

数字喷泉码及其应用的研究进展与展望

慕建君, 焦晓鹏, 曹训志

(西安电子科技大学计算机学院, 陕西西安 710071)

摘 要: 基于 TCP 协议的大部分网络通信都将接收到的数据包当作一个有序序列进行处理. TCP 协议的这种有序序列模式限制了其在大量数据传输或将数据分布到大量用户的应用. 基于数字喷泉码的数据传输模式不需要有序的数据序列, 从而简化了网络中数据的传输方式, 使得数字喷泉码成为一类适用于可靠通信的有效编码技术, 而具有广阔的应用前景. 本文综述了几类典型数字喷泉码的原理及其优缺点; 探讨了这几类数字喷泉码的具体应用; 指出了数字喷泉码研究中需要解决的一些关键问题. 最后, 对数字喷泉码的发展前景及研究方向进行展望.

关键词: 数字喷泉码; 纠删码; 删除信道; LT 码; Raptor 码

中图分类号: TN919.3 **文献标识码:** A **文章编号:** 0372-2112 (2009) 07-1571-07

A Survey of Digital Fountain Codes and Its Application

MU Jian-jun, JIAO Xiao-peng, CAO Xun-zhi

(School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: Most network communication based on TCP treats data as an ordered sequence of packets. This ordered-sequence paradigm of TCP is too restrictive when the data is large or is to be distributed to a large number of users. In the data transmission paradigm based on digital fountain codes, obviating the need for ordered data simplifies data delivery. Thus, digital fountain codes are a class of efficient codes which are suitable for reliable communication, and have been widely used in many respects. This paper surveys the principles of a few types of digital fountain codes and their advantages and disadvantages, and investigates their engineering applications. Some key problems needed to be solved in the study of digital fountain codes are pointed out. Finally, the future development and research directions of digital fountain codes are discussed.

Key words: digital fountain codes; erasure codes; erasure channel; LT codes; Raptor codes

1 引言

互联网(Internet)中数据是以数据包的形式传输. 每个数据包都有一个标志其源与目的的头标识符和一个描述此数据包在给定数据流中的绝对或相对位置的序列号. 数据包通过网络传输时由于中间路由器上的缓冲溢出等原因一些数据包可能丢失而永远不会到达目的地; 另外一些数据包由于其校验和不能匹配也可认为被丢失. 因此, 互联网是二元删除信道(Binary Erasure Channel, BEC)的典型真实信道模型^[1].

网络中数据传输的可靠性主要利用合适的协议来确保. 互联网中普遍使用的 TCP/IP (Transmission Control Protocol/ Internet Protocol) 协议通过重传传输窗口中发送方未确认的数据包来确保所传数据的可靠性. 众所周知在多播传输以及诸如无线或卫星连接等严重失真的信道中 TCP/IP 协议的性能并不好. 而且对于基于确认的

TCP/IP 协议, 当接收方与发送方之间的距离较远时会产生发送方必须等待确认而不能发送数据这样的闲时段而导致大的时延, 所以这时 TCP/IP 协议的性能也不好. 这些缺点对于互联网中大容量数据的实时传输是不可接受的.

克服这些缺点的一种方法是利用纠删编码技术^[2~4], 即在发送端把要传输的 k 个源数据包编码为 n ($n > k$) 个编码包后通过网络发送出去, 只要接收方接收到 n 个编码包中的任何 k 个编码包, 则运用适当的译码方法就可重构 k 个源数据包. 对于实际应用而言, 我们希望构造出具有非常快速编译码算法, 同时又能纠尽量多删除错误的纠删码.

对于删除概率为 p 的删除信道, Elias 证明了用最大似然译码算法进行译码时随机码的码率能以指数小的错误概率任意逼近删除信道的容量限 $1 - p$ ^[1]. 而且, 这时线性分组码的最大似然译码等价于求解线性方程

组. 而求解线性方程组可用高斯消去法在多项式时间内完成. 然而, 高斯消去法并不足够快.

RS 码可克服随机码的这一缺点. RS 码能够在平方时间内纠最大数量的删除错误^[5]. 然而, 对许多实际应用而言, RS 码平方时间译码算法仍然不够快^[5,6]. 在文献[7]中, Luby 等构造了一种具有线性时间编译码算法的 Tornado 码可任意逼近删除信道的容量限. 由于 Tornado 码具有关于码长的线性时间复杂度编译码算法, 因此, 对于低码率 Tornado 码的编译码算法太慢. 这一缺点对于许多要求低码率的实际应用而言并不适用^[8].

