

软件定义光网络中基于最小点覆盖的控制平面跨层生存性设计

熊余* 董先存 李圆圆 吕翊 王汝言
(重庆邮电大学光通信与网络重点实验室 重庆 400065)

摘要: 为降低软件定义光网络对单控制器的依赖, 并避免多控制器冲突, 有效提升控制平面的生存性, 该文提出基于最小点覆盖的控制平面生存性设计策略。该策略结合集中控制约束条件, 以最小点覆盖理论为基础, 建立可靠的分级管控模型, 设定控制器的管控优先级: 全局控制器具有最高管控优先级, 对全网进行集中管控; 本地控制器次之, 只对本地业务进行集中管控; 权威交换机的管控优先级最低, 用于完成局部波长粒度的光层快速管控。在此基础上, 基于跨层信息模型为控制信道路由和资源分配进行生存性冗余设计。仿真表明, 该策略能够满足网络对控制时延的要求, 使控制平面的故障概率降低了30%, 有效提升了网络在恶劣环境下的生存性。

关键词: 软件定义光网络; 生存性; 控制平面; 分级管控; 最小点覆盖

中图分类号: TN915.07

文献标识码: A

文章编号: 1009-5896(2016)05-1211-08

DOI: 10.11999/JEIT150815

The Cross-layer Survivable Design of Control Plane Based on Minimum Point Covering in Software Defined Optical Network

XIONG Yu DONG Xiancun LI Yuanyuan LÜ Yi WANG Ruyan
(Key Laboratory of Optical Communication and Networks, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: In order to lower the reliance on single controller in Software Defined Optical Network (SDON), avoid the conflict of different controllers, and improve effectively the survivability on the control plane, the survivable design of SDON control plane based on minimum point covering is proposed. Combined with the constraint of centralized control, the algorithm based on the minimum point covering establishes reliable hierarchical control model, and sets control priority to controllers. The master controller with the highest priority centralized controls the whole net, and the next comes the regional controller, which only intensively controls regional traffic, besides the authority switch which is introduced to control local wavelength in the optical layer owns the lowest level. Meanwhile, the model based on cross-layer information designs survivability redundancy for routing and allocates resource for the control channel. Simulation results show that the proposed strategy can satisfy the request on control delays, and lower the failure probability in the control plane by 30%, thus promoting the network survivability under critical environment.

Key words: Software Defined Optical Network (SDON); Survivability; Control plane; Hierarchical control; Minimum point covering

1 引言

随着网络规模的不断扩大和业务种类的不断丰富, 扮演“大动脉”角色的光网络基础设施面临着前所未有的挑战: 一方面要求网络传输容量变得更

大; 另一方面要求对网络的管控更加地动态、灵活。为了达到对大容量光网络基础设施灵活管控的目的, 软件定义光网络(Software Defined Optical Network, SDON)应运而生, 它是继基于通用多协议标签交换(Generalized MultiProtocol Label Switching, GMPLS)和基于路径计算单元(Path Computation Element, PCE)光网络之后的第3代智能光网络^[1]。SDON是软件定义网络(Software-Defined Networking, SDN)技术在光网络控制层面上的具体应用。网络的节点被分为两类, 一类为哑的转发设备, 专注于转发; 另一类为软件驱动的控制平面, 专注于控制, 它动态地管控光调制、光层路由、波长分配、波长转换等光层任务, 从而实现大

收稿日期: 2015-07-08; 改回日期: 2015-12-25; 网络出版: 2016-02-26

*通信作者: 熊余 xiongyu@cqupt.edu.cn

基金项目: 国家自然科学基金(61401052), 重庆市教委科学技术研究项目(KJ1400418, KJ1500445), 重庆邮电大学博士启动基金(A2015-09)

Foundation Items: The National Natural Science Foundation of China (61401052), The Science and Technology Project of Chongqing Municipal Education Commission (KJ1400418, KJ1500445), The Ph. D. Start-up Fund of Chongqing University of Posts and Telecommunications (A2015-09)

规模光网络中海量业务的灵活动态资源调度。

SDON 将快速、高效、基于包的分布式本地转发机制和全局、最优、基于流的集中式控制决策机制完美地结合起来,网络的逻辑控制和数据的转发分离是其主要特征^[2,3]。故 SDON 的控制平面承担着网络状态信息收集、路由决策、光通路动态建立与拆除等管控任务,并且为了快速传输控制信令,在转发设备和控制器之间特别增加了由控制信道组成的拓扑。可见,SDON 的控制平面由于面向全网进行集中调度,其承载了大量重要的信令业务,任何一条控制信道的故障都可能造成大面积的数据平面失去转发能力,从而带来无法估量的损失。

