软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(11):2833−2842 [doi: 10.3724/SP.J.1001.2011.03938]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于匹配理论的无线广播编码重传算法*

戴 彬[1+], 曹志刚[1,2], 杨 军[1], 黄 辰[1], 王芙蓉[1]

[1](华中科技大学 电子与信息工程系,湖北 武汉 430074)
[2](空军第一航空学院 基础部,河南 信阳 464000)

Coded Retransmission Algorithm Based on Matching Theory for Wireless Broadcasting

DAI Bin[1+], CAO Zhi-Gang[1,2], YANG Jun[1], HUANG Chen[1], WANG Fu-Rong[1]

[1](Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)
[2](Department of Base Course, First Aeronautic Institute of Air Force, Xinyang 464000, China)
+ Corresponding author: E-mail: daibin@hust.edu.cn

Dai B, Cao ZG, Yang J, Huang C, Wang FR. Coded retransmission algorithm based on matching theory for wireless broadcasting. *Journal of Software*, 2011,22(11):2833−2842. http://www.jos.org.cn/1000-9825/3938.htm

Abstract: This paper considers the application of network coding to reduce the number of retransmission by combining the lost packets from different receivers with network coding in a wireless broadcast network. In particular, the study propses an algebraic expression to define the coding conditions of retransmission packets by matrix and vector operations. According to the graph constructed by correlation matrix, an algorithm is presented to find more coding opportunity based on maximum matches in graph theory. The proposed algorithm has the ability to find the maximum pairwise coding opportunities, meanwhile, it takes into account the probability of coding the lost packets as possible to more than one retransmission packet to minimize the number of retransmissions by coding optimization, and increase the network bandwidth efficiency and throughput.
Key words: network coding; wireless broadcasting; retransmission; matching algorithm; bandwidth efficiency

摘 要: 针对成批数据在无线广播过程中发生丢包后的重传策略,利用网络编码技术,研究尽可能多地减少重传次数的方法.首先通过矩阵及向量的运算给出了待重传数据满足编码条件的代数表达形式,再根据关联矩阵构造相应的图,最后通过图论中的最佳匹配理论给出了一种寻找编码机会的优化算法.此算法中不但能够找出最多的两两编码机会,而且还考虑了将尽可能多的数据包编在一起的可能性,从而尽可能地减少了重传次数,实现了编码的优化,有效提高了网络带宽效率和吞吐量.
关键词: 网络编码;无线广播;重传;匹配算法;带宽效率
中图法分类号: TP393 **文献标识码**: A

在无线网络传输中,为实现无线链路的可靠传输,对于信宿上的丢包采用自动重传(automatic retransmission request,简称 ARQ)是一种常用的实现差错控制的方式.但由此会导致在差错信道的传输中需要较多的重传次数,极大地降低了无线信道的有效容量.2000 年,Ahlswede 等人基于网络信息流的概念提出了网络编码的思

* 基金项目: 国家自然科学基金(60803005); 国家科技重大专项课题(2009ZX03004-004, 2010ZX03003-003)
收稿时间: 2010-03-11; 修改时间: 2010-07-06; 定稿时间: 2010-08-27

想[1],并证明其是可以逼近网络容量理论传输极限的有效方法.随着网络编码理论的逐步完善,网络编码被广泛地应用到包括提高网络吞吐率、保障链路的可靠性、安全性及能量效率等各个方面.其中,有研究者提出结合网络编码提高网络吞吐率的特点,可以利用网络编码的方法来减少无线广播重传的次数,从而提高无线网络的吞吐率[2−6].

由于无线信道不可靠,一旦在传输过程中发生丢包,就需要信源进行广播重传.如何在相同的信道质量下减少广播重传次数,提高网络利用率,一直是相关学者探究的目标.其中,利用编码方法来提升广播重传效率是近年来的研究热点.文献[7]提出了在发送原始数据包的同时发送若干个冗余的编码包做差错恢复,当发生少量原始数据包丢失时,可以通过冗余的编码包来保证信宿能够以高概率解码.但不同信宿的数据需求率的差异性、网络链路状态的不稳定性等因素会导致冗余量的确定是十分困难的,而不恰当的数据冗余往往会造成整体网络吞吐率的严重下降.如何确定冗余量从而保障可靠传输,在文中并未涉及.文献[8]指出,网络编码的增益可以无限大,并给出了几种特殊网络拓扑的增益值.而对一般拓扑结构,则无法给出通用的编码算法.文献[6]指出,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.文献[9]讨论了在多个信宿需要一批相同数据的特殊情况下,提出了编码重传时采用动态编码的思想,并给出了用来衡量网络传输增益的一种表示方法,但对编码的最优化问题并未进行讨论.文献[10]提出了一种适用于无线广播的基于网络编码的重传方法——NCWBR.该方法针对文献[9]中的思想给出了具体的算法,但同样未给出优化方法.文献[11]指出,文献[9]中的机会网络编码重传算法(opportunistic network coding retransmission,简称 ONCR)效率并不总是优于文献[12]给出的随机网络编码重传算法(random network coding retransmission,简称 RNCR),它会受到信宿数目、数据包数及平均需求率的影响,并且给出了在何种情况下采用何种编码重传方法的判决依据,并提出了自适应编码重传算法(adaptive network coding retransmission,简称 ANCR).文献[13]简要回顾了网络编码的理论研究,阐述了网络编码优化问题研究的重要意义.在介绍网络信息流模型的基础上,针对优化问题的陈述、特点和解法,结合最新的研究成果进行了综述.提出根据优化目标的不同,优化问题可分成 4 类:最小花费组播;无向网络的最大吞吐率;最小编码节点、编码边;基于网络编码的网络拓扑设计.归纳了问题的求解性质,对其中的(线性或凸)规划问题总结了求解的一般方法,对 NP 完全问题讨论了最新的启发式算法及其设计难点.