传统的分组码用于数据传输时也有一些缺点. 当从一个发送者向多个接收者并发传输数据时具有单一删除概率 p 的删除信道也不适用. 因为这时从发送者向每个接收者的删除信道可能具有不同的删除概率. 在实际应用当中典型的做法是发送者或接收者可以通过深入研究其删除信道的特性以便对所采用信道的删除概率做一个合理的估计, 同时相应地调整码率. 但是如果对于极其众多的接收者, 或者像如接收信道容易发生突变的卫星或无线传输这样的情况, 要确定或跟踪每个接收者的数据损失率是不现实的. 这就迫使发送者采用对所有接收者来说最坏情况的数据损失率, 然而当实际的损失率小一些时, 这样做就会增加不必要的网络负担.

由此可见, 为了设计适用于互联网中大容量数据的实时传输系统 Tornado 码以及传统的纠删编码技术并不适合. 近年来受到普遍关注的数字喷泉码 (Digital Fountain Codes) 成为构造这种可靠传输方案的最佳技术^[7~10], 这类码可以解决前面遇到的所有问题.

本文首先介绍纠删码的纠删原理和数字喷泉码的概念, 然后综述几类典型数字喷泉码的原理及其优缺点, 探讨这些数字喷泉码的具体应用, 指出数字喷泉码研究中需要解决的一些关键问题. 最后, 对数字喷泉码的发展前景及研究方向进行展望.

2 纠删码与数字喷泉码

2.1 纠删码及其纠删原理

纠删码与纠错码的差别在于纠删码的译码器知道编码数据流中错误发生的位置. (n, k) 纠删码把 k 个源数据包编码为 n ($n > k$) 个数据包, 使得接收方用这 n 个编码包中的任何 k 个编码包均可恢复原来的 k 个源数据包^[3,4]. 称 $R = \frac{k}{n}$ 为 (n, k) 纠删码的码率. 一个 (n, k) 线性纠删码可表示为: $y = xG$, 其中 $y = (y_0, y_1, \dots, y_{n-1})$, $x = (x_0, x_1, \dots, x_{k-1})$ 是源数据包向量, G 为 $k \times n$ 矩阵, 称 G 为此 (n, k) 线性纠删码的生成矩阵, 容易得到如下定理:

定理 1^[3,4] 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, 则接收方利用接收到的任意 k 个数据包均可恢复原来的 k 个源数据包.

2.2 数字喷泉与数字喷泉码

1998 年, Luby 等首次提出了用于分布数据的数字喷泉 (Digital Fountain) 技术, 一个数字喷泉具有类似于水喷泉的特性: 当你在水喷泉下给水杯接水时, 你只想接到足够量的水来解渴, 而不必关心是那一点水流入你的杯中^[8]. 类似地, 借助数字喷泉编码技术一个客户端从一个或多个服务器接收编码包, 一旦收到足够的编码包, 那么该客户端就可以重构原文件. 而具体接收到编码包序列中的哪些编码包却并没有关系.

一个理想数字喷泉应该具有如下特性^[8,11]:

- (1) 一个源能够利用原数据产生无限编码包序列;
- (2) 对于被分割为 k 个数据包的一消息, 一旦接收到编码包流中任意 k 个编码包, 接收者就能够重构这一消息. 这种重构算法应该非常快.

数字喷泉可通过放松如下的一些要求来近似实现: (1) 发送者编码后的编码包数可以是有限的; (2) 编译码算法可以慢一些; (3) 接收者为了重构原消息而需要接收的编码包数目可略大于原数据包数目 k .

数字喷泉码是一类码率不受限的纠删码 (rateless erasure codes), 即从源符号编码产生的编码符号序列是无限的, 而且可以在线产生这些编码符号序列. 数字喷泉码产生的符号称为输出符号 (output symbols), 用来计算输出符号的 k 个符号称为输入符号 (input symbols). 对于给定的 k 个输入符号 $(x_1, x_2, \dots, x_k)$, 数字喷泉码可产生无限的输出符号流 $z_1, z_2, \dots$. 这些输入和输出符号可以是任意长度的向量. 每个输出符号等于若干随机独立选取的输入符号的异或和.

数字喷泉码和分组码的差别为:

- (1) 对于传统的分组码, 码的结构在信息传输之前确定; 而数字喷泉码可“在线 (online)”编码产生.

(2) 当采用传统的分组码传输信息时, 发送者与接收者之间都知道所采用的编码方法; 而对于数字喷泉码情况并不是这样. 由于编码的输出符号与数据的传输并发产生, 因此, 为了从输出符号恢复原信息, 必须将采用的编码方法与输出符号一并传输. 在符号对应于包的网络环境中, 可以给每个传输包增加用来表示该输出符号是由那些输入符号产生的头信息, 于是, 描写输入与输出符号之间关系的信息在接收端就可通过数据包头来获得, 或者可通过发送端与接收端之间依赖于应用的同步方法来获得.