为此,基于文献[4,5]提出的 SANE 和 Ethane 两种集中控制网络模型,在原来网络的链路层和 IP 层之间加入了专有保护层,以专门负责信令分发。在小型网络情况,控制平面的生存性得到了很好的保障,但随着网络规模的扩大,性能便趋于恶化。为此,文献[6]中提出具有划时代意义的 OpenFlow 技术来解决对大规模网络集中管控的问题,最新版本 OpenFlow 规约(v1.3.4)^[7]指出 SDON 中应对控制信道故障的算法一般是 1+1 保护算法和快速重路由恢复算法,但这两种算法存在冗余资源预留大或故障恢复延时长的问题。为了降低保护冗余,文献[8]通过对 256 种网络拓扑控制平面的分析,证明在网络中尽管只部署了一个控制器来集中管控整个网络,在合理安排控制器位置的情况下,依然能满足对控制时延的要求,并且能有效降低资源开销,但这造成了网络的运行对该单控制器过度依赖。为了减轻网络运行对单控制器的依赖,文献[9]在大规模网络中部署了多个控制器,控制器之间相互提供保护。但文中并未设定控制器的优先级,当多个控制器同时向同一转发设备发送控制指令时,转发设备执行的指令存在不确定性,造成多控制器冲突。文献[10]对控制器设定管控优先级,定义了低优先级的控制器——权威交换机(Authority Switch, AS),用于缓解对高优先级单控制器的依赖,但该方案中 AS 的部署只适用于静态网络环境。文献[11]证明了控制器的部署对控制平面生存性有很大的影响,并由此提出了基于最小割的控制器分簇部署算法(Min Cut-Centroid, MCC)。但该算法只增强了簇内的控制平面生存性,忽略了跨簇的控制平面生存性。文献[12]以“最大化节点与控制器之间满足链路分离的路径数”为目标建立 ILP 模型,求解控制器部署方案(SurViVoR, SVVR),但没有考虑控制信道的长度,造成较大的控制时延。此外,对于大型网络,ILP 很难求得最优解。

为了进一步减轻 SDON 运行对单控制器的依赖,同时避免多控制器冲突,最小化控制平面的故障概率,本文提出基于最小点覆盖的控制平面(Control Plane based on Minimum Point Covering, CP-MPC)生存性设计策略。该策略分为基于最小点覆盖的控制器部署算法(Controllers deployment algorithm using Minimum Point Covering, C-MPC)和基于跨层信息模型的控制信道生存性算法(Survivability against Control Channel failure based on Cross-Layer Information Model, SCC-CLIM)。C-MPC 结合集中控制约束条件,引入最小点覆盖理论,对网络分区,并构建分级管控的控制平面。在全网部署一个全局控制器(Master Controller, MC),进行全网、跨层、多粒度的集中管控;在每个分区部署一个本地控制器(Regional Controller, RC),对网络进行局部、跨层、多粒度的集中管控,分担控制平面的故障风险;同时,为了保证控制平面故障的快速恢复,在每个分区部署多个权威交换机 AS,以进行更小范围、光层、波长粒度的管控。SCC-CLIM 根据跨层信息模型,为控制信道计算跨层生存性路由和分配网络资源,使基于最小点覆盖设计的控制平面真正嵌入 SDON。

2 分级管控网络模型

在满足 SDON 集中控制约束条件的前提下,为降低 SDON 运行对单控制器的过度依赖,增强控制平面的生存性,在控制平面部署多个控制器,同时为了避免多控制器冲突,通过对网络进行分区来建立控制平面的分级管控结构,如图 1 所示。MC 为第 1 级的控制器,用于获取全网信息,掌握跨层、跨域、多粒度的全局信息(包括拓扑和网络状态等),对全网进行集中管控;RC 为第 2 级的控制器,用于缓解控制平面对 MC 的依赖,并保证可扩展性和局部自愈能力;AS 为第 3 级的控制器,负责控制信息的下发和波长粒度的管控,保证控制平面故障的快速恢复。避免多控制器冲突的机制为:3 种控制器被设定为不同的管控优先级,当同一转发设备同时收到不同控制器的指令时,只选择性地执行优先级高的控制器发出的指令。

此外,定义状态信息收集交换机(State information collection Switch, SS)负责收集本地拓扑和网络状态等信息,并将其和 RC 部署在一起,保证 RC 能掌握局部的网络状态。每个 OpenFlow 交换机与一个首选 AS 和一个(或多个)备选 AS 邻接,从而降低 OpenFlow 交换机与控制平面失联的概率。

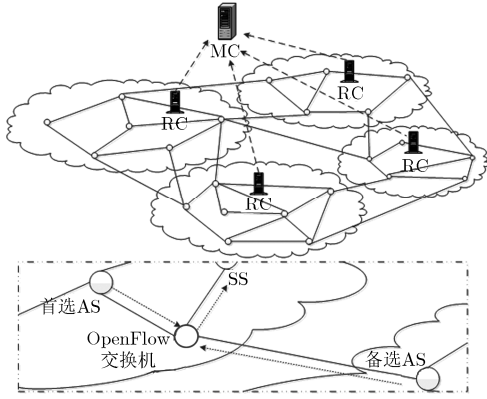


图1 分级管控SDON网络模型

控制器的部署情况决定着控制信道的路由。SS和AS被部署在不同的位置，这决定了网络状态的收集路由与流表信息的下发路由独立设计，使控制平面上行信道和下行信道的故障分离，有利于提高控制平面的生存性。

3 控制器部署算法(C-MPC)

由前两节的分析可知，控制器的数量和位置直接影响控制平面的生存性。设SDON网络拓扑 $G = (V, E)$ ，其中 V 为网络中的节点集， E 为网络中的链路集。MC所在节点用 v_{MC} 表示，RC和SS所在节点用 v_{SS} 表示，AS所在节点用 v_{AS} 表示。

3.1 部署目标

为了使SS快速收集网络状态信息和AS快速下发流表信息，并由RC形成区域性的低优先级管控，结合集中控制的约束条件，C-MPC算法应满足式(1)~式(4)目标。式中的数学符号含义示于表1。