综合分析可以看出,ONCR 算法是基于贪婪算法的编码方法,在大多数情况下优于 RNCR 算法.但贪婪的编码方案也并不是最优的编码方案,在编码复杂度、网络效率等方面还有进一步提升的空间.本文立足于图论中的最优匹配理论来分析编码的最优化问题,在此基础上提出了一种基于匹配理论的网络编码重传算法 MNCR (matching based network coding retransmission),并给出了一般条件下的编码实现方法.MNCR 算法的设计主要基于以下两个原则:一是信宿收到相应编码包后能够立即解码;二是编码包数尽可能少,即重传次数尽可能的少.文中利用需求矩阵描述了各信宿对同批数据的需求情况,同时,利用缓存矩阵将数据的接收状态进行了代数表示.在此基础上,对未收齐的原始数据包能进行编码重传的条件,通过矩阵及向量的运算给出了明晰的数学表达式:关联矩阵.最后,通过关联矩阵构造对应的图,并用图论中的最佳匹配理论确定了相应的编码方案,从而实现了利用网络编码将原始包进行重传的一种优化方法.而且,此算法是 P 难的.通过理论推导和仿真分析,验证了 MNCR 算法与 ONCR 算法、RNCR 算法以及 ANCR 相比,性能都有较为明显的提升.

本文第 1 节提出需要解决的主要问题.第 2 节对问题进行详细描述,并给出解决该问题时必要的一些假设条件.第 3 节是本文的核心,指出本文的数据传输方式并构造重传的具体匹配算法.第 4 节对算法的效率进行讨论,并利用仿真结果与其他算法的带宽效率进行比较.最后对本文工作进行总结,并对下一步的工作加以展望.

1 问题描述

假设信源需要广播一组数据,每个信宿需要其中的某些或者全部数据.由于在传输过程中可能会存在丢包现象,这时就需要信源进行重传.比如,信源 S 需要把数据 x_1 广播给 R_1,把 x_2 广播给 R_2,但结果是 R_1 收到了 x_2,R_2 收到了 x_1.按照传统方法,信源 S 就需要把两个数据重新发送,这需要广播两次.而如果采用网络编码,则只需要把

$x_1 \oplus x_2$ 广播 1 次,R_1 就可以根据 $x_1=(x_1\oplus x_2)\oplus x_2$ 从中解出所需要的 x_1.同理,R_2 可以从中解出所需要的 x_2,这样就可以把重传次数减少 1 次.对于 n 个信宿接收 m 个数据包的情况,研究者们利用网络编码的方法提出了各种各样的传输机制来减少重传次数.如文献[12]给出了将数据包进行随机组合的方法.该方法在组合前并不保证信宿能够立即解码,当收到所有信宿反馈的数据收全信息后才停止组合重传.文献[9]给出了动态网络编码的方法,其思想是,当收到反馈信息后,如果信宿已经不需要某个数据包,则在下一个时段重传时就不再将此包进行编码,而是将其余未收到的包重新组合重传.但该文并未给出具体的优化编码方法.基于此,我们讨论了在保证信宿收到相应编码包后能够立即解码的条件下,如何组合编码才能使包数尽可能地少.由于最优组合问题是 NP 难的,所以本文采用最佳匹配的方法来逼近其最优解.

2 符号及假设

假定信源 S 需要向 n 个信宿 $R_i(i=1,2,\ldots,n)$广播 m 个数据包 $x_i(i=1,2,\ldots,m)$,每个信宿需要 m 个数据包中的某些或者全部.每个信宿 $R_i(i=1,2,\ldots,n)$需要的数据用矩阵表示如下:

$$X=\begin{array}{c} \begin{array}{cccc} R_1 & R_2 & \ldots & R_n \end{array} \\ \left[\begin{array}{cccc} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{array}\right] \end{array} \begin{array}{c} \\ x_1 \\ x_2 \\ \ldots \\ x_m \end{array},$$

其中,$x_{ij}=\begin{cases}1, & \text{如果信宿}R_j\text{需要数据}x_i \\ 0, & \text{否则}\end{cases}$, $i=1,2,\ldots,m;j=1,2,\ldots,n$.

由于矩阵 X 中的元素取值表示各信宿对数据的需求情况,故称其为需求矩阵.