总之, 数字喷泉码可以用纠删码来近似实现. 用 RS

码近似实现的数字喷泉码不实用^[10],1997 年 Luby 等首次提出了一种适用于网络环境数据传输的纠删码,也称为“复损码(Loss-Resilient Code)”^[7],随后给出了这一类码中度分布的设计算法^[12,13],并对此类码进行了大量的研究^[14~19].在此基础上 1998 年 Luby 等提出的 Tornado 码适合于近似实现数字喷泉码^[8],同时给出了数字喷泉码的一些实用实现^[20~21].2002 年 Luby 提出的 LT (Luby transform) 码^[22]以及 2004 年 Shokrollahi 提出的 Raptor 码也可以用来近似实现数字喷泉码^[23],而 Raptor 码是迄今为止最有效的数字喷泉码.下面介绍这几类典型数字喷泉码并对其优缺点进行分析与总结.

3 几类数字喷泉码

3.1 RS 码

数字喷泉码可以用纠删码来近似实现.若选取纠删码的生成矩阵为范德蒙矩阵和柯西矩阵,则可得相应的范德蒙纠删码(Vandermonde Erasure Code)和柯西纠删码(Cauchy Erasure Code)^[3,4],它们都属于 RS 码类.因此,理论上 RS 码可以用来设计数字喷泉码:对于 k 个数据包组成的一信息,只要接收到这一信息编码产生的编码包中的任何 k 个不同的编码包就可以恢复的原信息.然而,即使对于中等大小的编码包个数 k ,RS 码标准译码算法的平方时间复杂度太慢,这对于实际应用而言是不可行的.因此,RS 码并不是数字喷泉码的最好候选码.

3.2 Tornado 码

1997 年 Luby 等发表了一篇里程碑性的论文,基于级联二部图的方法首次提出了一种适用于网络环境数据传输的纠删码,也称为“复损码(Loss-Resilient Code)”^[7].这种码具有线性时间的编译码算法,而且能以任意逼近删除信道容量限的速率进行传输.即有如下结论.

定理 2^[7] 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高的概率和 $O(n \ln(1/\varepsilon))$ 的运行时间从 $(1-R)(1-\varepsilon)$ 的随机损失率中恢复源数据.

为了介绍(级联型)复损码,现定义一有 k 个信息比特和 $\beta k (0 < \beta < 1)$ 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 k 个结点,右边集有 βk 个结点)的两结点集中的结点相对应.

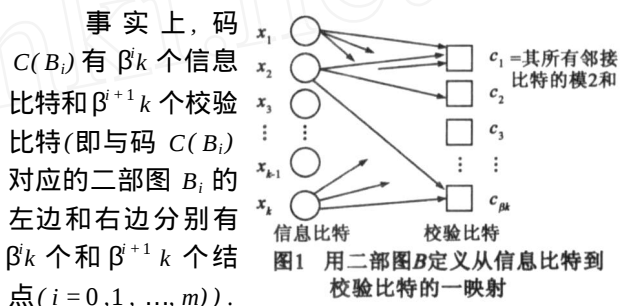
$C(B)$ 型码的编码方式为每个校验比特等于二部图 B 中此校验比特的所有邻接比特的异或和(如图 1 所示).若称二部图 B 中左边的结点为信息结点,右边的结点为校验结点,则 $C(B)$ 型码的删除错误恢复算法(erasure recovery algorithm)为:

(1) 初始化:按照相邻关系,把所有已接收到的信

息比特值和其邻接校验结点上已接收的校验比特值进行异或操作,在这些校验结点上生成新的比特值,并去掉这些已知值的信息比特结点和由这些信息比特结点引出的所有边,并令 $i = 1$,并设所得的二部图 B 的子图为 G_i .

(2) 迭代(恢复信息比特的值):在以上所得子图 G_i 中寻找一度数为 1 的校验比特,将此校验比特的值传给它唯一的邻接信息比特 ℓ ,于是恢复了信息比特 ℓ 的值,然后将信息比特 ℓ 的值和其所有邻接校验比特的当前值上进行异或操作,同时去掉从信息比特 ℓ 所在结点和由此结点引出的所有边,称所得 G_i 的子图为 G_{i+1} .

