基于混合优化算法的网络流量有效测量点选择

葛洪伟，彭震宇，岳海兵
（江南大学 信息工程学院，江苏 无锡 214122）

摘　要：提出一种基于禁忌搜索和蚁群算法的求解最小弱顶点覆盖问题的混合优化算法，用于解决网络流量有效测量点的选择问题。仿真结果表明，比较现有算法，本算法能够找到更小的弱顶点覆盖集，且具有更好的可扩展性和实用性。

关键词：蚁群优化算法；禁忌搜索算法；最小弱顶点覆盖
中图分类号：TP301.6　　**文献标志码**：A　　**文章编号**：1001-3695(2009)04-1480-04

Hybrid optimization algorithm for efficient monitor-nodes selection in network traffic

GE Hong-wei, PENG Zhen-yu, YUE Hai-bing
(*School of Information, Jiangnan University, Wuxi Jiangsu* 214122, *China*)

Abstract: This paper proposed a new hybrid optimization algorithm for solving minimum weak vertex cover set problem. The experimental results show that the proposed hybrid optimization algorithm is more expansibility and practicability, and can find smaller weak vertex cover set than other algorithms.

Key words: ant colony optimization; tabu search; minimum weak vertex cover set

随着 Internet 重要性的日益提高和网络结构的日益复杂，网络管理越来越成为人们关注的焦点。现代的网络管理系统注重于服务级、应用级的管理，如主动式和被动式的资源管理、流量工程、端到端的服务质量保证等。所有这些网管业务均以了解网络流量等网络运行参数为基础。为此，有必要对网络流量进行测量和分析，以利于发现网络瓶颈，优化网络配置，并进一步发现网络中可能存在的潜在危险。

由于 Internet 是一个不断变化的庞大网络，网络流量的测量要具有一定的实时性。目前，网络流量的测量方法要求针对特定的感兴趣的链路，人工合理地规划网络的观测节点，并安装特定的监测软件或硬件。但这种方法难以扩展，不便于动态适应网络的变化，且可能因为设置过多的观测节点而增加网络的额外负担。

测量方法关键的步骤是既要准确获取网络流量参数，又要尽量减少数据收集对实际网络数据传输造成的额外负荷。因此，观测节点的选择尤其重要。研究中一般将网络流量有效测量点选择问题转换为求给定无向图中的最小弱顶点覆盖问题，由于这个问题已被证明是 NP-hard 问题，因此一般采用近似算法求解，如文献[1]给出了一个求解弱顶点覆盖问题的贪婪秩算法；文献[2]又提出了一个弱化的贪心算法。但所有这些算法均需知道网络拓扑的全局信息，因而可扩展性较差，而且问题规模较大时，算法求解质量不高，或根本求不出最优解。于是本文提出了一种求解最小弱顶点覆盖问题的混合优化算法，即基于禁忌搜索的蚁群优化算法 TSBACO（tabu search based ACO），对网络流量有效测量点的选择问题进行研究和实验，取得了良好的成效。

1　网络流量测量模型描述

在网络中某一节点设置测量器（如 SNMP agent），可以得到与这一节点相连的所有链路上的流量。因此，为了得到网络中所有链路的网络流量，一般可以通过在某些交换节点（路由器）配置测量器实现。要考虑的问题是：在网络哪些节点上设置测量器，才能使得在可以得到每一条链路流量的条件下，所需测量器数目最小。这一问题自然可以归纳为图论中的最小顶点覆盖问题。

为了方便问题模型的描述，参照文献[2]给出以下定义：

定义 1　给定无向图 $G=(V,E)$。其中：V 是顶点集；E 是边集；S 是 V 的子集，若根据与 S 中顶点相关联的各条边的流量，可确定 E 中任意边的流量，则称 S 是图 G 关于流量的测量集。

