

基于可靠性与负载优化的多控制器弹性部署算法

陆 杰, 张 震, 胡 涛

(国家数字交换系统工程技术研究中心, 郑州 450002)

摘 要: 现有多控制器部署方案中存在控制路径可靠性差、控制器负载不均衡的现象。为此, 提出一种基于可靠性与负载优化的多控制器弹性部署算法。基于谱聚类思想将多控制器部署问题转化为网络拓扑矩阵的行向量分类问题后, 采用基于模拟退火的 k-medoids 算法对行向量进行分类, 从而实现多控制器的弹性部署。仿真结果表明, 与 KCBP 和优化 K-means 算法相比, 该算法在保证负载均衡和网络低时延的基础上, 可使控制路径可靠性平均提高高达 17%。

关键词: 软件定义网络; 控制器部署; 可靠性; 负载均衡; 谱聚类

中文引用格式: 陆 杰, 张 震, 胡 涛. 基于可靠性与负载优化的多控制器弹性部署算法[J]. 计算机工程, 2018, 44(8): 135-141.

英文引用格式: LU Jie, ZHANG Zhen, HU Tao. Multi-controllers elastic placement algorithm based on reliability and load optimization[J]. Computer Engineering, 2018, 44(8): 135-141.

Multi-controllers Elastic Placement Algorithm Based on Reliability and Load Optimization

LU Jie, ZHANG Zhen, HU Tao

(National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, China)

[Abstract] In order to solve the problem of poor control path reliability and unbalanced load of controllers in the existing multi-controllers placement scheme, a Multi-Controllers Elastic Placement (MCEP) algorithm based on reliability and load optimization is proposed. Based on spectral clustering, the problem of multi-controllers placement is transformed into a row vector classification problem of network topology matrix. The k-medoids algorithm based on simulated annealing is used to classify line vectors, thus the flexible placement of multi-controllers is realized. Simulation results show that, compared with the KCBP and the optimized K-means algorithm, the proposed algorithm can make the reliability of the control path up to 17% on the basis of guaranteeing the load balancing and low network delay.

[Key words] Software Defined Network (SDN); controller placement; reliability; load balancing; spectral clustering

DOI: 10.19678/j.issn.1000-3428.0050521

0 概述

随着软件定义网络 (Software Defined Network, SDN)^[1] 的引入, 传统网络中逻辑集中式架构开始朝着控制和数据平面分离的方向转变, 这种转变能够方便网络的管理和维护, 因此, 受到学界和工业界的广泛关注。随着 SDN 网络规模的不断扩大, 单个控制器已经难以满足现有的网络需求。为提高网络的可扩展性, 避免单点失效, 研究人员在集中式控制的基础上提出将网络划分为多个域, 通过部署多个控制器来实现逻辑上集中、物理上分布的 SDN 网络^[2]。多

控制器的部署主要有 3 个关键: 控制器的部署数量; 控制器的部署位置; 交换机与控制器间的映射关系, 即哪些交换机由哪些控制器控制。一个网络所需的控制器数量由实际场景所决定, 因此, 本文主要研究和分析控制器的部署位置和交换机与控制器间的映射关系。

在实际中, 要实现控制器的合理部署, 会面临较多挑战。控制层的失效概率要满足特定的要求以保证网络的可靠性。同时, 要利用控制器的负载优化来保证负载的合理分配, 并减少交换机到控制器的时延, 以提高网络性能。

基金项目: 国家自然科学基金 (61372121、61309019); 国家重点基础研究计划项目“可重构信息通信基础网络体系研究” (2012CB315900); 国家重点研发计划项目 (2017YFB0803200)。

作者简介: 陆 杰 (1994—), 男, 硕士研究生, 主研方向为新型网络体系结构、软件定义网络; 张 震, 讲师、博士; 胡 涛, 硕士研究生。

收稿日期: 2018-02-27 **修回日期:** 2018-04-10 **E-mail:** luwangli@hotmail.com

目前,针对多控制器部署的研究主要分为两类:

1) 考虑时延、负载均衡等因素,通过多目标优化来求解控制器部署的位置。文献[3]考虑控制报文的平均时延和最差时延后确定控制器的最佳部署位置。文献[4]针对控制器的负载过大导致时延显著增加的问题,提出在满足控制器负载均衡的条件下,寻求平均时延最优部署。文献[5]考虑额外的目标,如控制器之间的负载平衡问题,以在不同目标相互竞争的情况下,在不同的位置选择之间进行权衡。文献[6]利用谱分类算法,综合考虑时延与容量限制下的多控制器负载均衡问题,然后提出 KCBP (K Controllers Balanced Placement) 算法,该算法是基于 Normalized Cut 的标准化分割方式。文献[7]在 K-means 算法的基础上提出优化 K-means 算法,为避免初始值的选取对分区的影响,其使用一种逐步递进分区的方式。这类方案均以提高网络通信质量、设备利用率为目标,但是忽略了设备故障对网络的影响。

