

采用启发式算法的无线自组网网络编码方法^{*}

焦贤龙⁺, 王晓东, 周兴铭

(国防科学技术大学 计算机学院 并行与分布处理国家重点实验室,湖南 长沙 410073)

Network Coding Method Using Heuristic Algorithms for Wireless Ad Hoc Networks

JIAO Xian-Long⁺, WANG Xiao-Dong, ZHOU Xing-Ming

(National Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: jiaoxianlong1982@gmail.com

Jiao XL, Wang XD, Zhou XM. Network coding method using heuristic algorithms for wireless ad hoc networks. Journal of Software, 2010,21(11):2892–2905. <http://www.jos.org.cn/1000-9825/3640.htm>

Abstract: This paper first proves that the optimizing problems concerning network coding are NP-complete, and then puts forward a network coding method called COMP, using heuristic algorithms. This network coding method adopts the basic ideas of the set packing heuristic algorithm and the set cover heuristic algorithm with the aim of digging up as many network coding opportunities as many as possible. The results of the simulation carried out by the NS-2 simulator show that in scenarios where there are a large number of nodes, a small maximum transmission range, and a large number of sessions, this network coding method effectively alleviates the impact of session concurrency, improves the performance of the existing broadcast algorithm, and its performance improvement is greater than that of the existing network coding method.

Key words: wireless ad hoc network; network coding; greedy set packing; greedy set cover; broadcast

摘要: 首先证明了与应用网络编码相关的最优化问题是 NP 完全的,然后提出了一种采用启发式算法的网络编码方法 COMP.该网络编码方法利用贪婪集合配置启发式算法和贪婪集合覆盖启发式算法的基本思想来尽可能多地挖掘网络编码机会.NS-2 仿真结果表明,该网络编码方法在应用到节点数比较多、最大传输范围比较小以及会话数比较多的场景中时,有效地减轻了并发会话的影响,提高了现有广播算法的性能,而且其性能提高超过了现有的网络编码方法.

关键词: 无线自组网;网络编码;贪婪集合配置;贪婪集合覆盖;广播

中图法分类号: TP393 **文献标识码:** A

无线自组网因其部署简单,采用分布式计算以及不依赖固定基础设施而被广泛应用到灾难救护和军事对抗等场合中.远距离节点之间的通信可以通过中继节点的多跳转来完成,从而降低了对单个节点信号发送设备的要求.但是,无线信道之间的竞争和无线信号的冲突等容易制约无线自组网的性能提高,而并发会话是导致无线自组网中信道竞争、信号冲突和冗余重传的关键因素之一.为此,许多研究者将注意力集中于减轻并发会

* Supported by the National Basic Research Program of China under Grant No.2006CB303000 (国家重点基础研究发展计划(973))

Received 2009-01-17; Accepted 2009-04-27

话的影响,从而提出了网络编码这种有效手段.

网络编码的概念首先是由 Ahlswede 等人于 2000 年提出来的^[1].其主要思想是,允许中间转发节点在发送报文之前对报文进行组合,然后将组合后的报文发送给目的节点,目的节点可通过解码操作获得所需要的报文,从而可以减少报文的传输次数,提高网络的性能.网络编码最初是应用在有线网络中^[1-3],近几年来,许多研究者也将网络编码方法应用到无线网络中来提高网络吞吐量、鲁棒性、安全性和能量效率等^[4-10].现有的研究工作都已证明网络编码是减轻并发会话对网络性能影响的有效手段,而与路由算法、队列调度以及流媒体技术等领域的结合也将是网络编码发展的重要趋势^[11].

研究无线网络中如何应用网络编码最著名的工作之一是文献[4],其中提出了一种基于异或操作的简单实用的网络编码方法 COPE,针对的是单播路由问题.文献[5]也是应用到单播路由中的网络编码方法,其中也提出了许多针对广播或者多播路由问题的网络编码方法^[6-9].但是,很少有研究者对网络编码方法的最优性进行过系统的分析,从而其提出的网络编码方法往往不能获得最优的性能.正如文献[9]中所说的,无论是针对单播路由问题还是针对广播路由问题,计算出满足网络编码条件的最佳数量的报文组合都是 NP 难的.然而,文献[9]并没有给出系统的证明,而且也没有采用针对这些 NP 难问题的启发式算法.

因此,本文首先证明了将网络编码方法应用到无线自组网的广播路由中所涉及的最优化问题是 NP 完全的,然后提出了一种采用启发式算法的网络编码方法 COMP.该网络编码方法利用贪婪集合配置启发式算法 GSPHA(greedy set packing heuristic algorithm)和贪婪集合覆盖启发式算法 GSCHA(greedy set cover heuristic algorithm)的基本思想来尽可能多地挖掘网络编码机会.本文将 COMP 网络编码方法应用到 Ingelrest 等人提出的 LBIP 广播算法上^[12].因为 COMP 网络编码方法只是基于局部的网络拓扑信息,所以可以应用到任何其他分布式广播算法上.LBIP 广播算法是著名的 BIP 广播算法^[13]的分布式实现版本,解决的是无线自组网中的广播能耗问题,而且其性能非常接近 BIP 广播算法的性能.最后,本文使用比较流行的 NS-2 模拟器实现了 COMP 网络编码方法,并与 COPE 网络编码方法进行了比较.实验结果表明,COMP 网络编码方法在应用到节点数比较多,最大传输范围比较小以及会话数比较多的场景中时有效地减轻了并发会话的影响,提高了现有广播算法 LBIP 的性能,而且其性能提高超过了现有的网络编码方法 COPE.

本文第 1 节对研究网络编码的相关工作进行简要地评述.第 2 节介绍本文所采用的网络模型和所分析的问题.第 3 节详细论述提出的 COMP 网络编码方法.第 4 节给出 NS-2 模拟实验的结果以及详细的性能分析.第 5 节对本文的工作进行总结并展望下一步的工作.

1 相关工作

文献[1]以最大流最小割定理来引入网络编码的概念,并证明了网络编码可以节省网络带宽,提高网络吞吐量.Li 等人针对多播路由提出了一种非常容易实现的线性网络编码方法^[2].文献[3]通过引入一种代数学框架来研究应用线性网络编码方法的网络容量问题.以上网络编码方法都是针对有线网络,而最近的一些研究工作也证明了网络编码方法也可以解决无线网络中的许多问题.例如,Katti 等人提出了无线网络中一种基于异或操作的简单实用的网络编码方法^[4].该网络编码方法针对单播路由,主要包括机会监听和机会编码两个部分.机会监听是指无线网络中的节点监听其周围的无线通信,存储收到的报文,并构建邻居接收表,存储邻居节点已收到的报文信息.机会编码则是指节点在发送报文之前根据邻居接收表和待发送的报文集合对报文进行编码,然后发送编码后的报文,而其邻居节点通过简单的异或操作就可以解码出所需要的报文.另外,文献[5]针对无线网络中的单播路由问题提出了一种节能的机会网络编码方法.

文献[6-9]以多播路由或者广播路由为研究问题,分别提出了一些有效的网络编码方法,目的是降低多播路由或者广播路由的能耗以及提高报文成功接收率等.例如,文献[6]提出了一些分布式算法来解决采用网络编码方法的最低开销多播问题,Li EL 等人在文献[9]中针对广播路由问题提出了两种网络编码方法.这两种网络编码都只依赖于局部的两跳拓扑信息,并且利用机会监听来减少传输次数.一种网络编码方法基于简单的异或操作,类似于 COPE 网络编码方法.另一种网络编码方法基于 Reed-Soloman 编码,性能比前一种方法更好.但是,基

于 Reed-Soloman 的网络编码方法需要对报文进行分组,从而会引入一定的开销并且增加解码延迟.另外,这种网络编码方法要求每个邻居节点缺少的报文数不能低于某个特定的值,从一定程度上减少了应用网络编码的机会.本文提出的 COMP 网络编码方法基于简单的异或操作,所以加、解码都非常容易实现,而且不需要对报文进行分组,因此开销也较小.

