

SDN 中基于生物启示的控制器负载均衡机制*

刘强, 张建辉, 胡涛, 赵伟



(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

通讯作者: 胡涛, E-mail: hutaondsc@163.com

摘要: 针对 SDN 多域部署中存在子域控制器负载不均衡现象, 基于生物领域物种遗传和种群迁徙的启示, 依据交换机迁移, 提出了 SDN 控制器负载均衡机制. 首先综合网络中主要开销, 应用生物遗传的思想, 设计了最优迁移域选择算法, 得到最优迁入/迁出域; 同时类比种群迁徙现象, 对交换机应用存活期和淘汰机制, 设计了交换机竞争迁移算法, 均衡了子域交换机数量. 仿真结果表明, 与现有算法相比, 迁移域的选取过程得到优化, 均衡子域控制开销, 有效保证了控制器负载的均衡分布.

关键词: 软件定义网络; 控制器; 负载均衡; 交换机迁移; 遗传算法

中文引用格式: 刘强, 张建辉, 胡涛, 赵伟. SDN 中基于生物启示的控制器负载均衡机制. 软件学报, 2017, 28(Suppl. (2)): 50-60. <http://www.jos.org.cn/1000-9825/17018.htm>

英文引用格式: Liu Q, Zhang JH, Hu T, Zhao W. Controller load balancing mechanism based on biological implications in SDN. Ruan Jian Xue Bao/Journal of Software, 2017, 28(Suppl. (2)): 50-60 (in Chinese). <http://www.jos.org.cn/1000-9825/17018.htm>

Controller Load Balancing Mechanism Based on Biological Implications in SDN

LIU Qiang ZHANG Jian-Hui HU Tao ZHAO Wei

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

Abstract: Aiming at the loading imbalance of subdomain controller on multi-domain deployment in SDN, based on genetic migrating population and the ideas of migrating switch, this paper has proposed the load balancing mechanism for multi-domain controller in SDN. It integrates main network overheads and designs the optimal migration domain selecting algorithm according to the biological genetic thought, getting the optimal immigration and emigration domain. At the same time, contrasting with the phenomenon of population migration, this paper designs the competition and migrating algorithm for switch based on survival time and elimination mechanism, which balances the number of switches for each subdomain. Compared with the existing algorithm, the simulation shows that the process of migration domain has been significantly optimized. Moreover, it balances the control overheads in subdomain and effectively ensures the balanced distribution of the controller loads.

Key words: software defined networking; controller; load balancing; switch migration; genetic algorithm

基于 OpenFlow^[1]的软件定义网络是近些年来新兴的网络技术, 该技术的主要特点是: 转发平面和控制平面实现了完全解耦, 控制平面采用集中式控制, 且具有灵活的可编程能力. 然而, 控制器的集中式控制使得网络可扩展性和灵活性降低. 因此研究者相继提出了一个逻辑上集中, 但物理上分布的多控制器部署策略, 例如, HyperFlow^[2], Kandoo^[3]和 Onix^[4]等人的方案, 将 SDN 网络划分成多个不相交的子域, 每个子域由独立的主控制器进行控制, 负责管理子域内所有交换机.

尽管控制器的多域部署解决了 SDN 网络可扩展性和灵活性问题, 但是由于流量传输在时间和空间上具有

* 基金项目: 国家网络空间安全专项(2017YFB0803204); 国家高技术研究发展计划(863)(2015AA016102); 国家自然科学基金(61521003)

Foundation item: National Network Space Safety Subject (2017YFB0803204); National High-Tech R&D Program of China (863) (2015AA016102); National Natural Science Foundation of China (61521003)

收稿时间: 2017-06-30; 采用时间: 2017-10-20

不均匀性,在某个时间段或者某个域内会出现流量的激增或者衰减.控制器间流量负载差异较大则很容易导致某些域控制器出现过载或轻载的状况,不利于整个网络的架构稳定和控制器负载均衡.

近年来,针对 SDN 多域网络中控制器负载不均衡问题,主要存在以下研究方案.Kandoo 等人^[3]提出了一种两层控制器架构,在底层,控制器群并不进行相互通信,上层采用一个逻辑上集中的控制器来维持全局网络状态.两层间的控制器需要不断进行交互,整个网络的控制开销较大.Eugen 等人^[5]提出了多控制器的部署优化方案,在该方案中,控制器问题的重点着重于控制器负载均衡和控制器间时延等因素,权衡了各个互相矛盾的优化目标,但是对于交换机的数量和部署未作要求.在利用率最低迁移策略中,负载均衡算法(load balancing algorithm,简称 LBA)^[6]的思想是将交换机从高负载控制器迁移至负载最低的控制器以实现负载均衡.Fu 等人^[7]提出了一种控制器休眠模型,在轻流量状态下允许闲置的控制器进入休眠状态以节省资源,利用数学理论定量分析系统性能,构造了全局消耗函数,寻求最优解.但在该方案下,需要部署一定数量的冗余控制器,难免会使网络的代价增大.Koerner 等人^[8]研究了一种负载均衡服务器,利用多个 Openflow 控制器处理多服务器的负载.例如,一个控制器处理邮件服务器负载,其余的处理网络服务器负载.提高了控制器的使用效率.但是设备的增加也带来了可扩展性问题.Lin 等人^[9]提出了一种高度可扩展性架构 MSDN,通过对初始流量进行分类,在整个网络视图下对于流量进行并行处理,利用负载均衡的思想解决了控制平面和数据平面间的静态配置问题.基于 OpenFlow 1.3, Dixit 等人^[10]提出了一种动态交换机迁移协议来确保网络的活跃度和安全性,同时产生较少的控制器负载.