有效测量问题的目标就是求给定图 G 关于流量的最小测量集。虽然采用求解最小顶点覆盖问题的算法可以求出一个测量集，但已证明，最小顶点覆盖问题是一个 NP-hard 问题，尚无多项式时间的求解算法，且求出的测量集也未必是最小测量集。如果测量器是路由器或交换机等交换设备，那么图还满足以下两条约束条件：

a）对图 G 的顶点集 V 中的任意顶点 v，其 degree$(v)\geqslant 2$；

b）对图 G 的顶点集 V 中的任意顶点 v，满足流守恒方程，即流进 = 流出。

尽管以下原因将会导致流守恒方程的失真，如：交换设备是数据的源或汇，而不仅仅是数据转发器；多播导致输出端口

收稿日期：2008-07-27；**修回日期**：2008-10-08

作者简介：葛洪伟（1967-），男，江苏无锡人，副教授，博士，主要研究方向为模糊系统、算法分析、图像处理、模式识别等（ghw8601@163.com）；彭震宇（1981-），男，江苏无锡人，硕士，主要研究方向为算法分析、人工智能、图像处理；岳海兵（1985-），男，内蒙古乌兰察布人，硕士研究生，主要研究方向为算法分析、无线网络、人工智能.

的数据复制;交换设备本身的数据包延迟或丢失。但是若干研究表明,流守恒方程所具有的相对误差小于 0.05%。因此,由以上两条约束条件,求图 G 关于流量的最小测量集问题可以转换为求定义 2 中图的弱顶点覆盖问题。

定义 2　假定无向图 $G=(V,E)$ 满足对任意 $v\in V$ 有 degree$(v)\geqslant 2$,称 $S\subset V$ 是图 G 的弱顶点覆盖集,当且仅当执行以下操作能使 E 中所有的边都可被标记:

a)标记所有与 S 中顶点相关联的边;

b)若某个顶点 v 的 degree$(v)-1$ 条相关联的边已被标记,则标记剩下的那条相关联的边;

c)重复第 b)步,直到不能再标记新的边为止。

图 G 的弱顶点覆盖 S 就是在流守恒约束条件下的图 G 关于流量的测量集。首先与集合 S 中顶点相关联的边的流量可以通过测量手段来获取;其次,如果 $v\notin S$,且 v 的 degree$(v)-1$ 条边的流量已获得,那么根据流守恒方程可以计算出另外一条边的流量。反复应用流守恒方程,可计算出图 G 中所有边的流量。网络流量测量中的有效测量点选择问题实际上就是求解无向图 G 的最小弱顶点覆盖问题。

2　最小顶点覆盖问题

设简单图 $G=(V,E)$ 是一无向图。其中:V 是图中顶点集合;E 是边集合。下面给出顶点覆盖的有关定义。如果图 G 中的一个顶点和一条边相互关联,则称其相互覆盖;称顶点集 V 的一个子集 V' 为顶点覆盖,如果图 G 的任意一条边都至少有一个端点属于 V';如果一个顶点覆盖是图 G 的其他任何顶点覆盖的真子集,则称该顶点覆盖为图 G 的极小顶点覆盖;包含顶点数最少的顶点覆盖就是图 G 的最小顶点覆盖,最小顶点覆盖的顶点数称为图 G 的覆盖数。例如,图 1 中黑色顶点所组成的顶点集就是相应图的最小顶点覆盖。

由上述定义可知,最小顶点覆盖(minimum vertex covering,MVC)问题是指,给定一个无向图 $G=(V,E)$,求顶点集 V 的一个最小子集 S,使得 $e=(u,v)\in E$,且 $u\in S$ 或 $v\in S$,即 E 中的任一边至少有一个顶点属于 S,即 S 中的顶点覆盖了边集 E。

由于最小顶点覆盖问题是 NP-hard 问题,到目前为止还不存在多项式时间算法来求解。虽然一些学者已经证实对于特殊的图,通过采用特殊的算法,可以在多项式时间内求得图的最小顶点覆盖,如互补图、圆弧图、立方图等。但本文不是对一些特殊的图求其多项式时间算法,而是要对一般的图进行最小顶点覆盖问题的求解。在已有的近似算法中,算法仅根据局部信息进行决策,因此在许多情况下,这类算法并不能确定最小顶点覆盖问题。而对于传统的精确算法如分支限界法、割平面法,对偶启发方法等,尽管这些方法能求得最小顶点覆盖的最优解,但其计算复杂度高、花费时间长,并不能应用到大规模问题中,所以必须使用启发式算法来求解图的最小顶点覆盖问题,使得在较短时间内求出问题的较优解甚至最优解。由上述最小顶点覆盖问题的定义可知,该问题是一个组合优化问题,也是一个子集类问题,所以非常适合采用蚁群优化(ant colony optimization,ACO)算法来求解。

