关于纠删码的研究与进展 [1]

慕建君 路成业 王新梅

(西安电子科技大学综合业务网国家重点实验室 西安 710071)

摘 要 该文简述了几类纠删码的纠删原理，系统地综合分析了各类纠删码的优缺点及其相互区别与联系。证明了若选取 MDS(Maximal Distance Separable) 码作为纠删码，只要接收者接收到源数据个数的数据，就能恢复原来的源数据。分析结果表明：复损码以及旋风 (Tornado) 码不仅能以线性时间可编码和可成功地译码，而且能以任意接近删除信道容量的速率进行传输。最后指出了目前复损码的研究中需要解决的一些问题。这些分析和结论为进一步研究纠删码提供了理论基础和新的思路。

关键词 纠删码，范德蒙码，柯西码，复损码，旋风码，二元删除信道

中图号 TN918.1

1 引 言

近年来，多址传输 (Multicast) 已成为互联网的一个重要组成部分。大型软件通过互联网传送给众多用户时要求用多址传输或广播 (Broadcast) 传输 [1] 。这些传输必须是完全可靠的，同时有小的网络开销和支持众多各类用户的随机访问。ARQ(Automatic Repeat reQuest) 技术和分层恢复等新技术可有效地提高数据传输的可靠性和避免网络拥塞的出现，但可能导致大的时延，这个缺点在大容量数据的实时传输 (如互联网等) 中是不可接受的。克服这一缺点的办法是利用编码的方法 [1] ，即把要传输的 k 比特的原数据编码为 $n(n>k)$ 比特的数据后发送出去，若接收方接收到足够量的数据，则运用适当的译码方法就可恢复 k 个比特的源数据，称这种码为前向纠错 (FEC， Forward Error Correcting) 码 [1], 或称复损码 (Loss-Resilient Code)[2] 。通常在接收到的编码数据流中数据的位置是已知的，所采用的信道模型是删除信道 [3], 此信道中每个编码符号丢失的概率均为 p, 且在传输中编码符号的丢失是相互独立的。我们的目的是构造具有较低编码和译码时间复杂度且性能良好的纠删码，即这种纠删码既能以线性时间可编码和可成功地译码又能以任意接近删除信道容量的速率进行传输。

2 纠删码及其纠删原理

在通信系统中前向纠错码纠正的误码所在的错误位置事先一般是不知道的，而在删除信道中错误的数据帧被遗弃，丢失的数据在数据流中的位置是知道的，这样纠删码比纠错码处理起来容易些。一个 (n,k) 纠删码是把 k 个源数据编码为 $n(n>k)$ 个数据，使得用这 n 个数据中任意 k 个编码数据均可重构原来 k 个源数据 [4,5](如图 1 所示) 。一个 (n,k) 线性纠删码是可以表示为 $\underline{y}=\underline{x}G$ ，其中 $\underline{x}=(x_0,x_1,\cdots,x_{k-1})$ 是源数据， $\underline{y}=(y_0,y_1,\cdots,y_{n-1})$, G 为 $k\times n$ 矩阵，称 G 为此 (n,k) 线性纠删码的生成矩阵，若 G 的任意 k 列组成的子矩阵 G' 均可逆，则有如下定理：

[1] 2001-02-26 收到， 2001-08-10 定稿

国家自然科学基金资助项目 (69972035), 重庆市 / 信息产业部移动通信技术重点实验室开放课题基金资助项目

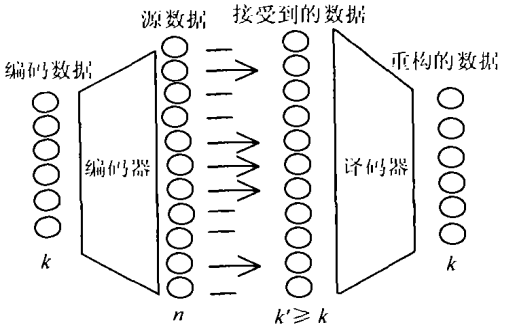

图 1　编译码过程的图形表示

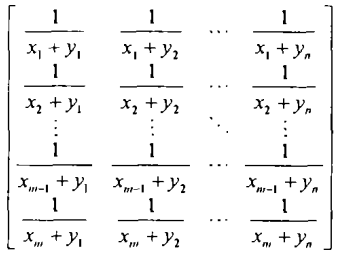

图 2　柯西矩阵的结构图

定理 1[4,5]　设 G 为一 (n,k) 线性纠删码的生成矩阵，若 G 的任意 k 列组成的子矩阵 G' 均可逆，则利用接收到的任意 k 个数据均可重构原来的 k 个源数据。