综合比较上述研究方案,主要存在以下不足,迁入/迁出域的选择容易陷入局部最优,未考虑整体性,部署过多的控制器会增加网络开销,同时交换机迁移机制也未考虑迁移门限的动态变化和过多交换机迁移带来的二次迁移问题.

针对上述问题,本文基于生物领域内物种遗传和种群迁徙的启示,提出了一种新的控制器负载均衡机制.首先对各子域内控制器的负载进行测量,将测量数据回传到数据中心,并设置控制器过载判定门限值,判定各 SDN 子域是否存在过载控制器;然后根据子域控制器负载情况进行迁移域(包括迁入/迁出域)选取.由于迁入/迁出域的选择过程中需要权衡多种因素(控制器负载,传输时延,通信流量等).因此本文融合生物遗传的思想,对备选域进行选择、交叉、变异等操作,设计了最优迁移域选择算法,从而在 SDN 网络的众多可行域中选择最优迁入/迁出域.同时将生物种群中的迁徙现象映射到交换机迁移过程中,类比于物种特征,对交换机设定存活期和淘汰机制,应用迁移率对迁移交换机进行量化,设计了交换机竞争迁移算法,将迁出域内高负载度交换机按一定的选取规则迁移至迁入域,从而实现各个子域控制器的负载均衡,提升了网络的稳定性.

本文的主要贡献和创新工作总结如下:

- (1) 融合生物遗传思想,设计了最优迁移域选择算法;
- (2) 类比种群迁徙现象,引入交换机存活期和淘汰机制,设计了交换机竞争迁移算法;
- (3) 动态设定迁移率来优化迁移交换机数目.

本文第 1 节建立 SDN 多控制器负载均衡模型.第 2 节设计相应的实施方案.第 3 节对本文所提机制的性能进行评估.第 4 节对全文进行总结,描述下一步的研究工作.

1 模型构建

SDN 网络的多域部署拓扑如图 1 所示,通过将 SDN 网络划分为多个子域,每个子域由特定的控制器实施控制转发,控制器间彼此交互网络状态和链路信息,从而做出转发决策^[11],且各个子域之间进行互联,有效增强了网络的可扩展性和可靠性.

然而扁平式分布架构^[12]设计会造成各个 SDN 子域内控制器负载不均衡,因此,本文在原有底层物理链路的基础上,对上层控制器的功能进行设定,增加了控制器过载判定模块、信息存储模块和交换机迁移模块,完成对迁出域内高负载度交换机的迁移任务.为了实现所设计的功能,对 SDN 子域中控制器进行改造,增加相应的实施模块,如图 2 所示.