表1 数学符号意义

符号	意义
K_i	第 i 个分区中的节点集
v_{AS_i}	第 i 个分区中 AS 的集合
v_{SS_i}	第 i 个分区中 SS 的集合
K'_i	K_i 中除 v_{AS_i} 以外的节点集， v'_i 为其中的节点
K''_i	K_i 中除 v_{SS_i} 以外的节点集， v''_i 为其中的节点
NUM_{link}^i	分区 i 内包含的链路数

$$\min \sum_{v'_i} D(v'_i, v_{AS_i}), v_{AS_i} \in K_i; v'_i \in K'_i \quad (1)$$

$$\min \sum_{v''_i} D(v''_i, v_{SS_i}), v_{SS_i} \in K_i; v''_i \in K''_i \quad (2)$$

$$\min \sum_i |K'_i| \quad (3)$$

$$\max NUM_{link}^i \quad (4)$$

式(1)表示在第 i 个分区内，最小化 AS 到其下

属节点的距离之和，以保证分区内各个 AS 下发流表信息的整体时延最小。式(2)表示在第 i 个分区内，最小化 SS 到其下属节点的距离之和，以保证分区内各个 SS 收集状态信息的整体时延最小。式(3)表示在整个网络中，最小化 SS 和 RC 的数量，使 RC 尽可能管控更大的范围，以提高分区内集中管控的效率。式(4)表示在第 i 个分区内，最大化包含的光纤链路数，以保证 RC 的局部生存性决策服务更大的范围，并使故障在分区内部被快速发现，快速恢复。

3.2 最小点覆盖的生存性优势

与文献[13]基于最小支配集分簇算法相比，基于最小点覆盖来部署控制器更适合SDON的集中控制网络场景，更能满足控制平面对高生存性的要求。具体原因通过对所提定理1和定理2的证明分析如下：

定理 1 设 $G = (V, E)$ 是简单无向图，且所有节点的度数不小于 2。 $S \subseteq V$ 为 G 的最小点覆盖，则对任意 S 以外节点 x ，至少有两个 S 内的节点与 x 邻接。

证明 在简单图 G 中 $S \subseteq V$ 为最小点覆盖，对任意确定的节点 $x \notin S$ ，与 x 关联的边集合设为 $F = \{(x, y) | (x, y) \in E\}$ ；对于 F 内的边，异于 x 的端点集合设为 $C = \{y | (x, y) \in F\}$ 。由 G 为简单图 $\Rightarrow |C| = |F|$ ；由节点的度数不小于 2 $\Rightarrow |F| \geq 2 \Rightarrow |C| \geq 2$ ；又， S 为最小点覆盖且 $x \notin S \Rightarrow$ 任意 $y \in S$ ，即 $C \subseteq S$ ；故证得：至少有两个 S 内的节点与 x 邻接。

证毕

根据以上定理，结合SDON网络场景可以得出以下结论：软件定义网络中，基于点覆盖部署控制器，会使每个转发设备至少与两个控制器邻接。显然，这将获得更高的控制平面生存性。

如图2(a)所示，黑色节点表示图的最小支配集，如图2(b)所示，灰色节点表示图的最小点覆盖。被着色的节点表示被选取的 v_{AS} ，负责下发流表业务，未被着色的节点为普通的OpenFlow交换机。可见，当图中所示光纤链路发生故障时，图2(a)中的节点1就无法获取控制平面下发的流表信息，而在图2(b)中，由于每个节点同时与两个或以上AS邻接，节点1仍然可以通过其它的AS获取流表信息。

由文献[13]可知，网络的支配集能快速收集网络状态信息。为了使RC能同时掌握本地状态信息和实施本地集中管控，可将SS和AS的部署进行联合设计。

定理 2 设 $G = (V, E)$ 是简单无向连通图， S 为 G 的最小点覆盖，则 S 为 G 的支配集，且有若干个 G 的极小支配集是 S 的子集。

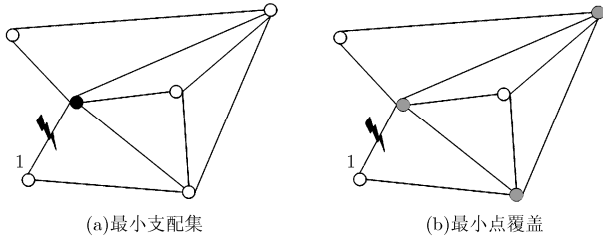


图2 最小支配集与最小点覆盖的对比

证明 由于 S 为 G 的最小点覆盖 $\Rightarrow S$ 与 G 的所有边关联; 又, G 为简单无向连通图 $\Rightarrow S$ 能支配 G 的所有节点 $\Rightarrow S$ 为 G 的支配集 \Rightarrow 有若干个 G 的极小支配集是 S 的子集。 证毕

为了降低 SDON 控制平面的故障概率和减少故障告警信息的收集时间, 所提算法首先基于定理 1 选取网络 G 的最小点覆盖部署 AS; 然后基于定理 2 选取该最小点覆盖包含的支配集, 以部署 SS; 接着在每个 SS 处部署 RC; 最后在中心位置的 RC 处部署 MC。