2) 考虑控制器的可靠部署。文献[8]引入失效路径百分比来评价 SDN 网络的可靠性,并用数学方法证明可靠控制器部署 (Reliable Controller Placement, RCP) 是 NP-Hard 问题。文献[9]建立 Survivor 模型,其考虑路径连通性、控制器容量和快速恢复性 3 个指标。文献[10]通过分析表明,控制器位置选择对 SDN 网络可靠性有重要影响,同时,过多的控制器会显著增加运营成本。这类方案在考虑控制层的可靠性时,没有考虑数据转发层对控制层可靠性的影响。由于转发层和控制层实际上是复用同一个网络,因此当数据流量超过链路带宽时,数据层的拥塞也会导致控制层的失效。同时,这类方案仅通过考虑网络可靠性指标来进行控制器部署,其对于交换机分配、网络时延等因素缺乏考虑,容易导致控制器负载失衡。

针对上述方案存在的不足,本文在谱聚类思想的基础上,对控制路径时延、控制路径可靠性和控制器负载均衡 3 个指标进行描述,提出一种基于可靠性和负载优化的多控制器弹性部署 (Multi-Controllers Elastic Placement, MCEP) 算法。首先,利用节点间的相似关系将网络拓扑表示为相似矩阵,并将多控制器部署问题转化为相似矩阵的行向量分类问题;然后,采用基于模拟退火的改进 k-medoids 算法对行向量进行分类,以实现多控制器弹性部署;最后,通过实验验证该算法的可行性。

1 问题重述与建模

1.1 问题重述

在 SDN 中,控制器部署位置和交换机与控制器之间的映射关系会对整个控制网络的可靠性和性能造成

影响。图 1 所示为在由 6 个节点和 2 个控制器组成的 SDN 网络中,3 种不同的控制器部署方式对控制网络的影响。其中,实线表示物理链路,虚线表示控制链路。

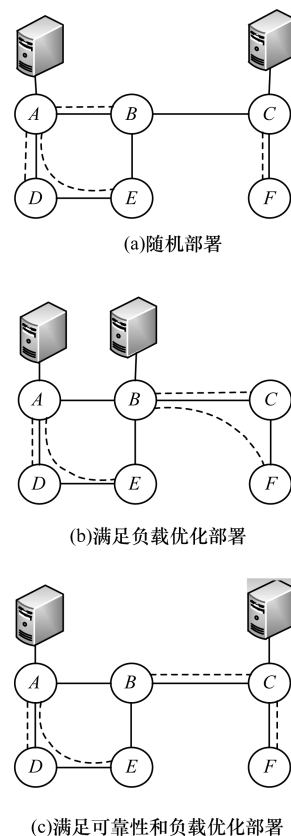


图 1 控制器部署方式对网络性能的影响

图 1(a)、图 1(c) 表示相同的控制器部署位置、不同的交换机与控制器映射关系对控制器负载的影响。从中可以看出,在控制器处理容量相同的情况下,图 1(a) 的控制器负载更均衡。图 1(b)、图 1(c) 表示网络分区相同、控制器部署位置不同对网络可靠性的影响。从中可以看出,假如 B 节点与 C 节点之间的物理链路发生故障,则图 1(c) 中有一条控制路径会失效,图 1(b) 中有 2 条控制路径会失效,即图 1(c) 的控制网络比图 1(b) 更可靠。

基于上述分析,本文所要研究的问题总结如下:

- 1) 在给定网络拓扑和控制器数量的情况下,确定控制器的部署位置。
- 2) 确定网络中交换机与控制器间的映射关系。

1.2 问题建模

将物理网络表示为一个无向图 $G(V, E, B)$ 。其中, V 是图中节点 (交换机) 的集合, $V = (v_1, v_2, \dots, v_n)$, $|V|$ 表示节点个数, E 是边 e_{ij} 的集合, e_{ij} 表示节点 i 和节点 j 之间直连链路的长度, 非直连链路 e_{ij} 取值为 0, B 表示直连链路 e_{ij} 的带宽。对于 SDN 控制层, 控制器的控制路径可以视为物理网络上的覆盖网络, 如图 1 所示。

V_c 表示控制器所在的位置, $V_c \subseteq V$, 且 $|V_c| = k$ 表示控制器的数量, V_c^j 表示第 j 个控制器的位置。每个控制器管理一个域, 则控制器 j 管理的交换机集合 $SC_j = (s_1, s_2, \dots, s_n)$, $|SC_j|$ 表示交换机个数, 对应的, SC_j^i 表示交换机 i 所属的控制器 j 。对于任意给定的节点 s 和 t , P_{st} 表示从节点 s 到节点 t 的最短路径, $P_{st} = (v_s, v_{s+1}, \dots, v_t)$, $v_s, v_{s+1}, \dots, v_t$ 表示最短路径依序经过的节点, $d_{st} = e_{s(s+1)} + e_{(s+1)(s+2)} + \dots + e_{(t-2)(t-1)} + e_{(t-1)t}$ 表示节点 s 和节点 t 之间的最短路径 P_{st} 的长度。若 $s \in V_c, t \in SC_s$, 则 P_{st} 表示一条控制路径。

为解决多控制器部署问题, 本文定义如下指标。

定义 1 (控制路径失效概率 (Control Path Failure Probability, CPFP)) 路径失效可分为 2 种情况:

1) 由物理原因所引起, 如光缆损坏, 物理中断的概率与链路有关。假设链路使用的介质单位长度失效的概率为 p_α , 则直连链路 e_{ij} 物理可用概率 L_{ij}^b 可表示为:

$$L_{ij}^b = (1 - p_\alpha)^{e_{ij}} \quad (1)$$

2) 由于控制路径实际上是与物理转发层使用同一个网络, 因此转发层数据流量过大会导致一些链路的拥塞, 也会影响控制链路的可靠性。假设直连链路 e_{ij} 带宽为 b_{ij} , 流量为 f_{ij} , 当流量处于带宽允许的范围内时, 不会导致链路拥塞, 而当流量超过一定的阈值时, 流量越大, 链路拥塞程度越高。当流量大于带宽时, 则认为该链路不可用, 即拥塞的概率为 1。因此, 链路拥塞的概率可表示为:

$$L_{ij}^c = \begin{cases} 0, & f_{ij} < \delta \\ \frac{f_{ij}}{b_{ij}}, & f_{ij} < b_{ij} \\ 1, & \text{else} \end{cases} \quad (2)$$

其中, δ 表示阈值。

综合考虑物理失效与链路拥塞, 直连链路 e_{ij} 可用概率 AP_{ij} 表示为:

$$AP_{ij} = L_{ij}^b (1 - L_{ij}^c) \quad (3)$$

$$\forall i, j \in n, AP_{ij} = AP_{ji} \quad (4)$$

式(3)、式(4)表示链路可用概率与方向无关。对于一条控制路径 P_{st} , 其可用的概率取决于路径上各条直连链路的状态。因此, 控制路径失效概率可表示为:

$$CPFP_{st} = 1 - \prod_{e_{ij} \in P_{st}} AP_{ij} \quad (5)$$

本文采用控制路径平均失效概率 (Control Path Average Failure Probability, CPAFP) 对网络的整体可靠性进行评估, 其表示为:

$$CPAFP = \frac{1}{|V|} \sum_{s \in SC_j, t \in V_c} CPFP_{st} \quad (6)$$

定义 2 (控制路径时延 (Control Path Delay, CPD)) 在 SDN 网络中, 当新流到达交换机时, 由交换机交付

给控制器处理, 该过程会产生时延, 且时延对于 SDN 网络的性能有较大影响。通常情况下, 时延可以分为 4 个方面: 传播时延, 处理时延, 排队时延, 发送时延。对于广域网而言, 传播时延大于其他 3 种时延。本文主要进行广域网的 SDN 研究, 因此, 使用传播时延近似表示网络时延。

设光纤中的信号传播速度是真空中光速 c 的 $\frac{2}{3}$ 。对于任意一条控制路径, 节点 i 表示控制器部署位置, 节点 j 作为交换机, 且交换机 j 隶属于控制器 i 。对于一条控制路径 p_{ij} , 其时延可表示为:

$$CPD_{ij} = \frac{d_{ij}}{\frac{2}{3}c} = \frac{3}{2c} \sum_{e_{st} \in p_{ij}} |e_{st}| \quad (7)$$

控制路径平均时延 (Control Path Average Delay, CPAD) 是所有控制路径时延的平均值, 可表示为:

$$CPAD = \frac{1}{|V|} \sum_{i \in SC_j, j \in V_c} CPD_{ij} \quad (8)$$

定义 3 (控制器负载标准差 (Controller Load Standard Deviation, CLSD)) 控制器轻载会导致控制器性能浪费, 而控制器过载时, 由于控制器处理性能跟不上流请求速度, 从而导致时延大幅增加, 网络性能下降。因此, 在多控制器位置部署中, 需要考虑控制器之间的负载均衡。

本文将 SDN 中控制器的负载视为控制器因管理交换机而担负的计算负载。为论述方便, 本文将一个控制器的负载视为该控制器所管理的交换机数目。控制器负载差异表示为各控制器实际管理交换机数量的差异, 可通过各控制器所管理交换机数量的标准差来表示:

$$CLSD = \sqrt{\frac{1}{k} \sum_{i=1}^k \left(|SC_i| - \frac{|V|}{k} \right)^2} \quad (9)$$

基于以上 3 个指标的定义, 低时延与高可靠的多控制器均衡部署问题即为找到一个合适的方法, 使 CPAFP、CPAD、CLSD 3 个指标均达到最小。可表示为:

$$\min [CPAFP, CPAD, CLSD] \quad (10)$$

$$\text{s. t. } CPFP_{ij} < P, i \in SC_j, j \in V_c \quad (11)$$

$$CPD_{ij} < D, i \in SC_j, j \in V_c \quad (12)$$

$$|SC_j| < N, j \in V_c \quad (13)$$

式(11)表示控制链路失效概率不能高于 P , 式(12)表示控制链路的最差时延不能高于 D , 式(13)表示单个控制器所管理交换机的个数不能多于 N 。

多控制器均衡部署问题是一个带约束的多目标优化问题, 其难以在多项式时间内求得最优解。因此, 本文采用谱聚类思想来求解。为实现对网络的合理划分, 在对图上的节点进行聚类时, 原则上将距离较近的节点聚为一类。而根据本文的需求, 可以用 2 个节点之间的通信时延和可靠性来定义节点距

