

## 喷泉码在物联网信道中的压缩方法研究

陈 旻<sup>1</sup>, 王开云<sup>2</sup>, 刘建平<sup>1</sup>, 李红梅<sup>1</sup>, 曹好顺<sup>1</sup>, 彭 宇<sup>1</sup>

(1. 云南警官学院 信息网络安全学院, 云南 昆明 650223; 2. 昆明学院 学报编辑部, 云南 昆明 650214)

**摘要:** 介绍在物联网传输信道中喷泉码的校验矩阵的生成办法, 以此平衡收发双方的通信代价. 然后使用两种熵编码压缩算法对校验矩阵进行压缩, 从而降低由于校验矩阵规模增大而带来的过高传输代价. 实验结果表明, 对校验矩阵的压缩, 确实能够降低传输代价, 同时针对不同信道, 使用熵编码压缩的效果较好.

**关键词:** 喷泉码; 校验矩阵; 熵编码; 物联网

**中图分类号:** TN911.2 **文献标识码:** A **文章编号:** 1674-5639(2016)06-0052-04

**DOI:** 10.14091/j.cnki.kmxyxb.2016.06.012

### Study the Compression for the Fountain Codes in the Channel of the Internet of Things

CHEN Min<sup>1</sup>, WANG Kaiyun<sup>2</sup>, LIU Jianping<sup>1</sup>, LI Hongmei<sup>1</sup>, CAO Haoshun<sup>1</sup>, PEN Yu<sup>1</sup>

(1. Information Security College, Yunnan Police College, Kunming, Yunnan, China 650223;

2. Editorial Department of Journal, Kunming University, Kunming, Yunnan, China 650214)

**Abstract:** The generation methods of the testing matrix of fountain codes in the channel of internet of things are introduced in order to balance the communication cost between the sender nodes and receiver nodes. Two compression algorithms of entropy coding were used to compress the testing matrix so as to decrease the cost when the scale of the testing matrix becomes larger. The result showed that the cost can be reduced consequently when the testing matrix is compressed, meanwhile the good effect is acquired when the entropy codes are compressed for the different channels.

**Key words:** fountain codes; testing matrix; entropy coding; internet of things

随着物联网技术的逐渐普及, 对相应的数据传输也提出了更高的要求. 一般情况下, 物联网数据传输为无线模式, 其传输载体相比有线网络而言, 更容易受到外界因素的干扰, 或者说, 噪声来源要更强. 这就要求物联网数据编码的纠错能力较强<sup>[1]</sup>. 此外, 物联网设备的计算能力相对较弱, 存储体容量较小, 这就意味着无法设置过高的缓存来缓解传输压力, 同时无法采用更加复杂的编码方法来提高其纠错能力. 因此, 针对物联网传输的编码方法具有以下两个特点: 1) 编码方法简单; 2) 容错性较好. 一般来说, 这两个特点是相互矛盾的, 但这基于一个条件, 即编码的传输不变性. 也就是说对于待传输的一个符号, 在编码端对其编码, 得到码字后, 在整个传输过程中, 这个码字都不会变. 一旦在传输中码字遭受

干扰而改变, 就意味着在接收端有可能解码为另一个符号, 也就是误码. 这就说明当一个符号被正确传输时, 其对应的码字在传输过程中是不可变的. 在此情况下, 要想得到的较高的纠错能力, 码字中校验位(冗余位)的数量必然得增加, 这也是分组码<sup>[2]</sup>的原理. 但是, 如果假设一个符号对应的码字并不是一开始就确定的, 相反却是在传输过程中由各个传输节点的某些运算结果的组合而得到, 那么这样的码字不仅能够降低受干扰的概率, 同时还能够将编码的运算复杂度分摊到传输过程中的各个节点上, 这无疑使得每个编码设备负担的运算量大大降低. 当然, 这样编码的方式带来的结果是使得接收端的运算复杂度增加. 但对于数字信号来说, 其中一些运算, 例如异或, 其运算的收敛性其实完全有可能使得接收

收稿日期: 2016-11-22

基金项目: 云南省自然科学基金资助项目(2016FB113).

作者简介: 陈旻(1982—), 男, 云南昆明人, 教授, 博士, IEEE 高级会员, 主要从事信息传输理论和物联网通信技术研究.

端的运算复杂度收敛在一个可以接受的范围内. 同时对物联网传输来说,接收端完全有可能配备一个运算性能较强的设备,以完成解码工作.

