以openflow协议为南向接口,可根据数据中心网络实时链路资源信息,灵活制定路由策略,下发到交换机等转发设备,完成流量调度,高效地实现网络链路负载均衡。

本文首先介绍了传统数据中心网络结构在流量调度过程中存在的链路负载不均衡问题和软件定义网络为数据中心可提供的技术优势,以及面向SDN的数据中心网络链路负载均衡相关研究工作,然后提出了一种SDN环境下基于粒子群优化的负载均衡机制,并最终通过搭建仿真平台验证了机制的性能。

1 面向SDN的数据中心网络相关研究工作

目前,针对传统ECMP路由策略未考虑流的带宽需求这一缺陷,文献[6]提出了名为Hedera的数据中心网络流量调度系统,该系统提出了全局优先匹配(Global first fit,GFF)和模拟退火(Simulated annealing,SA)两种流量调度器,且将占据链路带宽容量的10%的数据流定义为大象流。GFF调度器先通过线性搜寻所有可能的路径来找到一条满足大象流带宽需求的路径,然后在该可用路径的对应边缘和聚合交换机中创建转发流条目。然而,随着网络负载的不断增加,多条链路逐渐饱和,此后的路径选择将变得困难,因此,GFF不能保证所有的数据流完

成流量调度;此外,由于GFF将遍历出的第一条满足带宽需求的路径作为调度流,这样易造成链路剩余带宽碎片化。针对上述问题,文献[7]提出了一种最大概率路径调度算法(Maximum probability path scheduling algorithm,MPP-SA),该算法在提高带宽利用率和减轻剩余带宽碎片化之间作出了平衡,但该算法仅仅是按照数据流的到达先后顺序将其调度到其他满足带宽需求的路径上,从而忽略了当网络链路拥塞时部分数据流需要被调度的迫切性。针对该问题,文献[8]提出了一种LABERIO的动态负载均衡机制,该算法将当前全网链路已用带宽方差的瞬时值作为负载均衡度参数,当链路已用带宽方差大于设定的阈值时,SDN控制器将网络中拥塞程度最高的链路上占据带宽最大的数据流调整到能够满足该数据流带宽需求且开销最小的路径上,从而满足了某些数据流需要被调度的紧迫性,但该机制没有区分大小流,因此没有考虑到因某个时间内小鼠流带宽可能出现极值,造成已用带宽瞬时方差偏大而持续性触发LABERIO算法,最终引起额外网络开销的情况。文献[9]提出了一种名为Freeway的自适应动态路径划分算法,针对大象流和小鼠流的需求特征,该算法自适应地将数据中心可用路径划分为高吞吐量路径和低时延路径,将大象流调度到高带宽路径上,采用ECMP方法将小鼠流调度到低时延路径上。虽然该算法不仅满足了小鼠流的低时延需求,也满足了大象流的高吞吐量需求,在一定程度上缓解了数据流冲突问题,避免了链路拥塞,但是该算法需要使用到路径隔离技术如VLAN标签标记,这显然增加了工程实现的难度。

针对数据中心网络流量特征及现有的数据中心负载均衡算法的不足,本文提出了一种基于粒子群优化的动态数据中心负载均衡机制(Dynamic load balancing mechanism based on particle swarm optimization,DLB-PSO)。首先将最小化最大链路利用率作为最优化目标,建立了相应的数学模型;因为该问题是一个NP完全问题,在对此类问题求解的众多算法中,粒子群算法^[10]仅根据自己的速度向量和位置向量来决定搜索,且具有较高的精度和较快的速度,因此本文采用动态权重方法对粒子群算法进行优化,以此来对该问题进行求解。此外,为减弱负载方差受短时间内链路已用带宽极值的影响,本文采用指数平滑法来预测链路已用方差,并用预测值作为负载均衡机制的触发条件,进而实现数据中心动态负载均衡机制。

2 DLB-PSO 机制实现

2.1 DLB-PSO 系统架构

本文提出的面向 SDN 的 DLB-PSO 机制通过在 SDN 控制器 Ryu 上部署相关功能模块来完成数据中心负载均衡。整体实现架构如图 1 所示,该架构由 Ryu 控制器各模块及数据中心 Fat-tree 网络拓扑关联组成。