2 问题描述

2.1 网络模型

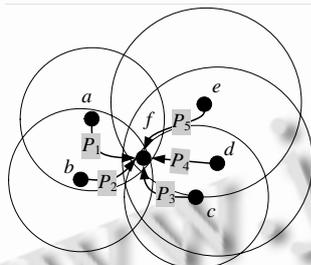
本文将无线自组网建模成一个无向图 $G=(V,E)$.其中: V 表示网络中的节点集合,这些节点随机地分布于一个矩形区域上,并且假设都是静止的; E 包含的是表示两个节点之间可以通信的边集.本文假设每个节点都知道自己的地理位置信息,因为这种信息可以很容易地通过某种定位技术来获取,例如比较普及且价格便宜的 GPS 定位技术.每个节点可以调整它的发送功率,从而可以调整其传输范围.为了简单起见,本文只考虑节点发送报文所消耗的能量.本文参照文献[13]来定义两个节点 i 和 j 之间通信所需的功率 P_{ij} :

$$P_{ij} = cr_{ij}^\alpha \tag{1}$$

其中, r_{ij} 表示节点 i 和节点 j 之间的欧式距离, α 是传播损失系数, c 是一个常数.不失一般性,本文将 α 规格化为 2, 并将 c 规格化为 1.

2.2 问题分析

本文以图 1(a)所示的包含 6 个节点的网络为例来介绍所分析的问题.假设节点 a,b,c,d 和 e 需要广播各自的报文 P_1,P_2,P_3,P_4 和 P_5 .节点 f 是这些节点的转发节点.图 1(b)给出了节点 f 的邻居接收表,该表描述了节点 f 的邻居节点的报文接收情况,表中的 1 表示邻居节点已收到对应的报文,而 0 表示未收到对应的报文.如果不采用网络编码方法,节点 f 需要传输 5 次报文.如果采用一种简单的基于异或操作的方法,例如类似于 COPE 的网络编码方法,则可以将 P_1 和 P_4 两个报文编码成一个报文 $Q_1=P_1\oplus P_4$,然后将编码后的报文 Q_1 发送出去.节点 a 和 b 收到 Q_1 报文后可以通过执行 $Q_1\oplus P_1$ 解码出 P_4 ,而节点 c,d 和 e 则可以通过 $Q_1\oplus P_4$ 解码出 P_1 .因为节点 f 只需要发送 Q_1,P_2,P_3 和 P_5 这 4 个报文,所以总的传输次数降为 4.但是,对报文的编码存在多种选择,例如也可以将报文 P_2 和 P_4 进行编码.如果按照 COPE 以最大可能地减少传输次数为目的进行编码,则需要计算数量最多的一组报文,从而所有邻居节点在收到对这组报文进行编码后的报文时都可通过一次异或操作解码出所需要的报文,即所有邻居节点最多缺少这组报文中的一个报文.然而,本文将在第 3.1 节中证明,该最优化问题是 NP 完全问题,所以只能给出解决该最优化问题的启发式算法.



(a) A six-node wireless ad hoc network
(a) 包含 6 个节点的无线自组网

	P_1	P_2	P_3	P_4	P_5
a	1	1	0	0	0
b	1	1	0	0	0
c	0	0	1	1	0
d	0	0	1	1	1
e	0	0	0	1	1

(b) Neighbor reception table of node f
(b) 节点 f 的邻居接收表

Fig.1 An example for adopting network coding method

图 1 采用网络编码方法的示例

另外,COPE 无法进一步对 P_2,P_3 和 P_5 这 3 个报文进行处理,因为对这些报文的任意组合都不满足 COPE 的报文编码条件.然而,可以考虑使用以下的方法来进一步减少传输的次数.节点 f 分别发送以下两个编码后的报

文: $Q_2=P_2\oplus P_3$ 和 $Q_3=P_3\oplus P_5$,另外,假设节点不立即丢弃无法解码的报文,而是存储一段时间,等待以后进行解码.首先,节点 a 和 b 在收到编码后的报文 Q_2 后可以通过一次简单的异或操作 $Q_2\oplus P_2$ 来解码出它们没有的报文 P_3 ,而节点 c 和 d 也可以通过 $Q_2\oplus P_3$ 得到报文 P_2 .节点 e 无法解码出任何报文,只是将编码后的报文 Q_2 存储一段时间.类似地,另一个编码后的报文 Q_3 被广播之后,节点 a,b 和 c 可以解码出报文 P_5 ,节点 e 则可以解码出报文 P_3 .通过再次分析节点 f 的邻居接收表可以发现,除了节点 e 之外,所有其他节点都解码出了它们没有的报文,而节点 e 只缺少报文 P_2 .另外,节点 e 还存储了一个没有解码的报文 Q_2 ,而报文 P_3 已经被它解码出,所以它只需要再执行一个简单的异或操作 $Q_2\oplus P_3$,就可以得出报文 P_2 .通过采用这种方法,总的传输次数可以减少为 3.本文将在第 3.2 节中讨论使用这种方法的条件,并且证明与该方法相关的最优化问题也是 NP 完全问题,并给出解决该问题的启发式算法.

3 网络编码方法 COMP

本节将详细描述网络编码方法 COMP.该网络编码方法利用类似于 COPE 的基于异或操作的方法,并且包括机会监听和机会编码两个组成部分.与 COPE 不同,COMP 网络编码方法包括两种启发式算法:一种启发式算法 GSPHA 基于贪婪集合配置算法,主要用于计算待发送的报文集中满足 COPE 网络编码条件的基数最大的报文子集,从而转发节点可以将报文子集中的报文直接编码成一个报文发送出去,而其邻居节点只需要通过一次异或操作就可以解码出所需要的报文;另一种启发式算法 GSCHA 利用贪婪集合覆盖算法的基本思想以及第 2.2 节中所介绍的方法,查找满足条件的基数最小的报文子集,从而转发节点可以对报文子集中的报文依次两两编码,而转发节点的所有邻居节点在收到这些报文后都可以解码出所需要的报文.

3.1 启发式算法 GSPHA

本节将证明将 COPE 的网络编码条件应用到广播路由中,其所涉及的最优化问题是 NP 完全的,其等价于最优集合配置问题.然后本文将给出一种启发式算法 GSPHA 来解决该问题.

定理 1. 给定节点 v 需要发送的一组报文集合 S ,邻居节点集 N 和邻居接收表 R ,查找 S 满足以下条件的基数最大的报文子集 S_x 是 NP 完全的:每个邻居节点至少拥有 S_x 的 $|S_x|-1$ 个报文.

证明:首先将这个最大化问题转换为一个判定问题,即对于给定的整数 k ,是否存在这样的子集 S_x ,它的基数为 k ,并且每个邻居节点都拥有 S_x 的 $k-1$ 个报文.定理 1 的证明分为两个部分:首先证明该最大化问题属于 NP 类,然后证明最优集合配置问题可以在多项式时间内规约为该问题.

