

基于快速图挖掘的网络拓扑局部调节区域算法^①

余 平¹, 胡 玲²

1. 重庆电子工程职业学院 人工智能与大数据学院, 重庆 401331; 2. 内江师范学院 教务处, 四川 内江 641100

摘要: 针对 IP 骨干网重新配置中繁重工作量的问题, 提出一种快速图挖掘算法来解决网络拓扑的局部调节区域问题, 解决了从网络拓扑中找到组件时子图同构的 NP-hard 问题, 减少了网络重构的操作工作量. 该文提出的启发式图挖掘算法顶点, 称为顶点目标搜索(vertex targeting search, VTS)算法, 通过考虑网络操作条件减少了搜索空间的大小. 实验结果表明, 该文方法可以快速得到搜索网络模式图, 与其他方法比较, 该文具有较少的搜索时间, 说明该文方法具有可行性和有效性.

关键词: 组件分析; 图形挖掘; 网络拓扑; 顶点目标搜索; 子图同构问题

中图分类号: TP393

文献标志码: A

文章编号: 1000-5471(2019)05-0121-05

运营商的 IP 骨干网络应根据环境变化(如流量分配变化、向新服务迁移、网络设备更新)定期重新配置为理想形式, 即到达终点的互联网协议(Internet Protocol, IP)路由器. 关于骨干网络重构的研究已经有很多报道^[1-5], 例如网络拓扑最佳形式的计算^[6], 网络有效监控以进行重新配置^[7], 以及网络设备的快速配置等.

将骨干网重新配置为任意形式导致网络运营商的工作量很大, 由于骨干网络庞大, 通常由边界网关划分为边界网关协议^[8](border gateway protocol, BGP)联盟或多区域开放最短路径优先^[9](open shortest path first, OSPF). 在传统的骨干网络分段中, 通过多个组件将边缘路由器的直通链路添加到网关路由器. 由于变化范围很广, 网络运营商应重新设计路由, 如 OSPF 区域、IP 寻址、策略路由和路由重新分配, 然后更新手册.

在骨干网络拓扑分析研究中, 学者们从宏观上已经进行了大而复杂的互联网拓扑分析^[10], 发现互联网拓扑具有无标度特征, 还具有节点度的各种特征, 如频谱、中心性和可能性. 已有研究者收集了网络拓扑的特征, 然后提出方法, 用于分析网络的更多实际特征^[11-12].

目前 IP 骨干网络配置的研究有: 文献[13]针对 IP 网络重配置引发的微环效应, 提出避免 IP 重配置的微环方法, 该方法利用区间重叠的特点直接计算中间值. 文献[14]通过 7 年运营 IP 骨干网数据, 用以研究网络配置的演变, 通过将每个任务与配置文件中的某组命令相关联来识别和构建配置模板, 为运营骨干网提供自动配置. 文献[15]对骨干网负责功率做了研究, 如网络重新配置, 给出了最小化骨干网络功耗的方法. 现有方法都没有对 IP 骨干网重新配置过多工作量做过研究.

本文在现有对 IP 网络重配置研究的基础上, 提出了使用快速图挖掘将网络拓扑解析到局部调节区域的算法, 来减少 IP 网络重配置过程中网络重构的操作工作量问题. 首先提出用于搜索组件的概述算法, 从网络拓扑中查找组件问题的核心被转换为子图同构问题. 由于该问题属于 NP-hard 类, 因此提出了启发式图挖掘算法, 称为顶点目标搜索(vertex targeting search, VTS)算法, 在搜索过程中 VTS 算法根据子图中的特定节点定向网络拓扑中的顶点, 具有网络操作的条件, 例如星形子图的中心被放置在特定建筑物上,

① 收稿日期: 2018-09-10

基金项目: 四川省教育厅自然科学重点科研项目(13ZA0001).

作者简介: 余 平(1969-), 女, 副教授, 主要从事软件技术研究.