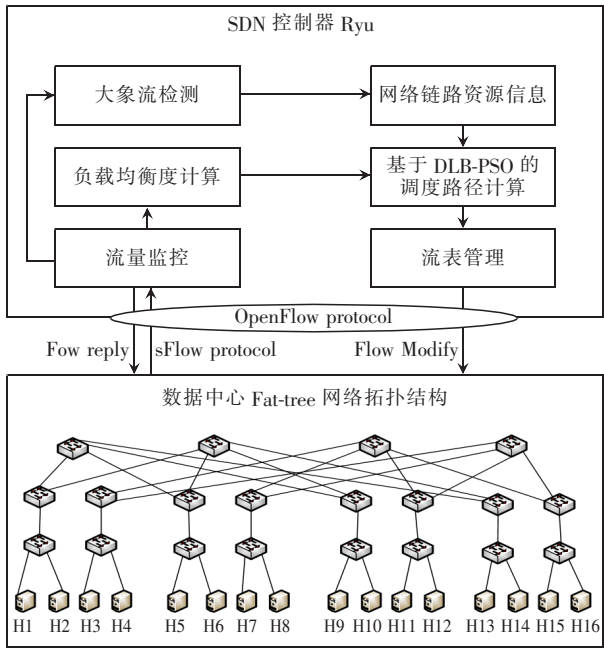


图 1 基于 DLB-PSO 的总体架构图

该架构中的主要模块包括流量监控模块、负载均衡度计算模块、大象流识别模块、基于粒子群优化算法的路径计算模块和流表管理模块,各模块功能描述如下:

(1) 流量监控模块:该模块主要有两个目的,第一,监控全网流量,为实现大象流的调度功能,首先需要对全局流量进行监控。作为一种可扩展的流量监控技术,sFlow^[11]可以持续性收集、存储和分析数据流,从而可实现全局网络负载计算。sFlow 监控系统主要由嵌入在交换机中的 sFlow 代理和控制器中的 sFlow 收集器组成,sFlow 代理仅将数据包打包成 sFlow 数据报并快速发送到 sFlow 收集器中,最大限度地减少对交换机 CPU 和内存的占用,从而达到在减少网络监控成本的同时最大程度避免因流量监控而降低交换机转发性能的目的;第二,判断监控到的数据流是否为新流,在 SDN 网络中,当一个交换机在某个端口接收到一个数据包时,如果交换机流表项中并没有该数据包的流表匹配项时,交换机采用

pack-in message 机制将捕获的新流转发到 Ryu 控制器,控制器根据当前网络信息向交换机中下发相应流表项,该部分由 OpenFlow 协议完成。

(2) 负载均衡度计算模块:sFlow 收集器接收 sFlow 代理发送过来的数据包并进行链路利用率的计算,根据各链路利用率计算全网的已用带宽方差,以此来评估当前全网负载均衡程度。

(3) 大象流识别模块:根据 sFlow 系统所得到的全局流量监控信息,参考 Hedera 中大象流标记方法,该模块将流的带宽高于链路容量 10% 的数据流标记为大象流,以便完成后续对大象流的调度。

(4) 路径计算模块:该模块基于全网链路资源和大象流需求信息,利用粒子群优化算法为大象流计算全局最优转发路径。

(5) 流表管理模块:该模块主要通过控制器向交换机发送 pack-out 消息实现,将路径计算模块得到的最优路径规则下发到数据中心交换机流表项中,从而完成大象流的重路由,实现数据中心网络链路负载均衡。

2.2 DLB-PSO 机制流程

本文提出了一种基于粒子群优化算法的动态负载均衡机制(DLB-PSO)。本文以链路负载方差来度量当前网络负载平衡度,当网络链路负载方差大于预设定的阈值时,控制器调度策略模块根据当前网络状态和调度流的带宽需求等信息,利用粒子群优化算法,为需要调度的大象流计算出最佳路径。该负载均衡算法流程如图 2 所示。