3.3 C-MPC 算法实现步骤

C-MPC 算法为离线算法, 即该算法预先配置在网络运行之前。

步骤 1 求最小点覆盖 由图论知识可知: “ S 是 G 的极小点覆盖 $\Leftrightarrow V - S$ 是 G 的极大独立集”。因此, 该算法采用布尔变量运算的方法, 求取 G 的所有极大独立集, 之后取逆, 获得 G 的所有极小点覆盖。

$$\varphi = \bigvee_{(v_i, v_j) \in E} (v_i \wedge v_j) \quad (5)$$

$$\bar{\varphi} = \varphi_1 \vee \dots \vee \varphi_i \vee \dots \vee \varphi_m \quad (6)$$

$$Q = \{\vartheta_1^{\min}, \dots, \vartheta_i^{\min}, \dots, \vartheta_m^{\min}\} \quad (7)$$

首先构造布尔函数式(5), 然后将其化成式(6)的形式, 其中 φ_i 为 $(v_x \wedge v_y \wedge \dots)$, 每个 φ_i 确定一个 G 的极大独立集, 进而求得最小点覆盖集合式(7), 其中 ϑ_i^{\min} 表示 G 的一个最小点覆盖的节点集合, m 表示求得 m 种最小点覆盖选取方案, 即: 有 m 种权威节点 v_{AN} 选取方案。此步骤的目的是为了满足部署目标式(3)和式(4)。

步骤 2 求极小支配集

$$\chi_i = v_i + \sum_{v_j \in \text{adj}(v_i)} v_j \quad (8)$$

$$\bar{\varphi} = \chi_1 \chi_2 \dots \chi_n = \prod_{n=|V|} \chi_i \quad (9)$$

$$\bar{\varphi} = \zeta_1 + \dots + \zeta_i + \dots + \zeta_p \quad (10)$$

$$P = \{\psi_1, \dots, \psi_i, \dots, \psi_p\} \quad (11)$$

首先对每个节点 v_i (共 $n = |V|$ 个节点) 构造布尔表达式(8), 作式(9), 式中 “+” 表示逻辑加, 乘表

示逻辑乘。然后通过布尔运算, 化为式(10)的形式, 其中 ζ_i 为 $v_x v_y \dots$ 的形式, 每个 ζ_i 确定一个 G 的极小支配集。最终求得极小支配集的集合式(11), 其中 ψ_i 表示 G 的一个极小支配集的节点集合, p 表示极小支配集的个数。可知 $p \geq m$, 这是因为: “对于同一简单无向图 G , 每个极大独立集是一个极小支配集, 而极小支配集不一定是独立集。”^[14]

步骤 3 选取 m 对 $(\vartheta_i^{\min}, \psi^i)$

$$P_i = \{\psi_1^i, \dots, \psi_j^i, \dots, \psi_{t_i}^i\} \quad (12)$$

$$\min |\psi_j^i|, \quad j = 1, 2, \dots, t_i \quad (13)$$

$$\max \sum_{v_i \in \psi_j^i} \deg(v_i) \quad (14)$$

由定理 2 可知, 网络 G 的最小点覆盖 ϑ_i^{\min} 包含若干个极小支配集, 设其集合为式(12)。若 $t_i = 1$, 则选取该极小支配集 ψ_1^i 为 ψ^i ; 若 $t_i > 1$, 则按照式(13), 取维数(包含节点数)最小的极小支配集 ψ_j^i 赋予 ψ^i , 若式(13)对应的 ψ_j^i 不唯一, 则按式(14)将包含节点度数之和最大的 ψ_j^i 赋予 ψ^i , 极端情况下, 式(14)对应的 ψ_j^i 还不唯一, 为了降低算法复杂度, 随机选取式(14)对应的 ψ_j^i 赋予 ψ^i 。输出并记录这 m 对 $(\vartheta_i^{\min}, \psi^i)$ 。

步骤 4 在 m 对 $(\vartheta_i^{\min}, \psi^i)$ 中筛选出 m' 对 $(\vartheta_k^{\min}, \psi_{\min}^k)$

$$r = \min |\psi^i|, \quad i = 1, 2, \dots, m \quad (15)$$

对于在步骤 2 中求得的 m 个 ψ^i , 由于其为不同 ϑ_i^{\min} 的子集, 故 ψ^i 包含的节点数存在差异。按照式(15), 筛选包含节点数最小的 ψ^i , 将 ψ^i 赋值于 ψ_{\min}^k , 并将对应的 ϑ_i^{\min} 记为 ϑ_k^{\min} , 从而获得 m' 对 $(\vartheta_k^{\min}, \psi_{\min}^k)$, m' 表示筛选的 ψ_{\min}^k 的个数; r 表示 ψ_{\min}^k 包含的节点数, 将其设定为分区数。

步骤 5 确定最优的最小点覆盖和极小支配集 (ϑ, ψ)

$$\min_k \left[\alpha \sum_{v_i} D_{\vartheta} + (1 - \alpha) \sum_{v_i} D_{\psi} \right], \quad v_i \notin \vartheta_k^{\min} \quad (16)$$

$$D_{\vartheta} = \min_{v_j} D(v_i, v_j), \quad v_j \in \vartheta_k^{\min} \quad (17)$$

$$D_{\psi} = \min_{v_d} D(v_i, v_d), \quad v_k \in \psi_{\min}^k \quad (18)$$

若步骤 4 中, $m' = 1$, 则将 $(\vartheta_1^{\min}, \psi_{\min}^1)$ 赋予 (ϑ, ψ) ; 否则, 利用式(16), 选出使其值最小的 $(\vartheta_k^{\min}, \psi_{\min}^k)$ 赋予 (ϑ, ψ) , 其中 v_i 表示不属于最小点覆盖 ϑ_k^{\min} 的节点。 D_{ϑ} 由式(17)求得, v_j 表示属于最小点覆盖 ϑ_k^{\min} 的节点, D_{ϑ} 表示节点 v_i 到 v_j 的最短距离的最小值。 D_{ψ} 由式(18)求得, v_d 表示属于极小支配集 ψ_{\min}^k 的节点, D_{ψ} 表示节点 v_i 到 v_d 的最短距离的最小值。