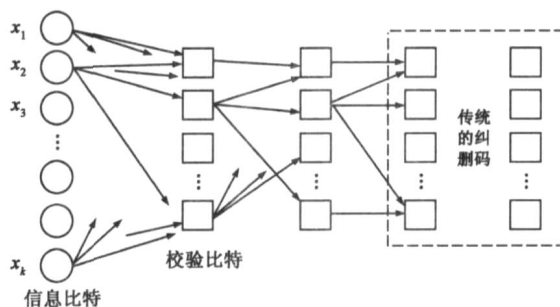
(3) 终止:对(2)所得的子图 G_{i+1} 重复进行步骤(2)直到其所得的子图中找不到度为 1 的校验比特,或所有的信息比特的值均已恢复.若所有的信息比特的值均已恢复,则称此删除错误恢复算法可成功译码.



若级联 $C(B)$ 型码 $C(B_0), C(B_1), \dots, C(B_m)$ 和传统复损码 C 并取扩展级数 m 使得 $\beta^{m+1} k \approx \sqrt{k}$, 最后一级码 C 是一有 $\beta^{m+1} k$ 个信息比特和码率为 $1 - \beta$ 的传统纠删码(如柯西码^[2]). 于是,得到一有 k 个信息比特和

$$\sum_{i=0}^{m+1} \beta^i k + \frac{\beta^{m+2} k}{1 - \beta} = \frac{k\beta}{(1 - \beta)}$$

个校验比特的码率为 $1 - \beta$ 的级联型复损码 $C(B_0, B_1, \dots, B_m, C)$ (如图 2 所示).



级联型复损码通常用度分布 (Λ, P) 刻画,其中 $\Lambda(x) = \sum_i \lambda_i x^{i-1}$, $P(x) = \sum_i \rho_i x^{i-1}$, $\lambda_i (\rho_i)$ 表示二部图中左边(右边)度数为 i 的边的比率^[7]. 文献[7]的主要结论为:对于初始删除概率为 δ 和具有度分布 (Λ, P)

的随机二部图,若对于所有 $x \in (0, \delta)$ 有 $\delta \Delta \left(1 - P(1-x) \right) < x$ 成立,则以上删除恢复算法可高概率成功译码。

若取复损码的度分布为 Heavy-Tail/ Poisson 度分布 $\left(\Lambda_n, P_n \right)$, 则称所得复损码为 Tornado 码^[8], 即

$$\Lambda_n(x) = \frac{1}{H(n)} \sum_{i=1}^n \frac{x^i}{i}, P_n(x) = e^{\mu_n(x-1)}$$

其中 μ_n 是满足 $\frac{1-e^{-\mu_n}}{\mu_n} = \frac{1}{H(n)} R \left(1 - \frac{1}{n+1} \right)$ 的唯一解,

$$H(n) = \sum_{i=1}^n \frac{1}{i}.$$

Tornado 码具有关于码长 n 的线性时间复杂度编译码算法,因此与 RS 码相比较, Tornado 码是一种更好的数字喷泉码的候选码。然而, Tornado 码还有一些致命缺陷^[10]。当我们利用 Tornado 码进行网络数据传输时,编码数据包由表示 Tornado 码的二部图来确定,这时必须提前确定所要产生的编码包数目(即 Tornado 码的码率)。理论上说可以产生任意大数目的编码包,然而,由于编译码中所需要的运行时间和存储空间与所产生的编码包数目成比例增长,因此这种方法在实际中并不可行。

由此可见, RS 码和 Tornado 码都具有固定码率这一缺陷。为了克服这一缺点 2002 年 Luby 提出了一种码率不受限的 LT 码(rateless codes)^[19]。由于不需要提前确定固定的码率,因此从给定的信息可产生无限长的编码包流。

3.3 LT 码

2002 年 Luby 提出了一类非常适用于网络数据分布的编码方案 - LT(Luby Transform) 码^[22]。LT 码是第一类码率不受限码的实用实现,即其码率不需要事先确定。LT 码是通用的(universal)数字喷泉码,也就是说对于具有不同删除概率的各种删除信道均是逼近最优的^[22]。

在数据传输时,将长为 N 的文件分割成 $k = N/l$ 个输入符号(即每个输入符号的长为 l),并称每个数据包为输入符号,称每个编码包为输出符号。

定义 1 LT 码的度分布 $\rho(d) (d \geq 1)$ 定义为一个输出符号结点的度为 d 的概率。

LT 码的编码算法:

- (1) 根据给定的度分布 $\rho(d)$ 随机选取一度 d ;
- (2) 随机选取 d 个不同的输入符号;
- (3) 编码后的输出符号为这 d 个不同输入符号的异或和。

LT 码的编码算法定义了一个连接输出符号到输入符号的二部图。假定 LT 码的译码器能够准确重构这个二部图,即译码器知道用哪 d 个不同输入符号来生成给定的输出符号,但不知道这些输入符号的具体值。当