3　基于禁忌搜索的蚁群优化算法

ACO 算法作为一种新型元启发式算法,最近被用来解决许多不同的组合优化问题,但 ACO 算法在搜索过程中容易出现停滞现象或陷入局部最优。而禁忌搜索是对局部邻域搜索的一种扩展,是一种全局逐步寻优算法,采用的是一种确定性的局部极小突跳策略。禁忌搜索算法通过引入一个灵活的禁忌表和相应的禁忌准则来避免重复迂回搜索,并通过藐视准则来赦免一些被禁忌的优良状态,进而保证多样化的有效探索以最终实现全局优化。禁忌策略最重要的思想是标记对应已搜索到的局部最优解的一些对象,并在进一步的迭代搜索中尽量避开这些对象(而不是绝对禁止循环)从而保证对不同的有效搜索途径的探索。但在搜索算法中,可能会出现候选解全部被禁忌,或者存在一个优于到目前为止最好解的禁忌候选解,此时藐视准则就将使某些状态解禁,以实现更高效的优化性能。藐视准则的设置就是要算法避免遗失优良状态,激励对优良状态的局部搜索,进而实现全局优化。对于非禁忌候选状态,算法无视它与当前状态的优劣关系,仅考虑它们中间的最佳状态,如此便可实现对局部极小的跳跃。正由于禁忌技术具有灵活的记忆功能和藐视准则,且在搜索过程中可以接受劣解,所以搜索时能够跳出局部最优解,转向搜索空间的其他区域,从而增强获得更好的全局最优解的概率。

为防止 ACO 算法出现停滞现象或陷入局部最优,本文将 ACO 算法与禁忌搜索算法结合起来,提出了一种基于禁忌搜索的蚁群优化算法 TSBACO。该算法可以在一定程度上有效解决扩大搜索空间和加快收敛速度之间的矛盾,从而使得 TSBACO 算法既能避免重复迂回搜索,跳出局部最优解,又能对不同搜索区域进行更有效探索,不仅提高了全局优化性能而且还提高了搜索效率。

4　基于 TSBACO 的最小弱顶点覆盖问题求解

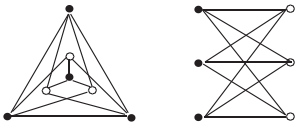
由于在网络环境中,所有顶点的度 $\deg(v)\geqslant 2$,所以求解最小顶点覆盖问题可以转换为求解最小弱顶点覆盖问题。

4.1　关联矩阵

首先给出最小弱顶点覆盖问题的关联矩阵定义,无向图 G 的关联矩阵 $\boldsymbol{A}=(a_{ij})$ 是指如下定义的 $n\times m$ 矩阵:

$$a_{ij}=\begin{cases}1,\text{如果顶点 } v_i \text{ 与边 } e_j \text{ 相关联}\\0,\text{其他}\end{cases}$$

图 2 为某网络拓扑图。



(a)　(b)

图 1　最小顶点覆盖问题实例图

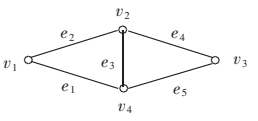


图 2　网络流量测量中有效测量点选择示意图

对于图 2 中的网络拓扑图,其关联矩阵 $\boldsymbol{A}$ 为

	e_1	e_2	e_3	e_4	e_5
v_1	1	1	0	0	0
v_2	0	1	1	1	0
v_3	0	0	0	1	1
v_4	1	0	1	0	1

由上面的网络拓扑图和其关联矩阵可知,对于任何一个有 n 个顶点和 m 条边的无向图 $G=(V,E)$,满足对任意 $v\in V$ 有 degree$(v)\geqslant 2$,则整个图 G 都可用一个 $n\times m$ 的二阶矩阵来表示,矩阵由{0,1}组成。其中每一行表示一个顶点,每一列表示一条边;图中每个顶点对应行中的各个值表示该顶点的邻边情况,1 表示顶点和边存在关联关系,0 则表示不存在关联关