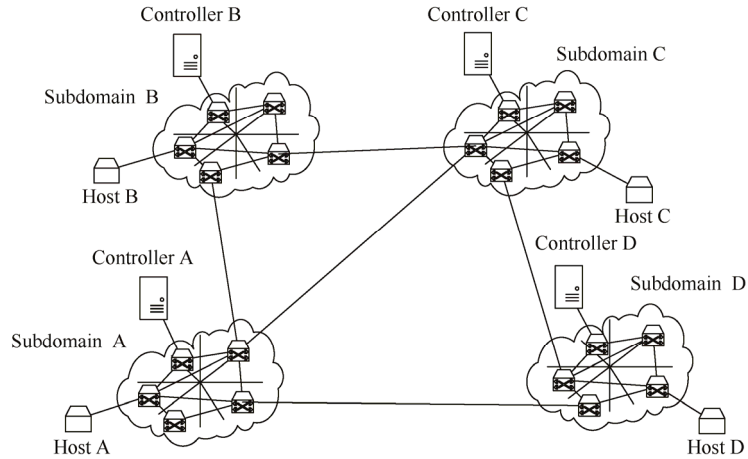


Fig.1 SDN multi-domain network topology

图 1 SDN 多域网络拓扑

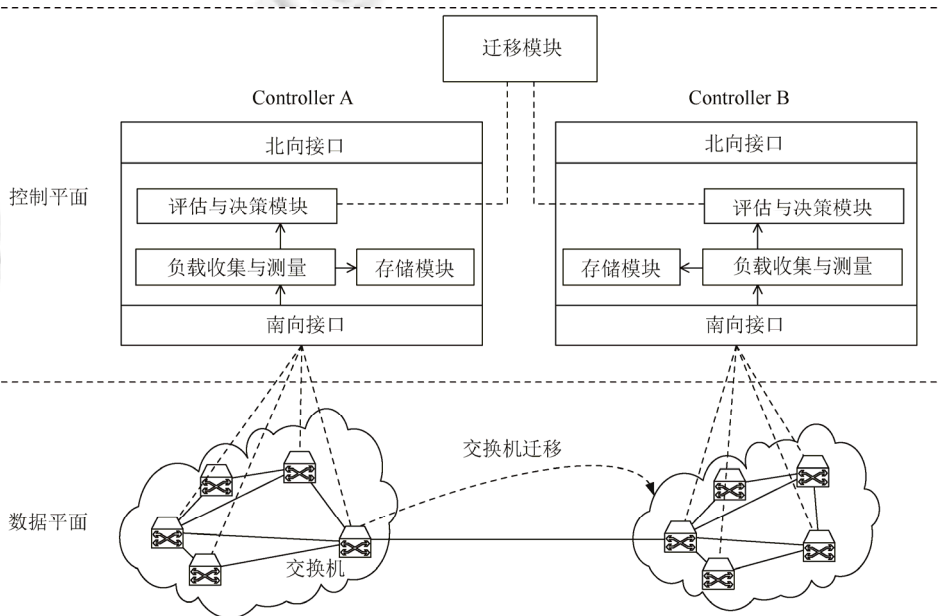


Fig.2 Load balancing architecture towards distributed SDN controller

图 2 分布式 SDN 控制器负载均衡架构

SDN 子域的数据平面由多个交换机通过物理链路连接构成.控制器部分主要包括:

- (1) 负载收集与测量模块,负责收集控制器负载信息和链路状态信息;
- (2) 评估与决策模块,负责评估控制器负载是否超过预设的门限值,根据各控制器负载情况进行决策;
- (3) 存储模块、存储子域内所有交换机的链路消息、拓扑消息和流量消息;

(4) 迁移模块位于整个网络的顶端,网络中的各个控制器都与迁移模块相连接,负责协调交换机在各 SDN 子域内进行有效迁移.

根据各个模块之间的功能,本文设计了如图 3 所示的流程图来说明该负载均衡机制的实现过程.

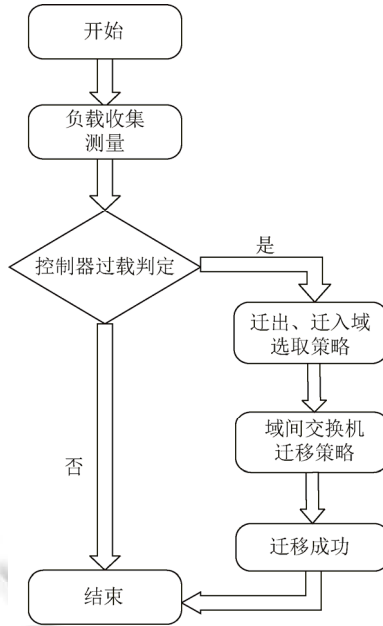


Fig.3 The flow chart of controller load balancing
图 3 控制器负载均衡流程图

2 设计实施

2.1 负载收集与测量

2.1.1 控制器负载收集

在 SDN 网络中,控制器的负载主要由以下 4 个部分构成:(1) 待处理的 Packet_In 事件数(P);(2) 维护其管理域和全局拓扑所需的流量(F);(3) 控制器间通信开销(K);(4) 安装上层应用产生的 Flow Entry(N).每个部分在不同控制器中所占的权值不同,设定对应的权值分别为 u_1, u_2, u_3 和 u_4 .因此,控制器负载的计算公式为

$$C = u_1 \cdot P + u_2 \cdot F + u_3 \cdot K + u_4 \cdot N \quad (1)$$

$$\sum_{i=1}^4 u_i = 1 \quad (2)$$

2.1.2 交换机负载度测量

SDN 多域部署将整个网络划分为多个子域,设整个网络拓扑为 $G = (V, E)$,其中 V 代表网络中的交换机, E 代表交换机之间的链路.将网络划分为 N 个子域,子域可表示为 $S_i(W_i, L_i)$, W_i 表示子域 S_i 中的交换机, L_i 表示 S_i 中的链路.

$$G = \bigcup_{i=1}^N S_i \quad (3)$$

定义 1. 交换机负载度.交换机向控制器请求的 Packet_In 事件数定义为交换机的负载度,设为 f_k .

SDN 子域 S_i 中交换机的个数为 M_i ,所有交换机的负载度总和 $Load_{S_i}$ 为

$$Load_{S_i} = \sum_{k=1}^{M_i} f_k \quad (4)$$

则该 SDN 子域 S_i 的平均负载度 $f_{avg}^{S_i}$ 为

$$f_{avg}^{S_i} = \sum_{k=1}^{M_i} f_k / M_i \quad (5)$$

在整个 SDN 域中,共有 N 个子域,则子域 S_i 的相对平均负载度 $F_{avg}^{S_i}$ 为

$$F_{avg}^{S_i} = f_{avg}^{S_i} / \sum_{i=1}^N f_{avg}^{S_i} \quad (6)$$

2.2 控制器过载判定

针对流量动态特性,本文设计了一种控制器过载判定机制,设置动态负载门限值为 L_A ,子域 S_i 中控制器负载值为 C_i ,判定过载门限值为 ε ,则 L_A 的计算公式为

$$L_A = \begin{cases} \varepsilon & \exists C_i \leq \varepsilon \\ \sum_{i=1}^N C_i / N & \forall C_i > \varepsilon \end{cases} \quad (7)$$

2.3 选择与迁移策略

选择与迁移策略主要包含:(1) 融合生物遗传思想的最优迁移域选择算法(optimal migration domain selecting algorithm,简称 OMDSA);(2) 类比种群迁徙的交换机竞争迁移算法(competition migration algorithm,简称 CMA).