为便于讨论,文中作如下假设:

(1) 每条链路都有可靠的反馈回路,信宿 R_i 能够及时将数据接收情况反馈到信源;

(2) 每个信宿都有足够的缓冲区,以保存数据 $x_i(i=1,2,\ldots,m)$;

(3) 各个链路上的丢包率相互独立,信源 S 到信宿 R_i 的丢包率以 p_i 表示.

3 编码重传的优化算法

3.1 首　播

信源在收到邻近的一批发包请求后,可以根据需求矩阵先求出各个信宿对每个数据的需求量 x^i,

$$x^i=\sum_{j=1}^{n} x_{ij}\,(i=1,2,\ldots,m);$$

然后按照从大到小排出顺序,不妨设为 $x^1\geqslant x^2\geqslant\ldots\geqslant x^m$;接下来将每个数据依此顺序广播一次,这样处理的目的是尽可能减少总时延.

3.2 编码分组的确定

原始数据包发送完毕,信源根据信宿反馈的收包信息确定是否需要广播重传以及如何进行广播重传.

所有信宿收到的数据情况用缓存矩阵 B 来表示:

$$B=\begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{m1} & b_{m2} & \ldots & b_{mn} \end{bmatrix},$$

其中,$b_{ij}=\begin{cases}1, & \text{如果信宿}R_j\text{收到数据}x_i \\ 0, & \text{否则}\end{cases}$, $i=1,2,\ldots,m;j=1,2,\ldots,n$.

令 $b_i=(b_{i1},b_{i2},\ldots b_{in})(i=1,2,\ldots,m)$.然后将矩阵 X 和矩阵 B 做 Hadamard 乘积,并记为 C,即

$$C = X \circ B = \begin{bmatrix} x_{11}b_{11} & x_{12}b_{12} & \dots & x_{1n}b_{1n} \\ x_{21}b_{21} & x_{22}b_{22} & \dots & x_{2n}b_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1}b_{m1} & x_{m2}b_{m2} & \dots & x_{mn}b_{mn} \end{bmatrix}.$$

如果有 $C=X$,则说明各信宿都收到了所需数据;否则,就需要对数据进行广播重传.

各个信宿需要广播重传的数据用矩阵 T 表示,即有

$$T = X - C = X - X \circ B = \begin{bmatrix} x_{11}(1-b_{11}) & x_{12}(1-b_{12}) & \dots & x_{1n}(1-b_{1n}) \\ x_{21}(1-b_{21}) & x_{22}(1-b_{22}) & \dots & x_{2n}(1-b_{2n}) \\ \dots & \dots & \dots & \dots \\ x_{m1}(1-b_{m1}) & x_{m2}(1-b_{m2}) & \dots & x_{mn}(1-b_{mn}) \end{bmatrix} \triangleq \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \dots & \dots & \dots & \dots \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{bmatrix},$$

其中,$t_{ij} = \begin{cases} 1, & \text{如果信宿}R_j\text{需要重传数据}x_i \\ 0, & \text{否则} \end{cases}$, $i=1,2,\ldots,m;j=1,2,\ldots,n$.

令 $t_i = (t_{i1}, t_{i2}, \ldots, t_{in}), s_i = \sum_{j=1}^{n} t_{ij} (i = 1, 2, \ldots, m)$. 由于矩阵 B 中元素只取 0 或 1,显然有,

$$t_i b_j^T = t_{i1}b_{j1} + t_{i2}b_{j2} + \ldots + t_{in}b_{jn} \leqslant t_{i1} + t_{i2} + \ldots + t_{in} = s_i.$$

在进行广播重传时,如果能够将多个包利用网络编码编码为 1 个包,则可减少相应广播重传次数.

通过网络编码进行广播重传,必须满足下列定理.

定理. 假定需要把 k 个数据包 $x_1,x_2,\ldots,x_k$ 分别重传给信宿 $R_1,R_2,\ldots,R_k$,则能够将它们编码为 1 个包进行广播重传的充要条件是,每一个信宿 $R_i(i=1,2,\ldots,k)$都已经收到了其余的 $k-1$ 个包 $x_j(j=1,2,\ldots,k;j\neq i)$.

对于两个原始包 x_i 和 x_j 来说,它们能够编码的条件是:需要 x_i 的那些信宿在此之前都已经收到了 x_j;同时,需要 x_j 的那些信宿在此之前已经收到了 x_i.对于前者,即要求重传矩阵 T 中第 i 行向量 t_i 的非零分量 t_{ik} 对应的缓存矩阵 B 中第 j 行向量 b_j 的分量 b_{jk} 必须全为 1,即公式(1)成立.

$$t_i b_j^T = \sum_{k=1}^{n} t_{ik} b_{jk} = s_i \tag{1}$$

同理,对于后者应有:

$$t_j b_i^T = \sum_{k=1}^{n} t_{jk} b_{ik} = s_j \tag{2}$$

这样,通过 $x_i=(x_i\oplus x_j)\oplus x_j$ 及 $x_j=(x_i\oplus x_j)\oplus x_i$,各信宿就可以得到所需原始数据.事实上,公式(1)、公式(2)同时成立当且仅当

$$t_i b_j^T + t_j b_i^T = s_i + s_j \tag{3}$$

数据包 x_i 之间是否满足编码条件,也可以用关联矩阵 $R=(r_{ij})$来表示:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mm} \end{bmatrix},$$

其中,$r_{ij} = \begin{cases} 1, & \text{当数据包}x_i, x_j\text{同时满足关系式 (1)及关系式(2)时} \\ 0, & \text{否则} \end{cases}$, $i=1,2,\ldots,m;j=1,2,\ldots,n$.