用于图挖掘搜索树的大小显著减小。

1 网络解析问题定义

在本节中解释基于子图同构问题的网络解析问题，该问题如图 1 所示。数据图（例如，IP 骨干网）和预定义模式图被给出作为输入，为了通过模式图解析数据图，目标函数被设置为数据图中模式图出现的最大化，数据图中出现的模式图不得与模式图的另一次出现重叠。

数据图和模式图集如下： $G_d = (V_d, E_d)$ 和 $G = (G_1, G_2, \dots, G_L)$ ，其中 $G_l = (V_l, E_l)$ ， $l = (1, 2, \dots, L)$ 。 V_d 和 V_l 是节点集， E_d 和 E_l 是链路集。 G_l 和 G_d 子图同构的条件是 G_l 中的节点和链接与 G_d 子图中的节点和链接一一对应：当 $v_i \in V_l, V \subseteq V_d, E \subseteq E_d$ 时，有一个方程 $f: V_l \rightarrow V$ ，其中

$$V \forall i, j, (\{v_i^l, v_j^l\} \in E_l \Rightarrow \{f(v_i^l), f(v_j^l)\} \in E)$$

(1)

将 G_l 中 G_d 的出现定义为 $S(G_l, G_d)$ ，并且出现次数表示为 $|S(G_l, G_d)|$ 。 为了避免计算模式图 $G_i, G_j (i \neq j)$ 的重叠， G_i 不包括在 $S(G_i, G_d)$ 中。 最后，如果 $w_i > w_j$ ，将模式图的权重定义为 $w_l, (l = 1, 2, \dots, L)$ ，其中 G_i 具有比 G_j 更高的优先级。

基于以上网络模型，网络解析表述如下：假设 $G_d, G_l l = (1, 2, \dots, L)$ 和 $w_l, (1, 2, \dots, L)$ ，输出问题 $S(G_l, G_d) l = (1, 2, \dots, L)$ 可以用最优公式(2) 表示为

$$\max \sum_{l=1}^L w_l |S(G_l, G_d)|$$

(2)

当给出一个数据图 $G_d = (V_d, E_d)$ 和模式图 $G_l = (V_l, E_l)$ ， f 的个数也就是解的空间变成 $P(V_d, V_l)$ ，NP-hard 问题。 本文提出 VTS 方法，解决了 IP 骨干网表示的数据图 NP-hard 网络解析问题，并提高了搜索速度。

2 本文算法

为了解决 NP-hard 网络解析问题，提出了一种快速图挖掘概述算法(算法 1) 和 VTS(算法 2)，算法 1 用于解决网络解析问题，VTS 用于提高搜索速度。 VTS 的关键思想是根据网络操作条件，从子图中的特定节点瞄准数据图中的顶点(节点)。

图 2 给出了 VTS 的过程图，因为搜索树的大小为 $P(V_d, V_l)$ ，当 $V_d = 10$ 和 $V_l = 4$ 没有修剪时，树的大小变为 $P(10, 4) = 5\,040$ 。 假设模式图中来自节点 A 的函数 f 只允许链接到数据图中的节点 a，如图 2(a) 所示，不必从节点 A 到节点 b-j 创建 f ：搜索树的深度缩小。 另外，由于模式图的形式是 1 级星，因此不应该创建 f 节点 B,C 和 D 到节点 h,i 和 j：搜索树的宽度也缩小。 搜索树大小缩小到如图 2(b) 所示的 $P(6, 3) = 120$ 。