系。

4.2 求解最小弱顶点覆盖问题的近似算法

以下给出一个求解最小弱顶点覆盖问题的近似算法,并把求解得出的弱顶点覆盖集作为蚁群初次循环的初始解,算法的思想为:

a)选取一个含有边数最多的顶点 v_i;

b)删除关联矩阵中顶点 v_i 对应的行及该行中值为 1 的元素所在列;然后在剩下的关联矩阵中再次删除所有行元素之和不超过 1 的其他行及这些行中值为 1 的元素所对应的列,直到不能再删除新的行和列为止;

c)重复以上步骤,直到所有的边都被包含到。

以图 2 为例,对上述近似算法思想进行形象说明:

a)计算关联矩阵 A 的每行元素之和 $\sum_{j=1}^{m} a_{ij}$,当 $i=1,2,3,4$ 时,每行元素之和分别为 2,3,2,3;可知当 $i=2$ 或 $i=4$ 时,对应行的元素之和最大,即对应顶点覆盖的边最多,所以在这里可以选取 v_2 或 v_4,假定本文选取的是 v_2。

b)删除关联矩阵中 v_2 对应的行及该行中值为 1 的边,即删除的是边 e_2,e_3 和 e_4,最后得到剩下的关联矩阵为

	e_1	e_5
v_1	1	0
v_3	0	1
v_4	1	1

c)在这个关联矩阵中再依次删除所有行元素之和不超过 1 的其他行,如 v_1 和 v_3 行,以及这两行中值为 1 的元素所对应的列,如 e_1 和 e_5 列,最后得到的剩余关联矩阵 $A=0$。于是得到该近似算法的最小弱顶点覆盖集 $S=\{v_2\}$,即只要测量一个节点即可得到各条链路的流量。

虽然在此图例中 S 是最优解,但是随着问题规模的扩大,通过此近似算法求解得出的最小弱顶点覆盖集将是一个近似解,而不是一个较优解更不可能是一个最优解。所以为了求解最小弱顶点覆盖问题的最优解,必须在此近似解的基础上,继续使用 TSBACO 算法对其进行全局寻优。

4.3 随机比例状态转移规则

ACO 算法在求解大多数问题时,主要是通过蚁群对边的集合进行搜索,因此信息素和期望启发信息都被设置在边上。而最小弱顶点覆盖问题的求解则是通过对顶点的集合进行的搜索,所以在本文提出的 TSBACO 算法中,信息素释放在顶点上,而不是在边上,同时还使用了期望启发信息来引导蚂蚁的搜索。在蚂蚁 k 构造弱顶点覆盖集的过程中,对顶点的选择将按如下概率选择式进行:

$$p_i^k = \tau_i^{\alpha}\eta_i^{\beta} / \sum_{j\in C_k} \tau_j^{\alpha}\eta_j^{\beta} \tag{1}$$

其中:τ_i 表示顶点 v_i 的信息素轨迹强度;η_i 表示顶点 v_i 的期望启发信息;C_k 表示蚂蚁 k 的候选顶点集;α 为信息启发式因子;β 为期望启发式因子,这两个参数都是常量,其值的选取需要人为经验的指定。

顶点的期望启发信息是静态的,其值在算法运行前已确定,即顶点的期望启发信息值在蚁群的寻优过程中保持不变。虽然这种静态定义方式不需要反复计算顶点的度,并提高了算法运算速度,但当图中各顶点的度相差过大或过小时,就会失去对蚁群搜索的指导意义。所以这里,对顶点的期望启发信息采用了动态定义的方式。由于顶点的期望启发信息值是顶点的度,当与该顶点相连的顶点被选择加入弱顶点覆盖集时,该顶点的度要减 1。顶点 v_i 的期望启发信息 η_i 表达式如下:

$$\eta_i = |v_i| \tag{2}$$

当顶点 v_i 被蚂蚁 k 选择后,笔者将顶点 v_i 加入蚂蚁 k 的部分解集合 S 中,同时将顶点 v_i 和与该顶点相连的所有边都删除;然后按照上述近似算法思想更新关联矩阵,继续删除顶点和边,直到没有任何可以删除的顶点和边。重复以上步骤直到关联矩阵 $A=0$,即所有边都已被部分解集合 S 中的顶点所覆盖,那么这个集合 S 就是蚂蚁 k 构造的弱顶点覆盖集。