类似地,若需重传多个原始包 $x_{i_1}, x_{i_2}, \ldots, x_{i_l}$,要保证每个收点都能够解码,依据上述定理则须满足:

$$\sum_{k=1}^{n} t_{i_1 k} b_{i_2 k} \ldots b_{i_l k} + \ldots + \sum_{k=1}^{n} t_{i_l k} b_{i_1 k} \ldots b_{i_{l-1} k} = s_{i_1} + s_{i_2} + \ldots + s_{i_l} \tag{4}$$

本文将满足公式(4)的一组数据包称为一个团,团中所包含元素数称为团的基数.

3.3 编码分组的确定

根据前面的运算可知,只要能将 L 个包编在一起,重传次数就可以减少 $L-1$ 次.因此,自然的想法是尽量寻找编码机会以获取最多的编码包.或者说,把需要重传的原始包进行分组,使得组数最少,并且每一组编码重传后能使各信宿进行解码.下面用图论的方法来研究这一问题.

将矩阵 T 中的每一非零行向量(对应需要广播重传的数据)和一个顶点相对应,所有顶点的集合记为 E.如果两个非零行向量满足公式(1)和公式(2),则将对应的两个顶点用一条线连起来,所有连线的集合称为边集,记为 V.这样可以得到一个连通图 $G=(V,E)$或者若干个连通子图,从而寻找编码机会问题就转化为把图的顶点进行分组问题.与公式(4)相对应,即要使图 $G=(V,E)$中顶点的分组数最少,并且每一组中任意两顶点之间有线连接.这个问题是 NP 难的,最坏情况下的算法时间复杂度为 $O(2^n)$,一般情况下只能给出近似解.其中可以采用贪婪算法:先找出满足分组条件的包含顶点数最多的分组,即基数最大的团;然后再依次在剩下的图中中继续寻找包含顶点数次少的组,直到所有的顶点都分组完成.但贪婪算法存在两个缺陷:首先,它实际上是在图中依次寻找最大团的问题,仍然是 NP 难的;其次,即使每次都能找到最大团,也不能保证总的团数最少.下面,我们将通过举例来说明每次找最大团的贪婪算法并不能得到最优解的情形.

对于如图 1 所示的顶点和连通性(图中的实线和虚线都表示连通),如果先找最大团,则可得到基数为 3 的团{2,3,6},如实线所标示.此时,剩余顶点无法再与其他顶点组合,故总的分组数即总团数为 4.由图 1 我们可以看出,如果简单地将顶点依次两两分组,分别可得到团{1,2},{3,4},{5,6},此分组下总团数仅为 3,小于用贪婪算法得到的总团数.类似地,在图 2 中如果先找最大团,则可得基数为 4 的团{2,3,5,6},而余下的顶点 1、顶点 4 也无法再组合,则得到总团数为 3.但如果将图 2 分为团{1,2,6}和{3,4,5},则总团数仅为 2.从以上的两个例子可以看出,在很多情况下,利用贪婪方法并不能得到最优解.

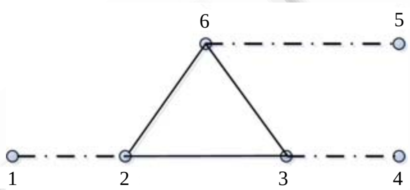


Fig.1
图 1

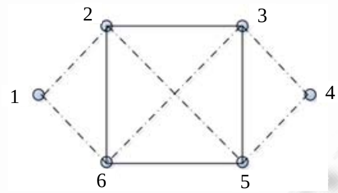


Fig.2
图 2

鉴于以上的分析,本文提出根据图论中的匹配理论,采用从小到大的分组思想来逼近问题的最优解.

如果图 $G=(V,E)$不连通,则可以对各个连通子图单独求解后再合并起来.所以,这里不妨设所得图为连通图.图 $G=(V,E)$的一个匹配 M 是边集 V 的一个子集,使得 M 中任何两条边没有公共端点.设 M 是图 G 的一个匹配,M 中的边称为匹配边,不在 M 中的边称为自由边.最大匹配问题是 P 难的,利用非二部图匹配算法[14]找到图 $G=(V,E)$的最大匹配后,所减少的广播重传次数就等于最大匹配中的匹配对数.