离。此处,本文引入节点相似度来计算节点间的相似程度。

定义 4(节点相似度) 相似度表示 2 个节点相似的程度,相似程度越高,表示 2 个节点归为一类的可能性越大。对应于本文的路径可靠性与时延,任意 2 个节点之间的时延越小,通信可靠性越大,其相似程度也越大。因此,首先定义包含可靠性和时延的节点相似度。因为 *CPEP* 是取值为 (0,1) 的概率值,而 *CPD* 是一个具体的数值,为方便两者的比较,对 *CPD* 做归一化处理:

$$CPD'_{ij} = \frac{CPD_{ij} - \min CPD}{\max CPD - \min CPD} \quad (14)$$

其中, $\max CPD$ 表示网络中链路的最长时延, $\min CPD$ 表示网络中链路的最短时延。

当 2 个节点之间的链路可靠性低于可容忍的可用概率 P 或时延超过可容忍的最大时延 D 时,将这 2 个节点的相似度设为 0。因此,2 个节点之间相似度 w_{ij} 表示为:

$$w_{ij} = \begin{cases} \frac{1}{\alpha CPD'_{ij} + (1-\alpha) CPFP}, & CPD < D, CPFP < P \\ 0, & \text{else} \end{cases} \quad (15)$$

其中, α 表示时延与可靠性之间的权值,当 $0 \leq \alpha < 0.5$ 时,聚类更偏向于可靠性,当 $0.5 \leq \alpha \leq 1$ 时,聚类原则上偏向于时延。由各节点相似度可以得到表示整个网络拓扑的相似度矩阵 W 。

对于一个网络拓扑 G ,将其切割为 2 个子图 A 、 B ,定义 A 和 B 之间的切割为:

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij} \quad (16)$$

由切割的定义可得,若将网络拓扑切割为 k 个子图,则最小割可以表示为:

$$Cut(A_1, A_2, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k W(A_i, \bar{A}_i) \quad (17)$$

其中, $\bar{A}_i$ 为 A_i 的补集, $\frac{1}{2}$ 是为避免每条边被计算 2 次而设置的数值。式 (17) 使相似度大的节点归为一类,属于不同类的节点相似度尽可能小。但是这种划分原则没有考虑子域之间节点个数的负载情况,有可能会产生图 2 所示的情况:在节点 2 和节点 3 之间进行切割,能够使 $Cut(A, \bar{A})$ 最小化,但是却导致 2 个子图之间的不平衡。

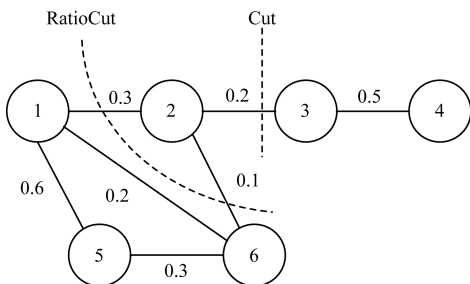


图 2 最小割与比率割的网络分区结果对比

基于负载均衡的考虑,本文采用比率割 (RatioCut) 作为图划分的原则。RatioCut 不仅考虑最小化切割 Cut,而且考虑最大化每个子图节点的个数。RatioCut 表示为:

$$RatioCut(A_1, A_2, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{|A_i|} \quad (18)$$

其中, $|A_i|$ 表示区域 A_i 的节点个数。当子图划分不均匀时,如某个区域只包含很少的节点,则该子图的 $|A_i|$ 会很小,导致总的 RatioCut 值很大。而整个图中的节点数是一个固定值,即 $|A_1| + |A_2| + \dots + |A_k| = |V|$,由均值不等式可知,当各区域节点个数趋向均匀时,RatioCut 最小。因此,式 (18) 的求解可以转化为求 RatioCut 的最小值,即:

$$\operatorname{argmin} RatioCut(A_1, A_2, \dots, A_k) \quad (19)$$

由于 W 是相似度矩阵,根据谱理论可知,度矩阵 D 表示为:

$$D = \operatorname{diag}(d_1, d_2, \dots, d_k) \quad (20)$$

其中, $d_i = \sum_{j=1}^n w_{ij}$ 表示列相加。则拉普拉斯矩阵 L 表示为:

$$L = D - W \quad (21)$$

对于任意 n 维向量 $f = [f_1, f_2, \dots, f_n]$,有:

$$\begin{aligned} f^T L f &= f^T D f - f^T W f = \sum_{i=1}^n d_i f_i^2 - \sum_{i,j=1}^n w_{ij} f_i f_j = \\ &= \frac{1}{2} \left(\sum_{i=1}^n d_i f_i^2 - 2 \sum_{i,j=1}^n w_{ij} f_i f_j + \sum_{j=1}^n d_j f_j^2 \right) = \\ &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \end{aligned} \quad (22)$$