证明　设 $\underline{y}'$ 为 $\underline{y}$ 的任意 k 个分量组成的向量，即 $\underline{y}'$ 为接收到的数据，则利用对应于 $\underline{y}$ 的 k 个分量 (即 $\underline{y}'$) 的 k 个方程组可以重构源数据。不妨设 $G'_{k\times k}$ 为表示这个方程组的系数矩阵，则 $\underline{y}' = \underline{x}G'$，由此得源数据 $\underline{x} = \underline{y}'(G')^{-1}$。　　证毕

对有限域 GF(p^r) (p 为素数，r 为正整数)，利用以下引理有更一般的定理：

引理 1[6]　设 C 为 GF(p^r) 上一线性分组码，则 C 为 MDS(Maximal Distance Separable) 码当且仅当其生成矩阵 G 的任意 k 列线性无关。

定理 2　在 GF(p^r) 上若取 (n,k) 线性纠删码为 MDS 码，则利用接收到的任意 k 个数据均可重构源数据。

证明　由于此 (n,k) 线性纠删码为 MDS 码，由引理 1 知生成矩阵 G 的任意 k 列线性无关，从而由这 k 列组成的子矩阵 G' 可逆，结合定理 1 知结论成立。　　证毕

若选取纠删码的生成矩阵为范德蒙矩阵和柯西矩阵，则可得相应的纠删码——范德蒙码 (Vandermonde code)[4,6] 和柯西码 [6](Cauchy code)，它们都属于 RS 码类。从基于随机二部图的校验关系出发可构造一种特殊的纠删码——复损码 [2]。

2.1 范德蒙码

定义 1　若选取编码生成矩阵为 $G_{k\times n}$，$G^T = (g_{ij})$，其中 $g_{ij} = x_i^{j-1}$，$x_i \in \mathrm{GF}(p^r)$ (p 为素数，r 为正整数)，则称所得纠删码为范德蒙码，G 的任意 k 列组成的子方矩阵 G' 的转置矩阵 $(G')^T$ 为范德蒙矩阵，若 $x_i \neq 0$ ($i = 1,2,\cdots,k$) 互不相同，则 $|(G')^T| \neq 0$，从而 $|G'| \neq 0$，即 G 的任意 k 列组成的子方矩阵 G' 为非奇异的，因此，这样得到的矩阵满足纠删码中生成矩阵的特性。

2.2 柯西码

定义 2　设 $\{x_1,x_2,\cdots,x_m\}$ 和 $\{y_1,y_2,\cdots,y_n\}$ 是有限域 F 中两个元素集，若对 (1)$\forall i \in \{1,2,\cdots,m\}$，$\forall j \in \{1,2,\cdots,n\}$ 有 $x_i + y_j \neq 0$；(2) 对 $\forall i,j \in \{1,2,\cdots,m\}(i \neq j)$ 有 $x_i \neq x_j$ 和 $\forall i,j \in \{1,2,\cdots,n\}(i \neq j)$ 有 $y_i \neq y_j$，则称如图 2 所示的矩阵为域 F 上的柯西矩阵。

在有限域 F 上，设 $I_{k\times k}$ 为单位矩阵，$C_{k\times(n-k)}$ 为柯西矩阵，若取生成矩阵 $G = (I|C)$，则称所得纠删码为柯西码 [6]，而且易见如下引理：

引理 2[7]　设 C 为某一有限域上的柯西矩阵，则 C 的任意子方阵为非奇异矩阵。

引理 3[7]　在有限域 F 上，设 $I_{k\times k}$ 为单位矩阵，$C_{k\times(n-k)}$ 为柯西矩阵，若取 (n,k) 线性分组码的生成矩阵为 $G = (I|C)$，则此码为 MDS 码当且仅当 C 的每一个子方阵为非奇异的。

利用引理 2、引理 3 和定理 2，我们有

定理 3 在有限域 F 上，设 $I_{k\times k}$ 为单位矩阵，$C_{k\times(n-k)}$ 为柯西矩阵，则对于取生成矩阵为 $G=(I|C)$ 的柯西码，利用接收到的任意 k 个数据均可重构原来的 k 个源数据。