通过最大匹配,求解广播重传分组的算法步骤如下:

1. 首先找出图 $G=(V,E)$的最大匹配.依据广播重传数据包是否满足公式(1)和公式(2)给出关联矩阵,再根据匹配算法找出最大匹配.根据匹配理论可知,这一步骤的时间复杂度为 $O(|V|^4)$($|V|$是图中的顶点数).
2. 进一步优化.当全部顶点的分组数最少时,所得解才是最优解.所以一般情况下,步骤 1 找到的最大匹配并非最优解,还需要进一步组合.自由边中未能配对的的顶点称为自由点.根据解码条件,图 $G=(V,E)$中三角形的 3 个顶点对应数据可以编码,四边形中如果存在两条对角线则 4 个顶点对应数据也可以编码,所以在步骤 1 之后,接下来继续构造新图:

   (1) 把匹配对的两个顶点合并为一个复合顶点,再与自由点一起作为新图的顶点.

   (2) 如果一个复合顶点对应的两个顶点和一个自由顶点在图 $G=(V,E)$中是一个基数为 3 的团,则将其相

连;如果两个复合顶点所对应的 4 个顶点在图 $G=(V,E)$中是一个基数为 4 的团,则将对应的两个复合顶点相连.

(3) 如果一个匹配对$\{U_1,U_2\}$是两个三角形(如$\triangle U_1V_1W_1$ 和$\triangle U_2V_2W_2$)的顶点,且在步骤 1 的结果中$\{V_1,W_1\}$,$\{V_2,W_2\}$也是匹配对,则将匹配对$\{U_1,U_2\}$保留,并将 U_1 和$\{V_1,W_1\}$对应的复合顶点相连, U_2 和$\{V_2,W_2\}$对应的复合顶点相连.

(4) 对不满足上述条件的顶点及匹配对不再放入新图中.

这样就得到一个新的图 $G'=(V',E')$,它由一个或若干个连通子图构成.对新图 G'重复步骤 1,继续寻找最大匹配.

步骤 2(1)中合并操作的执行次数小于$|V|$;步骤 2(2)是确定复合顶点与自由顶点的关系,执行次数为复合顶点数和自由顶点数的乘积,不超过$|V|^2$;同理,步骤 2(3)的执行次数也不超过$|V|^2$.故第 2 步总的时间复杂度为$O(|V|^2)$.

3. 根据第 2 步的匹配结果将所对应数据包编码重传,未能复合的顶点对应数据则直接重传.

对于编码包,仍然存在丢失的可能.在重传之后,如果仍有信宿未收全所需数据,则可以按照以下两种方式处理:

(1) 根据新的反馈矩阵及缓存矩阵再次构造相应的图重复上述算法,直到所有信宿都收全为止;

(2) 与下一批数据进行编码组合.

算法说明:在步骤 1 中,利用最优匹配理论把信宿进行初步分类,首先可以保证有一个次优解.在步骤 2 中,在此次优解的基础上进一步优化,使分组数目再次减少,最终可以使所得解接近问题的最优解,从而避免了求解 NP 难的最优分组问题.

同样是对于图 1 和图 2 所给出的情景,我们利用本文提出的匹配算法对图中的顶点进行分组,针对图 1 给出的例子,利用上述的匹配,由步骤1可以得到最大匹配{1,2},{3,4},{5,6},如图3所示.此时无自由顶点,也无满足条件的四边形,得到的分组是最优解.

对应于图 2 的例子,由步骤 1 可以得到最大匹配{1,2},{3,6},{4,5}(如图 4 所示,其他最大匹配解不再列出).因为{3,6}连接两个三角形{1,2,6}和{3,4,5}(如图 5 所示),根据匹配算法可得最大匹配{1,2,6}和{3,4,5},复原后即为图 6,此分组也为最优解.

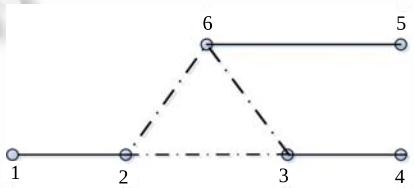


Fig.3
图 3

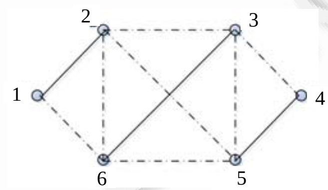


Fig.4
图 4

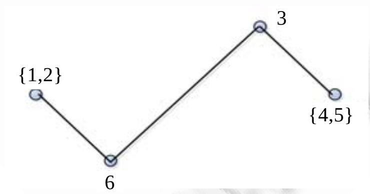


Fig.5
图 5

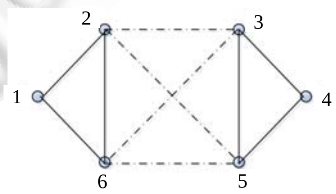


Fig.6
图 6

在图 2 中,虽然{2,3,5,6}所构成的四边形中有两条对角线,但是在步骤 1 的结果中,其 4 个顶点并未构成两

个匹配对,所以在接下来的算法中不能将它们一起处理.

以上例子说明,采用贪婪算法即使能找到一系列最大团,也不能保证总的团数最少.而本文提出的匹配算法不仅可以避开找最大团这个 NP 难问题,而且比贪婪算法更接近最优解.