为对网络节点进行聚类,需要为每个节点赋予一个向量来指示该节点的类别。引入指示向量 $H_{n \times k} = [h_1, h_2, \dots, h_k]$ 。对于任意一个 n 维向量 h_i ,定义其元素 h_{ij} 为:

$$h_{ij} = \begin{cases} 0, & v_i \notin A_i \\ \frac{1}{\sqrt{|A_i|}}, & v_i \in A_i \end{cases} \quad (23)$$

则对于 $h_i^T L h_i$,有:

$$\begin{aligned} h_i^T L h_i &= \frac{1}{2} \sum_{m=1}^n \sum_{n=1}^k w_{mn} (h_{im} - h_{in})^2 = \\ &= \frac{1}{2} \left\{ \sum_{m \in A_i, n \notin A_i} w_{mn} \left(\frac{1}{\sqrt{|A_i|}} - 0 \right)^2 + \sum_{m \notin A_i, n \in A_i} w_{mn} \left(0 - \frac{1}{\sqrt{|A_i|}} \right)^2 \right\} = \\ &= \frac{1}{2} \left\{ Cut(A_i, \bar{A}_i) \frac{1}{|A_i|} + Cut(\bar{A}_i, A_i) \right\} = \\ &= \frac{Cut(A_i, \bar{A}_i)}{|A_i|} = RatioCut(A_i, \bar{A}_i) \end{aligned} \quad (24)$$

式 (24) 是将网络拓扑切分成 2 个子图的结果,因此,对于 k 个子图切分的 RatioCut,可表示为:

$$RatioCut(A_1, A_2, \dots, A_k) = \sum_{i=1}^k h_i^T L h_i = \sum_{i=1}^k (H^T L H)_{ii} = \operatorname{Tr}(H^T L H) \quad (25)$$

因此,求解 RatioCut 最小值问题就转化为求解矩阵 $\mathbf{H}'\mathbf{L}\mathbf{H}$ 迹的问题,即:

$$\operatorname{argmin} \operatorname{Tr}(\mathbf{H}'\mathbf{L}\mathbf{H})$$

$$\text{s. t. } \mathbf{H}'\mathbf{H} = \mathbf{I}_k \quad (26)$$

由 Rayleigh-Ritz 定理^[11]可知,若 $\mathbf{L}$ 为实对称矩阵,其特征值 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$,且其对应的特征向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 正交,则:

$$\min \operatorname{Tr}(\mathbf{H}'\mathbf{L}\mathbf{H}) = \lambda_1 + \lambda_2 + \dots + \lambda_k \quad (27)$$

其中,最小值对应的矩阵 $\mathbf{H} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$ 。

由于 $\mathbf{H}$ 中向量的取值受到限制条件的影响,其取值是离散的,因此该问题的求解是 NP-Hard。为求解该最小值,可将其进行松弛。将 h_{ij} 的取值放宽到实数范围^[6],则最小实数解出现在矩阵 $\mathbf{L}$ 的前 k 个最小特征值对应的特征向量组成的矩阵 $\mathbf{H}$ 中。此时, $\mathbf{H}$ 矩阵由 n 个 k 维行向量组成,其分别表示 n 个节点的 k 维坐标,每个节点的 k 维坐标分别代表该节点的 k 维属性,越相似的节点,其 k 维属性越相近。

在获得 $\mathbf{H}$ 矩阵后,谱聚类通常采用 K-means 算法对 k 维坐标系中的 n 个点进行分类。由前述推论可知,谱聚类的实质是一个降维过程,其将 $n \times n$ 维矩阵的分类问题转化为 $n \times k$ 维矩阵的分类问题,通常 $k \ll n$,通过该转化可以有效降低时间复杂度。考虑到 RatioCut 中原有的 K-means 算法对极值比较敏感,同时有可能陷入局部最优解,因此,本文采用模拟退火的 k-medoids 算法替代 K-means 算法。

2 算法描述与分析

2.1 算法描述

针对现有多控制器部署方案中存在控制路径可靠性差和控制器负载不均衡的问题,本文提出一种基于可靠性和负载优化的多控制器弹性部署算法 MCEP。MCEP 算法基于改进的 RatioCut 算法,并使用模拟退火的 k-medoids 算法对其聚类过程进行优化。MCEP 算法的主要改进体现在:

1) 采用 k-medoids 代替 RatioCut 中原有的 K-means 聚类,将中心点的选取限制在当前聚类所包含的数据点集合中,防止孤立点对聚类的影响。

2) 初始聚类后,增加一个模拟退火的调整步骤,避免陷入局部最优。

设节点 i 的坐标为 x_i ,所在类的中心点(即控制器位置)为 SC_j^i ,其坐标为 $xc_{SC_j^i}$ 。常用的 k-medoids 分类算法的目标函数表示为:

$$\operatorname{obj} = \sum_{i=1}^n (x_i - xc_{SC_j^i})^2 \quad (28)$$

对由矩阵 $\mathbf{H}_{n \times k}$ 表示的 n 个节点进行分类,任意选取 k 个节点,按照 k-medoids 算法迭代,直至稳定,

记录稳定状态时的分类结果并更新最优结果,用式(28)计算此时目标函数 obj 。然后将类中某个节点归入其邻居节点的类别,按照 k-medoids 算法迭代,直至能够稳定计算目标函数 obj ,若调整后,新 obj 的值减小,则保留调整之后的结果作为新的分类,并更新最优结果;若调整后 obj 的值增加,则以一定的概率值接受新结果。同时,按照退火速率降低该概率值。在连续多次退火后,结束调整,以最优结果作为最后的分类结果,输出。