4.4 信息素更新规则

信息素更新规则有多种形式,本文提出的求解最小弱顶点覆盖问题的 TSBACO 算法采用的是 MMAS[3] 的信息素更新规则,其主要特点在于:

a)在每次蚁群循环后,只有循环最优蚂蚁可以进行信息素释放;

b)各顶点的信息素轨迹强度被限制在$[\tau_{\min},\tau_{\max}]$范围内;

c)各顶点的信息素轨迹强度被初始化为 $\tau_{\max}$。

MMAS 通过这种信息素轨迹强度范围控制,可以防止个别顶点上的信息素轨迹强度过度增加或减少,从而可以有效避免算法出现早熟停滞现象。

下面针对求解最小弱顶点覆盖问题,详细给出 TSBACO 算法的信息素更新规则:

首先需要为网络拓扑图中的每个顶点 v_i 设置一个信息素轨迹强度初始值 $\tau_{\max}$,而信息素轨迹强度的更新则是在蚁群完成一次循环之后。为了充分利用循环最优解,每次循环后只有循环最优蚂蚁可以进行信息素释放。为了模拟信息素挥发,图中所有顶点的信息素轨迹强度都要乘以一个信息素挥发系数 ρ;然后,循环最优蚂蚁才会释放信息素。所以顶点信息素的更新分为两步进行:

a)顶点信息素的挥发

$$\tau_i(t+1) = (1-\rho)\tau_i(t) \tag{3}$$

b)顶点信息素的增加

假定 WVC_e 是某次循环中蚁群找到的包含顶点数最少的弱顶点覆盖,即循环最优解;而 WVC_g 则是从算法运行以来蚁群找到的包含顶点数最少的弱顶点覆盖,即全局最优解。在蚁群完成每次循环后,循环最优蚂蚁把一定量的信息素释放在 WVC_e 中的各个顶点上。通过循环最优蚂蚁信息素的释放,循环最优解 WVC_e 中的顶点将得到大量的信息素增强。顶点信息素增量表达式如下:

$$\Delta\tau_i = Q/(1+\|WVC_e\|-\|WVC_g\|) \tag{4}$$

其中:Q 为信息素增量系数,其值由经验所定;$|WVC_e|$ 和 $|WVC_g|$ 分别表示 WVC_e 和 WVC_g 中的顶点个数。

4.5 禁忌表的使用

禁忌搜索策略中禁忌表的使用就是要使算法尽量避免重复迂回搜索,这样不仅能加强搜索的多样性而且能使算法跳出局部最优解。在禁忌搜索框架下,多样性的加强是通过一些移动的暂时禁止获得的。针对最小弱顶点覆盖集问题,笔者将蚂蚁完成一次最小弱顶点覆盖集的构造称为一次移动。只要一次移动完成,那么在随后的 L 次循环中,这次移动将被禁止,即在以下连续的 L 次循环中,具有相同结构的弱顶点覆盖集将被视为禁忌对象。禁忌搜索策略的实现则是要通过禁忌表的使

用,其禁忌对象的替换是采用先进先出的方式。蚂蚁每完成一次移动,这次移动就被放入禁忌表,并将刚放入禁忌表的移动禁忌初始值记为 L,这样每经过一次循环,禁忌表中每个移动的禁忌值均减 1,当移动的禁忌值变为 0 时,这个移动就要从禁忌表中删除并被解禁。禁忌表长度是一个很重要的参数,它决定禁忌对象的任期,其大小直接影响整个算法的搜索进程和行为。所以在算法求解过程中,如果某些弱顶点覆盖集的结构重复频繁出现,那么就应该增加禁忌表的长度;如果禁忌表的长度已经影响到了算法的性能,那么就减少禁忌表的长度。

4.6　藐视准则

TSBACO 中的藐视准则定义如下:当候选顶点集中没有可选非禁忌点,或者存在某个特殊的禁忌点,它的选择会比其他非禁忌点的选择有更高效的优化性能,那么这个顶点就会在本次移动中被选择,并放入所求最小弱顶点覆盖集中。为了判断顶点的禁忌状态,在蚂蚁构造新弱顶点覆盖集时加入了一个顶点禁忌状态核对函数 $f(v_i)$。如果顶点 v_i 的返回值是 1,那么这个顶点就被禁止,否则就是可选的。