此步骤的目的是为了满足部署目标式(1)和式(2)。调节因子 α 是为了权衡这两种时延，同时作为标记因子，将步骤5的决策倾向引入步骤8，以平衡MC到 v_i 的整体时延。

步骤6 在 ϑ 内部，按 ψ 分区，部署RC

$$\min_{v_y \in \text{adj}(v_x)} d(v_x, v_y), v_x \in (\vartheta - \psi); v_y \in \psi \quad (19)$$

由步骤4确定网络的分区数为 r ，即 ψ 中的每个节点确定一个分区，故将这些节点设为SS，并在每个SS处部署RC。 ϑ 中 ψ 以外的每一个节点 v_x ，按式(19)找出对应的链路 (v_x, v_y) ，将节点 v_x 加入对应 v_y 所在的分区。

此步骤的目的是为了使一个分区内同时拥有一个SS和一个(或多个)AS，以保证在RC形成分区内部的生存性决策，从而实现快速保护倒换。

步骤7 确定网络中 ϑ 以外的节点 v_z 的分区

$$\min_s [\alpha D_\vartheta^s + (1 - \alpha) D_\psi^s], s = 1, 2, \dots, r \quad (20)$$

$$D_\vartheta^s = \min_{v_x^s} D(v_z, v_x^s), v_z \in (G - \vartheta); v_x^s \in (\vartheta - \psi)_s \quad (21)$$

$$D_\psi^s = D(v_z, v_y^s), v_z \in (G - \vartheta); v_y^s = (\psi)_s \quad (22)$$

α 为在步骤5中引入的调节因子，这里是为了进一步扩大对分区内两种时延的决策倾向。

(1)确定 v_z 所属的第1分区：网络 G 中 ϑ 以外的每一个节点 v_z ，按式(20)选取所加入的分区 s ，并标记为 v_z 的首选分区 $s_{v_z}^1$ 。 D_ϑ^s 由式(21)求得，其中 v_x^s 表示分区 s 中，属于最小点覆盖 ϑ 而不属于极小支配集 ψ 的节点， D_ϑ^s 表示节点 v_z 到 v_x^s 最短距离的最小值。 D_ψ^s 由式(22)求得，其中 v_y^s 表示分区 s 中属于极小支配集 ψ 的节点，即 v_{SS} ， D_ψ^s 表示节点 v_z 到 v_y^s 的最短距离。

(2)确定 v_z 所属的第2，第3分区：对于每个节点 v_z ，选择使式(20)取次最小值的分区 s ，设为 v_z 的第2分区 $s_{v_z}^2$ 。类似地，获得 v_z 的第3分区 $s_{v_z}^3$ (若 $\text{deg}(v_z)=2$ ，则只求 $s_{v_z}^2$)。

此步骤将 v_z 划分到所属的多个分区，并确定次序。其目的一方面是确定在无故障情况下，MC的控制信息下发路径；另一方面是保证链路故障出现时(尤其是在多链路故障情况下)，依然有分区为 v_z 提供控制信息和收集 v_z 的状态信息，保证控制平面的生存性。

步骤8 确定MC节点

$$\min_{v_f \in \vartheta} \left[(1 - \alpha) \sum_{\vartheta} D(v_f, v_\vartheta) + \alpha \sum_{\psi} D(v_f, v_\psi) \right] \quad (23)$$

选取使式(23)最小的节点为工作MC的位置，同时在使式(23)次小的节点上放置备用MC设备，用于为MC提供冗余。式(23)中第1个求和表示备

选节点 v_f 到 ϑ 中所有节点的最短距离之和，第2个求和表示备选节点 v_f 到 ψ 中所有节点的最短距离之和。

为了弥补步骤5和步骤7的决策倾向导致的决策不合理性，将两种时延的调节因子 α 与 $1 - \alpha$ 对换，使得算法能在两个部署目标(及时下发MC的控制信息与及时收集网络状态信息)之间找到平衡。

4 跨层生存性算法(SCC-CLIM)

第3节中，C-MPC算法解决了控制器生存性部署的问题，但为了保证控制平面的生存性，还需要对控制信道进行生存性路由与资源分配设计。控制信道负责传输待下发的流表业务和待上传的状态信息业务，它们作为控制信令业务，具有信令业务的一般性特征：业务量小，交互频繁。在OpenFlow规约中，这两种业务的路径被称为OpenFlow信道，它具有最高等级的生存性要求。

根据跨层信息模型(Cross-Layer Information Model, CLIM)^[15]来分配网络资源进行生存性冗余设计。按图3所示，CLIM将SDON分为多种粒度的资源层，自下而上，资源的粒度越来越小，资源利用率越来越高，但资源分配所需的时间越来越长。