2.3.1 最优迁移域选择算法

最优迁移域选择算法是基于生物遗传的思想,将交换机迁移域的选择问题转化为多因素下求解最优可行解.多因素主要包括 SDN 子域的控制器的负载 C ,子域间传输时延 D ,交换机和控制器的通信流量 T 类主要影响因素.在这里,设定所求解的目标函数为 Q ,其中 w 是 3 种代价的权值,见公式(9).

$$Q = w_1 \cdot C + w_2 \cdot D + w_3 \cdot T \quad (8)$$

$$\sum_{i=1}^3 w_i = 1 \quad (9)$$

算法的流程是,首先检查各子域网络中控制器负载值,当子域控制器负载超出设定门限值时,遍历搜索整个网络中的子域,考虑控制器负载、传输时延和通信流量等因素,选择合适的子域作为可行域,然后基于可行域进行选择、交叉、变异等遗传操作,以目标函数作为适应度函数,最终得到最优迁入/迁出域.算法伪代码见算法 1.

算法 1. 最优迁移域选择算法.

输入:SDN 子域: $\{S_1, S_2, \dots, S_N\}$ (N 为子域的个数)

控制器负载: C

目标函数: Q

迭代次数: I

输出:迁出域: S_h

迁入域: S_l

无须选择迁入/迁出域:0

1: 计算 SDN 子域的平均负载度 $\{f_{avg}^{S_1}, f_{avg}^{S_2}, f_{avg}^{S_3}, \dots, f_{avg}^{S_N}\}$

2: **IF** $f_{avg}^{S_h} = \text{Max}\{f_{avg}^{S_1}, f_{avg}^{S_2}, f_{avg}^{S_3}, \dots, f_{avg}^{S_N}\}$

3: **THEN** 选择 $f_{avg}^{S_h}$ 所在的 SDN 子域作为迁出域 S_h

4: **ELSE RETURN 0**

5: **ENDIF**

6: S_h 的相邻子域个数为 r ($r \leq N$);

7: 相邻子域集合为 $A = \{S_{h-1}, S_{h-2}, S_{h-3}, \dots, S_{h-r}\}$ $A \neq \emptyset$

8: $x=x+1$

9: 将目标函数 Q 作为适应度函数,生成初始群体

10: 计算 S_h 相邻子域的适应度 q_i ,得到 q_{avg} 和 q_{max}

11:Parent 1:Select $S_{h-\alpha}$ $\alpha \in \{1,2,3,\dots,r\}$ $q_\alpha \leq q_{avg}$

Parent 2:Select $S_{h-\beta}$ $\beta \in \{1,2,3,\dots,r\}$ $q_\beta > q_{avg}$

12:BEGIN Crossover

IF $I < x/r$

THEN $P_c = a$ (a 为固定交叉概率)

ELSE $I \geq x/r$

THEN $P_c = \begin{cases} k_1, & q_c \leq q_{avg} \\ k_2(q_{max} - q_c), & q_c \geq q_{avg} \end{cases}$ (q_c 交叉个体适应度)

13:BEGIN Mutation

IF $I < x/r$

THEN $P_m = b$ (b 为固定变异概率)

ELSE $I \geq x/r$

THEN $P_m = \begin{cases} k_3, & q_m \leq q_{avg} \\ k_4(q_{max} - q_m), & q_m \geq q_{avg} \end{cases}$ (q_m 变异个体适应度)

14:计算交叉和变异产生新个体的适应度,构成新一代群体

15:IF $x=I$ 解码得到 Q_{min} ,

16:选择具有 Q_{min} 的子域作为迁入域 S_i

17:ELSE RETURN 6

OMDSA 算法的时间复杂度分析:行 1 计算 SDN 子域的平均负载度,其时间复杂度为 $O(N)$;行 2~行 4 选择迁出域,其时间复杂度为 $O(N)$;行 6~行 8 得到迁出域的相邻子域,其时间复杂度为 $O(r)$;行 12~行 16 通过遗传算法得到迁入域,其时间复杂度为 $O(N \cdot I)$.因此,OMDSA 总的时间复杂度为 $O(N \cdot I)$.由于子域和控制器数量有限,因此 OMDSA 的算法复杂度较低,且皆为简单线性运算,算法快速收敛.