首先,可以很容易判断出该问题属于 NP 类.因为给定 S_x 和邻居接收表 R ,可以非常容易地判断出 S_x 的基数是否为 k ,以及每个邻居节点是否拥有 S_x 的 $k-1$ 个报文.显然,这种判断可以在多项式时间内完成.最优集合配置判定问题描述如下:

定义 1. 给定包含 n 个元素的集合 X 和包含 m 个 X 的子集的族集合 F ,并且 X 中的每一个元素都属于 F 中的至少一个元素.最优集合配置判定问题分析,对于给定的整数 k , F 是否存在一个子集 U ,满足 U 的基数为 k ,并且 U 中的元素两两互不相交,即相交为空集.

最优集合配置问题是一个公认的 NP 完全问题,下面给出从最优集合配置问题到本定理所证明的最大化问题的转换.首先创建一个转发节点 v .如图 2(a)所示,对于 X 中的每个元素 $x_i(1\leq i\leq n)$,创建节点 v 的一个邻居节点 v_i .对于 F 中的每个元素 $f_j(1\leq j\leq m)$,创建一个待发送的报文 P_j .如果 f_j 包含 x_i ,则假设与 x_i 对应的邻居节点 v_i 没有与 f_j 对应的报文 P_j ,否则假设该邻居节点拥有该报文.基于这种转换规则,可以得到节点 v 的邻居接收表 R ,如图 2(b)所示.显然,这种转换可以在多项式时间内完成.下面将证明这种转换是从最优集合配置问题到本定理所证明的最大化问题的规约.

如图 3 所示,一方面,假设 F 存在一个基数为 k 的集合配置 $U=\{u_1,u_2,\dots,u_k\}$.考虑与该集合配置对应的报文子集 $S_x=\{P_{u_1},P_{u_2},\dots,P_{u_k}\}$.因为 U 是一个集合配置,所以没有两个元素存在非空交集.即对于 X 中的每一个元素 $x_i(1\leq i\leq n)$,如果 x_i 属于 U 中的某个元素 $u_j(1\leq j\leq k)$,则 x_i 不应该再属于 U 中的其他元素.所以,对于与 x_i 对应的

邻居节点 v_i , 如果 v_i 没有报文 P_{u_j} , 则它不应该再缺少 S_x 的任何其他报文; 否则, 假设 v_i 还缺少了报文 P_{u_t} ($1 \leq t \leq k, t \neq j$), 则可以基于转换规则推断出 x_t 同时属于 u_j 和 u_t , 与开始所作的假设 U 是集合配置矛盾. 从而根据 i 的任意性可以推断出, 每个邻居节点都至少拥有 S_x 的 $k-1$ 个报文.

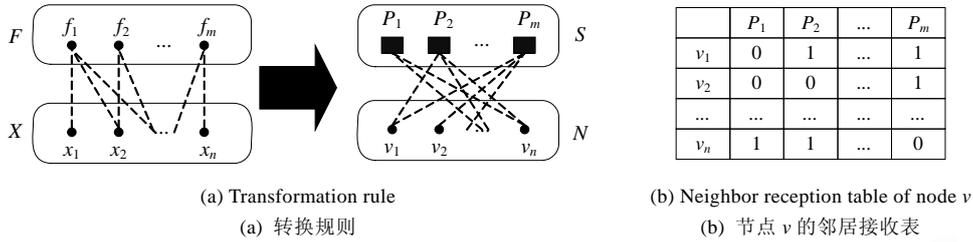


Fig.2 Transformation from the maximum set packing problem to the maximizing problem

图2 最优集合配置问题转换为最大化问题

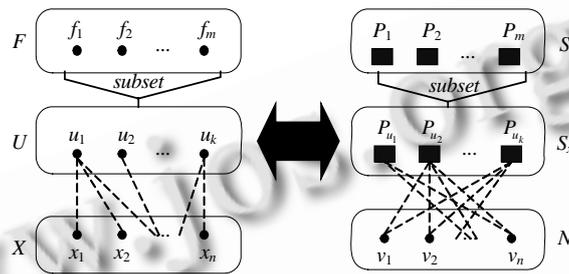


Fig.3 An example to explain the reduction process

图3 解释规约过程的示例

另一方面, 假设报文集合 S 存在一个子集 $S_x = \{P_{u_1}, P_{u_2}, \dots, P_{u_k}\}$ 满足: 该子集的基数为 k , 并且每个邻居节点都至少拥有 S_x 的 $k-1$ 个报文. 考虑与 S_x 对应的 F 的子集 $U = \{u_1, u_2, \dots, u_k\}$. 下面用反证法证明 U 是一个集合配置. 假设 U 的两个元素 u_i 和 u_j ($1 \leq i \neq j \leq k$) 交集不为空, 并且假设 X 中的元素 x_r ($1 \leq r \leq n$) 属于该交集. 根据转换规则, 与 x_r 对应的邻居节点 v_r 将同时缺少报文 P_{u_i} 和 P_{u_j} , 所以该邻居节点拥有 S_x 的报文数将少于 $k-1$, 这与假设矛盾. 综上所述, 定理 1 成立. \square

因为最优集合配置问题是 NP 完全问题, 而文献[14]中的贪婪集合配置算法获得比较小的近似比系数 $O(\sqrt{|X|})$, 其中, $|X|$ 表示集合 X 的基数, 所以本节提出的启发式算法 GSPHA 利用该贪婪集合配置算法的基本思想来解决定理 1 中所讨论的最大化问题. 因此, 对于该最大化问题来说, 近似比系数为 $O(\sqrt{|N|})$, 其中, $|N|$ 表示邻居节点数. GSPHA 算法和贪婪集合配置算法的伪代码如算法 1 和算法 2 所示. GSPHA 算法的基本思想是, 计算报文集合 S 的一组满足以下条件的子集: 转发节点可以直接对这些子集的报文进行异或操作, 而其邻居节点在收到编码报文后可以直接解码出所需要的报文.

算法 1. GSPHA(S, N, R).

1. $F = \emptyset$;
2. $M = \{1, 2, \dots, |S|\}$;
3. **foreach** $P_i \in S$ ($1 \leq i \leq |S|$)
4. creat the element f_i of F according to the transmission rule;
5. **endfor**
6. $X = N$;

7. $A=\emptyset$;
8. **repeat**
9. $D=Greedy-Set-Packing(X,F,M)$;
10. **if** $|D|>1$ **then**
11. add the packets corresponding to the numbers in D into a set, and then insert the set into A ;
12. remove these packets from F , and also remove the corresponding numbers from M ;
13. **endif**
14. **until** $|D|\leq 1$
15. return A

算法 2. $Greedy-Set-Packing(X,F,M)$.

1. $F'=F$;
2. $U=\emptyset$;
3. **repeat**
4. find $f_k\in F'$ with the minimum cardinality, where k is the corresponding number in M ;
5. $U=U\cup\{k\}$;
6. $Z=\{f\in F'|f\cap f_k\neq\emptyset\}$;
7. $F'=F'-Z$;
8. **until** $|F'|=0$
9. return U

在 GSPHA 算法的伪代码中,第 1 行将 F 初始化为空集.第 2 行中的集合 M 表示 F 中的元素与 S 中的报文的对应关系.第 3~5 行中的 for 循环计算出与每个报文 P_i 对应的 F 中的每个元素 f_i .第 6 行将集合 X 初始化为邻居节点集 N .第 7 行中作为结果返回的集合 A 包含 S 的满足条件的所有报文子集,初始为空集.第 8~14 行中的 repeat 循环利用 Greedy-Set-Packing 算法来查找满足条件的所有报文子集.需要注意的是, Greedy-Set-Packing 算法返回的是一组序号,这组序号与各个报文在集合 S 中的序号对应. Greedy-Set-Packing 算法的基本思想是,每一次从 F' 中选择基数最小的元素,然后移除与该元素相交不为空集的元素,直到 F' 中没有元素为止.下面将证明 GSPHA 算法的时间复杂度为 $O(|M||S|^2\log_2|S|)$.