2.3 复损码

定义一有 n 个信息比特和 βn 个校验比特的码 $C(B)$，使得它们分别与给定二部图 B(左边集有 n 个结点，右边集有 βn 个结点) 的两结点集中的结点相对应。

$C(B)$ 型码的编码方式为每个校验比特等于二部图 B 中此校验比特的所有邻接比特的和(如图 3 所示)。若称二部图 B 中左边的结点为变量结点，右边的结点为校验结点 (如图 4 所示)，则 $C(B)$ 型码的删除错误译码算法为 [2]:

(1) 初始化　初始化校验结点为 0，把非删除错误变量结点的值加到校验结点的当前值上，同时去掉这些变量结点和由这些变量结点引出的所有边，并令 $i=1$，称所得的 B 的子图为 G_i;

(2) 恢复变量结点的值　在以上所得子图 G_i 中寻找一度数为 1 的校验结点，将此校验结点的值传给它唯一的邻接变量结点 ℓ，于是恢复了结点 ℓ 的值，然后将结点 ℓ 的值加到其所有邻接校验结点的当前值上，同时去掉从变量结点 ℓ 引出的所有边，称所得的 G_i 的子图为 G_{i+1};

(3) 终止条件　令 $i:=i+1$，对 (2) 所得的子图 (这时即 G_i) 重复进行步骤 (2)，直到其所得的子图中找不到度数为 1 的校验结点，或所有的变量结点均已恢复。若所有的变量结点均已恢复，则称此删除错误译码算法可成功译码。

若级联 $C(B)$ 型码 $C(B_0),C(B_1),\cdots,C(B_m)$ 和传统复损码 $C(C(B_i)$ 有 $\beta^i n$ 个信息比特和 $\beta^{i+1}n$ 个校验比特，$C(B_i)$ 的校验比特即 $C(B_{i+1})$ 型码的信息比特)，与码 $C(B_i)$ 对应的二部图 B_i 的左边和右边分别有 $\beta^i n$ 个和 $\beta^{i+1}n$ 个结点 $(i=0,1,\cdots,m)$，并取 m 使得 $\beta^{m+1}n\approx\sqrt{n}$，最后一级码 C 是一有 $\beta^{m+1}n$ 个信息比特和码率为 $1-\beta$ 的传统纠删码 (如柯西码 [5])。于是，得到一有 n 个信息比特和 $\sum_{i=1}^{m+1}\beta^i n+\beta^{m+2}n/(1-\beta)=n\beta/(1-\beta)$ 个校验比特的码率为 $1-\beta$ 的复损码 $C(B_0,B_1,\cdots,B_m,C)$(如图 5 所示)。

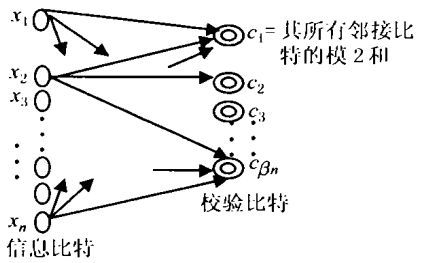

图 3　用二部图 B 定义从信息比特到校验比特的一个映射

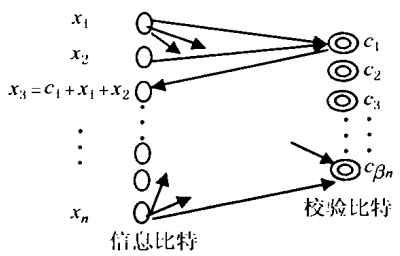

图 4　码 $C(B)$ 中用比特 x_1, x_2 和 c_1 来求解 x_3

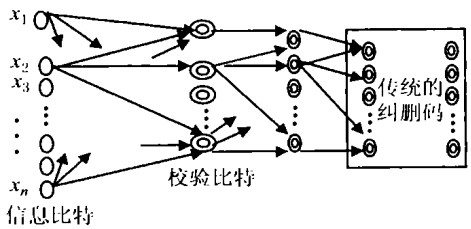

图 5　基于二部图的复损码的结构示意图

从复损码 $C(B_0,B_1,\cdots,B_m,C)$ 的构造知其编码思想由 $C(B)$ 型码的编码方式易得，其译码思想为：若码 C 的译码算法可恢复其所有损失，当然译出 $C(B_m)$ 的所有校验比特，由此就能恢复 $C(B_m)$ 的所有信息比特，如此逐步递推译码直到用 $C(B_0)$ 的校验比特来恢复原来的 n 个信息比特。