基于上述考虑,1997年杨伟豪教授提出网络编码<sup>[3]</sup>的概念. 随后网络编码开始被应用到各个信息传输领域. 在P2P流媒体传输方面,基于自适应编码窗的网络编码方式用于缓解播放延时问题<sup>[4]</sup>. 然而网络编码发展的理论意义远大于应用,因为它的结构很难在实际应用中加以实现. 前人<sup>[5]</sup>经过研究,提出采用喷泉码来改进信道传输过程中发送端与接收端的误码问题. 目前,喷泉码已经得到广泛的应用,然而其存在的问题在于发送端会一直广播其数据,这样会占用信道,形成冗余. 一种直观的想法是设置校验矩阵来平衡收发双方,以降低冗余的存在,但是校验矩阵会增加传输负担,有可能无法抵消由于发送端减少发送次数而带来的收益. 不过校验矩阵完全可以看作一种二进制信源,因此可采用熵编码技术对其进行压缩,从而保证校验矩阵发挥作用的同时,又不至于引入过高代价.

本文中,首先给出校验矩阵的生成办法,然后使用熵编码对其进行压缩.

## 1 网络编码原理

网络编码的一种基本网络结构形态是蝶形网络编码. 编码过程中,发送端S要向其他节点发送信息,其经过的路由节点对S端发出的信息进行运算后,再发送到路由中的下一级节点,直到到达接收端节点,其过程如图1所示.

通信要求:发送端S将向两个接收端Y和Z发送信息 $b_i$ 和 $b_j$ . 即Y端和Z端均需要收到这两个信息.

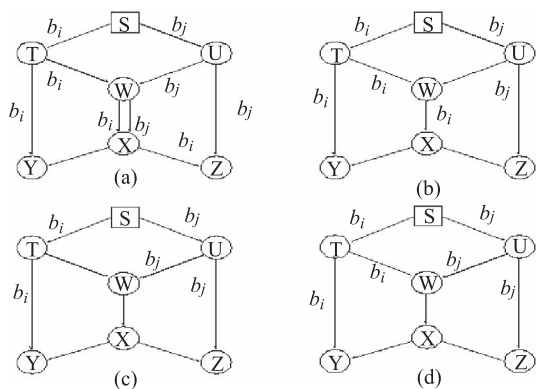


图1 蝶形网络编码示意图

图1(a)是传统网络发送信息的方式,在信道W-X上,有可能同时出现待发送的两个信息 $b_i$ 和 $b_j$ ,此时一般的路由算法就会发挥作用. 但是对于传输的数据量来说,并不会因此而改变. 图1(b)和图1(c)描述的是单路径数据传输时的情况,作为接收端的Y和Z一次实际上只能收到一个属于自己端的信息,例如 $b_i$ 到Z端, $b_j$ 到Y端. 然而采用网络编码后,如图1(d)所示,由于节点W对传输的数据 $b_i$ 和 $b_j$ 做了异或处理,经由X端广播到Y端和Z端时,在这两个接收端能够收到信息,这就意味着使用两个接收端本来就已经收到的信息,完全可以将 $b_i$ 和 $b_j$ 进行解码得到. 这样一来,无疑使得网络的负载降低. 如果将信息 $b_i$ 和 $b_j$ 的大小(bit)考虑进去,完全可以计算出网络中负载的能力量化值. 前人研究已经对此进行过讨论,证明采用蝶形网络编码确实能够以更小的传输代价实现通信过程.

然而蝶形网络仅是理论上的模型,在真实通信中不可能构建这样的通信信道. 一种替代或者改进的方法是使用喷泉码.

## 2 喷泉码原理

喷泉码保留了网络编码思想中广播的特点,但不同之处在于其发送端随机编码,然后将数据包分解为K个分组(编码分组),每个分组的长度不一,也不确定. 在编码过程中,发送端并不知道接收端是否已经收到信息,因此,其一直持续发送信息. 但是对于接收端而言,只需要收到略大于K个分组,就可以以极低的误码概率恢复出所有数据包. 喷泉码原理(采用预编码方式)如图2所示.

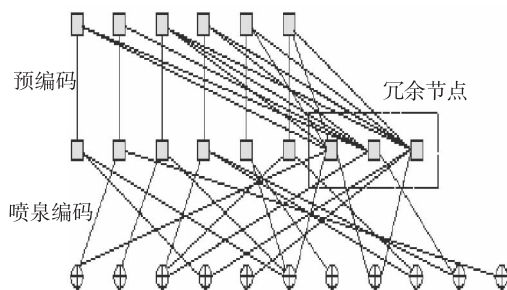


图2 喷泉码原理

在使用喷泉码进行弱带宽(例如无线信道)信息传输时,为了控制发送端发出信息包的数量,可以考虑使用一个发送矩阵来描述(或记录)接收端是否已经收到. 换言之,虽然喷泉码发送和接收端不通信,但是依然可以控制其同步通信. 对此可以用以下

数学方法对其进行描述.