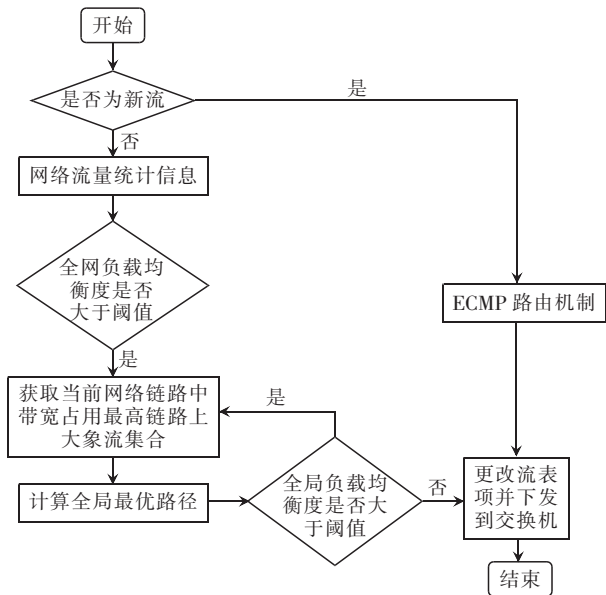


图 2 负载均衡算法流程图

图 2 演示了本文提出的负载均衡机制流程。在

网络初始阶段,当网络中的新数据流进入交换机后,SDN 控制器默认采用 ECMP 的路由方法,基于流的 10 元组值(包括输入端口、MAC 源地址、MAC 目的地址、以太网类型、VLANID、IP 源地址、IP 目的地址、IP 端口、TCP 源端口、TCP 目的端口在内的 10 个关键字)完成数据流的转发。DLB-PSO 负载均衡算法模块周期性监控网络负载状态,当全局网络链路负载平衡度大于阈值时,控制器根据网络链路负载情况,将拥塞度最高的链路上的大象流调度到满足其带宽需求的路径上;反之,即使当前链路上检测到大象流,也不需要对其进行调度,这样可以避免多次触发流量调度机制,以降低控制器负担和减少资源消耗。

2.3 网络负载均衡度

本文提出的动态负载均衡机制中,大象流是否被调度取决于由当前网络统计信息计算出来的全网链路负载均衡度。基于 SDN 具有全局网络资源感知的技术优势,为衡量该负载均衡度,本文采用文献[8]中方法,以全局链路已用带宽方差来衡量当前网络各链路已用带宽的离散程度,并将其作为负载均衡机制的触发条件。 t 时刻的网络链路负载方差为:

δ(t) = 1/M ∑ [load_a(t) - load(t)]^2

其中,δ(t)表示所有链路负载的方差,与网络稳定度成反比。其中,load_a(t)表示 t 时刻链路 a 上已用带宽,load(t)表示全网所有链路已用带宽的平均值,M 表示 Fat-tree 网络拓扑中所有的链路数目。

由于现代数据中心网络中数据流数目庞大且链路已用带宽随时间波动较大,为减弱负载方差受短时间内链路已用带宽极值的影响,根据统计学思想,本文采用时间序列预测法(Time series forecasting method)中的指数平滑法,来预测下一时刻的网络负载均衡度。相比于文献[8]中将过去时间的网络负载方差全部同等处理的简单移动平均法,指数平滑法将上一时刻的预测值和实际值进行指数加权来计算下一时刻的预测值,它的优点是在不舍弃远期预

测值的同时给与了近期实际值更大的权重,从而有效地消除预测中的随机波动,得到的平滑值更符合数据中心实际网络带宽波动情况。预测公式如下:

S_t = λ · δ_t + (1 - λ) · S_{t-1}

其中,S_t 为加权计算得到的 t 时刻全网负载平滑值,δ_t 为 t 时刻的真实链路负载方差,S_{t-1} 为前一时刻的平滑值,λ 为平滑常数,其取值范围为[0,1]。将式(1)带入式(2)中得到:

S_t = λ · (1/M ∑ [load_a(t) - load(t)]^2) + (1 - λ) · S_{t-1}

其中,平滑常数 λ 越接近 1,远期全网链路已用带宽方差实际值对本期的平滑值影响的程度下降越快;反之,远期实际值对本期平滑值的影响程度下降越缓慢。根据经验估计,平滑常数 λ 基本判断标准如表 1 所示:

表 1 时间序列模型与平滑常数 λ 的关系

时间序列模型	λ
序列平稳	0.05 ~ 0.20
序列波动,但长期变化趋势不明显	0.10 ~ 0.40
序列波动较大,有明显上升或下降	0.40 ~ 0.60
时间序列为上升或下降序列,满足加性模型	0.60 ~ 1.00

2.4 路径选择数学模型

本文设数据中心网络 Fat-tree 拓扑为图 G = (V,E),其中,V/E 分别表示网络交换机节点集合和链路集合;c_{i,j}表示节点 i→j 间链路容量,且网络拓扑结构不发生改变;K = {k_1,k_2,...,k_r},r = 1,2,...,R 表示 t 时刻网络负载均衡参数 S_t 大于阈值 δ* 时大象流集合;y_{k_r} ∈ Y,z_{k_r} ∈ Z,b^{k_r} 分别表示第 r 个大象流 k_r 的源节点、目的节点以及带宽需求;x_{i,j}^{k_r} 表示满足第 r 个大象流 k_r 带宽需求下链路(i,j)上的流量。以最小化最大链路利用率 u 为目标函数,负载均衡问题的最佳路由数学模型如下:

目标函数: Minimize: u

约束条件:

s. t. { ∑_{j: (y_{k_r},j) ∈ E} x_{y_{k_r},j}^{k_r} - ∑_{j: (j,y_{k_r}) ∈ E} x_{j,y_{k_r}}^{k_r} = b^{k_r}, r = 1,2,...,R ∑_{j: (z_{k_r},j) ∈ E} x_{z_{k_r},j}^{k_r} - ∑_{j: (j,z_{k_r}) ∈ E} x_{j,z_{k_r}}^{k_r} = -b^{k_r}, r = 1,2,...,R ∑_{j: (i,j) ∈ E, i ≠ y_{k_r}, z_{k_r}} x_{i,j}^{k_r} - ∑_{j: (j,i) ∈ E, i ≠ y_{k_r}, z_{k_r}} x_{j,i}^{k_r} = 0, r = 1,2,...,R

$$\sum_{1 \leq r \leq R} x_{i,j}^{k_r} \leq u \cdot c_{i,j}, (i,j) \in E, r = 1, 2, \dots, R \quad (6)$$

$$x_{i,j}^{k_r} \geq 0, (i,j) \in E, r = 1, 2, \dots, R \quad (7)$$

其中,式(4)表示目标函数,即最小化最大链路利用率 u ;式(5)是流量调度过程中的需要满足的流量守恒定律,即除了源节点和目的节点之外,任意大象流 k_r 进入某节点的流量与流出该节点的流量相等;式(6)考虑了大流调度过程中需要满足的带宽约束条件;式(7)定义了每条链路上的流量必须满足非负性;由于数据中心大象调度问题可转化为多商品流问题,本文将采用粒子群优化算法对这个问题进行最优路径求解。

2.5 最优路径选择算法

2.5.1 粒子群优化算法

本文优化目标函数为最小化最大链路利用率,相比于遗传算法,粒子群优化算法在对此类 NP 完全问题的最优解求解过程中不需要进行遗传操作,如交叉和变异等,因此本文采用粒子群优化算法来求解上述问题的最优解。假设在一个 D 维搜索空间中,由 n 个粒子组成的种群为 $X = (X_1, X_2, \dots, X_n)$,第 i 个粒子在 D 维空间中的坐标向量为 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, $i = 1, 2, \dots, n$,飞行速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$,个体极值为 $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$,全局极值为 $P_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ 。群算法首先将一组随机解作为该算法的初始解,然后初始化粒子的位置和速度,并且计算每个粒子的适应值,通过多次迭代求解出最优值。在每一次的迭代中,粒子通过跟踪全局极值和个体极值来更新自己的位置,粒子根据如下公式更新自身的速度和位置。

$$v_{id}^{k+1} = w \cdot v_{id}^k + c_1 \cdot r_1 \cdot (p_{id}^k - x_{id}^k) + c_2 \cdot r_2 \cdot (p_{gd}^k - x_{id}^k) \quad (8)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (9)$$