4.7　TSBACO 算法的伪代码

```
初始化信息素轨迹
repeat
  for 每只蚂蚁 k
    蚂蚁 k 随机选择一个起始顶点 v_i ∈ V,并把起始顶点加入最小弱顶点覆盖集 WVC_k
    repeat
      构造一个非禁忌候选顶点集 C_k
      按照式(1)的概率选择函数 p_j^k
      从 C_k 中选择下一个顶点 v_j ∈ C_k
      把顶点 v_j 加入最小弱顶点覆盖集 WVC_k
    until 最小弱顶点覆盖集 WVC_k 被构造完成
  end for
  保存循环最优解 WVC_c
  如果|WVC_c| < |WVC_g|,那么|WVC_g| = |WVC_c|,即更新全局最优解 WVC_g
  把禁忌初始值为的 WVC_c 加入禁忌表
  修改禁忌表中各对象的任期
  删除禁忌值为 0 的禁忌对象
  按照更新规则进行信息素轨迹更新
until 最优解被找到或已完成最大循环次数
返回全局最优解,即|WVC_g|
```

4.8　算法正确性分析

定理 1　若图 $G=(V,E)$ 是简单连通无向图,且满足对任意 $v\in V$ 有 degree$(v)\geqslant 2$,则 TSBACO 算法所得的顶点集是图 G 的一个弱顶点覆盖集。

证明　TSBACO 算法中每只蚂蚁在每次循环中进行的搜索实际相当于如下:

a)令 $G'=G$;

b)从图 G'中选择状态转移概率最大的顶点,记录该顶点,并且删除该顶点和该顶点的所有邻边,记生成的图为 G'';

c)在图 G''中,对于所有度为 1 的顶点,删除该顶点及该顶点的一条邻边,仍记生成的图为 G'';

d)重复 c),直到图 G''中所有顶点的度都大于 1 或者图 G''中没有任何顶点;

e)令 $G'=G''$,转回 b),直到图 G'中没有任何顶点。

由前面定义 2 可知,图 G 也是其本身的弱顶点覆盖集。

由上述的 c)可知,当一个顶点只剩一条邻边时,该条边也将被删除,这与定义 2 中的 b)相符,当且仅当该顶点的所有邻边均被删除后,该顶点才被删除,即该顶点的所有邻边都已被记录的所有顶点覆盖。

因此所有已经记录了的顶点和图 G'中的剩余顶点构成了图 G 的一个弱顶点覆盖集。由于每次蚁群循环结束时,图 G'中没有任何顶点,因此每次循环结束时,每只蚂蚁记录的顶点集合便构成了图 G 的一个弱顶点覆盖集。算法最终是从每次循环每只蚂蚁产生的顶点记录列表中选取顶点个数最少的顶点记录集,因此该顶点记录集也是图 G 的一个弱顶点覆盖集。

证明完毕。

5　仿真实验

为验证 TSBACO 算法在实际网络流量有效测量点选择中的优越性能,本文参考文献[1]进行了仿真实验,并比较了几种有效测量点选择算法。实验中所用到的网络拓扑图是通过 Waxman 网络模型生成的,Waxman 网络模型是网络研究中一种比较流行的拓扑生成模型。在 Waxman 模型中,n 个网络节点随机分布在笛卡尔空间,考虑节点间所有可能的连接,网络链路以一定概率方式加入网络拓扑图中。节点 i 和 j 之间的边按概率 $p(i,j)$ 生成,其概率表达式如下:

$$p(i,j)=\lambda e^{-d/(\gamma L)} \tag{5}$$

其中:参数 $\lambda>0$,用于控制拓扑图中边的密度;$\gamma\leqslant 1$,用于控制拓扑图中节点的平均度数;e 为自然指数;d 为节点 i 和 j 之间的距离;L 为各个节点间的最长距离。

由此可见,通过设置 Waxman 模型的三个参数 n、λ、γ,即可得到各种不同的网络拓扑结构图。下面设置四组参数,生成四种不同的网络拓扑结构图,其参数设置如表 1 所示。

表 1　Waxman 模型参数设置

	n	λ	γ
1	500	0.4	0.02
2	500	0.4	0.04
3	500	0.4	0.06
4	500	0.4	0.08

基于以上参数设置生成的四种网络拓扑结构图,比较了六种算法,分别是:文献[1]中提到的三种算法,即生成简单顶点覆盖的最大匹配启发算法(VC)、生成弱顶点覆盖的最大匹配启发算法(WVC)、生成弱顶点覆盖的贪婪秩算法(greedy rank,GR);文献[4]中提出的 SMN 算法;标准 ACO 算法以及本文提出的 TSBACO 算法。