定理 2. GSPHA 算法的时间复杂度为 $O(|M||S|^2\log_2|S|)$.

证明:因为在该算法中,第 3~5 行的 for 循环和第 8~14 行中的 repeat 循环所花费的时间最多,所以只分析这两个循环的时间复杂度.首先,因为构造 F 中的每一个元素 f_i 都要遍历所有邻居节点,而 F 中总共包含 $|S|$ 个元素,所以可以分析出第 3~5 行中的 for 循环其时间复杂度为 $O(|M||S|)$.因为第 8~14 行中的每一次 repeat 循环都至少从 S 中选出两个元素,而最坏的情况是每一次都选出两个元素,因此,该循环最多迭代 $\log_2|S|$ 次.另外,该 repeat 循环体中 Greedy-Set-Packing 算法的时间复杂度最高.分析 Greedy-Set-Packing 算法可以得出,该算法中的 repeat 循环体的第 6 行需要 $O(|M||S|)$ 时间来实现,这也是该循环体中耗时最多的操作,而 repeat 循环最多迭代 $O(|S|)$ 次,所以可以判断出 Greedy-Set-Packing 算法的时间复杂度为 $O(|M||S|^2)$,从而可以计算出 GSPHA 算法的时间复杂度为 $O(|M||S|^2\log_2|S|)$.因此,定理 2 成立. \square

3.2 启发式算法 GSCHA

本节详细介绍在前面第 2.2 节中所讨论的对报文依次两两编码所需要的条件,所涉及的最优化问题以及启发式算法 GSCHA.重新回到图 1(b)中可以发现,每个邻居节点至少拥有报文 P_2, P_3 和 P_5 中的一个报文.下面将证明这就是可以对这些报文依次两两编码的条件.

定理 3. 对于节点 v 和待发送的报文集合 S ,如果每个邻居节点都拥有 S 的某个子集 $S_b = \{P_{b_1}, P_{b_2}, \dots, P_{b_k}\}$ 中的至少 1 个报文,则可以使用所讨论的方法将 S_b 编码成 $\{P_{b_1} \oplus P_{b_2}, P_{b_2} \oplus P_{b_3}, \dots, P_{b_{k-1}} \oplus P_{b_k}\}$,然后发送出去,所有的邻居节点在收到所有这些编码报文后都可以解码出它们所需要的报文.

证明:以最坏的情况来考虑,假设某个邻居节点 u 只有 S_b 中的一个报文 $P_{b_i} (1 \leq i \leq k)$. 它在收到所有的编码报文后,可以通过 $P_{b_i} \oplus (P_{b_{i-1}} \oplus P_{b_i})$ 和 $P_{b_i} \oplus (P_{b_i} \oplus P_{b_{i+1}})$ 分别解码出 $P_{b_{i-1}}$ 和 $P_{b_{i+1}}$. 类似地,可以继续将 $P_{b_{i-1}}$ 与 $P_{b_{i-1}} \oplus P_{b_{i-2}}$ 进行异或操作得到 $P_{b_{i-2}}$, 然后再使用 $P_{b_{i-2}}$ 得到 $P_{b_{i-3}}$. 以此类推,最终可以得到的报文有 $\{P_{b_{i-1}}, P_{b_{i-2}}, P_{b_{i-3}}, \dots, P_{b_1}\}$. 另外,利用 $P_{b_{i+1}}$ 和类似的方法也可以获得报文 $\{P_{b_{i+2}}, P_{b_{i+3}}, \dots, P_{b_k}\}$. 最后,通过这种方法可以解码出所有缺少的报文. 因为这是最坏的情况,所以其他情况自然也成立,因此定理 3 成立. \square

因为每次找到一个满足条件的子集时就可以减少一次传输次数,所以,为了最大可能地减少传输次数,需要找出满足条件的基数最小的子集 S_b . 这样,在剩余的报文中就可以查找出更多满足条件的子集. 下面将证明计算出 S 的满足条件的基数最小的子集 S_b 是 NP 完全问题.

定理 4. 给定节点 v 待发送的报文集合 S , 邻居节点集 N 和邻居接收表 R , 计算出满足条件的基数最小的报文字集 S_b 是 NP 完全问题.

证明:首先将该最小化问题重新描述为一个判定问题,即对于给定的整数 k , 是否存在满足条件的基数为 k 的子集 S_b . 类似地,首先证明该问题属于 NP 类,然后证明最优集合覆盖问题可以在多项式时间内规约为该问题.

证明该最小化问题属于 NP 类所选择的证据是 S_b 和邻居接收表 R . 给定 S_b 和 R , 可以非常容易地使用一个多项式时间算法判断出 S_b 的基数是否为 k , 以及每个邻居节点是否拥有 S_b 的至少 1 个报文. 另外,将最优集合覆盖判定问题描述如下:

定义 2. 给定包含 n 个元素的有限集合 X 和包含 m 个 X 的子集的族元素 F , 并且 X 中的每个元素都属于 F 中的至少 1 个元素. 最优集合覆盖判定问题询问, 对于给定的整数 k , F 是否存在一个子集 C 满足基数为 k , 并且对于任意的 $x \in X, C$ 中都存在一个元素 c 使得 $x \in c$.

最优集合覆盖问题是一个公认的 NP 完全问题. 下面将给出从最优集合覆盖问题到本定理所证明问题的转换规则, 如图 4(a) 所示. 该转换规则类似于证明定理 1 时所采用的转换规则, 但是也有所区别. 首先创建一个转发节点 v . 对于 F 中的每个元素 $f_j (1 \leq j \leq m)$, 创建一个待发送的报文 P_j . 对于每个元素 $x_i \in X$, 创建一个邻居节点 v_i . 与定理 1 的转换规则不同, 如果 f_j 包含 x_i , 则假设与 x_i 对应的邻居节点 v_i 拥有与 f_j 对应的报文 P_j ; 否则, 该邻居节点缺少该报文. 基于这种转换规则, 可以得到节点 v 的邻居接收表, 如图 4(b) 所示. 显然, 这种转换可以在多项式时间内完成.

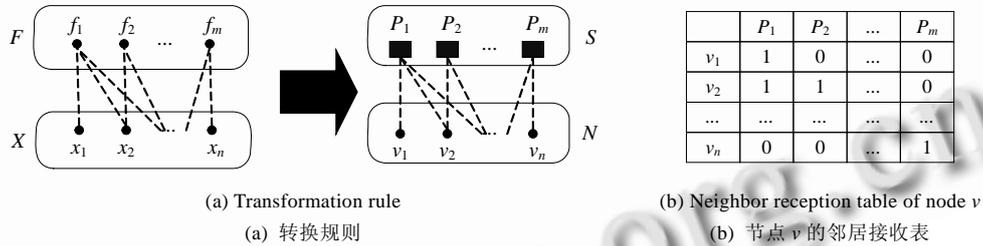


Fig.4 Transformation from the minimum set cover problem to the minimizing problem

图 4 最优集合覆盖问题转换为最小化问题

下面将证明这种转换是从最优集合覆盖问题到本定理所证明问题的规约, 即证明 F 存在基数为 k 的集合覆盖 C 当且仅当 S 存在一个满足条件的基数为 k 的子集 S_b . 图 5 描述了该规约过程.