MCEP 算法具体描述如下:

算法1 MCEP 算法

输入 控制器个数 k ,权重值 α ,链路流量矩阵 $\mathbf{F}$,离开温度 $T_{\min}$,初始温度 T_{org} ,降温速度 r

输出 控制器位置 bestV_c ,交换机与控制器之间的映射关系 bestSC

```

1. for each  $i, j \in V$  do
2.  $\text{CPFP}_{st} \leftarrow$  式(5),  $\text{CPD}_{ij} \leftarrow$  式(7)
3. end for
4.  $\mathbf{W} \leftarrow$  式(16),  $\mathbf{D} \leftarrow$  式(20),  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 
5.  $[\mathbf{H}] = \text{eigs}(\mathbf{L}, k)$ ,
6.  $\text{tempV}_c \leftarrow \text{initKCenter}(\mathbf{H}, k)$ ,  $T = T_{\text{org}}$ 
7. 调用 k-medoids, 得  $\text{tempV}_c, \text{tempSC}$ 
8.  $\text{bestV}_c \leftarrow \text{V}_c \leftarrow \text{tempV}_c$ ,  $\text{bestSC} \leftarrow \text{SC} \leftarrow \text{tempSC}$ 
9. while(  $T > T_{\min}$  )
10. 随机选择一个节点, 将其归入邻居节点类别; 调用 k-medoids, 得  $\text{tempV}_c, \text{tempSC}$ 
11.  $\Delta = \operatorname{obj}(\text{tempSC}) - \operatorname{obj}(\text{SC})$ 
12. if (  $\Delta \leq 0$  )  $\text{SC} \leftarrow \text{tempSC}$ 
13. if (  $\operatorname{obj}(\text{SC}) < \operatorname{obj}(\text{bestSC})$  )
14.  $\text{bestSC} \leftarrow \text{SC}$ ,  $\text{bestV}_c \leftarrow \text{V}_c$ 
15. end if
16. else
17.  $\text{SC} \leftarrow$  以  $\exp(-\Delta/T)$  的概率将  $\text{tempSC}$  赋值
18. end if  $T = T * r$ 
19. end while
20. return  $\text{bestSC}, \text{bestV}_c$ 
```

在 MCEP 算法中:第1行~第5行,得到矩阵 $\mathbf{L}$ 的特征值和特征向量,并从中选取 k 个最小特征值对应的特征向量组成矩阵 $\mathbf{H}$;第6行~第8行,使用传统的 k-medoids 算法对表示节点的 k 维行向量进行第一次分类;第9行~第19行表示基于模拟退火对分类结果进行调整;第20行输出最优的分类结果。

2.2 算法复杂度分析

MCEP 算法首先利用 Johnson 算法^[12]计算任意2点之间的最短路径,然后通过式(5)和式(7)得到相应的 CPFP 和 CPD ,其时间复杂度为 $O(V^2 \lg V + VE)$ 。计算矩阵 $\mathbf{W}, \mathbf{D}, \mathbf{L}$ 的时间复杂度为 $O(V)$,计算特征向量组成的矩阵 $\mathbf{H}$,其时间复杂度为 $O(V^3)$,

k-medoids 算法的时间复杂度为 $O(V^2kt)$, 其中, t 为 k-medoids 算法的迭代次数, 因为在每次模拟退火的步骤中都需调用 k-medoids 进行分类, 所以其调整时间为 $O(V^2kt \log_r(T_{\min}/T_{\text{org}}))$ 。因此, MCEP 算法总的时间复杂度为 $O(V^3 + V^2 \log_r T)$ 。

3 实验结果与分析

3.1 实验环境

为评估 MCEP 算法的有效性和可行性, 本文以实际网络拓扑 OS3E^[13] 作为仿真进行分析。OS3E 拓扑结构如图 3 所示, 其由 34 个节点、42 条边组成。仿真中控制器部署方法使用 Matlab 模拟实现, 实验机配置为 Intel Core i5 2.4 GHz, 8 GB RAM。

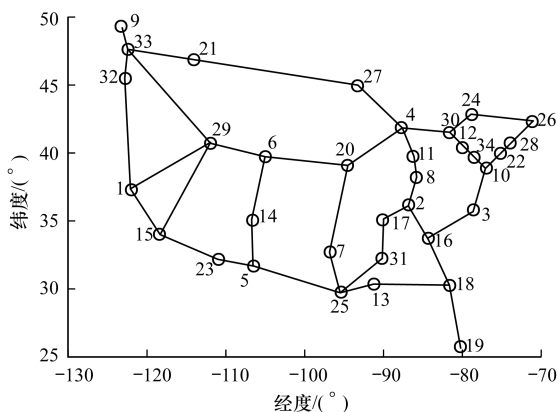


图 3 OS3E 网络拓扑

通常链路流量处于动态范围内, 且每条链路的带宽也并不相同。为使实验易于实现, 在不影响结果的前提下, 假定每条物理链路的带宽 $b_{ij} = 50$ Mb/s。所有链路有相同的流强度, 且流量 f 服从强度为 20 Mb/s 的泊松分布^[14]。链路物理失效的概率取决于链路材料和外部环境, 在本实验中, 假设所有的链路具有相同的物理失效概率, 每 100 km 失效概率为 0.01^[15]。