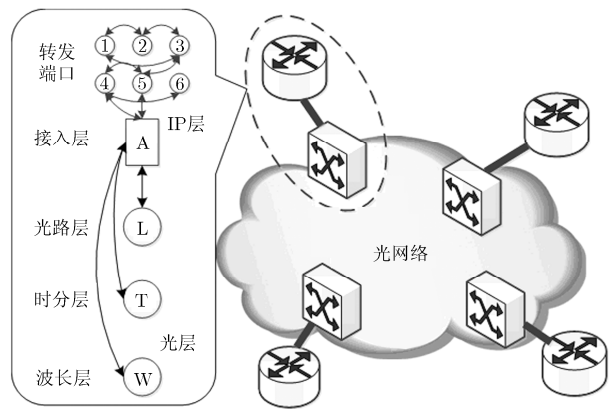


图3 CLIM节点抽象

SCC-CLIM算法将原本分别在波长层选路的信令业务，在光路层汇聚在一起之后，再下达到波长层，减少了承载信令业务的光纤链路数量，进而达到降低控制平面故障概率的目的。

SCC-CLIM算法步骤如下：

步骤1 执行C-MPC算法，部署AS、RC和MC，同时对网络分区。

步骤2 定义并初始化节点集合 ϑ' 和边集合 ξ ， ϑ' 只包含节点MC， ξ 为空集。

步骤3 对网络拓扑 $G = (V, E)$ ，将 ϑ' 以外的AN中，与 ϑ' 内节点距离最短的AN加入节点集合

ϑ' , 将它们之间的链路(可能为路径)加入边集合 ξ , 并在网络图谱 G 中将该边权值设为 0。

步骤 4 判断 ϑ' 是否与 ϑ 相等。若不等, 返回步骤 3; 若相等, 记录边集合 ξ 。

步骤 5 在边集合 ξ 内, 按照最短路由算法, 确定以下两种路由: MC 分别与各分区 RC 之间的路由(为双向通信, MC 下发 RC 对应分区内全部节点的流表信息, RC 向 MC 上传其局部视野的网络状态信息); RC 到分区内各个 AS 的路由(为单向通信, 下发与该 AS 邻接的所有节点的流表信息)。

步骤 6 易证明 ξ 是一个以 MC 为根的树, MC 的流表下发路径 ξ 是一个多播路径。使用多播保护算法^[16], 在光网络中寻找一条哈密顿圈(Hamiltonian cycle), 使用圈上链路为 ξ 内的工作路由和保护路由。

步骤 7 按图 3 示意, 从光路层分配网络资源, 若该层资源不足, 则依次向时分层、波长层申请更大粒度的网络资源, 为步骤 5 和步骤 6 确定的 ξ 内工作路由和保护路由分配工作资源和保护资源。

步骤 8 当控制平面无故障时。对于每个普通节点 v_i , v_i 从其第 1 分区邻接的 AS, 获取下发给自己的流表信息; v_i 向其邻接的 RC, 上传自己收集的状态信息。

步骤 9 当控制平面发生故障时, 在局部启动快速恢复机制。对于受故障影响的 v_i , 依次向其第 2, 第 3 分区邻接的 AS 发送请求, 获取下发给自己的流表信息; v_i 沿其获得流表的路径, 上传自己收集的状态信息至 RC。

SCC-CLIM 算法为控制信道建立生存性路由和分配网络资源, 依靠保护机制保障了 MC 与 RC 之间的双向通信和 RC 到 AS 的下行单向通信; 依靠快速恢复机制保障了 AS 到 OpenFlow 交换机的下行单向通信和 OpenFlow 交换机到 RC 的上行单向通信。

5 仿真与数值分析

在 NSF 和 COST239 两个经典网络拓扑中, 对比文献[11]所提 MCC 算法和文献[12]所提 SVVR 算法, 进行数值分析和仿真验证。假设网络的节点故障可以通过热备份策略恢复, 光纤链路故障独立出现, 不存在共享风险链路组。

可靠性是衡量网络生存性的重要指标。衡量 SDON 控制平面的可靠性可以分为两个具体的指标:

定义 1 故障告警出现概率 P_1 为控制信道工作路由发生故障的概率, 如式(24):

$$P_1 = \text{ave}_{i,j} \left[1 - (1-p)^{W(s_i, c_j)} \right] \quad (24)$$

定义 2 故障恢复概率 P_2 为依靠控制信道保护路由使控制信道快速恢复的概率, 如式(25):

$$P_2 = \text{ave}_{i,j} \left\{ 1 - \left[1 - (1-p)^{W(s_i, c_j)} \right] \cdot \left[1 - (1-p)^{B(s_i, c_j)} \right] \right\} \quad (25)$$

其中, p 表示光纤百公里故障概率; s_i 表示第 i 个 OpenFlow 交换机节点; c_j 表示第 j 个控制器节点; $W(s_i, c_j)$ 和 $B(s_i, c_j)$ 分别表示控制信道工作路由和保护路由的长度; $\text{ave}(\cdot)$ 为取均值。

图 4 给出的是在 NSF 网络中, 3 种控制平面结构故障告警出现概率 P_1 的对比。如图 4(a)所示, P_1 随光纤百公里故障概率 p 在 0~0.10 范围内变化的情况。结果表明, 所提的分级控制结构与 MCC 管控模型和 SVVR 管控模型相比, 故障告警出现概率分别改善了 30%和 10%。这是由于在基于分级控制结构的 SDON 中, MC 和 AS 被分散地部署在整个网络, 有效减少了控制信道的长度。如图 4(b)所示, 在 P_1 随着光纤百公里故障概率 p 从 0 逐渐增大到 1.0 的整个过程中, 本文所提结构的性能一直优于另外两种。这是由于 C-MPC 算法使每个转发设备至少得到两个 AS 的服务, 在恶劣环境下(如战争、自然灾害等)AS 完成了最基本的控制任务。