一方面, 假设 F 存在一个基数为 k 的集合覆盖 C . 记 $C = \{c_1, c_2, \dots, c_k\}$, 考虑与其对应的报文字集 $S_b = \{P_{c_1}, P_{c_2}, \dots, P_{c_k}\}$. 因为 C 是一个集合覆盖, 所以对于 X 中的任意元素 $x_j (1 \leq j \leq n)$, C 中至少存在 1 个元素 $c_i (1 \leq i \leq k)$, 满足 $x_j \in c_i$. 所以, 根据转换规则, 与 x_j 对应的邻居节点 v_j 拥有与 c_i 对应的报文 P_{c_i} . 根据 j 的任意性, 可以推断出 S_b 满足, 每个邻居节点至少拥有该集合的 1 个报文, 并且该集合的基数等于 k .

另一方面, 假设 S 存在一个满足条件的基数为 k 的子集 $S_b = \{P_{c_1}, P_{c_2}, \dots, P_{c_k}\}$, 考虑与 S_b 对应的 F 的子集 $C = \{c_1, c_2, \dots, c_k\}$. 对于任意的邻居节点 $v_j (1 \leq j \leq n)$, 根据假设, 它拥有 S_b 的至少 1 个报文 $P_{c_i} (1 \leq i \leq k)$. 所以, 与 v_j 对

应的 X 中的元素 x_j 属于与 P_{c_j} 对应的元素 c_j . 根据 j 的任意性, 可以推断出 C 是 F 的一个集合覆盖, 并且基数等于 k . 综上所述, 定理 4 成立. □

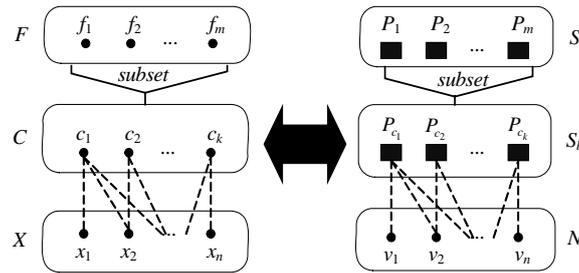


Fig.5 An example to explain the reduction from the minimum set cover problem to the minimizing problem

图 5 解释最优集合覆盖问题规约为最小化问题的示例

根据文献[15], 在 $P \neq NP$ 的情况下无法设计出一个近似比系数小于 $c \ln m$ 的多项式时间近似算法来解决最优集合覆盖问题(其中, c 为一个常数), 因而采用文献[16]中的贪婪集合覆盖算法的基本思想来解决定理 4 所讨论的最小化问题, 并提出启发式算法 GSCHA. 所采用的贪婪集合覆盖算法的近似比系数为 $H(\max\{|f|: f \in F\})$, 其中, $H(\lambda) = \sum_{i=1}^{\lambda} (1/i)$ 是一个调和函数. GSCHA 算法和贪婪集合覆盖算法的伪代码如算法 3 和算法 4 所示. GSCHA 算法的主要步骤类似于 GSPHA 算法的主要步骤, 区别在于, GSCHA 算法采用贪婪集合覆盖算法查找满足条件的基数最小的子集, 并且 F 中的元素构造规则也不同. 贪婪集合覆盖算法的基本思想是, 每一次循环选择覆盖剩余元素最多的集合 f_k , 直到 X 中的所有元素都被覆盖. 如果循环结束, 而 X 中的元素没有被完全覆盖, 则返回一个空集, 表示没有找到集合覆盖.

算法 3. GSCHA(S, N, R).

1. $F = \emptyset$;
2. $M = \{1, 2, \dots, |S|\}$;
3. **foreach** $P_i \in S (1 \leq i \leq |S|)$
4. creat the element f_i of F according to the transmission rule;
5. **endfor**
6. $X = N$;
7. $A = \emptyset$;
8. **repeat**
9. $D = \text{Greedy-Set-Cover}(X, F, M)$;
10. **if** $|D| > 1$ **then**
11. add the packets corresponding to the numbers in D into a set, and then insert the set into A ;
12. remove these packets from F , and remove the corresponding numbers from M ;
13. **endif**
14. **until** $|D| \leq 1$
15. return A

算法 4. Greedy-Set-Cover(X, F, M).

1. $X' = X$;
2. $C = \emptyset$;
3. $F' = F$;

4. **repeat**
5. select an $f_k \in F'$ that maximizes $|f_k \cap X'|$, where k is the corresponding number in M ;
6. **if** $|f_k \cap X'| > 0$ **then**
7. $X' = X' - f_k$;
8. $C = C \cup \{k\}$;
9. $F' = F' - \{f_k\}$;
10. **endif**
11. **until** $X' = \emptyset$ or $F' = \emptyset$ or $|f_k \cap X'| = 0$
12. **if** $X' = \emptyset$ **then** return C ;
13. **else** return \emptyset ;
14. **endif**

定理 5. GSCHA 算法的时间复杂度为 $O(|N||S|\min(|N|, |S|)\log_2|S|)$.

证明:与定理 2 类似,因为在该算法中,第 3~5 行的 for 循环和第 8~14 行中的 repeat 循环所花费的时间最多,所以只分析这两个循环的时间复杂度.第 3~5 行中的 for 循环与 GSPHA 算法中的 for 循环类似,可以在 $O(|N||S|)$ 时间内实现.考虑第 8~14 行中的 repeat 循环,最差的情况是每一次都发现两个元素的集合覆盖.所以迭代次数最多为 $\log_2|S|$.下面分析该 repeat 循环体的时间复杂度.根据文献[16],第 9 行中 Greedy-Set-Cover 算法的时间复杂度为 $O(|N||S|\min(|N|, |S|))$,这也是 repeat 循环体中最耗时的操作.所以,repeat 循环的时间复杂度为 $O(|N||S|\min(|N|, |S|)\log_2|S|)$,从而可以计算出 GSCHA 算法的时间复杂度为 $O(|N||S|\min(|N|, |S|)\log_2|S|)$.因此,定理 5 成立. \square

3.3 网络编码方法 COMP 的应用

本文将提出的网络编码方法 COMP 应用到 LBIP 广播算法上^[12].LBIP 广播算法是著名的 BIP 广播算法的分布式实现版本.LBIP 广播算法的基本思想是,每个节点基于父节点提供的信息在局部的 k 跳范围内构造局部广播树,然后根据局部广播树进行广播,直到所有节点都被覆盖为止.图 6 描述了 $k=2$ 时的 LBIP 广播算法的基本步骤.源节点 s 首先构造两跳范围内的局部广播树,然后将报文广播给下一跳转发节点,如图 6(a)所示.转发节点 a 收到来自源节点 s 的报文后,将它两跳邻居节点 d, e 和 f 加入到源节点 s 构造的以转发节点 a 为非叶节点的子树中,以完成其局部广播树的构造,如图 6(b)所示.最后,节点 b 也以类似的方法将节点 g 和 h 加入到其局部广播树中,如图 6(c)所示.