3.2 算法性能分析

实验 1 验证在不同参数设置下 MCEP 算法的控制器部署和分区情况。对 MCEP 算法进行不同参数仿真, 网络划分和控制器部署结果如图 4 所示。其中, 相同的符号表示处于同一个子域, 黑色实心符号表示控制器部署位置。图 4(a)、图 4(b) 表示将网络分为 2 个区域时, α 值不同对网络分区的影响。其中, $\alpha = 1$ 表示只考虑链路可靠性指标, 此时长度较长的链路更有可能被切断。从图 4(a)、图 4(b) 可以看出, 尽管时延和可靠性都与链路的长度相关, 但是链路失效的概率与链路长度成指数关系, 而时延与链路长度成线性关系。因此, 可靠性对链路长度更

敏感。图 4(c)、图 4(d) 的参数分别为 $k = 3, \alpha = 0.5$ 和 $k = 4, \alpha = 0.5$, 其分别将网络划分成 3 个和 4 个区域, 且划分出的子网大小较均匀。这表明可以通过改变 MCEP 的参数来调整 SDN 网络的可靠性和负载均衡之间的权衡关系。

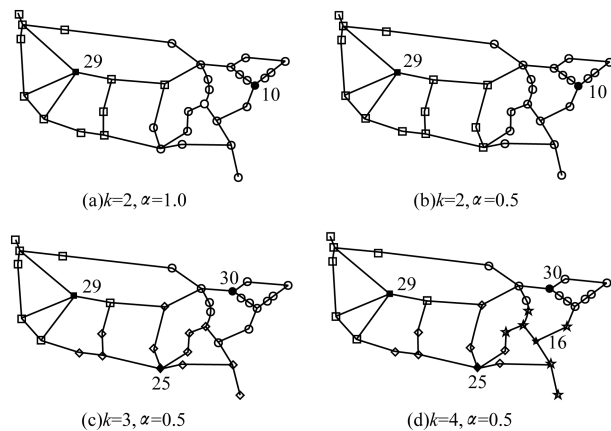


图 4 不同参数时 MCEP 算法的部署情况

实验 2 将 MCEP 算法与 KCBP 算法^[6]、优化的 K-means 算法^[7] 在可靠性、时延与负载均衡 3 个指标上进行对比。实验中 MCEP 算法的权重 α 默认取 0.5。图 5 所示为 3 种算法的控制路径可靠性对比结果, 从中可以看出, 随着控制器个数的增加, 3 种算法的可靠性均有所提高, 原因是控制器数量增加, 可以使每个子域更紧密, 从而使可靠性得到提升。但在控制器数量相同时, MCEP 算法的可靠性明显大于其他 2 个算法, 与另外 2 个算法相比, MCEP 算法的可靠性平均提升了 17%。

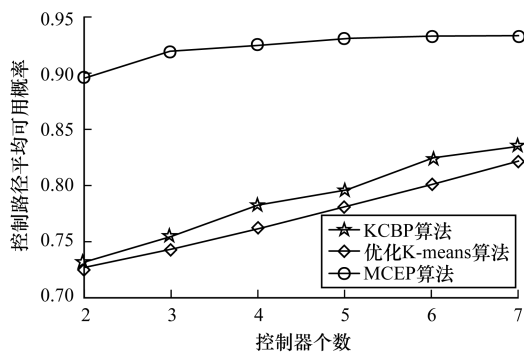


图 5 3 种算法控制路径可靠性对比结果

图 6 所示为 3 种算法的平均时延情况, 从中可以看出, 随着控制器数量的增加, 3 种算法的控制路径平均时延均有所下降。其中, MCEP 和 KCBP 算法在数值和下降趋势上都很接近, 而优化 K-means 算法在时延指标上相对较差。这是由于优化 K-means 算法仅考虑负载均衡, 没有考虑其他因素的约束, 而 MCEP 和 KCBP 算法将时延作为控制器部署的一个参考因素。