$$F_{\text{fitness}} = \min(u) \quad (10)$$

在每次的迭代过程中,粒子通过式(8)和式(9)更新自己的速度和位置。式(8)中 k 是迭代次数; w 为惯性权重值; c_1, c_2 分别表示加速因子,即学习率,表示粒子向自身或群体学习的能力,为非负值; r_1, r_2 是介于 $[0, 1]$ 的均匀随机分布数以保持群体的多样性。式(10)表示适应度函数,本文用目标函数构造使用度函数,通过该式计算每个粒子的个体极值和全局极值。实验表明,在标准的粒子群算法中,惯性权重值 w 越大,越有利于全局收敛,但如果粒子一直处于较高速度,很容易超出可解空间的最优区域,导致发散现象,最终无法收敛到最优值;而较小的权

重值 w 使得算法拥有较强的局部搜索能力。故如果选择固定权重值,算法并不会得到很好的收敛效果^[12]。本文采用动态权重,同时对惯性权重 w 进行优化处理,以平衡全局搜索能力与局部搜索能力:

$$w(k) = w_{\text{start}} - \frac{(w_{\text{start}} - w_{\text{end}})}{K_{\text{max}}} \cdot t \quad (11)$$

式(11)定义了惯性权重值的优化方法,即随着算法迭代次数的增加,惯性权重值逐渐减小。其中 w_{start} 、 w_{end} 分别表示初始迭代权重值和迭代次数增加到最大值 K_{max} 时的权重值, k 为当前时刻的迭代次数。本文通过优化权重,使算法在早期具有较高的全局搜索能力以提高收敛速度,在后期具有较高的局部搜索能力以提高算法收敛精度,从而达到既可避免算法早熟又能较快完成收敛的效果。依据经验可得, $w_{\text{start}}, w_{\text{end}}$ 的值分别为0.9和0.4时,算法性能最佳。

2.5.2 算法实现描述

在核心交换机数为 k 的Fat-tree拓扑结构中,不同pod的主机间有 $(k/2)^2$ 条路径,但如果为某条流指定核心交换机的标号值,就能唯一确定该对主机间的通信路径。设网络负载均衡度大于阈值时大象流数量为 n ,用向量 $C = [c_1, c_2, \dots, c_n]$ 表示 n 条大象流对应的核心交换机标号, $c_i (i = 1, 2, \dots, n) = (1, 2, \dots, (k/2)^2)$ 表示第 i 条大象流选择的核心交换机标号值,根据选择结果将路径映射到各条链路上,并以式(4)为目标函数计算链路利用率,确定大象流的最佳传输路径。具体过程如下。

(1) 定义粒子编码方式。本文用 $X_s = \{x_{s1}, x_{s2}, \dots, x_{sn}\}$ 表示维度为 n 的第 s 个粒子的位置,即每一个 X_s 表示 n 条大象流调度问题的一个可行解 $C = [c_1, c_2, \dots, c_n]$,其中 x_{si} 的取值对应可行解 C 中的 c_i 。

(2) 对各个参数赋值,包括粒子群规模、粒子维度、学习率、 $w_{\text{start}}, w_{\text{end}}$ 。

(3) 随机初始化所有粒子的位置和速度,随机选择 n 个值作为当前粒子初始位置,通过式(10)计算每个可行解的最大链路利用率值,并将拥有该最小值的粒子的位置设置为个体最优解,通过搜索比较每个粒子的最优解得到整个粒子群全局最优解。

(4) 更新迭代次数,根据式(11)更新惯性权重值,从而使得所有的粒子进入到下一代。

(5) 根据适应度函数式(10)计算该代每个粒子适应值,并将该适应值分别与它所经历的最好位置值和全局最好位置值相比较,如果该粒子的适应

值小于它所经历的最好位置值,则将该适应值更新为当前该粒子的历史最好位置;如果该适应值小于全局最优值,则将其更新为全局历史最好位置。

(6) 判断算法是否搜索到理论最优值或到达最大迭代值,如果否,返回步骤(4);反之,停止迭代并输出最优解,得到全局最优调度路径。

3 实验与分析

3.1 实验环境设计

为验证本文提出算法的可行性,本文在 Ubuntu16.04 系统上采用 Ryu 控制器和 Mininet^[13] 搭建网络仿真环境,如图 3 所示。本文以 $K=4$ 的 Fat-tree 拓扑作为当前数据中心网络结构,链路带宽为 10 Mbit/s,并使用 Iperf 工具产生 UDP 数据流,流的到达模型服从泊松分布。