2.3.2 交换机竞争迁移算法

基于最优迁移域选择算法,得到实施交换机迁移的最优迁入/迁出域.接下来要在迁移域内完成交换机的迁移任务.本文基于种群迁徙的启示,将交换机迁移类比为生物种群迁徙现象,迁移域类似于种群迁徙过程中的迁徙域,交换机类似于迁徙过程中的生物个体.在迁徙过程中,生物个体具有一定的生命周期,且迁徙数量约束遵循物竞天择、适者生存的淘汰法则.因此,对迁移域内交换机也设定存活期和淘汰机制,设计了交换机竞争迁移算法.

首先对本节中迁移率、存活期和淘汰机制的概念进行介绍.

(1) 迁移率

迁移率是对子域迁入/迁出能力的定量描述.通过计算与比较迁移率可以得出迁移交换机的最佳数量.

子域 S_p 和 S_q 的相对平均负载度分别为

$$F_{avg}^{S_p} = f_{avg}^{S_p} / \sum_{i=1}^N f_{avg}^{S_i} \quad (10)$$

$$F_{avg}^{S_q} = f_{avg}^{S_q} / \sum_{i=1}^N f_{avg}^{S_i} \quad (11)$$

若 $F_{avg}^{S_p} < F_{avg}^{S_q}$,则迁移率

$$M(S_p, S_q) = u(F_{avg}^{S_p} - F_{avg}^{S_q}) + v \quad (12)$$

式(12)中两个子域,若一个子域的平均负载度为 0 时, $M(p, q) = 1$ 需要满足条件 $u + v = 1$.

若 $F_{avg}^{S_p} \geq F_{avg}^{S_q}$, 则迁移率

$$M(S_q, S_p) = d(F_{avg}^{S_q} - F_{avg}^{S_p}) + e \quad (13)$$

式(13)中两个子域,若一个平均负载度为 0,另一个为极值时, $M(q, p) = 0$ 需要满足条件 $d + e = 0$.

(2) 存活期和淘汰机制

类比生物种群中的迁徙现象,本文对迁移域内的交换机设存活期和淘汰机制,即对参与迁移的交换机计算其存活期,若迁入个体年龄大于它的存活期就要被淘汰,不再参与迁移过程.

存活期基于两个原则:

- (1) 低负载度交换机的存活期大于高负载度交换机的存活期;
- (2) 当某个 SDN 子域的交换机规模过大时不再接受新的交换机.

在此,为交换机设定两个参数:年龄 $Y(X_j)$ 和寿命 $L(X_j)$. 交换机 X_j 每经过一次迁移流程,称为一代,其年龄加 1,其存活的代数不能超过其存活期 $L(X_j)$. 为了防止子域交换机的数量过多或过少,在这里引入规模控制门限值 R , $f(X_j)$ 为交换机的负载度函数值,当前域的最大、最小和平均负载度为 $f_{max}, f_{min}, f_{avg}$, 设置 SDN 子域中交换机的最大数目为 M_{max} . 规模控制门限值 R 为

$$R = \begin{cases} 0, & size \geq M_{max} \\ 1, & size < M_{max} \end{cases} \quad (14)$$

$$L(X_j) = \min \left[\frac{1}{2} (f_{max} - f_{min}) \frac{f(X_j)}{f_{avg}}, \max LT \right] \cdot R \quad (15)$$

其中, $\min LT$ 是允许最小寿命, $\max LT$ 是允许最大寿命.

算法的流程主要是:首先计算迁入/迁出两个 SDN 子域的相对平均负载度,得到对应迁移率,根据迁移率确定迁移交换机的数目. 然后在迁移过程中,对迁移交换机加入存活期和淘汰机制,选择符合迁移条件的交换机实施迁移. 最后根据交换机迁移情况,构建新的 SDN 网络拓扑,见算法 2.

算法 2. 交换机竞争迁移算法(CMA).

输入: 迁出的 SDN 子域: S_h

 迁入的 SDN 子域: S_l

 年龄计数器 t

 可供迁移的交换机数量: λ

S_h 域内交换机数量: M

 迁移成功的交换机数量: SUM

输出: 新的 SDN 多域控制器网络拓扑没有可供选择的交换机: 0

1: 计算 S_h 域内交换机 $X_h, h \in \{1, 2, 3, \dots, M\}$ 的负载度 f_h ,

$$\Omega = \{f_{1-h}, f_{2-h}, \dots, f_{M-h}\} \quad (f_{1-h} \geq f_{2-h} \geq \dots \geq f_{M-h})$$

2: 计算 S_h, S_l 两个域的迁移率 $M(h, l)$ 和 $M(l, h)$

3: **IF** $M(h, l) - M(l, h) \geq \delta$

4: **THEN** $\lambda = INT[M(h, l) / M(l, h)]$

5: **ELSE RETURN 0**

6: **ENDIF**

7: 在集合 Ω 中选取前 λ 个交换机记为集合 Γ

$$\Gamma = \{X_{1-h}, X_{2-h}, \dots, X_{\lambda-h}\} \quad 0 < \lambda < M$$

8: 计算集合 Γ 内存活期 $L(X_j), j \in \{1-h, \dots, \lambda-h\}$

9: **IF** $t < L(X_j)$

10: **THEN** 放弃迁移交换机 X_j

11: **ELSE** $t \geq L(X_j)$ 成功迁移交换机 X_j

```

12:  $t=t+1$ 
13:ENDIF
14:IF  $SUM < \lambda$ 
15:THEN Return 9
16:    ELSE  $SUM \geq \lambda$  结束迁移
17:ENDIF
    
```