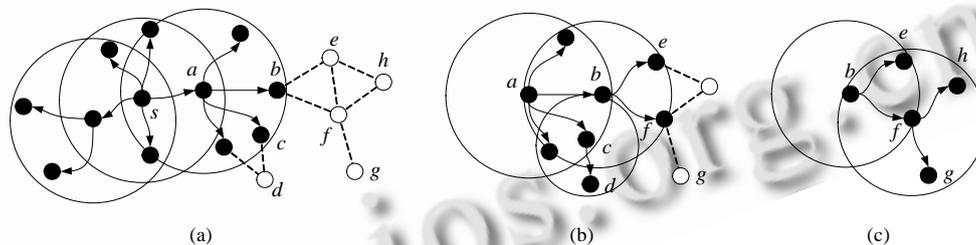


Fig.6 An example to explain the LBIP algorithm

图 6 解释 LBIP 算法的例子

因为文献[12]已证明了两跳是最佳的跳数选择,所以本文采用两跳作为局部广播树的构造范围.将 COMP 网络编码方法应用到 LBIP 广播算法后,主要包括以下 5 个部分:

(1) 获取局部拓扑信息

在网络初始化时,节点通过两轮信息交换来获取两跳范围内的邻居节点信息,包括邻居节点的标识和地理位置信息.在第一轮中,邻居节点之间通过 HELLO 报文交换彼此的信息;然后在下一轮中,通过 HELLO 报文交换彼此的一跳邻居信息.

(2) 更新邻居接收表

节点在网络初始化时构造邻居接收表,记录一跳邻居节点接收报文的信息.另外,节点只有在收到或者解码出新的原始报文时,才通过 HELLO 报文通知邻居节点已收到的新报文的标识信息.邻居节点在收到该 HELLO 报文后,更新其邻居接收表.

(3) 构造局部广播树

在广播报文的初始阶段,源节点 s 首先利用获取的局部拓扑信息在两跳邻居范围内构造一棵局部广播树.非叶节点被选择为转发节点,并被指派一个发送功率.这些信息包含在广播报文中,以便其子节点构造局部广播树.

(4) 利用 COMP 编码报文

因为节点 s 可能是其他报文的转发节点,所以节点 s 在发送自己的广播报文之前将待发送的报文集合 S 、一跳邻居节点表 N 和邻居接收表 R 作为参数调用 GSPHA 算法得到集合 A .集合 A 中的每一个元素都是 S 的满足 COPE 网络编码条件的子集 S_x .因此,节点 s 分别对这些子集的元素进行编码,然后再将 $S - \bigcup_{S_x \in A} S_x$ 、 N 和 R 作为参数调用 GSCHA 算法,获得集合 A' .集合 A' 的每一个元素都是 $S - \bigcup_{S_x \in A} S_x$ 的满足可依次两两编码的子集.节点 u 对这些子集的元素分别依次两两编码.最后,节点 s 按顺序将编码后的报文发送出去.

(5) 报文解码

当源节点的子节点 u 收到报文后,首先判断该报文是否是编码报文:如果是,则利用已接收的报文对该编码报文进行解码;如果不能解码则先存储起来,等待以后利用新获取的报文对它进行解码.节点 u 在获取未编码的报文或者解码出所需要的报文后,通过报文中的信息判断它是否是转发节点,如果不是就不再转发,否则它就在其父节点 s 构造的局部广播树的基础上构造自己的局部广播树.节点 u 在广播报文中记录局部广播树信息,然后按其父节点类似的方法对广播报文进行编码,最后发送给其子节点.以此类推,直到所有节点都被覆盖为止.

4 性能评估

4.1 实验设置

本文采用比较流行的 NS-2 模拟器实现了网络编码方法 COMP,并将其应用到 LBIP 广播算法中,记为 LBIP+COMP.另外,为了与 COPE 网络编码方法进行比较,本文采用启发式算法 GSPHA 来实现 COPE,并将它应用到 LBIP 广播算法中,记为 LBIP+COPE.模拟实验评估与分析不同的节点数、节点最大传输范围和广播会话数对这些算法性能的影响.模拟实验所使用的场景为 $1000\text{m} \times 1000\text{m}$ 的矩形区域,所有节点随机地分布于该区域中,并且保持静止状态.另外,实验中采用 Lucent 的 WaveLAN 模型和 Two-ray ground 无线传播模型,其位速率为 2Mb/s .每个广播会话的源节点都不同.其他的实验参数见表 1.模拟实验测试以下 4 种性能指标,并取 20 次不同实验的平均值作为实验结果:

- (1) 报文成功接收率(packet delivery ratio):接收到广播报文的节点数与网络中的所有节点数之比.
- (2) 信号冲突次数(number of signal collision):网络中所有节点上发生的信号冲突次数之和.
- (3) 报文传输次数(number of packet transmission):所有节点传输报文的次数之和.
- (4) 报文传输能耗(energy consumption of packet transmission):定义为所有节点消耗的能量之和.

Table 1 Simulation parameters

表 1 实验参数设置

Parameter	Value
MAC protocol	IEEE 802.11
Size of rectangle area	$1000 \times 1000\text{m}^2$
Number of nodes	100~400
LAN model	WaveLAN
Maximum transmission range	200~500m
Number of sessions	10~40
Number of run	20
Propagation loss coefficient	2

4.2 实验结果

模拟实验首先测试 LBIP, LBIP+COPE 和 LBIP+COMP 这 3 种算法在不同节点数下的性能变化情况, 结果如图 7 所示. 实验场景选择节点的最大传输范围为 200m, 会话数为 10. 从图 7(a) 中可观察到, 节点数对报文成功接收率的影响不大, 所有 3 种算法的报文成功接收率都保持在 95% 以上. 原因是, 会话数比较少, 转发节点同时发送报文导致信号冲突的次数比较少, 如图 7(b) 所示. 另外, 节点数比较少时, 转发节点数目也比较少, 所以对报文进行网络编码的机会不多, 因此, 网络编码的作用不明显. 但是, 当节点数不断增加时, 转发节点数目也随着增加, 因此网络编码的作用明显. 图 7(c) 和图 7(d) 描述了报文传输次数和报文传输能耗随节点数增加的变化情况. 从图中可以看出, LBIP 的传输次数随节点数的增加而增加, 但是传输能耗却不断减少. 原因是, 节点数目的增加导致了转发节点数目的增加, 从而传输次数不断增加, 但是每个转发节点的传输范围也不断减少, 所以总的能耗也不断减少. 同时, 因为网络编码方法的利用减少了转发节点的传输次数, 所以降低了 LBIP 算法的能耗. 而 COMP 网络编码方法比 COPE 网络编码方法挖掘的网络编码机会更多, 所以性能提高更为显著.