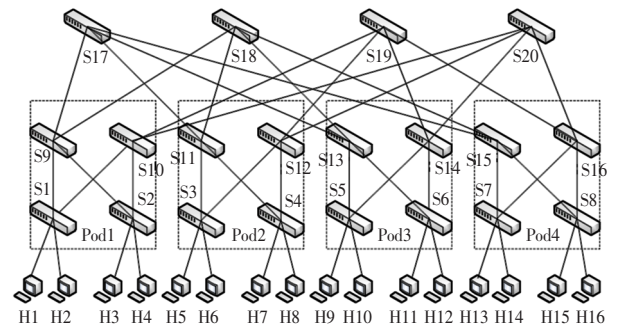


图 3 Fat-tree 网络拓扑

由于在 Fat-tree 网络结构中,位于不同 pod 内主机间的通信增加了核心层链路拥塞的可能性,因此本实验采用 staggered (EdgeP, PodP)^[6] 的流量模式测试不同算法的性能,即主机分别以概率 EdgeP、PodP 和 $1 - \text{EdgeP} - \text{PodP}$ 向与同一个边缘交换机下直连主机、同一个 Pod 内的主机以及不同 Pod 内的主机发送数据流,本文实验中 EdgeP、PodP 分别取 0.2 和 0.3,即以 $1 - 0.2 - 0.3 = 0.5$ 的概率向其他 Pod 内主机发送数据,以增加 Fat-tree 核心链路层拥塞的可能性。主要仿真参数设置见表 2。

表 2 主要仿真参数设置

参数	数值
网络拓扑模型	Fat-tree
粒子群规模	20
粒子维度	n
学习率 c_1	2
学习率 c_2	2
最大迭代次数 $K_{\max}$	100
初始权重 w_{start}	0.9
终止权重 w_{end}	0.4
平滑常数 λ	0.5

3.2 实验结果与分析

图 4 表示在 staggered(0.2,0.5) 通信模式下链路负载均衡度阈值 δ^* 与数据流平均吞吐量之间的关系。从图中可以看出当负载均衡度阈值 δ^* 设置为 $1.0 \times 10^{10} \text{ byte}^2$ 时,平均流吞吐量达到最大值。

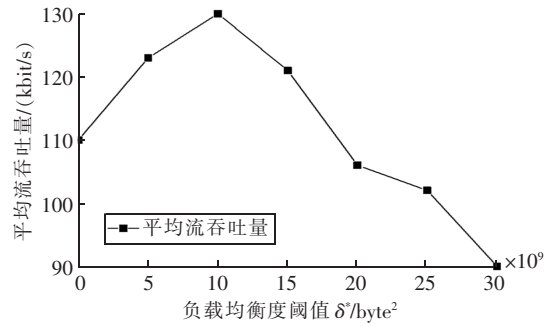


图 4 不同负载均衡度与平均吞吐量之间的关系

本文将数据流的平均吞吐量作为性能指标,并与 LABERIO、ECMP、GFF 机制作对比,以验证 DLB-PSO 机制的可行性。图 5 表示负载均衡度阈值 δ^* 为 $1.0 \times 10^{10} \text{ byte}^2$ 时,各机制在 staggered(0.2,0.3) 通信模式下的平均流吞吐量对比。