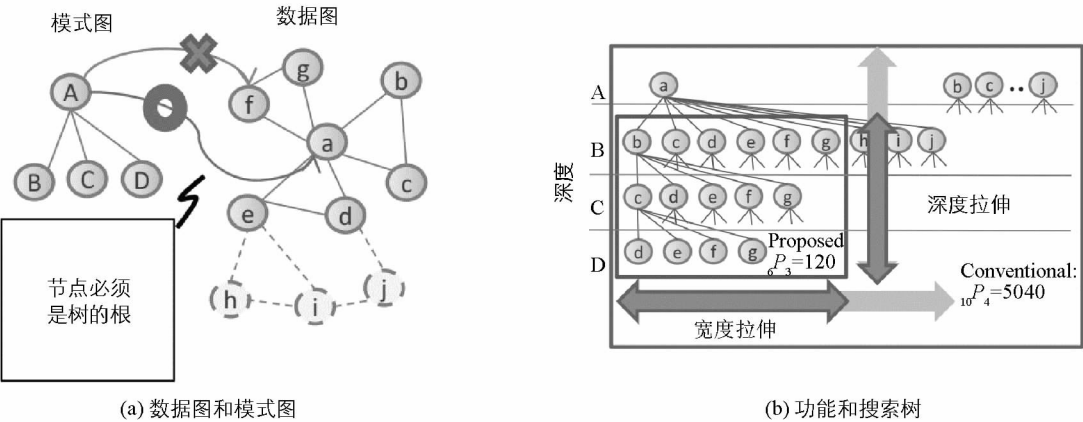


图 2 VTS 主要过程图

如图 2 所示, 如果可以基于特定的条件限制 f , 那么搜索树的大小就会大大减小. 根据网络运行情况进行估计, 对于星型图, 其中心位于特定的建筑上, 为其他使用者提供内容或接口点. 2 链图形用于建筑物的直线连接, 常称为 IP 卸载链路, 环型图一般可以放置为冗余.

2.1 用于搜索组件的概述算法

用表 1 中伪代码对搜索算法进行阐述. 算法的输入是模式图 G_l , $l = (1, 2, \dots, L)$ 的集合, 模式图的权重表示为 w_l , $(1, 2, \dots, L)$ 和数据图 G_d . 算法的输出表示为 $S(G_l, G_d)$, $l = (1, 2, \dots, L)$.

该算法首先根据 w_l 对 G_l 进行排序, 因为算法基本上覆盖了 G_l 的出

现在 G_l 发生后避免重叠, 高优先级的 G_l 发生最大化. 接下来, 使用算法 2 计算临时出现集合 $S_{tmp}(G_l, G_d)$, 这将在下一个小节中详细讨论. 在以下过程中, $S_{tmp}(G_l, G_d)$ 中的 S'_i 和 $S(G_m, G_d)$ 中的 S_j^m 进行比较, S'_i 适当地注册, 考虑到它们的重叠避免元素.

如果 S'_i 与 S_j^m 不过度重合, S'_i 直接注册到 $S(G_l, G_d)$, 算法 1 第一次运行对应于这种情况. 如果 S'_i 的子集 a 是 S_j^m 的子集 b 的子集, 从 $S(G_m, G_d)$ 中删除 S_j^m , S'_i 被注册到 $S(G_l, G_d)$, S_j^m 是 a 相对于集合 b 的相对补数, 被注册到 $S(G_m, G_d)$.

如果 S'_i 的子集 a 是 S_j^m 的子集, S_j^m 被从 $S(G_m, G_d)$ 中删除, S'_i 被注册到 $S(G_l, G_d)$. 在搜索完所有的图形之后, 算法 1 输出 $S(G_l, G_d)$, $l = (1, 2, \dots, L)$.

2.2 VTS 算法

本文用表 2 伪代码对 VTS 算法进行说明. 算法的输入是模式图 G_l 和数据图 G_d , 算法的输出是一组临时出现的事件 $S_{tmp}(G_l, G_d)$. 如果 G_l 是 2 链图, 用一个链接连接 2 个节点, 不需要创建搜索树. 然后, 将连接它们的每个节点对和链接注册为 $S_{tmp}(G_l, G_d)$ 作为 2 链图, 算法完成. 对于星型图, 算法有 2 种搜索模式: loose search 和 normal search. 在 loose search 模式, 由节点组成的子图 $r \in V_d$, 被标记为根节点, 所有的链接连接 r 注册到 $S_{tmp}(G_l, G_d)$, 算法完成. 在 normal search 模式下, 算法首先从根节点开始修复, f 在 G_l 到 G_d 的根节点上, 并创建搜索树.

此过程会减小搜索树的大小, 如图 2 所示. 算法从缩减树中搜索给定的模式图, 将结果注册到 $S_{tmp}(G_l, G_d)$, 如果在上述条件下模式图不匹配, 则使用 Ullmann 图同构算法得到 $S_{tmp}(G_l, G_d)$.