LT 码的译码器接收到 K (通常 K 比 k 略大一点)个输出符号和编码时所采用的二部图后就可以开始译码。LT 码的译码算法^[22]:

- (1) 在二部图中寻找度为 1 的输出符号 y_i ;
- (2) 令 $x_j = y_i$;
- (3) 对于与输入符号 x_j 邻接的每个输出符号 $t_s (s \neq i)$, 令 $t_s = t_s \oplus x_j$;
- (4) 移去与输入符号 x_j 相连接的所有边;
- (5) 重复以上所有步骤,直到所有的输入符号 x_j 都被恢复,这时就可以恢复原文件。

可以设计这样的 LT 码使得对于平均度数为 $O(\ln k)$ 的编码符号,只需要接收到其中 $k + O(k)$ 个编码符号就可高概率成功译码^[22]。

度分布的设计对 LT 码的构造来说是至关重要,因为度分布直接决定 LT 码的译码是否成功。同时也决定产生编码包所需要的异或运算次数。文献^[22]已给出了 LT 码高概率成功译码的 Robust Soliton 度分布,产生每个编码包所需要的平均异或运算次数为 $\ln k$, 其中 k 为原来的数据包数目。因此, LT 码编译码 k 个数据包大概需要 $k \ln k$ 次异或运算,这就是说 LT 码的译码并不具有线性时间的译码复杂度。

因此,为了将 LT 码的译码复杂度降低为线性时间复杂度,即对于被分割为 k 个数据包的一消息,一旦接收到编码包流中任意 k 个编码包,接收者就能够利用 LT 码的线性时间译码算法重构这一消息。从而 LT 码可作为数字喷泉码的候选码。这样的纠错码可通过增加预编码的方法实现,这种码就是 Raptor 码^[23]。

度分布的设计是 LT 码构造中最关键的问题。Hytiä 等对 LT 码度分布的设计和优化进行了研究,并给出了优化 LT 码度分布的迭代算法^[24,25]。Harrelson 等证明了随机性受限的 LT 码和完全随机 LT 码的性能几乎一样,这就提供了一种通过分析有限随机 LT 码来研究 LT 码的一理论依据^[26]。Maneva 等提出了分析 LT 码译码错误概率的新模型^[27]。

3.4 Raptor 码

Raptor 码是由 Shokrollahi 提出的迄今为止最有效的一类数字喷泉码^[23]。Raptor 码将 LT 码的思想向前推进了一步。对于 LT 码,为了确保以高概率恢复所有的输

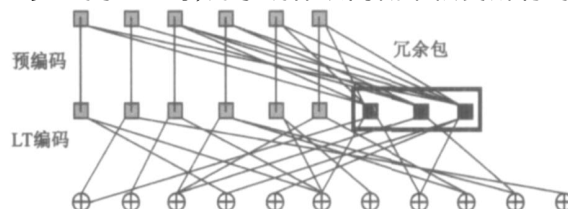


图3 Raptor码:原来的数据包附加了一些冗余包,利用一个适当的LT码可从预编码包来产生编码包

入结点(符号),其译码图中需要 $k \ln k$ 多条边. Raptor 码的基本思想是:通过放松这一条件而要求只要能够恢复部分输入符号即可(如图 3 所示). 具体地说, Raptor 码首先预编码原来的数据包,使得即使编码后的数据包仅有一部分可恢复,也可以利用这部分被恢复的编码包来恢复原来的所有数据包. 这样将一个经过预编码的码与一个适当选取的 LT 码级联即得具有线性时间译码复杂度的 Raptor 码.

4 数字喷泉码的网络应用

数字喷泉码在网络通信传输中具有广泛的应用. 目前,数字喷泉码在可靠多播传输^[8]、多源下载^[28]、数据存储^[29~32]和无线协作传输^[36]等方面的研究受到了普遍关注.

4.1 多播传输

数字喷泉码可用于诸如新软件产品发布等可靠多播环境中^[8]. 假设有众多用户要求在不同的时间下载同样的文件;组成文件的数据包通常分布在面向接收者的一个或多个多播树中,而接收者仅仅需要从与其处在同一分支的多播树结点中下载相关的数据包. 这就避免了需要向网络中每个用户发送所有数据包而产生的开销.