下面,我们举例说明对于任意给定的图,匹配算法的求解步骤.图 7 是任意给定的一个图,根据第 1 步的算法,可得最大匹配如实线所示.顶点{3,4,5,6}构成一个团,匹配对{7,8}连接两个三角形,且{3,6}和{9,10}也是匹配对.于是忽略其他不满足条件的顶点及匹配对,可得图 8.根据第 2 步可以得到优化结果如图 8、图 9 或者图 10 中实线所示.所以,{1,2},{3,4,5,6},{7,8},{9,10},{11,12},{13}是最优解之一.

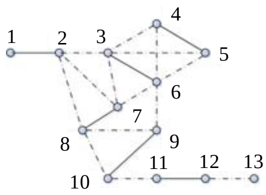


Fig.7

图 7

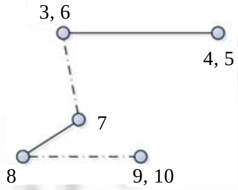


Fig.8

图 8

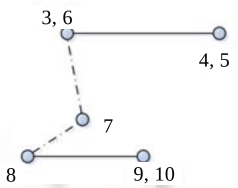


Fig.9

图 9

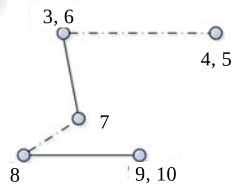


Fig.10

图 10

本文算法中把最大团的基数设为 4,即对于基数为 4 的团将不会再进行匹配步骤.这主要基于以下的考虑:由于要使多个信宿都能解码,所以满足公式(4)的团数将随其基数的增加急剧减少.模拟结果显示,存在基数为 5 以上的团的概率不超过 $10^{m\left(1-\frac{1}{\gamma}\right)\left(\gamma-\frac{m^2-m-2}{m^2-m}\right)}$(其中,$\gamma$是重传矩阵 T 中主对角线上方 1 的个数与此处元素个数 $\frac{m^2-m}{2}$ 的比值).在 $m>4$ 时,这个概率是很小的.随着最大团基数的增加,算法迭代次数逐步递增,算法的复杂度将逐渐增大,相应的求解时间将会大大增加.而与此同时,当 $m>4$ 时,迭代次数的增加对于匹配结果的影响并不大,反而会得不偿失.因此,本文算法设定的最大团的基数为 4,对于基数为 4 的团将不会再执行迭代匹配操作.在最大团的基数限定为 4 的前提下,本文算法的时间复杂度远小于贪婪算法.

4 仿真和结果分析

由于各信宿对各个数据的需求是随机的,即需求矩阵 X 的取值是随机的;各个信宿的丢包率也是不相同的,即缓存矩阵 B 中元素的取值是不相同的,因此,正如文献[11]所指出的,在此情况下无法给出重传次数的准确表达式.为此,我们需要通过仿真实验来验证该编码优化算法在一般情况下的性能.

为了有效地衡量编码算法对于重传性能的影响,我们给出以下几个定义:

- 需求率 u_i:信宿 R_i 需求的包数与总包数 m 的比值.用 u 表示 u_i 的平均值,即 $u=\sum u_i/n$,可以用需求矩阵 X 中 1 的个数与 mn 的比值来反映.
- 重传率:首播后,未被信宿收全的数据包数与 m 的比值,即重传矩阵 T 中非零向量个数与 m 的比值.它与需求率 u_i、丢包率 p_i 和信宿数 n 有关,并且是它们的增函数,记为 $RR(u_i,p_i,n)$.

- 编码率:编码包数与待重传的原始数据包数的比值,它与 u_i,p_i,n 及 m 有关,记为 $NCR(u_i,p_i,n,m)$.

于是,根据大数定律,经过首播及重传后,能够被成功接收的原始数据包数为

$$m\prod_{i=1}^{n}(1-p_i)^{u_i}+m\prod_{i=1}^{n}(1-p_i)^{u_i p_i}\left[1-\prod_{i=1}^{n}(1-p_i)^{u_i}\right].$$

信源的两轮广播次数为 $m+m\cdot RR\cdot NCR$.

如果重传的数据包仍存在丢包,则将待发送的数据包放在下一批再编码重传.于是,可以得到平均带宽效率为

$$AVE=\frac{m\prod_{i=1}^{n}(1-p_i)^{\mu_i}+m\prod_{i=1}^{n}(1-p_i)^{\mu_i p_i}\left[1-\prod_{i=1}^{n}(1-p_i)^{\mu_i}\right]}{m+m\cdot RR\cdot NRC}=\frac{\prod_{i=1}^{n}(1-p_i)^{\mu_i}+\prod_{i=1}^{n}(1-p_i)^{\mu_i p_i}\left[1-\prod_{i=1}^{n}(1-p_i)^{\mu_i}\right]}{1+RR\cdot NRC}.$$

由上式可知,平均带宽效率和信宿数、数据包数、需求率、丢包率等参数密切相关.当固定其中一些参数值,让其他参数变化时,平均带宽效率的变化情况如图 11~图 14 的仿真结果所示.仿真是由一个信宿通过 MBS (multicast broadcast services)向相邻的 n 个信宿广播 m 个数据包.每个信宿丢包率为在一定范围内随机选取.同时,每个信宿的需求率 u_i 在平均需求率 u 为常数的条件下随机变化.图中每个点的平均带宽效率是通过计算 2 000 次迭代结果的平均值得到.在上述场景下,图 11~图 13 中 ONCR,RNCR 以及 ANCR 的曲线是根据文献[11]中提供的算法进行仿真测试得到的结果,MNCR 的曲线为本文提出的基于匹配理论的网络编码重传算法进行仿真测试得到的结果.图 14 为 MNCR 算法在平均丢包率 p 为常数的条件下,平均带宽效率随着 p 的不同而随之变化的情况.