设发送端将待发送的数据包分为  $K$  个分组, 为了方便说明, 先限定一个分组的数据为 1 bit (真实情况要远比此数值大). 根据喷泉码传输过程接收端需要接收到  $K(1+\varepsilon)$  个分组数据, 此处即为  $K(1+\varepsilon)$  个 bit. 则此时码率  $\eta$  可由下式计算得到:

$$\eta = \frac{K}{K(1+\varepsilon)} = \frac{1}{1+\varepsilon}. \quad (1)$$

显然, 信道的码率跟分组个数无关, 只与信道本身的误码参数  $\varepsilon$  有关. 事实上,  $\varepsilon$  受信道质量影响. 在无差错信道中,  $\varepsilon = 0$ . 但是对于物联网环境,  $\varepsilon$  的取值将很大. 我们首先结合 (1) 式讨论无差错信道中的检查矩阵构建方法.

对无差错信道, 喷泉码的  $K$  个分组将无差错的发送到接收端. 设  $X$  代表发送端,  $Y$  代表接收端. 对发送端  $X$ , 数值 0 代表数据还未发送, 1 代表数据已发送. 对接收端  $Y$ , 数值 0 代表数据未收到, 1 代表数据已收到. 则期待的正值表如 (2) 式所示:

$$\begin{bmatrix} X & Y \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad (2)$$

在 (2) 式中, 第二行数值看似不符合逻辑, 即发送端未发送数据, 但接收端却收到数据, 事实上, 这意味着误码. 虽然在无差错信道中, 这种情况不会出现, 但对于一般信道而言, 这是有可能出现的情况, 因此不可忽略. 对信道传输来说, 真值表 (2) 前三行代表的 3 种情况均是需重发信息分组的. 令  $Z$  表示是否发送数据, 则很明显:

$$Z = X \cap Y, \quad (3)$$

对有  $K$  个分组数据的过程, 变量  $Z$  扩展成二进制矩阵  $Z$  (只取两个值). 其计算由同样扩展成向量的  $X$  和  $Y$  相乘得到, 即:

$$Z = X \times Y, \quad (4)$$

二值矩阵  $Z$  的规模在无差错信道时为  $K \times K$ , 但对于一般信道, 需要引入误码参数  $\varepsilon$ . 因为每次重发, 同样需要在矩阵  $Z$  中增加校验位. 这就意味着, 此时的矩阵  $Z$  的尺寸应当为:  $K(1+\varepsilon) \times K(1+\varepsilon)$ . 值得注意的是, 这只是理想尺寸, 在实际应用中, 这个尺寸其实并不能以此简单的计算得到, 相反, 需要重新发送的次数往往大于  $K(1+\varepsilon) \times K(1+\varepsilon)$ . 一种直观的想法是将此时的矩阵  $Z$  看作是无差错信道

时的矩阵  $\bar{Z}$  的增广矩阵. 这种想法的确符合信息传输中的实际情况, 同时也能够更好地反映出一个分组是否该重新传递. 但是这就无形中增加了传输代价. 换言之, 矩阵  $Z$  在通信过程中是需要传输的. 在给定的分组数  $K$  的情况下, 传输该矩阵的代价其实是可以接受的. 但由于信道质量, 参数  $\varepsilon$  有可能使得校验矩阵的规模变得很大, 使得传输代价超出可接受的范围. 因此考虑到传输矩阵的构成, 可以考虑使用压缩算法对其进行压缩.

### 3 校验矩阵的压缩

对二进制信源的压缩, 使用熵编码的方法可以获得较好的压缩效果. 本文中, 我们使用两种方法对其进行压缩, 一是文献 [6] 中使用的二维映射方法, 二是使用加权 Context 建模<sup>[7]</sup> 的方法进行压缩. 值得说明的是, 校验矩阵本身就是二维信号, 因此无需使用希尔伯特映射来实现一维到二维的映射. 相反, 对于加权的方法, 对矩阵应采用逐行编码的方式进行.