CMA 算法的时间复杂度分析如下:行 1 计算迁出域交换机的负载度,其时间复杂度为 $O(M)$;行 2~行 6 进行迁移率比较,其时间复杂度为 $O(2M)$;行 7~行 8 计算存活期,其时间复杂度为 $O(\lambda)$;行 9~行 17 执行交换机迁移任务,其时间复杂度为 $O(\lambda)$.因此,CMA 的总体时间复杂度为 $O(M)$.由于迁移交换机数量有限,因此 CMA 具有快速收敛的特性.

3 性能评估

本文选择 OpenDaylight^[13]作为实验控制器,OpenDaylight 采用 Java 语言编写,运行在 JVM 上,它提供了开放的北向应用接口,同时南向支持包括 OpenFlow 在内的多种 SDN 协议.底层同时支持混合模式的交换机和经典的 OpenFlow 交换机,并提供了一个模块化的开放 SDN 应用层.在此基础上通过编程添加了收集与测量模块、评估决策模块和存储模块,并编写了迁移模块负责整个域内交换机的迁移.同时为了使实验更具说服力,实验中的拓扑模型取自 Internet 2 OS3E^[14].OS3E 网络是对美国的实际骨干网络的抽象,由美国多所大学、研究机构和公司为建设下一代互联网共同创建.

实验环境设定为:(1) 各个子域内控制器的容量和流处理能力基本相同,在这里设定控制器的容量为 10M,平均流产生速率为 300KB/S;(2) 迁移过程中,不会出现链路失效问题;(3) 网络中各个子域不可能同时都处于高负载状态.

3.1 实验1

当 SDN 网络中出现控制器过载时,可以选择合适的迁移域来实施过载交换机迁移.对于如何进行最优迁移域选取,基于 OS3E 网络,应用本文所设计的最优迁移域选择算法和距离域最短算法(nearest domain algorithm,简称 NDA)^[15],负载均衡算法(load balancing algorithm,简称 LBA)^[6]进行对比实验,比较 3 种算法的迁移域选择性能.OS3E 网络共有 34 个节点,42 条链路,利用 K-Center AP 聚类算法^[16]对网络中节点进行聚类划分,划分为 3 个子域,不同子域的交换机分别用不同颜色进行表示,且数量基本一致,其中用正方形节点表示控制器位置,如图 4 所示.

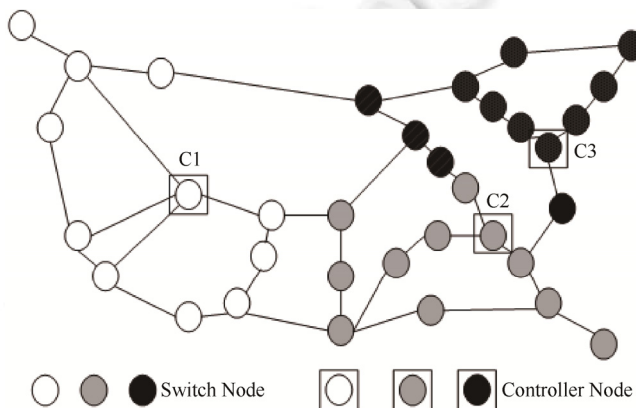


Fig.4 OS3E network topology and subdomain partition

图4 OS3E 网络拓扑及子域划分

在本实验中,逐步增加注入网络流量的速率,变化范围是控制器总容量的 1%~90%,观察各个子域的网络开销随输入流量的变化情况,如图 5 所示。

由图 5 可以看出,总体趋势是网络总开销随着输入流量的增加呈非线性增长.当输入流量较低时,3 类算法得到的网络总开销基本相同,但是随着输入流量的继续增加,3 类算法在不同子域得到的开销峰值相差较大.当输入流量在 3 个子域同时达到 90%时,NDA,LBA 和 OMDSA 算法得到的各个子域总开销统计值见表 1.