通过省略搜索树创建或减小搜索树的大小, 该算法提高了其计算速度. 如果模式图的形式与预定义条件不匹配, 则其速度与 Ullmann 算法相同, 本文算法不是用于通用图挖掘工具, 而是用于特定分析, 例如

表 1 算法 1 伪代码

输入: $G = (G_1, G_2, \dots, G_L)$, w_l , $l = 1, 2, \dots, L$, $G_d = (V_d, E_d, label)$
按照 w_l 的升序排序 G
for ($l = 1; l \leq L; l++$)
设 $S_{tmp}(G_l, G_d) = \{S'_1, S'_2, \dots, S'_k\}$, (G_l, G_d) 从算法 2 得到
for ($\begin{matrix} S'_i \in S_{tmp}(G_l, G_d), S'_j \in S(G_m, G_d), \\ m = 1, 2, \dots, l-1 \end{matrix}$)
if ($S'_i \not\subseteq S_j^m$)
设 S'_i 属于 $S(G_l, G_d)$
if ($a \subset b, a \subset S'_i, b \subset S_j^m$)
从 $S(G_m, G_d)$ 中删除 S_j^m
设 S'_i 属于 $S(G_l, G_d)$
设 $S_j^m = b \setminus a$ 属于 $S(G_m, G_d)$
if ($a \subset S_j^m, a \subset S'_i$)
从 $S(G_m, G_d)$ 中删除 S_j^m
设 S'_i 属于 $S(G_l, G_d)$
输出 $S(G_l, G_d)$, $l = (1, 2, \dots, L)$

表 2 算法 2 伪代码

输入: $G_l = (V_l, E_l)$, $G_d = (V_d, E_d, label)$
if G_l 是链图
设置链接 $(i, j) \in E_d$, 节点 $i \in V_d$ 来自 $S_{tmp}(G_l, G_d)$
输出 $S_{tmp}(G_l, G_d)$
if G_l 是星状图
if loose search
每个根节点 $r \in V_d$, $(r, i) \in E_d, i \in V_d$, $S_{tmp}(G_l, G_d)$
输出 $S_{tmp}(G_l, G_d)$
else
G_l 的映射根节点来自根节点 $r \in V_d$, 创建搜索树 T
从 T 中搜索 G_l , 设置为 $S_{tmp}(G_l, G_d)$
输出 $S_{tmp}(G_l, G_d)$
else
从 G_l, G_d 创建搜索树 T
从 T 中搜索 G_l , 设置为 $S_{tmp}(G_l, G_d)$
输出 $S_{tmp}(G_l, G_d)$

骨干网络分析.

3 实验结果与分析

本文评估目的有 2 个, ①可视化小型网络的组件分析; ②使用本文算法评估计算时间. 计算时间定义为计算一个模式图的出现时间. 本文使用传统子图同构算法比较了 VTS 的 2 种模式(normal search 减小了搜索树的大小, loose search 没有创建搜索树). 每种算法都是在 2.5 GHz 和 250 Gb 内存环境笔记本电脑下的 C++ 中实现的. 图 3 给出了物理网络解析组件分析的可视化.

图 3(c)是使用图 3(a)中的环形图和图 3(b)中的数据图可视化物理网络解析的组件分析. 通过使用环形图解决物理网络拓扑, 设计一个在物理区域中具有两条不相交路径的局部调节区域. 对于维护能力, 运营商的网络运营商通常采用分层 IP 拓扑, 其中 IP 分组从边缘 IP 路由器通过核心 IP 路由器转发. 由于每个层次结构(由 IP 边缘路由器和核心路由器对组成)应具有冗余, 因此 IP 网络运营商通过考虑物理布局来设计拓扑. 该物理网络解析可用于这种情况. 设计 IP 拓扑的层次结构也表明它还用于间接地设计诸如 BGP 或 OSPF 的 IP 逻辑区域, 由 IP 核心路由器划分的每个层次结构可以被视为 IP 逻辑区域.