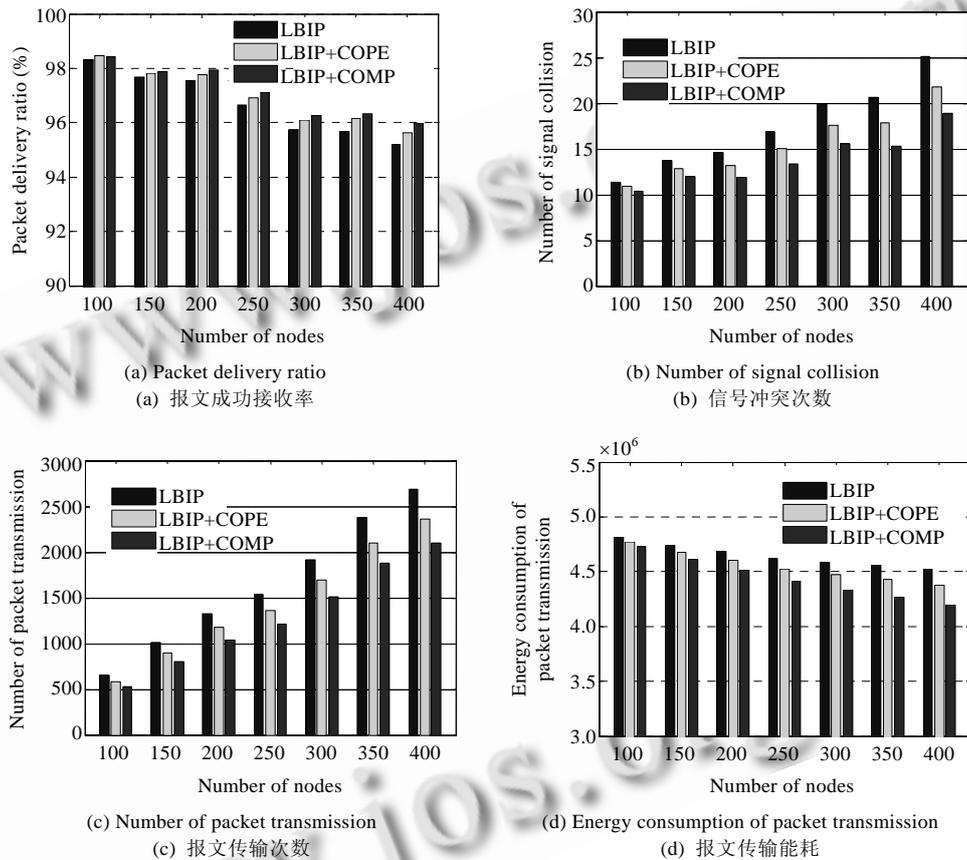


Fig.7 Performance variety with different numbers of nodes

图 7 不同节点数下的性能变化

图 8 描述了节点数为 400 以及会话数为 10 的场景下, 不同最大传输范围对 3 种算法性能的影响. 从图 8(a) 中可以看出, 报文成功接收率随着最大传输范围的增加而减少, 原因是最大传输范围的增加导致了信号冲突次数的增加, 如图 8(b) 所示, 从而影响了报文成功接收率. 另外, 最大传输范围增大以后, 局部最低能耗广播树考虑的覆盖区域更大, 所以转发节点的数目也不断减少, 从而报文传输次数不断减少, 其结果如图 8(c) 所示. 图 8(d) 描述的是报文传输能耗随最大传输范围增加而发生变化的情况. 从该图中可以观察到, LBIP 的能耗并没有发生显

著变化,而且随着最大传输范围的增加而减少.原因是,最大传输范围越大,转发节点的两跳邻居节点越多,从全局的角度看,其构造的局部广播树越节能,所以总的能耗不断减少.另外,从这 4 个图中也可以观察到,随着最大传输范围的扩大,网络编码的作用越来越弱,当最大传输范围增加到 500m 时,应用网络编码几乎不会带来任何性能提高.其原因是,随着最大传输范围的增加,转发节点的数目不断减少,转发节点的邻居节点也不断增多,所以进行网络编码的次数和满足网络编码条件的报文组合也不断减少,最终导致网络编码的机会微乎其微.

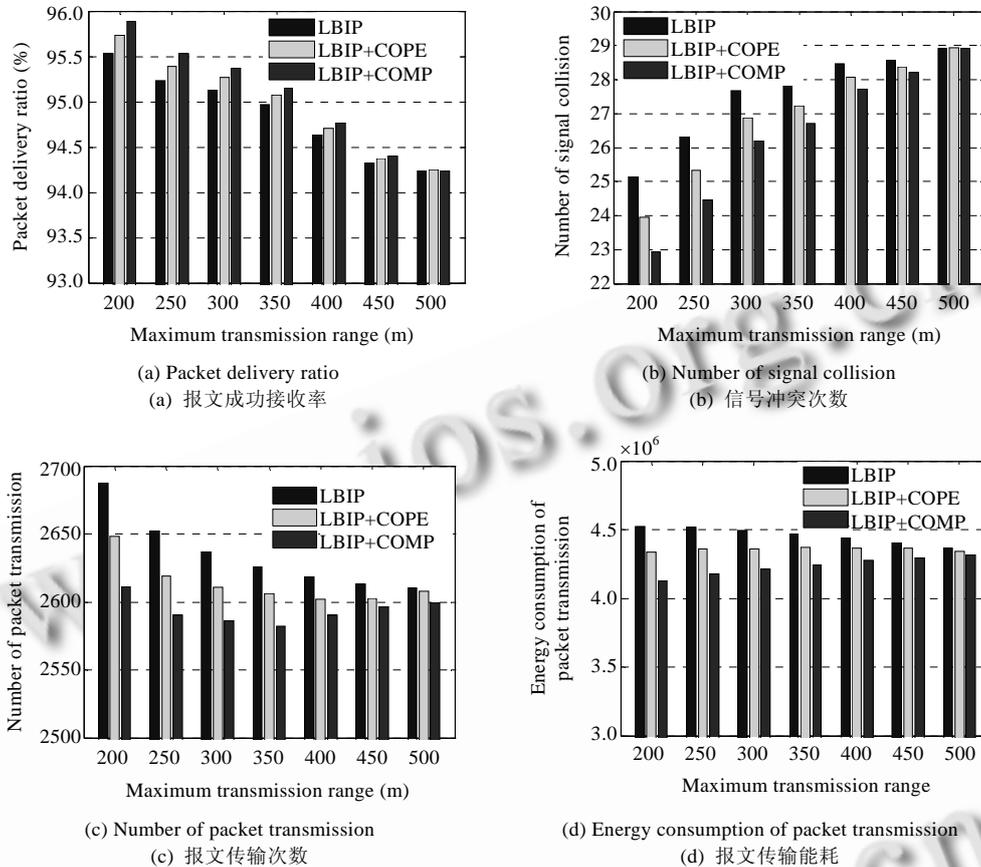


Fig.8 Performance variety with different maximum transmission ranges

图 8 不同最大传输范围下的性能变化

3 种算法在不同会话数下的性能变化情况如图 9 所示.实验场景为 400 个节点和 200m 的最大传输范围.从图 9(a)中可以看出,LBIP 算法的报文成功接收率随着会话数的增加而显著下降.主要原因是,并发会话的增多导致信号冲突的次数不断增加,如图 9(b)所示,许多节点都收不到广播报文,如果转发节点也收不到广播报文,则会影响广播的进一步转发,所以最终导致报文成功接收率的下降.另外,由于会话数的增加,转发节点需要转发的报文也不断增多,而由于转发节点的数目基本上不变,所以总的报文传输次数显著增加,从而其总的能耗也不断增加,其结果如图 9(c)和图 9(d)所示.从这 4 个图中也可以看出,网络编码的作用很明显,特别是会话数比较多的场景下.因为转发节点需要发送的报文数增加了,而邻居数并没有增加,所以满足编码条件的报文组合也会增多,从而利用网络编码的次数也不断增多.每一次利用网络编码都能减少传输次数,因此减少了信号冲突的次数,提高了报文成功接收率,降低了报文传输能耗.类似地,COMP 网络编码方法挖掘的网络编码机会比 COPE 网络编码挖掘的要多,所以其性能提高更为显著.