TSBACO 算法的参数设置如下:蚂蚁数量 $a=30$,信息素挥发系数 $\rho=0.02$,信息启发因子 $\alpha=2$,期望启发因子 $\beta=1$,信息素上界 $\tau_{max}=10$,信息素下界 $\tau_{min}=0.1$,信息素增量系数 $Q=0.5$,最大循环次数 $c=500$,禁忌长度 $L=10$。

六种网络流量测量点选择算法的比较如表 2 所示。

表 2　六种网络流量测量点选择算法的比较

avg degree	N_{match}^{vc}	N_{match}^{WVC}	N_{rank}^{WVC}	N_{sun}^{WVC}	N_{aco}^{WVC}	N_{tsbaco}^{WVC}
4.4	387	255	165	129	124	121
8.6	441	372	254	221	216	213
12.6	453	408	307	296	292	287
16.9	466	431	334	327	322	318

表 2 中的第一列为网络拓扑图中节点的平均度数,它随着 γ 参数值的增大而增大;N_{match}^{vc},N_{match}^{WVC},N_{rank}^{WVC},N_{sun}^{WVC},N_{aco}^{WVC} 和 N_{tsbaco}^{WVC} 表示由上述六种算法得到的测量点个数。由表中的实验结果可以看出,TSBACO 算法在四种网络拓扑图中选取的节点个数比前五种算法都少,即 TSBACO 算法可以生成更小的弱顶点覆盖集,而 VC、WVC、GR、SMN 这些近似算法往往求得的是次优解而不是较优解和最优解。再看 N_{aco}^{WVC} 和 N_{tsbaco}^{WVC} 列,(下转第 1486 页)

个比特,能够大大缩减报文长度。

3　理论分析及模拟实验

3.1　理论分析

首先假设教育资源网格中转发查询请求的节点数为 N 即覆盖率 $C=1$,共存在 E 条边,节点的平均度为 d 且 $d=2E/N$。根据文献[7]中的引理2,图中边的条数 E,可以由一个关于节点数 N 和指数 e 的函数计算得出:

$$E=[1/2(e+1)](1-1/N^{e+1})N \tag{4}$$

将式(4)代入 $d=2E/N$,得到

$$d=[1/(e+1)](1-1/N^{e+1})\rightarrow 1/(e+1)(N\rightarrow\infty) \tag{5}$$

从上可以看出 d 与 N 无直接关系。

显然,flooding 机制中所转发的查询请求个数为

$$FMsg_{total}=\sum_{i=1}^{N}(d_i-1)=N(d-1) \tag{6}$$

基于路由标注的检索策略中,每一次新加入目标集合的节点都在下一轮的转发中被删除,检索过程中转发的查询请求个数为

$$LMsg_{total}=\sum_{i=1}^{N}(d_i-v_i)=Nd-\sum_{i=1}^{N}v_i \tag{7}$$

其中:v_i 为节点 i 的邻节点中已加入目标集合的节点个数。

3.2　模拟实验

利用VC ++开发模拟工具,验证本文算法的有效性。首先由模拟器产生出教育资源网格的参与转发查询请求的节点分布(图5),然后在节点个数分别为 200 ~ 1 000 内,依据冗余信息数和系统传输开销两个指标,比较本文算法 L_flooding 和 flooding 算法的性能。为了提高实验数据的精确度,每次结果数据均取 10 次相同实验的平均值。实验结果如图6、7 所示。

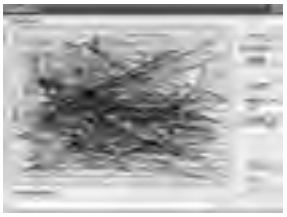


图 5　Node=200 模拟节点分布图

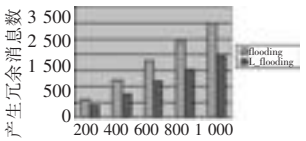


图 6　冗余消息数与系统规模

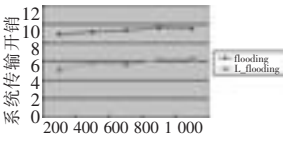


图 7　传输开销与系统规模

从实验结果得出,在同等系统规模下,本文策略产生的冗余消息数比 flooding 机制减少近 1/3,且其传输开销明显优于 flooding,即可节省大量网络带宽,能够适用于分布式系统的信息检索。