图 5 给出的是 NSF 和 COST239 网络拓扑中, 在故障告警出现之后, 不依靠网络层重新路由, 只通过控制平面的保护倒换, 故障恢复概率 P_2 随光纤百公里故障概率 p 在 0~0.10 和 0~1.0 范围内变化的情况。结果表明在 $p = 0.1$ 时, CP-MPC 与 MCC 和 SVVR 相比, 在 NSF 网络中分别提升了 18%和 5%, 在 COST239 网络中分别提升了 25%和 15%。这是由于 CP-MPC 使用保护+恢复的混合机制, 为控制信道的工作路由寻找保护路由。在 COST239 网络中获得更大收益是因为网络的节点连通度大, 更有利于控制信道保护路由的寻找。

控制时延是衡量控制平面性能的另一个重要指标。为了验证 CP-MPC 生存性策略能够满足控制平面对控制时延的要求, 使用 mininet+OpenDaylight 仿真工具, 在 NSF 和 COST239 两种网络环境中进行仿真验证。如图 6 所示, 100 ms, 50 ms 和 10 ms 分别为网状网恢复机制的触发时延、环形保护机制的触发时延和控制平面高效运行的容忍时延^[8]。结果表明, 本文所提策略 CP-MPC 与 MCC 和 SVVR 相比, 平均控制时延分别减少了 9%和 46%。所提策略控制时延的最大值也小于 10 ms, 能够很好地满足对控制时延的要求, 可以在实际的网络中运行。SVVR 算法性能最差, 是因为以“最大化节点与控

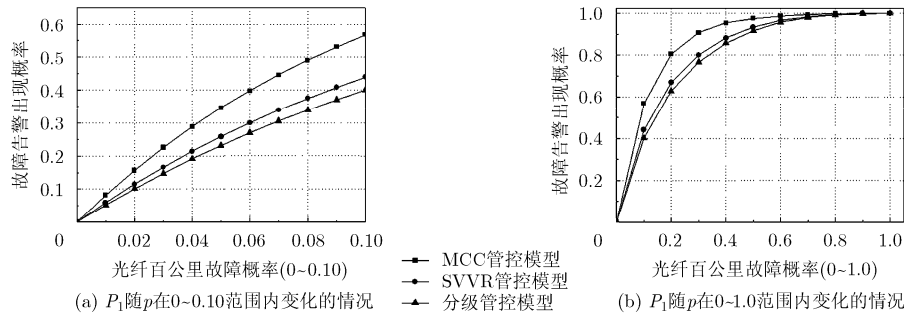


图 4 3 种控制结构的可靠性比较

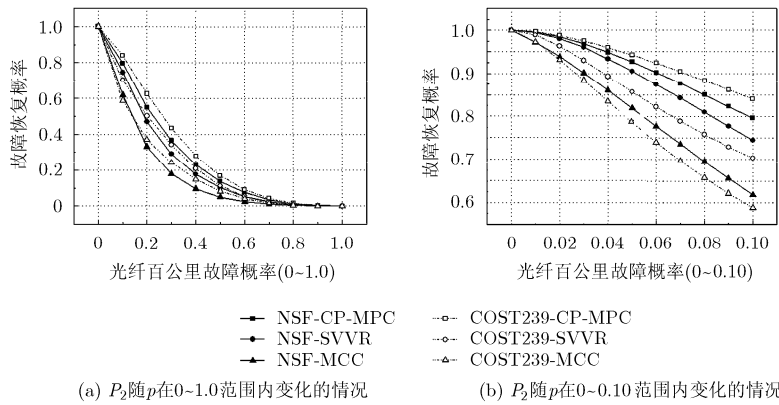


图 5 3 种控制平面设计策略的可靠性比较

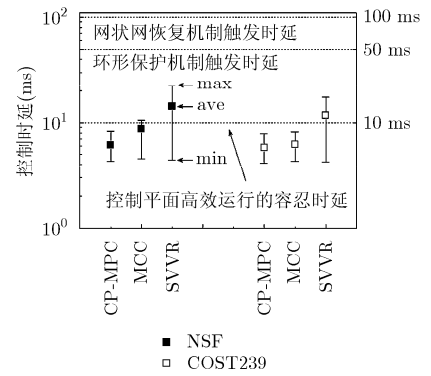


图 6 NSF 和 COST239 网络中 3 种策略控制时延的最大值、均值和最小值对比

制器之间满足链路分离的路径数”为目标部署控制器，控制器被部署在度数较大的节点，而这些节点往往都集中在网络的中心位置，控制器的位置相对集中，容易导致控制器到网络边缘 OpenFlow 交换节点的控制信道过长，控制时延过大。

6 结束语

为了缓解 SDON 网络运行对个别控制器的过度依赖，提高控制平面的生存性，本文提出了一种基于最小点覆盖的控制平面生存性设计策略(CP-MPC)。首先，为 SDON 建立分级的管控网络模型，该模型设定了控制器的管控优先级，结合了全局集中控制和局部集中控制的优点，避免了多控制器冲突。其次，在理论上证明了最小点覆盖在提高网络生存性上的优势，C-MPC 子算法为网络部署全局控制器(MC)、本地控制器(RC)和权威交换机(AS)。接着，SCC-CLIM 子算法为控制信道计算工作路由和保护路由，使分级的管控网络模型真正嵌入 SDON。通过对可靠性指标的数值分析和控制时延的仿真验证，结果表明，所提策略使 SDON 控制平面的可靠性提高了 30%，有效提升了网络在恶劣环境下的生存性，并能很好地满足网络对控制时延的要求。目前，针对软件定义光网络控制平面的生存性设计，