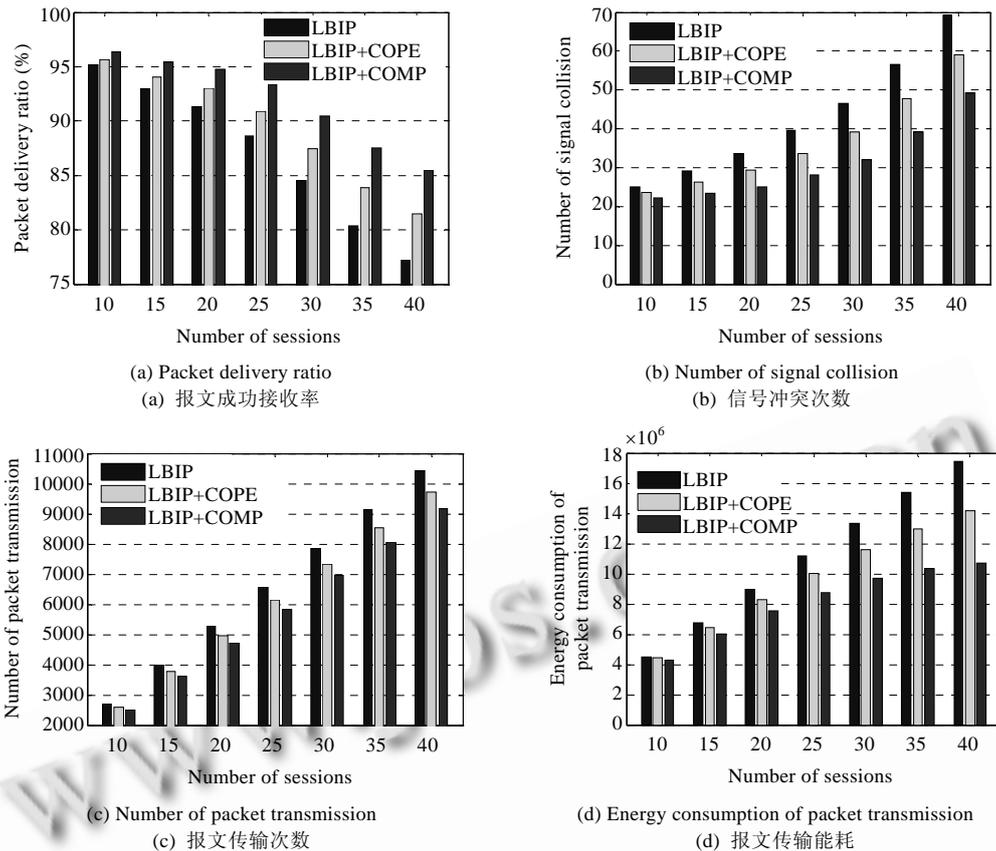


Fig.9 Performance variety with different numbers of sessions

图9 不同会话数下的性能变化

5 总结及下一步的工作

本文证明了与应用网络编码相关的最优化问题是 NP 完全的,并且提出了一种采用启发式算法的网络编码方法 COMP.该网络编码方法包括两种启发式算法 GSPHA 和 GSCHA,分别是针对不同的最优化问题,目的是尽可能多地挖掘网络编码机会.NS-2 仿真结果表明,该网络编码方法在应用到节点数比较多、最大传输范围比较小以及会话数比较多的场景中时有效地减轻了并发会话的影响,提高了现有广播算法 LBIP 的性能,而且其性能提高超过了现有的网络编码方法 COPE.

下一步将考虑网络编码方法 COMP 的可靠性问题,例如,邻居接收表的实时性以及是否能够准确地反映邻居节点的接收情况等.另外,准备研究如何在实际的环境中实现网络编码方法 COMP 并且应用到各种场景中.最后,将 COMP 应用到节点可移动的场景中也是未来准备进行的研究工作.

致谢 在此,向对本文提出宝贵建议的专家和教授表示衷心的感谢,并对参与本文研究工作讨论的老师和同学致以真诚的谢意.

References:

- [1] Ahlswede R, Cai N, Li SYR, Yeung RW. Network information flow. IEEE Trans. on Information Theory, 2000,46(4):1204-1216. [doi: 10.1109/18.850663]

- [2] Li SYR, Yeung RW, Cai N. Linear network coding. *IEEE Trans. on Information Theory*, 2003,49(2):371–381. [doi: 10.1109/TIT.2002.807285]
- [3] Koetter R, Medard M. An algebraic approach to network coding. *IEEE/ACM Trans. on Networking*, 2003,11(5):782–795. [doi: 10.1109/TNET.2003.818197]
- [4] Katti S, Rahul H, Hu W, Katabi D, Medard M, Crowcroft J. XORs in the air: Practical wireless network coding. In: *Proc. of the ACM SIGCOMM*. New York: ACM Press, 2006. 243–254.
- [5] Cui T, Chen L, Ho T. Energy efficient opportunistic network coding for wireless networks. In: *Proc. of the IEEE INFOCOM*. Piscataway: IEEE Inc., 2008. 361–365.
- [6] Lun DS, Ratnakar N, Koetter R, Medard M, Ahmed E, Lee H. Achieving minimum-cost multicast: A decentralized approach based on network coding. In: *Proc. of the IEEE INFOCOM*. Piscataway: IEEE Inc., 2005. 1607–1617.
- [7] Wu Y, Chou PA, Kung SY. Minimum-Energy multicast in mobile ad hoc networks using network coding. *IEEE Trans. on Communications*, 2005,53(11):1906–1918. [doi: 10.1109/TCOMM.2005.857148]
- [8] Fragouli C, Widmer J, Boudec JYL. A network coding approach to energy efficient broadcasting: From theory to practice. In: *Proc. of the IEEE INFOCOM*. Piscataway: IEEE Inc., 2006. 1–11.
- [9] Li EL, Ramjee R, Buddhikot M, Miller S. Network coding-based broadcast in mobile ad hoc networks. In: *Proc. of the IEEE INFOCOM*. Piscataway: IEEE Inc., 2007. 1739–1747.
- [10] Li SS, Liao XK, Zhu PD, Xiao N. A method for multipath routing based on network coding in wireless sensor network. *Journal of Software*, 2008,19(10):2638–2647 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/2638.htm> [doi: 10.3724/SP.J.1001.2008.02638]
- [11] Yang L, Zheng G, Hu XH. Research on network coding: A survey. *Journal of Computer Research and Development*, 2008,45(3): 400–407 (in Chinese with English abstract).
- [12] Ingelrest F, Simplot-Ryl D. Localized broadcast incremental power protocol for wireless ad hoc networks. *Journal of Springer Wireless Networks*, 2008,14(3):309–319.
- [13] Wieselthier J, Nguyen G, Ephremides A. On the construction of energy-efficient broadcast and multicast trees in wireless networks. In: *Proc. of the IEEE INFOCOM*. Piscataway: IEEE Inc., 2000. 585–594.
- [14] Halldorsson MM, Kratochvil J, Telle JA. Independent sets with domination constraints. In: *Proc. of the 25th Int'l Colloquium on Automata, Languages and Programming*. Berlin: Springer-Verlag, 1998. 176–187.
- [15] Raz R, Safra S. A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP. In: *Proc. of the 29th ACM Symp. on Theory of Computing*. New York: ACM Press, 1997. 475–484.
- [16] Cormen TH, Leiserson CE, Rivest RL. *Introduction to Algorithms*. Cambridge: MIT Press, 1990. 974–978.

附中文参考文献:

- [10] 李姗姗,廖湘科,朱培栋,肖依.基于网络编码的无线传感网多路径传输方法. *软件学报*,2008,19(10):2638–2647. <http://www.jos.org.cn/1000-9825/19/2638.htm> [doi: 10.3724/SP.J.1001.2008.02638]
- [11] 杨林,郑刚,胡晓惠.网络编码的研究进展. *计算机研究与发展*,2008,45(3):400–407.



焦贤龙(1982—),男,江西遂川人,博士生,CCF 学生会员,主要研究领域为无线自组网,无线传感器网络.



周兴铭(1938—),男,教授,博士生导师,主要研究领域为高性能计算技术,网络与分布式计算技术.



王晓东(1973—),男,博士,副研究员,主要研究领域为移动计算技术.