4　结束语

本文在研究教育资源网格架构的基础上,为了减少检索过程中冗余消息的产生,通过在消息报文中添加节点信息,提出了一种基于路由标注的信息检索策略。通过实例分析和仿真实验表明,该策略不降低算法覆盖度的情况下,能够有效减少由于 flooding 引起的冗余传输开销,节省网络带宽,能够适用于教育资源网格中的信息检索。

参考文献:

[1] 窦文,王怀民. 模拟 Rumor 传播机制的无结构 P2P 网络中广播机制的研究[J]. 计算机研究与发展,2004,41(9):1460-1465.

[2] 刘弘,刘希玉. 一种 Web 信息的启发式检索方法[J]. 小型微型计算机系统,2003,24(3):427-429.

[3] 张龙,李巍,李云春. 基于改进 DHT 算法的分布式资源发现模型的研究[J]. 计算机应用研究, 2007,24(12):313-316.

[4] RIPEANU M. Peer-to-peer architecture case study: Gnutella network [C]//Proc of the 1st International Conference on Peer-to-Peer Computing. Linkoping: IEEE Computer Society,2001:99-100.

[5] 谢鲲,张大方,谢高岗,等. 基于轨迹标签的无结构 P2P 副本一致性维护算法[J]. 软件学报,2007,18(1):105-116.

[6] MITZENMACHER M. Compressed Bloom filters [J]. IEEE/ACM Trans on Networking,2002,10(5):604-612.

[7] FALOUTSOS M, FALOUTSOS P, FALOUTSOS C. On Power-law relationships of the Internet topology[C]//Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. New York: ACM Press,1999:251-262.

(上接第 1483 页)由于 TSBACO 算法中禁忌表的使用,使得算法既能避免重复迂回搜索,跳出局部最优解,又能对不同搜索区域进行更有效地探索;相比 ACO 算法易陷入局部最优和早熟收敛的缺点,TSBACO 算法能以更大的概率收敛到全局最优解,求解出节点数更少的弱顶点覆盖集。

6　结束语

本文提出了一种求解网络流量有效测量点选择问题的混合优化算法——基于禁忌搜索的蚁群优化算法 TSBACO。该算法利用 ACO 算法的正反馈机制加快了收敛速度,利用 TS 算法的全局搜索能力避开了局部最优。仿真实验结果表明,TSBACO 算法是求解网络流量测量中的一种有效的测量点选择算法,该算法求解质量较高,而且拥有较好的可扩展性和实用性。在 TSBACO 算法中,禁忌表长度 L 是一个很重要的参数,其大小直接影响真个算法的性能,如何选择 L 使之能更好的求解最小弱顶点覆盖问题,则是今后的研究重点之一。

参考文献:

[1] BREITBART Y, CHAN Che-yong, GAROFALAKIS M, *et al*. Efficiently monitoring bandwidth and latency in IP networks [C]//Proc of IEEE INFOCOM. Anchorage, Alaska: IEEE Press,2001:933-942.

[2] 刘湘辉,殷建平,唐乐乐,等. 网络流量的有效测量方法分析[J]. 软件学报,2003,14(2):300-304.

[3] STUTZLE T, HOOS H. Max-min ant system[J]. Future Generation Computer Systems,2000,16(8):889-914.

[4] 蒋红艳,林亚平,黄生叶. 网络流量有效监测点的设置模型及求解算法研究[J]. 电子与信息学报, 2006,28(4):753-756.

[5] BATTITI R, PROTASI M. Reactive local search for the maximum clique problem[J]. Algorithmica,2001,29(4):610-637.

[6] FENET S, SOLNON C. Searching for maximum cliques with ant colony optimization[J]. Applications of Evolutionary Computing, 2003, 26(11):236-245.

[7] HUA Zhen, FAN Hui, LI Jin-jiang, *et al*. Solving point covering problem by ant algorithm [C]//Proc of International Conference on Machine Learning and Cybernetics. Shanghai: [s. n.],2004:3501-3504.

[8] 刘湘辉,殷建平,唐乐乐,等. 基于弱顶点覆盖的网络链路使用带宽监测模型[J]. 软件学报,2004,15(4):545-549.