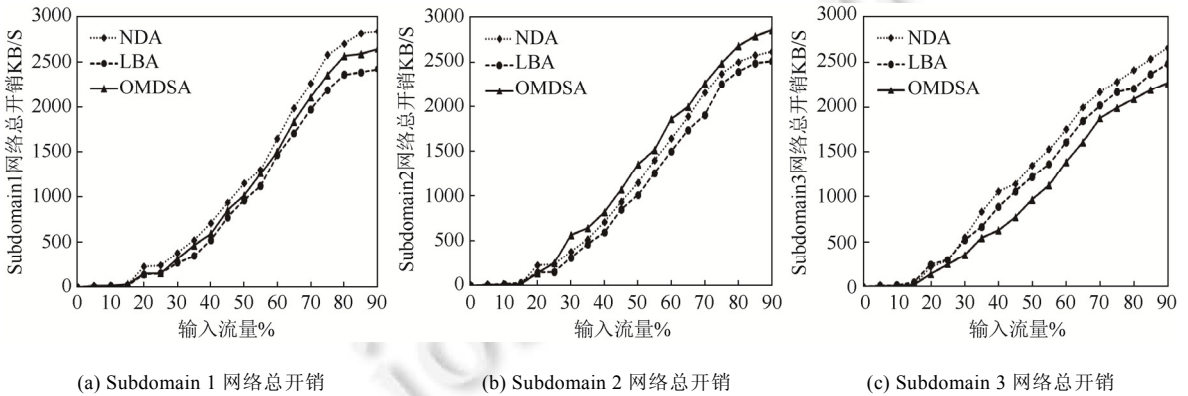


Fig.5 The total network overhead of subdomains

图 5 子域网络总开销

Table 1 The subdomain overhead when the input traffic reaches 90% (KB/S)

表 1 输入流量为 90%时的子域开销

名称	NDA	LBA	OMDSA
Subdomain 1	2 847	2 433	2 660
Subdomain 2	2 630	2 553	2 869
Subdomain 3	2 665	2 469	2 277
最大子域开销差	217	120	592

对表 1 中的结果进行分析,由于实验初始条件设定 Subdomain 1,Subdomain 2,Subdomain 3 包含的交换机数量相同,因此,在子域注入同等流量的情况下,当应用距离域最短算法时,由于 Subdomain 1 节点之间的距离较大,其余两个子域节点较为接近,因此 Subdomain 1 网络总开销较大,Subdomain 2 和 Subdomain 3 的开销基本相同,但仍维持较高水平.负载均衡算法得到的 3 个子域开销较为平均,且数值较小,这是由于该算法在流量的传输过程中使用了负载均衡策略,平衡了各子域的开销.应用本文设计的最优迁移域选择算法,由于该算法考虑到控制器负载,域间传输时延和通信流量 3 类参数,根据子域条件的不同自适应设置参数权值,并对权值进行优化.在 3 类算法中,子域开销的最大差值分别为 217KB/S,120KB/S,592KB/S.由于最优迁移域的选择与域间开销差值有关,开销差值越大,则交换机迁移的效果越好.因此,相对于 NDA 和 LBA,OMDSA 能够在全局网络中选出具有最大开销差值的两个子域作为交换机迁移的迁入/迁出域,实现了最优迁移域的选取。

3.2 实验2

当迁入/迁出域确定之后,就可以实施交换机迁移.为了验证本文所提出的交换机竞争迁移算法,继续使用实验 1 中 OS3E 拓扑进行实验,由于 Subdomain 2 的网络总开销最大,超出了控制器过载判定门限值(2700KB/S),因此将该子域设为迁出域,同时将总开销最小的 Subdomain 3 作为迁入域.在这里,交换机竞争迁移算法、交换机完全迁移算法(FMA)和文献[8]中提出的弹性迁移算法(EMA)进行实验对比.子域控制器负载随时间变化的结果如图 6 所示.在图 6(a)中,使用完全迁移算法将过载交换机从迁出域迁移至迁入域中,由于完全迁移算法对迁移交换机的数量未作限制,可以看出,在迁移过程中虽然降低了迁出域控制器开销但容易导致迁入域控制器负载再度升高,迁入/迁出域控制器负载值相差较大,整个网络仍处于负载不均衡状态.在图 6(b)中,使用弹性迁移算法,根据迁移域的状态分阶段迁移交换机,图中两个曲线较为接近且上下波动,这是由于弹性策略中分段迁移交换

机造成的,不利于网络状态的稳定.相对于以上两种算法,在图 6(c)中,使用交换机竞争迁移算法得到迁入/迁出域的控制开销较为接近,迁移完成后仍处于稳定状态,且都低于控制器过载门限值 2700KB/S,这是由于该算法依据控制器负载情况对交换机的迁移数量进行动态设定,利用淘汰机制选择处于存活期内的交换机进行动态迁移,有效地避免了迁移过程中的控制器负载波动.