其中,对于图 11~图 13 的仿真,各信宿的丢包率在 0.1~0.3 之间随机取值.下面我们结合仿真结果对算法性能进行定性分析.

图 11 是数据包数 m 和平均需求率 u 固定时,平均带宽效率随信宿数 n 的变化情况.从仿真结果可以看出,当信宿较多时,MNCR 算法的效率和 ANCR 算法接近.这是由于编码包必须满足解码条件,当信宿较多时,能将多于 3 个以上原始包编在一起的概率很小;而当信宿较少时,相应的效率要优于 ANCR 算法.这是由于此时可以将多于 3 个以上原始包编在一起的概率增大,从而优势更加明显.

图 12 是数据包数 m 和信宿数 n 固定时,平均带宽效率随需求率 u 的变化情况.从仿真结果可以看出,随着平均需求率的降低,平均带宽效率逐步增大.从理论上分析,这是由于需求率低时,需要重传的数据包也较少,即重传矩阵 T 中元素 1 比较稀少,将多于 3 个以上原始包编在一起的概率也比较大,所以平均带宽效率相应地大一些.

图 13 是信宿数 n 和平均需求率 u 固定时,平均带宽效率随数据包数 m 的变化情况.从仿真结果可以看出,当数据包数增加时,平均带宽效率也随之增大.这是由于随着数据包数的增加,重传矩阵 T 的行向量个数增加,而维数不变,行向量的线性相关性增强,对应的每个需要重传的数据包能参与编码的机会增大,从而平均带宽效率也随之增大.

由于本文中的算法最多可以将 4 个原始包编在一起,所以比文献[11]中提到的将两个包编码重传的方法减少的次数要多,从而平均带宽效率相应地也有显著提高,特别是当数据包数增大或者需求率低时优势更加明显.

图 14 是当数据包数 m 和信宿数 n 固定,在不同平均丢包率 p 的条件下,平均带宽效率随平均需求率 u 的变化情况.从仿真结果可以看出,带宽效率随着 p 增大而逐步降低.从理论上分析,这是由于当 p 较小时,需要重传的包数少,即重传矩阵 T 中元素 1 比较稀少,从而可以获得比较大的编码机会,取得较高的带宽效率.

从图 13 可以看出,随着原始数据包数的增大,编码增益也随之增大,所以应该尽量采用较大批量的数据包数 m;但同时不得不考虑另一问题,即等待的包越多,耗费的发包和反馈时间越长,同时算法的求解时间也将随之增加,从而会带来平均时延的增加.所以,选择多大批次需要综合考虑.关于此问题,将在以后的工作中进一步加以研究.

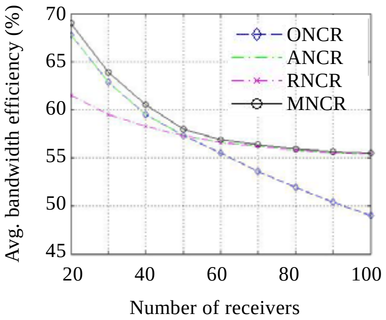


Fig.11 Impact of *n* to average bandwidth efficiency under *m*=20 and *u*=0.3

图 11 平均带宽效率随 *n* 的变化情况(*m*=20,*u*=0.3)

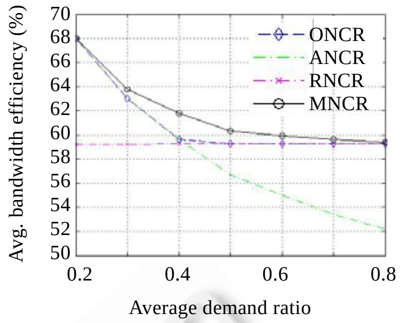


Fig.12 Impact of *u* to average bandwidth efficiency under *m*=20 and *n*=30

图 12 平均带宽效率随 *u* 的变化情况(*m*=20,*n*=30)

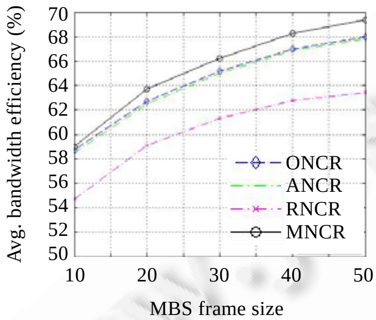


Fig.13 Impact of *m* to average bandwidth efficiency under *n*=30, *u*=0.3

图 13 平均带宽效率随 *m* 的变化情况(*n*=30, *u*=0.3)