复损码 $C(B_0,B_1,\cdots,B_m,C)$ 译码的核心是解一个线性方程组，若所采用的随机方程组是稀疏的，即每一个方程所包含变量的平均个数较少，则称所得复损码为 Tornado 码 [1]。这种稀疏矩阵使得 Tornado 码具有快的编译码速度，其优点是 Tornado 码能使众多用户以高的效率及时地获得数据。Tornado 码的局限性是为了得到更快的编译码速度必须接收到略大于源数据个数 k 的数据 (即付出的代价)。

为了设计和分析复损码的译码算法，对给定的二部图 Luby[2] 用一个微分方程为工具将译码过程模型化后给出了复损码可成功译码的一个分析性条件。利用这个条件，Luby[8,9] 和 Mackay[10] 等分析了基于正则二部图和非正则二部图所得的复损码后得出结论：正则二部图不可能产生接近最优的码。因此，非正则二部图一定是我们构造复损码时选取随机二部图的一类重要的图，同时基于这个条件，Luby[2] 设计了一个二部图，并且指出若构造复损码的 $C(B)$ 型码具有这种结构，则所得复损码能以大的概率成功译码，即有以下定理：

定理 4[2]　对任意码率 R 和给定参数 $\varepsilon>0$，若码长 n 足够大，一定存在一复损码能以大的概率和 $O(n\ln(1/\varepsilon))$ 的运行时间从 $(1-R)/(1-\varepsilon)$ 的随机损失率中恢复源数据。

由定理 1 可见：(1) 通过如上级联得到的复损码能以线性时间进行编码和成功译码，而传统的复损码 (如范德蒙码和柯西码) 的编译码时间至少为平方时间数量级的。于是，如此基于 Luby[2] 设计的二部图而构造的复损码的编译码时间可大大降低。这就是要采用级联的方式构造新的复损码的原因。(2) 将 k 比特源数据编码为 $n(n>k)$ 比特数据后发送出去，对任意的 $\varepsilon>0$ 和码长 n 足够大，若接收方接收到 $k'=k(1+\varepsilon)$ 个数据，则一定存在一复损码能以大的概率和 $O(n\ln(1/\varepsilon))$ 的运行时间恢复源数据。

3　几类纠删码编译码时间复杂度的分析

由文献 [4] 知范德蒙码的编码时间复杂度为 $O(n^2)$，译码时间复杂度高于 $O(n^2)$，后者主要由于需要空间和时间来求大矩阵的逆，而柯西码的编译码时间复杂度均为 $O(n^2)$[6]，它优于范德蒙码，这是由于柯西码译码时不需要求大矩阵的逆，而且把乘法和除法运算分别转化为有限域上的加法和减法运算，从而可用 "异或" 运算实现。Elias 已证明删除信道的信道容量为 $1-p$，而且随机线性码可在删除信道下以任意的速率 $R(R<1-p)$ 传输，其编码和译码的时间复杂度分别为 $O(n^2)$ 和 $O(n^3)$[3]。标准的 MDS 码 (如范德蒙码和柯西码) 类由 RS 码来给定的，理论上 RS 码的编码和译码的时间复杂度分别为 $O(n\log n)$ 和 $O(n\log^2 n\log\log n)$[11]，但是这类码不能与删除信道的信道容量很接近。于是一方面得到基于 RS 码的 MDS 码具有较低的编译码复杂度，但是不能与删除信道的信道容量任意接近；另一方面有随机线性码可与删除信道的信道容量任意接近，但具有高的编译码复杂度。

文献 [12] 首次设计一纠删码，它不仅能以线性时间可编码和可成功地译码而且能以任意接近删除信道的容量的速率进行传输。Luby[2] 等人用不同的方法给出了一具有线性时间编译码算法的复损码，而且此码能以任意接近删除信道容量的速率进行传输，即对任意 $\varepsilon>0$，这种码可得到码率为 $R=1-p(1+\varepsilon)$ 的码，其译码算法能以大的概率和 $O(n\ln(1/\varepsilon))$ 的运行时间恢复所传符号的比率为 p 的随机错误，它同样能以时间 $O(n\ln(1/\varepsilon))$ 进行编码 (n 为码长)。这种复损码属于低密度校验码 [8,13] 类。