将数字喷泉码的编码方法应用于多址传输中具有如下优点:(1) 由于每个用户的数据损失率可能不同,如果众多用户同时要求从相同的数据源下载相同的软件,就会导致网络阻塞. 而利用数字喷泉码的编码方法就能避免重传和反馈;(2) 数字喷泉码的使用便于处理众多用户进行数据传输中遇到的问题. 例如网络传输中两个接收者除了最后一跳外具有相同的路径,但具有不同的下载速率,可将数据包在路径中以较高的速率发送,这样可保证较快的接收者有高的下载速率. 这时较慢的接收者会因为低的下载速率而丢弃一些数据包,由于接收到的每个数据包都可用于数字喷泉码的译码,所以丢掉的数据包并不会影响较慢接收者的接收.

4.2 多源下载

利用数字喷泉码技术可以简化一个接收者的多源并行下载^[28]. 每个数据源可利用数字喷泉码独立产生无限长的编码包数据流,接收者从多个数据源下载时就不会接收到相同的数据包. 当接收者接收到足够的数据包后就可以关闭所有的连接,而不需要关心这些数据来自何方与每个数据源具有的数据损失率和发送速率.

4.3 数据存储

在 peer-to-peer 和大型分布式文件系统中常常利用备份的方法来提高文件访问的性能和容错性^[29,30]. 对

于一个广域网中由分布式存储结点组成的文件系统,假如多个客户端要求访问这一文件,文件系统使用的传统技术是将这一文件分为大小相同的 k 个数据包,并将这些数据包的备份分布式存储在系统中. 然而,这种方法的代价较高. 首先,由于每个数据包必须复制 m 份才能防止 $m-1$ 个服务器中数据包的损坏. 其次,客户端在访问文件时寻找该文件的数据包的最近备份比较困难,而且一个特定数据包的损坏或对这一数据包的较慢访问会导致整个文件访问性能的降低.

利用纠删编码技术就可以改善文件复制系统的容错和下载性能. 借助 Tornado 码,将文件的 k 个数据包编码为 $n(n > k)$ 个编码数据包并存储在多个服务器上,客户端只要下载其中任何 k 个编码数据包就可恢复原来的文件^[8].

目前有不少学者将纠删码应用于包括网络存储在内的分布式系统和 RAID 磁盘阵列中^[29~31].

4.4 无线协作传输

2007 年, Molisch 首次提出了数字喷泉码在协作多中继无线网络中的应用方案,同时指出基于数字喷泉码的协作传输方法可以看作一种逼近多中继信道信息论容量限的实用方案^[32]. 与传统的编码技术相比,数字喷泉码的一个最大优点是利用该种编码技术可将源信息包编码为一无限长编码包流. 对于协作多中继无线网络,在基于数字喷泉码的异步传输协议中,每个中继节点一旦对接收到的编码包完成译码,该中继节点便开始向其目的节点传输通过译码所得的源数据包. 由于无线信道的广播特性和数字喷泉码无限长编码包流特点,这种传输方式不仅可对目标节点提供有用的信息,而且也有助于未完成译码的中继节点进行其译码.

另外,纠删编码技术还可应用到广播传输^[10,33,34]、多媒体流^[35]、无线网络^[33]等网络通信系统中.

5 存在的问题和发展展望

尽管数字喷泉码的研究已取得一些可喜成果,但是目前所进行的研究大部分停留在理论研究和工程应用仿真方面,数字喷泉码的理论研究和实际实现有待人们进一步的深入研究.

在 Tornado 码、LT 码和 Raptor 码的构造中度分布的设计是非常关键的问题^[7,12,13,22,24,25]. 数字喷泉码的研究主要存在以下问题:

(1) 结合现有构造 Tornado 码中二部图的 PEG 算法^[36,37],研究构造具有较大围长(girth)和最优度分布的二部图的实用算法;

(2) 借助概率论工具从理论上深刻分析 LT 码中选用 Robust Soliton 度分布的本质原因;

(3) 当数据包数目 k 较大(约几千)时,选用 Robust

Soliton 度分布的 LT 码具有良好的性能;而设计一种适合于数据包数目 k 较小时 LT 码的最优度分布仍是一个公开问题;

(4) 分析 LT 码和 Raptor 码译码错误概率,并给出计算 LT 码和 Raptor 码译码错误概率的实用算法;

(5) 研究和分析有限长 LT 码和 Raptor 码的构造;

(6) 探索协作多中继无线网络中如何利用 LT 码和 Raptor 码等数字喷泉码技术设计可靠的异步传输协议,进而加快所有中继节点的译码和传输速度,这一研究有助于解决协作多中继无线网络的可靠传输问题.

近几年,关于数字喷泉码的研究已经取得了不少令人振奋的结果,数字喷泉码被认为是适用于可靠多播传输、多源下载、数据存储和无线协作传输等应用的最有前途的编码技术.在国际上掀起了一场数字喷泉码的热潮,人们对数字喷泉码的应用前景充满信心.我们相信随着人们对数字喷泉码研究的重视,一定会推动数字喷泉码的理论研究,同时也为数字喷泉码的具体工程应用提供新的思路.