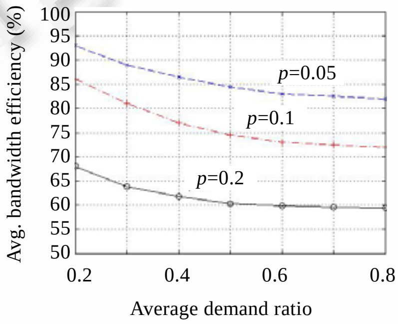


Fig.14 Impact of *u* to average bandwidth efficiency under *p*=0.05, *p*=0.1 and *p*=0.2

图 14 平均带宽效率随 *u* 的变化情况(*p*=0.05,*p*=0.1,*p*=0.2)

5 结论和进一步的研究

本文通过图论中的最佳匹配理论给出了一种寻找编码机会的优化算法——MNCR 算法.MNCR 算法不但能够找出最多的两两编码机会,而且还考虑了将尽可能多的数据包编在一起的可能性,从而可以最大限度地减少重传次数.仿真测试结果表明,MNCR 算法与 ONCR 算法、RNCR 算法以及 ANCR 算法相比,有效地减少了广播源发送次数,提高了传输效率.尤其在数据包数较大或者需求率低时,与其他算法相比优势更加明显.另外,本文还给出了 MNCR 算法在一般情况下的具体实现方法.

本文在算法设计时,限于篇幅,没有整体考虑时延因素,下一步的工作将在算法中考虑这一问题.可以将时延问题转化为权值问题,给出基于权值的匹配优化算法.

致谢 在此,我们向对本文的工作给予支持和建议的所有评审专家和同行表示由衷的感谢.

References:

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

[2] 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. [doi: 10.1109/INFCOM.2005.

1498443]

[3] 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]

[4] Widmer J, Fragouli C, Boudec JYL. Low-Complexity energy-efficient broadcasting in wireless ad-hoc networks using network coding. In: Proc. of the 1st Workshop on Network Coding. 2005. http://www.netcod.org/indexold.htm

[5] 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. [doi: 10.1109/INFOCOM.2006.45]

[6] Li L, 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. [doi: 10.1109/INFCOM.2007.203]

[7] Aly SA, Kamal AE. Network protection codes against link failures using network coding. In: Proc. of the Global Telecommunications Conf. 2008 (IEEE GLOBECOM 2008). IEEE, 2008. 1−6.

[8] Katti S, Rahul H, Hu WJ, Katabi D, Medard M, Crowcroft J. XORs in the air: Practical wireless network coding. IEEE/ACM Trans. on Networking, 2008,16(3):497−510. [doi: 10.1109/TNET.2008.923722]

[9] Nguyen D, Tran T, Nguyen T, Bose B. Wireless broadcasting using networking coding. IEEE Trans. on Vehicular Technology, 2009,58(2):914−925. [doi: 10.1109/TVT.2008.927729]

[10] Xiao X, Wang WP, Yang LM, Zhang S. Wireless broadcasting retransmission approach based on network coding. Journal on Communications, 2009,30(9):69−75 (in Chinese with English abstract).

[11] Sorour S, Valaee S. Adaptive network coded retransmission scheme for wireless multicast. In: Proc. of the ISIT 2009. Seoul, 2009. 2577−2581. [doi: 10.1109/ISIT.2009.5205995]

[12] Ho T, Koetter R, Médard M, Karger DR, Effros M. The benefits of coding over routing in a randomized setting. In: Proc. of the IEEE Int'l Symp. on Information Theory (ISIT 2003). 2003. 442.

[13] Huang Z, Wang X. Research on the optimization problems in network coding. Journal of Software, 2009,20(5):1349−1361 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3503.htm [doi: 10.3724/SP.J.1001.2009.03503]

[14] Papadimitriou CH, Steiglitz K, Wrote; Liu ZH, Cai MC, Trans. Combinatorial Optimization: Algorithms and Complexity. Beijing: Tsinghua Univeristy Press, 1988. 279−309 (in Chinese).

附中文参考文献:

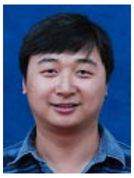
[10] 肖潇,王伟平,杨路明,张帅.基于网络编码的无线网络广播重传方法.通信学报,2009,30(9):69−75.

[13] 黄政,王新.网络编码中的优化问题研究.软件学报,2009,20(5):1349−1361. http://www.jos.org.cn/1000-9825/3503.htm [doi: 10.3724/SP.J.1001.2009.03503]

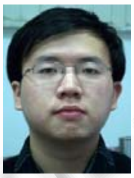
[14] Papadimitriou CH, Steiglitz K,著.刘振宏,蔡茂诚,译.组合最优化:算法和复杂性.北京:清华大学出版社,1988.279−309.

戴彬(1977－),男,江苏镇江人,博士,副教授,主要研究领域为网络编码,P2P 技术,协作中继,宽带移动业务.

曹志刚(1973－),男,讲师,主要研究领域为网络编码,网络拓扑测量.



杨军(1982－),男,博士生,主要研究领域为网络编码,P2P 网络拓扑发现.



黄辰(1982－),男,博士生,主要研究领域为网络编码,协作通信.

王芙蓉(1966－),女,博士,教授,博士生导师,主要研究领域为网络优化,宽带移动通信.