值得进一步深入研究。本文所提策略增加了控制器的种类和数量，虽然通过设定控制器的管控优先级，避免了多控制器的冲突，但是在一定程度会增加网络的建设成本。未来可以根据具体的控制要求做相关优化，以进一步提高控制平面的生存性。

参考文献

- [1] CHOI J S. Design and implementation of a PCE-based software-defined provisioning framework for carrier-grade MPLS-TP networks[J]. *Photonic Network Communications*, 2014, 29(1): 96-105.
- [2] 姚琳元, 陈颖, 宋飞, 等. 基于时延的软件定义网络快速响应控制器部署[J]. *电子与信息学报*, 2014, 36(12): 2802-2808. doi: 10.3724/SP.J.1146.2014.00211. YAO Linyuan, CHEN Ying, SONG Fei, et al. Delay-aware controller placement for fast response in software-defined network[J]. *Journal of Electronics & Information Technology*, 2014, 36(12): 2802-2808. doi: 10.3724/SP.J.1146.2014.00211.
- [3] 左青云, 陈鸣, 赵广松, 等. 基于 OpenFlow 的 SDN 技术研究 [J]. *软件学报*, 2013, 24(5): 1078-1097. ZUO Qingyun, CHEN Ming, ZHAO Guangsong, et al. Research on openflow-based SDN technologies[J]. *Journal of Software*, 2013, 24(5): 1078-1097.
- [4] CASADO M, GARFINKEL T, et al. SANE: a protection

- architecture for enterprise networks[C]. Usenix Security, Vancouver, 2006: 137-151.
- [5] CASADO M, FREEDMAN M J, PETTIT J, *et al.* Ethane: taking control of the enterprise[J]. *ACM SIGCOMM Computer Communication Review*, 2007, 37(4): 1-12.
- [6] MCKEOWN N, ANDERSON T, BALAKRISHNAN H, *et al.* OpenFlow: enabling innovation in campus networks[J]. *ACM SIGCOMM Computer Communication Review*, 2008, 38(2): 69-74.
- [7] Open Networking Foundation. version 1.3. 4. OpenFlow switch specification[S]. 2014.
- [8] HELLER B, SHERWOOD R, and MCKEOWN N. The controller placement problem[C]. Proceedings of the First ACM Workshop on Hot Topics in Software Defined Networks, Helsinki, 2012: 7-12.
- [9] JAIN S, KUMAR A, MANDAL S, *et al.* B4: experience with a globally-deployed software defined WAN[J]. *ACM SIGCOMM Computer Communication Review*, 2013, 43(4): 3-14.
- [10] YU M, REXFORD J, FREEDMAN M J, *et al.* Scalable flow-based networking with DIFANE[J]. *ACM SIGCOMM Computer Communication Review*, 2011, 41(4): 351-362.
- [11] ZHANG Y, BEHESHTI N, and TATIPAMULA M. On resilience of split-architecture networks[C]. IEEE Global Telecommunications Conference (GLOBECOM). Houston, 2011: 1-6.
- [12] MÜLLER L F, OLIVEIRA R R, *et al.* Survivor: an enhanced controller placement strategy for improving SDN survivability[C]. IEEE Global Communications Conference (GLOBECOM), Austin, 2014: 1909-1915.
- [13] 熊余, 张鸿, 王汝言, 等. 基于光通路状态感知的分簇式故障定位机制[J]. 电子与信息学报, 2014, 36(1): 41-47. doi: 10.3724/SP.J.1146.2013.00214.
- XIONG Yu, ZHANG Hong, WANG Ruyan, *et al.* Fault location mechanism based on lightpath status aware using cluster allocation[J]. *Journal of Electronics & Information Technology*, 2014, 36(1): 41-47. doi: 10.3724/SP.J.1146.2013.00214.
- [14] 张清华, 陈六新, 等. 图论及其应用[M]. 北京: 清华大学出版社, 2013: 126-137.
- ZHANG Qinghua, CHEN Liuxin, *et al.* Graph Theory and Its Application[M]. Beijing: Tsinghua University Press, 2013: 126-137.
- [15] ZHANG D, SONG X, CHEN C, *et al.* Software defined synergistic IP+optical resilient transport networks[J]. *Journal of Optical Communications and Networking*, 2015, 7(2): A209-A217.
- [16] DIN D R and HUANG J S. Multicast backup reprovisioning problem for Hamiltonian cycle-based protection on WDM networks[J]. *Optical Fiber Technology*, 2014, 20(2): 142-157.
- 熊余: 男, 1982年生, 博士, 副研究员, 研究方向为宽带网络可靠性理论及抗毁技术.
- 董先存: 男, 1990年生, 硕士生, 研究方向为软件定义光网络抗毁技术.
- 李圆圆: 女, 1992年生, 硕士生, 研究方向为弹性光网络抗毁技术.
- 吕翊: 男, 1963年生, 教授, 研究方向为下一代光网络理论与技术.
- 王汝言: 男, 1969年生, 博士, 教授, 研究方向为通信网络可靠性与故障管理.