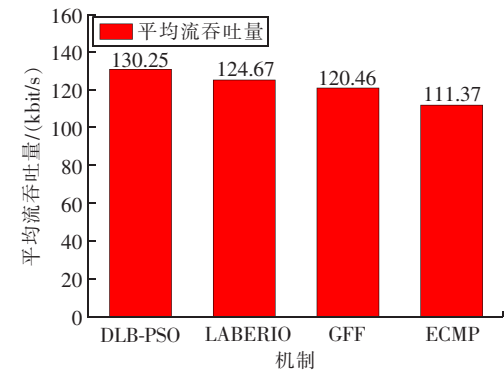


图 5 不同机制下的平均流吞吐量

实验结果表明,相比于 LABERIO、GFF 和 ECMP 机制的平均流吞吐量,DLB-PSO 机制分别提高了约 5%、8% 及 18%,故该算法具有可行性。为验证算法在不同流量负载下仍然具有可行性,本文以平均链路利用率及平均时延抖动为指标,测量各算法在流量负载由 0.1 递增到 0.9 时的网络性能。实验结果如图 6、图 7 所示。

图 6 对比了流量负载由 0.1 递增到 0.9 时 4 种算法的平均链路利用率。如图中所示,DLB-PSO 算法下的平均链路利用率明显高于其他算法。由于在流量负载较小时,Fat-tree 拓扑中许多链路处于空闲状态,故各算法下的链路平均利用率较低且差别不大;但当流量负载逐渐增加时,由于 ECMP 无法对大象流实时调度,易造成链路资源分配不合理,导致链

路利用率最低。相比 ECMP,其他 3 种算法都考虑了当前网络资源状态,但 GFF 算法随着流量负载的增加,当链路已用带宽接近饱和时,路径选择变得困难;LABERIO 算法并未考虑短时间内的网络波动情况;而本文提出的 DLB-PSO 算法,既考虑了全局链路资源,并在 LABERIO 算法基础上利用时间序列预测法平滑网络波动。故相比于 LABERIO、GFF 及 ECMP 算法,DLB-PSO 算法在链路平均利用率上表现更好。

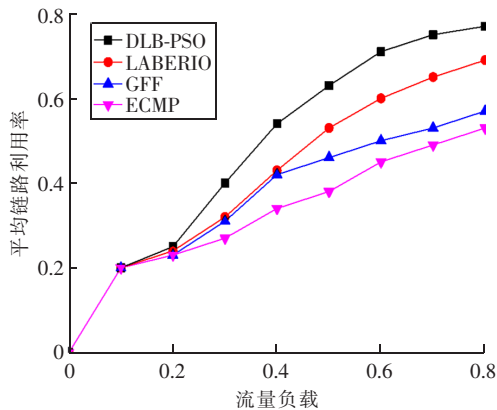


图 6 不同流量负载下的平均链路利用率

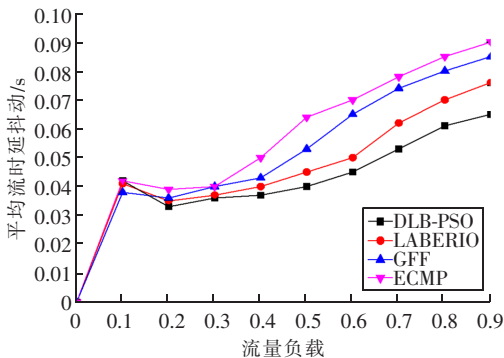


图 7 不同流量负载下平均时延抖动

图 7 对比了流量负载由 0.1 递增至 0.9 时 4 种算法的平均时延抖动。在实验开始阶段,由于 Ryu 控制器需要利用 openflow 南向协议与 openflow 交换机进行频繁通信以完成流表初始化,故在链路负载为 0.1 时,时延较为明显;在此之后,各算法的流量调度工作使得时延抖动有所下降,但随着流量负载的增加,链路带宽资源逐渐紧张,因此延时抖动呈上升趋势,但相比于其他 3 种流量调度机制,本文提出的机制 DLB-PSO 中时延抖动上升幅度最小。

4 结束语

针对数据中心网络流量特征及现有负载均衡算法的不足,本文提出了一种 SDN 环境下基于粒子群

优化的动态负载均衡机制(DLB-PSO)。首先,该机制通过监测全网负载已用带宽方差来刻画当前网络负载均衡程度,以将其作为该机制的触发条件,并利用时间序列预测方法尽可能减小在流量调度过程中受网络波动的影响;然后为大象流调度建立数学模型,并通过对粒子群进行优化来对该数学模型求解;最后利用 Ryu 控制器和 Mininet 网络仿真器实现该机制。实验对比表明,在 staggered 通信模式下,相比于 LABERIO、GFF 及 ECMP 算法,本文提出的 DLB-PSO 机制具有较高的平均吞吐量 and 平均链路利用率,及较低的平均流时延抖动。由于实验参数的设置直接影响到算法的性能,本文中的实验参数是通过查询文献和根据经验进行设置的,而在不同的网络场景及约束条件下粒子群算法参数的设置对于寻找最优路径的效率存在一定的影响,如何根据不同的网络场景实现各参数的自适应调整有待进一步的探究。