为了体现本文算法的有效性, 将其与传统算法与文献[16]中子图同构算法进行比较. 图 4 说明了模式图中节点数量变化时的网络搜索计算时间.

从图 4 中可以看出, 对于从 48 节点数据图中找到 7 节点模式图, 传统子图同构方法的时间最长, 大约需要 10 h, 其次是文献[16]中子图同构的方法, 而本文中具有 normal search 模式的 VTS 算法的搜索时间大约为 0.2 s. 作为参考, 具有 loose search 模式的 VTS(其不创建搜索树)可以用 10^{-5} 秒的数量级完成计算, 说明本文方法的有效性.

4 结 论

为了灵活地重新配置 IP 骨干网, 本文提出了使用快速图挖掘来解决网络拓扑到组件的算法. 该算法对于网络拓扑中子图同构 NP-hard 问题提出了启发式图挖掘算法, 称为顶点目标搜索(vertex targeting search, VTS)算法, 在搜索过程中 VTS 算法根据子图中的特定节点定向网络拓扑中的顶点, 大大减少搜索时间. 实验表明本文算法具有较少计算时间对模式图进行搜索, 从 48 节点网络中找到 7 节点子图, 本文算法可以在 1 s 内完成 NP-hard 子图同构问题的计算. 文本研究未来工作将开发本地重新配置架构以实现灵活操作, 并评估其与当前 IP 骨干网络全局重新配置相比的性能.

参考文献:

- [1] LI F L, YANG J H, ZHANG H J, et al. A Study of Configuration Evolution of an Operational IP Backbone Network [J]. China Communications, 2017, 14(6): 88-97.
- [2] LIU W, HONG A, OU L, et al. Prediction and Correction of Traffic Matrix in an IP Backbone Network [C]//2014 IEEE 33rd International Performance Computing and Communications Conference (IPCCC). Austin: IEEE, 2014.
- [3] NIE L S, JIANG D D, GUO L, et al. Traffic Matrix Prediction and Estimation Based on Deep Learning in Large-Scale