4 结束语

本文简述了范德蒙码、柯西码、复损码和 Tornado 码这几类纠删码的纠删原理，系统地综合分析了各类纠删码的优缺点及相互区别与联系。证明了若选取 MDS 码作为纠删码，只要接收者接收到源数据个数的数据，就能恢复原来的源数据。分析结果表明：复损码 (如 Tornado 码) 不仅能以线性时间可编码和可成功地译码，而且能以任意接近删除信道的容量的速率进行传输，即对任意 $\varepsilon > 0$ 和给定的码率 R，当码长 n 充分大时，一定存在某一复损码，当接收到 $k' = k(1+\varepsilon)$ 个数据时，此复损码就能以高的概率和 $O(n\ln(1/\varepsilon))$ 的运行时间恢复源数据 (k 为源数据的比特数)。其缺点是为了得到更快的编译码速度必须接收到略大于源数据个数 k 的数据 (即付出的代价)。这些分析和结论为进一步研究纠删码提供了理论基础和新的思路。关于复损码的研究目前主要有如下几个方面的问题：

(1) 利用最优化理论的知识研究构造获得最优二部图的度分布的方法，从而构造良好性能的复损码；

(2) 进一步从理论上分析并设计复损码的迭代译码算法，同时寻找此译码算法稳定收敛的条件。

参 考 文 献

[1] J. W. Byers, M. Luby, M. Mitzenmacher, A. Rege, A digital fountain approach to reliable distribution of bulk data, available at http://www.icsi.berkeley. edu/~luby/, 1998.

[2] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, V. Stemann, Practical loss-resilient codes, available at http://www.icsi.berkeley.edu/~luby/, 1998.

[3] P. Elias, Coding for two noisy channels, Information Theory, Third London Symposium, 1955, 61–67.

[4] L. Rizzo, Effective erasure codes for reliable computer communication protocols, ACM Computer Communication, Review, 1997, 27(2), 24–36.

[5] L. Rizzo, On the feasibility of software FEC, available as http://www.iet.unipi.it/'luigi/softfec.ps.

[6] J. Blomer, M. Mitzenmacher, A. Shokrollahi, An XOR-based erasure-resilient coding scheme, ICSI Technical Report, No.TR-95048[R], August 1995, Available at http://www.icsi.berkeley.edu/~luby/.

[7] F. J. MacWilliams, N. J. A. Sloane, The Theory of Error-Correcting Codes, North Holland, Amsterdam, 1977, Chapter 11.

[8] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, Improved low-density parity-check codes using irregular graphs and belief propagation, In Proc. of IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, 1998, 16–21.

[9] M. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. Spielman, Analysis of random processes via And-Or tree evaluation, In Proc. of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, California, 1998, 364–373.

[10] D. J. C. MacKay, S. T. Wilson, M. C. Davey, Comparison of constructions of irregular Gallager codes, IEEE Trans. on Communications, 1998, COM-47(10), 1449–1454.

[11] R. E. Blahunt, Theory and Practice of Error Control Codes, Reading, MA, Addison Wesley, 1983, Chapter 11.

[12] N. Alon, M. Luby, A linear time erasure-resilient code with nearly optimal recovery, IEEE Trans. on Information Theory, 1996, IT-42(6), 1732–1736.

[13] R. G. Gallager, Low Density Parity-Check Codes, Cambridge, MA. MIT Press, 1963, Chapter 1.

RESEARCH AND DEVELOPMENT ON ERASURE CODES

Mu Jianjun Lu Chengye Wang Xinmei

(*National Key Lab. of Integrated Service Networks, Xidian University, Xi'an* 710071, *China*)

Abstract This paper describes the principles of a few types of erasure codes, and analyzes their merits, drawbacks and relationships systematically. It is shown that if MDS codes are chosen as erasure codes, a receiver can reconstruct the original source data once it receives any portion of the encoding data equal to the length of the message. The analysis shows that loss-resilient codes and Tornado codes can not only be both encoded and decoded successfully in linear time, but also can come arbitrarily close to the channel capacity. Finally, some problems on loss-resilient codes which remain to be solved are presented. These analyses and conclusions provide theoretical base and new ideas for further studying erasure codes.

Key words Erasure code, Vandermonde code, Cauchy code, Loss-resilient code, Tornado code, Binary erasure channel

慕建君： 男，1965 年生，博士生，目前的研究兴趣为编码、信息论与应用数学.
路成业： 男，1977 年生，硕士生，研究兴趣为信息论、编码 / 调制理论.
王新梅： 男，1937 年生，教授，博士生导师，长期从事信息论、编码和密码学的教学与研究.