参考文献:

- [1] P Elias. Coding for two noisy channels[A]. Proc. Third London Symp. Information Theory[C]. London, U. K. :Buttersworth's Scientific Publications, 1955. 61 - 76.
- [2] J Bl mer, M Mitzenmacher, A Shokrollahi. An xor-based erasure resilient coding scheme[DB/OL]. ftp://ftp.icsi.berkeley.edu/pub/techreports/1995/tr-95-048.pdf, ICSI Technical Report, No. TR-95048, 1995.
- [3] L Rizzo. Effective erasure codes for reliable computer communication protocols[J]. ACM Computer Communication Review, 1997, 27(2) : 24 - 36.
- [4] L Rizzo. On the feasibility of software FEC[DB/OL]. DEIT Technical Report LR-970131, http://www.iet.unipi.it/~luigi/softfec.ps, 1997.
- [5] F J MacWilliams, N J A Sloane. The Theory of Error-Correcting Codes[M]. North Holland:Amsterdam, 1977.
- [6] S Reed, G Solomon. Polynomial codes over certain finite fields[J]. Journal of the Society for Industrial and Applied Mathematics, 1960, 8:300 - 304.
- [7] M Luby, M Mitzenmacher, M Shokrollahi, Daniel Spielman. Practical loss-resilient codes[A]. Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing[C]. El Paso: Texas, USA, 1997. 150 - 159.
- [8] J W Byers, M Luby, M Mitzenmacher. A digital fountain approach to reliable distribution of bulk data[A]. Proceedings of the ACM SIGCOMM '98 conference on Applications, technologies, architectures, and protocols for computer communication[C]. Canada: Vancouver 1998, 28(4) : 56 - 67.
- [9] M Luby, M Mitzenmacher, M Shokrollahi. Analysis of random processes via and/or tree evaluation[A]. Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms[C]. California: Francisco, 1998. 364 - 373.
- [10] D J C MacKay, Fountain Code[J]. IEE Proceedings Communications, 2005, 152(6) : 1062 - 1068.
- [11] M Mitzenmacher. Digital fountain: a survey and look forward[A]. Proceedings of the 2004 IEEE Information Theory Workshop[C]. San Antonio: IEEE Press, 2004. 271 - 276.
- [12] M Luby, M Mitzenmacher, M Shokrollahi. Efficient erasure correcting codes[J]. IEEE Transactions on Information Theory, 2001, 47(2) : 569 - 584.
- [13] Mu Jian-jun He Yu-cheng, et al. Design method of sequences of degree distribution for low-density erasure codes[J]. Chinese Journal of Electronics, 2002, 11(3) : 372 - 376.
- [14] A Shokrollahi. Capacity-achieving sequences[A]. Codes, Systems and Graphical Models (IMA Volumes in Mathematics and its applications) [C]. B. Marcus and J. Rosenthal, Berlin, Germany: Springer-Verlag, 2000, 123: 153 - 166.
- [15] A Shokrollahi. New sequences of linear time erasure codes approaching the channel capacity[A]. Proc. 13th Conf. Applied Algebra, Error Correcting codes, and Cryptography (Lecture Notes in Computer Science) [C]. Berlin, Germany: Springer Verlag, 1999. 65 - 76.
- [16] T Richardson, M A Shokrollahi, R L Urbanke. Design of capacity approaching irregular low-density parity-check codes[J]. IEEE Transactions on Information Theory, 2001, 47(2) : 619 - 637.
- [17] P Oswald, M A Shokrollahi. Capacity-achieving sequences for the erasure channel[J]. IEEE Transactions on Information Theory, 2002, 48(12) : 3017 ~ 3028.
- [18] 慕建君, 王鹏, 王新梅. 基于改进型右边正则度分布序列的低密度纠错码[J]. 计算机学报, 2003, 26(12) : 1734 - 1738.
MU Jian-Jun, WANG Peng, WANG Xin-Mei. Low-Density Erasure Codes Based on Improved Right Regular Sequences of Degree Distribution[J]. Chinese Journal of Computers, 2003, 26(12) : 1734 - 1738. (in Chinese)
- [19] Di Changyan, D Proietti et al. Finite length analysis of low-density parity-check codes on the binary erasure channel[J]. IEEE Transactions on Information Theory, 2002, 48(6) : 1570 - 1579.
- [20] M Luby. Information additive code generator and decoder for communication systems[P]. U. S. Patent 6 307 487, Oct. 23, 2001.
- [21] M Luby. Information additive code generator and decoder for communication systems[P]. U. S. Patent 6 373 406, Apr. 16, 2002.
- [22] M Luby. LT Codes[A]. In: Proceedings of the 43rd Annu.