参考文献:

- [1] AL-FARES M, LOUKISSAS A, VAHDAT A. A scalable, commodity data center network architecture[C]//Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. 2008:63-74.
- [2] GUO C, LU G, LI D, et al. BCube: a high performance, server-centric network architecture for modular data centers[J]. ACM SIGCOMM Computer Communication Review, 2009,39(4):63-74.
- [3] HOPPS C. Analysis of an equal-cost multi-path algorithm: RFC 2992[R/OL]. [2019-03-20]. <http://www.rfc-editor.org/info/rfc2992>.
- [4] BENSON T, AKELLA A, MALTZ D A. Network traffic characteristics of data centers in the wild[C]//10th ACM SIGCOMM Conference on Internet Measurement. 2010:267-280.
- [5] KREUTZ D, RAMOS F M V, ESTEVES VERISSIMO P, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2015, 103(1):14-76.
- [6] AL-FARES M, RADHAKRISHNAN S, RAGHAVAN B, et al. Hedera: dynamic flow scheduling for data center networks[C]//USENIX Symposium on Networked Systems Design and Implementation. 2010:19.
- [7] 陈琳,张富强. 面向 SDN 数据中心网络最大概率路径流量调度法[J]. 软件学报, 2016, 27(S2):254-260.
CHEN Lin, ZHANG Fuqiang. Maximum probability path scheduling algorithm for elephant flow in data center networks based on SDN[J]. Journal of Software, 2016, 27(S2):254-260. (in Chinese)

[8] LONG H, SHEN Y, GUO M, et al. LABERIO: dynamic load-balanced routing in OpenFlow-enabled networks [C] // IEEE International Conference on Advanced Information Networking & Applications. 2013:290 – 297.

[9] WANG W, SUN Y, SALAMATIAN K, et al. Adaptive path isolation for elephant and mice flows by exploiting path diversity in datacenters [J]. IEEE Transactions on Network and Service Management, 2016, 13(1) :5 – 18.

[10] FIGUEIREDO E M N, LUDERMIR T B, BASTO-FILHO C J A. Many objective particle swarm optimization [J]. Information Sciences, 2016, 374:115 – 134.

[11] GIOTIS K, ARGYROPOULOS C, ANDROULIDAKIS G, et al. Combining OpenFlow and sFlow for an effective and

scalable anomaly detection and mitigation mechanism on SDN environments [J]. Computer Networks, 2014, 62(5) :122 – 136.

[12] 陈曦, 傅明. 改进粒子群算法在动态交通分配问题中的应用 [J]. 计算技术与自动化, 2006, 25(2) :60 – 62.
CHEN Xi, FU Ming. Application on improved particle swarm optimization in dynamic traffic assignment [J]. Computing Technology and Automation, 2006, 25(2) :60 – 62. (in Chinese)

[13] LANTZ B, HELLER B, MCKEOWN N. A network in a laptop: rapid prototyping for software-defined networks [C] // ACM Workshop on Hot Topics in Networks. 2010: 1 – 6.

声 明

为适应我国信息化建设的需要,扩大作者学术交流渠道,实现期刊编辑、出版工作的网络化,本刊已加入《中国学术期刊(光盘版)》、《中国期刊网》全文数据库、《万方数据——数字化期刊群》和《中文科技期刊数据库》,并已许可《中国学术期刊(光盘版)》电子杂志社在中国知网及其系列数据库产品中,以数字化方式复制、汇编、发行、信息网络传播本刊全文,作者著作权使用费随本刊稿酬一次性给付。如不同意将文章编入相关数据库,请在来稿时声明,本刊将做适当处理。

本刊编辑部