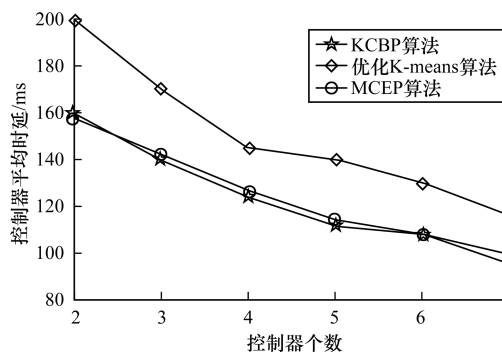


图6 3种算法控制器平均时延对比结果

图7所示为3种算法的控制器负载标准差对比情况,从中可以看出,当控制器个数较少时,与优化K-means算法相比,MCEP算法的负载均衡性相对较差,但随着控制器个数的增加,MCEP算法负载均衡性可以得到显著优化。这是由于MCEP算法在部署控制器时,会同时考虑负载均衡、时延和可靠性3个指标。如图3所示,OS3E拓扑节点分布不均匀,东北区域节点较密集。当控制器数量较少时,MCEP算法会将东北区域置于一个控制器之下,这样虽然导致负载不均衡,但是可以保证低时延和高可靠性。而随着控制器数量的增加,MCEP算法通过在密集区域部署多个控制器,可以在不降低时延和可靠性的基础上,明显减少各控制器的负载差异。

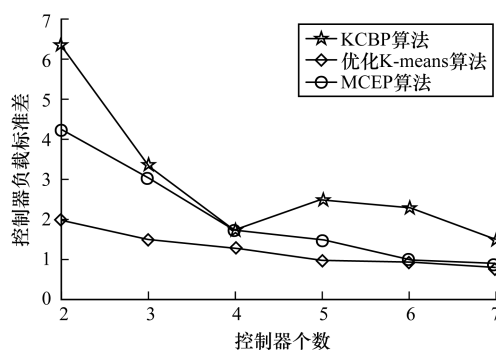


图7 3种算法控制器负载标准差对比结果

4 结束语

为提高控制路径的可靠性并优化控制器负载,本文提出控制路径可用概率、控制路径平均时延和控制器负载标准差3个度量指标,在此基础上,设计一种基于可靠性和负载优化的多控制器弹性部署算法MCEP。仿真结果表明,相比KCBP和优化K-means算法,MCEP算法能够在保证低时延和负载均衡的基础上,使控制路径的平均可靠性提高17%。今后将进一步研究如何利用谱聚类算法来提高本文多控制器部署算法的鲁棒性。

参考文献

- [1] KREUTZ D, RAMOS F M, VERISSIMO P E, et al. Software-defined networking: a comprehensive survey [J]. Proceedings of the IEEE, 2014, 103(1): 10-13.
- [2] YEGANEH S H, TOOTOONCHIAN A, GANJALI Y. On scalability of software-defined networking [J]. IEEE Communications Magazine, 2013, 51(2): 136-141.
- [3] HELLER B, SHERWOOD R, MCKEOWN N. The controller placement problem [J]. ACM SIGCOMM Computer Communication Review, 2012, 42(4): 473-478.
- [4] YAO G, BI J, LI Y, et al. On the capacitated controller placement problem in software defined networks [J]. IEEE Communications Letters, 2014, 18(8): 1339-1342.
- [5] HOCK D, HARTMANN M, GEBERT S, et al. Pareto-optimal resilient controller placement in SDN-based core networks [C]// Proceedings of the 25th International Teletraffic Congress. Washington D.C., USA: IEEE Press, 2013: 1-9.
- [6] 覃匡宇, 黄传河, 王才华, 等. SDN网络中受时延和容量限制的多控制器均衡部署 [J]. 通信学报, 2016, 37(11): 90-103.
- [7] WANG G, ZHAO Y, HUANG J, et al. A K-means-based network partition algorithm for controller placement in software defined network [C]// Proceedings of IEEE International Conference on Communications. Washington D.C., USA: IEEE Press, 2016: 1-6.
- [8] 胡延楠, 王文东, 龚向阳, 等. 软件定义网络中可靠性优化控制器的位置研究 [J]. 中国通信, 2014, 11(2): 38-54.
- [9] MÜLLER L F, OLIVEIRA R R, LUIZELLI M C, et al. Survivor: an enhanced controller placement strategy for improving SDN survivability [C]// Proceedings of 2014 IEEE Global Communications Conference. Washington D.C., USA: IEEE Press, 2014: 1909-1915.
- [10] NENCIONI G, HELVIK B E, GONZALEZ A J, et al. Impact of SDN controllers deployment on network availability [EB/OL]. [2018-02-20]. <http://people.item.ntnu.no/~gianfran/SDNctrlDep.pdf>.
- [11] GALLIER J. Geometric methods and applications: for computer science and engineering [M]. Berlin, Germany: Springer, 2011.
- [12] JOHNSON D B. Efficient algorithms for shortest paths in sparse networks [J]. Journal of the ACM, 1977, 24(1): 1-13.
- [13] KNIGHT S, NGUYEN H X, FALKNER N, et al. The interent topology zoo [J]. IEEE Journal on Selected Areas in Communications, 2011, 29(9): 1765-1775.
- [14] ZHAO Y, HE R, CHEN H, et al. Experimental performance evaluation of software defined networking (SDN) based data communication networks for large scale flexi-grid optical networks [J]. Optics Express, 2014, 22(8): 9538-9547.
- [15] ROS F J, RUIZ P M. Five nines of southbound reliability in software-defined networks [C]// Proceedings of ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking. New York, USA: ACM Press, 2014: 31-36.

编辑 吴云芳