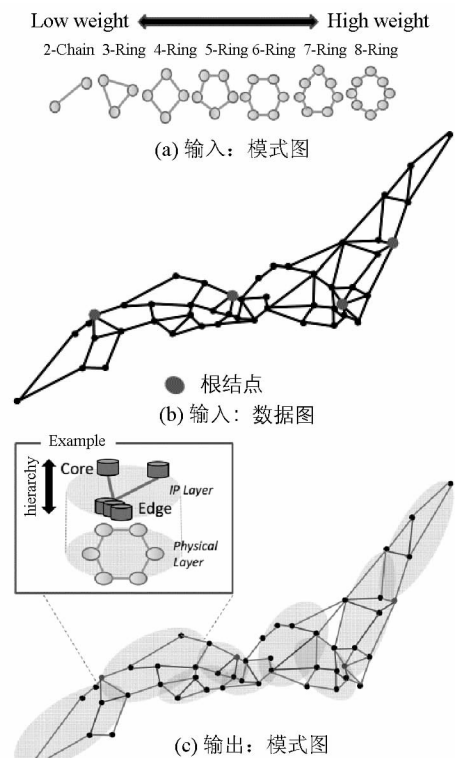


图 3 物理网络解析组件分析的可视化

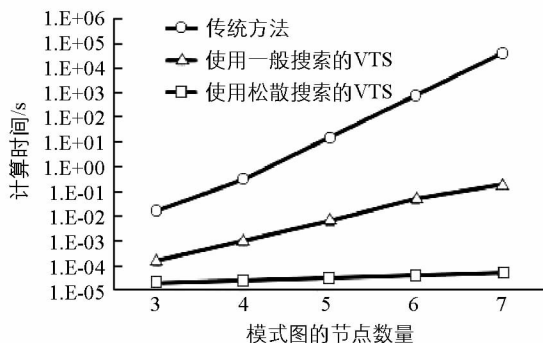


图 4 模式图中节点数变化时的计算时间

- IP Backbone Networks [J]. Journal of Network and Computer Applications, 2016, 76: 16-22.
- [4] DHARMAWEERA M N, PARTHIBAN R, SEKERCIOGLU Y A. Toward a Power-Efficient Backbone Network: The State of Research [J]. IEEE Communications Surveys and Tutorials, 2015, 17(1): 198-227.
- [5] NENCIONI G, HELVIK B E, HEEGAARD P E. Including Failure Correlation in Availability Modeling of a Software-Defined Backbone Network [J]. IEEE Transactions on Network and Service Management, 2017, 14(4): 1032-1045.
- [6] SUMMERS T, SHAMES I, LYGEROS J, et al. Topology Design for Optimal Network Coherence [C]//2015 European. Control Conference(ECC). Naples: IEEE, 2015.
- [7] BACCARELLI E, CORDESCI N, MEI A, et al. Energy-Efficient Dynamic Traffic Offloading and Reconfiguration of Networked Data Centers for Big Data Stream Mobile Computing: Review, Challenges, and a Case Study [J]. IEEE Network, 2016, 30(2): 54-61.
- [8] WEITZ K, WOOS D, TORLAK E, et al. Scalable Verification of Border Gateway Protocol Configurations with an SMT Solver [J]. ACM SIGPLAN Notices, 2016, 51(10): 765-780.
- [9] LIU W Y, HUA N, ZHENG X P, et al. Intelligent Inter-Domain Connection Provisioning for Multi-Domain Multi-Vendor Optical Networks [J]. Journal of Optical Communications and Networking, 2015, 7(3): 176-192.
- [10] CETINKAYA E K, ALENAZI M J F, PECK A M, et al. Multilevel Resilience Analysis of Transportation and Communication Networks [J]. Telecommunication Systems, 2015, 60(4): 515-537.
- [11] KAMAMURA S, SHIMAZAKI D, MORI H, et al. Optimization of Light-path Configuration Order in IP over WDM Networks using Fast Traffic Matrix Estimation [C]//Optical Fiber Communications Conference and Exhibition. San Diego: IEEE, 2014.
- [12] ZHANG X, MILLER-HOOKS E, DENNY K. Assessing the Role of Network Topology in Transportation Network Resilience [J]. Journal of Transport Geography, 2015, 46: 35-45.
- [13] 罗 龙. 网络更新过程中的微环避免技术和拥塞避免技术研究 [D]. 成都: 电子科技大学, 2015.
- [14] LI F L, YANG J H, ZHANG H J, et al. Evolution of Network Configurations: High-Level Analysis of an Operational IP Backbone Network [C]//The 16th Asia-Pacific Network Operations & Management Symposium. Hsinchu: IEEE, 2014.
- [15] DHARMAWEERA M N, PARTHIBAN R, SEKERCIOGLU Y A. Toward a Power-Efficient Backbone Network: The State of Research [J]. IEEE Communications Surveys & Tutorials, 2015, 17(1): 198-227.
- [16] 刘彩霞, 李凌书, 汤红波, 等. 基于子图同构的 vEPC 虚拟网络分层协同映射算法 [J]. 电子与信息学报, 2017, 39(5): 1170-1177.

Network Topology Local Adjustment Region Algorithm Based on Fast Graph Mining

YU Ping¹, HU Ling²

1. School of Artificial Intelligence and Big Data, Chongqing College of Electronic Engineering, Chongqing 401331, China;

2. Education Administrator Office, Neijiang Normal University, Neijiang Sichuan 641100, China

Abstract: Aiming at the heavy workload in the reconfiguration of IP backbone network, a fast graph mining algorithm has been proposed to solve the problem of local adjustment area of network topology, and solve the NP-hard problem of subgraph isomorphism when finding components from network topology, and to reduce the workload of network reconfiguration. In this paper, a heuristic graph mining algorithm called vertex targeting search (VTS) algorithm has been proposed, which reduces the size of search space by considering network operation conditions. The experimental results show that this method can quickly obtain the search network pattern. Compared with other methods, this method has less search time, which shows the feasibility and validity of this method.

Key words: component analysis; graph mining; network topology; vertex targeting search; sub-graph isomorphism problem