- IEEE Symp. Foundations of Computer Science (FOCS) [C]. Canada:Vancouver,Nov.2002.71-280.
- [23] A Shokrollahi. Raptor codes[J]. IEEE Transactions on Information Theory,2006,52(6):2551-2567.
- [24] E Hytiä, T Tirronen, J Virtamo. Optimizing the degree distribution of LT codes with an importance sampling approach [A]. Proceedings of the 6th International Workshop on Rare Event Simulation[C]. Germany:Bamberg,2006.
- [25] E Hytiä, T Tirronen, J Virtamo. Optimal degree distribution for LT codes with small message length[A]. Proceedings of IEEE Infocom mini-symposium [C]. Alaska, USA: Anchorage,2007.
- [26] C Harrelson, Ip Lawrence, Wei Wang. Limited randomness LT codes[A]. Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing [C]. USA:Urbana Champaign,2003.
- [27] E Maneva, A Shokrollahi. New model for rigorous analysis of LT-codes[A]. Proceedings of the IEEE International Symposium on Information Theory [C]. (ISIT06), USA: Seattle, 2006.2677-2679.
- [28] P Maymounkov, D Mazoeres. Rateless codes and big downloads[A]. Proceedings of the 2nd Int. Workshop Peer-to-Peer Systems (IPTPS '03) [C]. Berlin: Springer-Verlag, Berkeley, CA, USA,2003.1-6.
- [29] J S Plank, M G. Thomason. A practical analysis of low-density parity-check erasure codes for wide-area storage applications [A]. Proceedings of the 2004 International Conference on Dependable Systems and Networks[C]. Italy:Florence,2004.101-110.
- [30] M K Aguilera, R Janakiraman, L Xu. Using erasure codes efficiently for storage in a distributed system[A]. Proceedings of the 2005 international Conf on Dependable Systems and Networks[C]. Pennsylvania, USA:IEEE Press,2005.336-345.
- [31] J L Hafner, V Deenadhayan, K Rao, J A Tomlin. Matrix Methods for Lost Data Reconstruction in Erasure Codes[A]. Proceedings of the Fourth USENIX Conference on File and Storage Technologies [C]. USA:San Francisco CA,2005.183-196.
- [32] A F Molisch, N B Mehta, J S Yedidia, J Zhang. Performance of fountain codes in collaborative relay networks[J]. IEEE-Transactions on Wireless Communications,2007,6(11):4108-4119.
- [33] M Luby, M Watson, T Gasiba, T Stockhammer, W Xu. Raptor Codes for Reliable Download Delivery in Wireless Broadcast Systems[A]. Proceedings of Consumer Communications and Networking Conference (CCNC) [C]. NV, USA:Las Vegas, 2006.1:192-197.
- [34] T Gasiba T Gasiba et al. Reliable and Efficient Download Delivery with Raptor Codes[A]. Proceedings of the 4th International Symposium on Turbo Codes & Related Topics[C]. Germany:Munich,2006.
- [35] D L Mahant, M K Eager, Vernon, D Sundaram Stukel. Scalable on-demand media streaming with packet loss recovery [J]. IEEE/ACM Transactions on Networking,2003,11(2):195-209.
- [36] Hu Xiao-Yu, E Eleftheriou, D M Arnold. Regular and irregular progressive edge-growth Tanner graphs[J]. IEEE Transactions on Information Theory,2005,51(1):386-398.
- [37] Hu Xiao-Yu, E Eleftheriou, D M Arnold. Progressive edge-growth tanner graphs [A]. Proceedings of IEEE Global Telecommunications Conf. (GLOBECOM '2001) [C]. Texas, USA:San Antonio,2001,25-29.

作者简介:



慕建君 男,1965年3月生于陕西省吴堡,2002年获西安电子科技大学通信与信息系统专业博士学位.现为西安电子科技大学教授.目前的研究方向为纠错码、LDPC码与信息安全.已发表相关论文20余篇。
E-mail:jjmu@xidian.edu.cn

焦晓鹏 男,1984年1月生于山西省芮城,2004年获西安电子科技大学计算机科学与技术专业学士学位,现为西安电子科技大学计算机科学与技术专业博士研究生.目前的研究方向为LDPC码与信息论. E-mail:jiaoxi1216@126.com

曹训志 男,1982年10月生于河南省信阳,现为西安电子科技大学计算机科学与技术专业硕士研究生.目前的研究方向为纠错码与LDPC码.