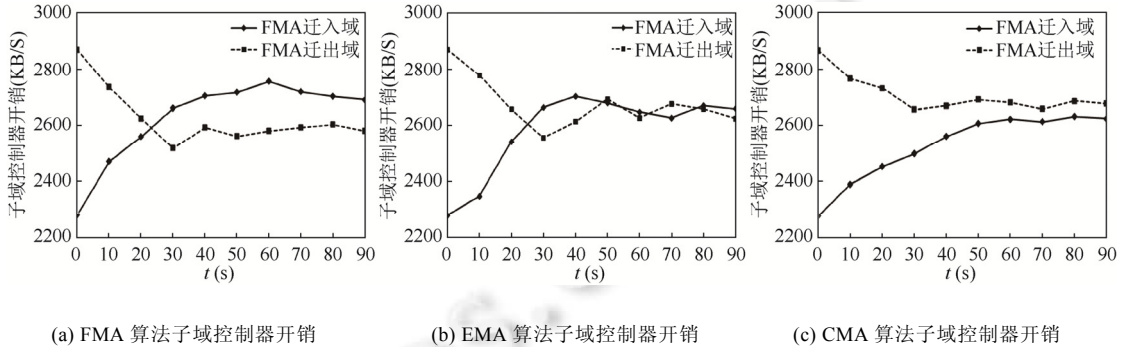


Fig.6 Subdomain controller overhead changes over time

图 6 子域控制器开销随时间变化

4 结论

本文针对 SDN 多域部署下各子域控制器的负载不均衡问题,基于生物领域物种遗传和种群迁徙的启示,依据交换机迁移思想,对 SDN 多域控制器负载均衡机制进行研究,包含了最优迁移域选择算法和交换机竞争迁移算法.在迁移域选择方面,最优迁移域选择算法应用生物遗传的思想,综合网络中控制器负载、传输时延、通信流量等因素,选择出最优迁入/迁出域.在进行交换机迁移过程中,类比种群迁徙现象,交换机竞争迁移算法对交换机应用存活期和淘汰机制,自适应实施交换机迁移.仿真结果表明,与现有算法相比,可以实现最优迁移域的选取,有效地均衡了各个子域控制器负载和交换机数目,增强网络的稳定性和可靠性.

下一步研究工作主要围绕以下 3 方面内容进行展开:(1) 根据启发式算法对控制器所能承载交换机的数量进行优化;(2) 优化多控制器的部署位置;(3) 考虑节点和链路失效情况.

References:

- [1] McKeown N, Anderson T, Balakrishnan H, Parulkar G, Peterson L, Rexford J, Shenker S, Turner J. OpenFlow: Enabling innovation in campus networks. *ACM Sigcomm Computer Communication Review*, 2008,38(2):69-74.
- [2] Tootoonchian A, Ganjali Y. HyperFlow: A distributed control plane for OpenFlow. In: *Proc. of the 2010 Internet Network Management Workshop on Research on Enterprise Networking*. Berkeley: USENIX Association, 2010.
- [3] Hassas Y S, Ganjali Y. Kandoo: A framework for efficient and scalable offloading of control applications. In: *Proc. of the 1st Workshop on Hot Topics in Software Defined Networks*. 2012. 19-24.
- [4] Koponen T, Casado M, Gude N, Stribling J, Poutievski L, Zhu M, Ramanathan R, Iwata Y, Inoue H, Hama T, Shenker S. Onix: A distributed control platform for large-scale production networks. In: *Proc. of the Usenix Conf. on Computer Systems Design and Implementation*. 2010. 351-364.
- [5] Borcoci E, Badea R, Obreja SG, Vochin M. On multi-controller placement optimization in software defined networking based WANs. In: *Proc. of the 14th Int'l Conf. on Networks (ICN)*. 2015. 261-266.
- [6] Yao G, Bi J, Li YL, Guo LY. On the capacitated controller placement problem in software defined networks. *IEEE Communications Letters*, 2014. 1339-1342.
- [7] Fu YH, Bi J, Wu JP, Chen Z, Wang K, Luo M. A dormant multi-controller model for software defined networking. *China Communications*, 2014,11(3):45-55.

- [8] Koerner M, Kao O. Multiple service load-balancing with OpenFlow. In: Proc. of the 13th Int'l Conf. on High Performance Switching and Routing. 2012. 210–214.
- [9] Lin PP, Bi J, Hu HY, Jiang XK. MSDN: A mechanism for scalable intra-domain control plane in SDN. Journal of Chinese Computer Systems, 2013,34(9):1969–1974 (in Chinese with English abstract).
- [10] Dixit A, Hao F, Mukherjee S, Lakshman TV, Kompella R. Towards an elastic distributed SDN controller. In: Proc. of the ACM HotSDN. 2013.
- [11] Cello M, Marchese M, Mongelli M. On the QoS estimation in an OpenFlow network: The packet loss case. IEEE Communications Letters, 2016,20(3):554–557.
- [12] Chen G, Pei D, Zhao Y. CQRD: A switch-based approach to flow interference in data center networks. Local Computer Networks. IEEE, 2014. 107–115.
- [13] OpenDaylight. <https://www.opendaylight.org/>
- [14] Internet 2. <https://www.internet2.edu/>
- [15] Hock D, Gebert S, Hartmann M, Zinner T, Tran-Gia P. Pareto-Optimal resilient controller placement in SDN-Based core networks. In: Proc. of the Network Operations and Management Symp. 2014. 1–2.
- [16] Eu JH, Kim YH, Chung CS, Park SJ. USN-Based architecture for traffic management. In: Proc. of Int'l Conf. on Advanced Communication Technology (ICACT). 2010. 1401–1406.

附中文参考文献:

- [9] 林萍萍,毕军,胡虹雨,蒋小可.一种面向 SDN 域内控制平面可扩展性的机制.小型微型计算机系统,2013,34(9):1969–1974.



刘强(1988—),男,河南新乡人,助教,主要研究领域为软件定义网络,网络协议.



胡涛(1993—),男,硕士生,主要研究领域为软件定义网络,网络协议.



张建辉(1977—),男,博士,副研究员,主要研究领域为宽带信息网,软件定义网络,网络安全.



赵伟(1990—),男,硕士生,主要研究领域为软件定义网络,网络协议.