在使用熵编码压缩方法过程中, 需要考虑的关键问题在于条件位的选择. 对校验矩阵而言, 代表的是一种分组数据的发送接收情况, 相邻数据块间的相关性本身就很强. 因此, 对二维方法来说, 只需要选取 1 位相邻位即可. 对于加权方式, 选取邻近的两个数据位作为条件即可. 两种方法的条件选取如下:

二维方法: 设当前信源的位置为  $(i, j)$ , 则其条件位选择为  $(i, j-1)$  位置上的数据.

加权方法: 设当前信源位置为一维化后的第  $i$  个, 则其条件位选择为第  $(i-1)$ 、第  $(i-2)$  个数据即可.

对加权方法, 采用基于描述长度的 Context 优化加权方法进行.

对以上两种方法, 采用 One-step 的方法对 Context 模型进行训练, 然后用训练好的条件概率分布驱动算术编码器进行编码.

### 4 实验

在本节中, 将对提出的压缩算法进行验证. 实验设计如下: 构建模拟数据, 数据总共分组 32 组, 即  $K = 32$ . 每个分组数据的大小从 8 bit 到 32 bit 不等. 模拟信道, 信道误码参数  $\varepsilon$  取值分别为 0.8, 1.2, 2.1, 其分别代表不同质量的信道模型.

在实验 1 中,首先对比 2 种方法的压缩效率.测试采用的误码参数  $\varepsilon = 1.2$ ,压缩结果如表 1 所示.在表 1 中的数据代表 bit/symbol. 2 种方法分别简称为二维和加权.

表 1 2 种方法的平均压缩比 bit/symbol

信道误码参数	二维	加权
1.2	0.312 3	0.332 8

从表 1 中可以看出两方面内容:一是对校验矩阵进行压缩后,基本上可以达到近 2/3 的压缩比.这就意味着能够将此部分的传输代价降低很多.从另一方面看,就算校验矩阵的规模增加,通过压缩,也有可能将其增加的代价控制在可以接受的范围内.二是在给定的误码参数前提下,并不足以看出究竟哪种算法更优,但不论哪种方法,其结果均能够达到较理想的压缩效果.

在实验 2 中,我们将 2 种方法应用到不同的误码参数中,对比在不同信道中 2 种方法的好坏.实验结果如表 2 所示.

表 2 不同误码参数下的平均压缩比 bit/symbol

误码参数	二维	加权
0.8	0.278 3	0.281 1
1.2	0.312 3	0.332 8
2.1	0.364 4	0.352 3

综合 2 个实验不难发现,在不同信道中,对校验矩阵的压缩结果,2 种方法各有优势.但在误码率较高的信道中,基于优化加权的 Context 编码方法要更优于基于二维的压缩方法.这主要是因为信道质量低,重发的次数就多,而有可能对一个分组,连续发送多次,接收端都未必能收到,这就使得参与加权的

条件概率分布因为新数据的统计而变得更加奇异,从而降低熵值.但这一点实际上并不能真实的反应通信中的实际情况.但目前我们尚未找到更好的方法去解决,有待下一步深入研究.

5 结论

在使用喷泉码进行物联网数据传输过程中,校验矩阵的引入确实能够控制发送端发送数据的频率,从而降低传输代价.但校验矩阵本身的传输也需要占用带宽.采用熵编码压缩后,确实能够获得较好的压缩效果,从而提升信道传输的利用率,降低传输代价.实验证明,2 种方法均能够获得较理想的压缩结果,达到设计目标.

[参考文献]

[1]王宜怀,朱仕浪,郭云. 嵌入式技术基础与实践:ARM Cortex-M0 + Kinetis L 系列微控制器[M]. 3 版. 北京:清华大学出版社,2013.

[2]YEUNG R W. Information theory and network coding[M]. New York:Springer,2008.

[3]YEUNG R W. Network coding theory:an introduction[J]. Frontiers of Electrical and Electronic Engineering in China, 2010,5(3):363-390.

[4]CHI H C,ZHANG Q. Deadline-aware network coding for video on demand service over P2P network[J]. Journal of Zhejiang University Science A,2006,7(5):755-763.

[5]祝开艳,王洪玉,孙文珠,等. 一种分布式喷泉码在协作通信中的应用[J]. 电子学报,2014,42(7):1249-1255.

[6]陈旻,王开云,吴建国,等. 基于希尔伯特分形的基因组序列压缩算法[J]. 昆明学院学报,2014,36(6):42-46.

[7]陈旻,王开云,贾学明,等. 基于加权 Context 建模的 DNA 序列压缩算法[J]. 昆明学院学报,2014,36(3):81-84.

