doi :10.3969/j.issn.1671-1122.2013.05.002

LT Code 在基于 DVB-S2 文件多播系统中的安全应用

张亚航[1]，程博文[1]，文伟平[1]，闫彦[2]

（1. 北京大学 软件与微电子学院信息安全系，北京 100084 ；2. 华中科技大学 软件学院，湖北武汉 430022）

摘 要 ：文章对欧洲卫星数据广播分发系统标准 DVB-S2 进行介绍，提出在气象卫星数据广播分发系统中，在应用层使用 LT 码进行前向纠错，以实现在恶劣的网络环境以及没有回传信道的情况下进行高速大文件分发、可靠的数据广播和完美的视频流服务。文章介绍 LT 码的核心理论，并对 LT 码的多个性能进行测试，得出了适用于该系统的一系列参数。

关键词 ：DVB-S2; LT 码 ；前向纠错

中图分类号 ：TP393.08 文献标识码 ：A 文章编号 ：1671-1122（2013）05-0005-03

The Safe Application of LT Code in Broadcasts in the System based on the DVB-S2 Document

ZHANG Ya-hang[1], CHENG Bo-wen[1], WEN Wei-ping[1], YAN Yan[2]

(1. *Department of Information Security, SSM, Peking University, Beijing* 100084*, China;*

2. *Department of Software, Huazhong University of Science and Technology, Wuhan Hubei* 430022*, China*)

Abstract: This article introduces the standard of Europe Satellite data broadcasting distribution system DVB-S2, and proposes to use LT code to realize FEC in application layer in Weather Satellite Data Broadcasting System, in order to implement the high speed big size file distribution, credible data broadcasting and perfect video frequency service in bad net work and unidirectional links. This article introduces LT code kernel theory, and a set of parameters which are gotten from LT code multi-performance testing.

Key words: DVB-S2; LT Code; forward error correction

0 引言

卫星数据广播分发是一种利用卫星天然广播特性的应用方式，把同一个 IP 数据包通过单播、组播或广播的方式同时分发给一个或多个接收者。DVB-S2 无疑是当前最具影响力也是性能最先进的卫星数据广播技术体制标准[1]。DVB-S2 作为一种单向链路，需要一种可靠的前向纠错编码技术保证其数据传输过程中的完整性[2]。

在卫星通信单向链路中，大多数系统所采用的前向纠错方案都是在链路层，以硬件编码的方式实现。作为一个大文件分发系统，气象卫星数据广播系统要求在文件层次实现文件的可靠性传播。传统的纠错编码如 RS Code、LDPC Code 主要是以 bit 为单位进行纠错，会在软件层次占用大量资源，使得系统运行的速度变慢，且无法处理整个 UDP 数据包丢失的情况[2-5]。

本文通过对众多类型编码的研究，提出了以一种喷泉码（Fountain Code）——LT 编码在软件层面实现应用层前向纠错（FEC）的方案，从而让系统能够在满足发送速度要求和处理资源占用较小的前提下，达到对丢失 UDP 数据包进行恢复的目的。

1 DVB-S2 单向链路大文件分发系统模型

基于 DVB-S2 单向链路的大文件分发系统包括软件层和硬件层，主要分为发送服务器和接收服务器两个部分。其中发送服务器包括 CRC 校验、文件分块、LT Code 编码器、组播 UDP。大文件首先在 CRC 编码区取得 CRC 值，然后将文件分割成多个块，以块为单位进行 LT Code 编解码。接着将 LT 校验节点同节点相关信息组合成为 UDP 包，通过 DVB-S2 卫星链路发送。与之相对应，接收服务器包括 CRC 校验、文件恢复、数据块组装、LT Code 解码器和 UDP 接收。当接收端从底层的 DVB-S2 卫星链

收稿时间 ：2013-04-11

作者简介 ：张亚航（1985-），男，硕士研究生，主要研究方向 ：网络安全 ；程博文（1984-），男，硕士研究生，主要研究方向 ：软件过程管理 ；文伟平（1976-），男，副教授，博士，主要研究方向 ：网络攻击与防范、恶意代码研究、信息系统逆向工程和可信计算技术等 ；闫彦（1985-），女，硕士研究生，主要研究方向 ：人工智能。

路上接收到 UDP 数据包之后，对该数据包进行数据分析，提取 LT Code 校验节点信息，然后进行 LT 解码，恢复成为一个文件数据块。最终，上层文件组装模块将所有恢复成功的文件数据模块化，并对文件进行 CRC 校验。系统结构如图1所示。

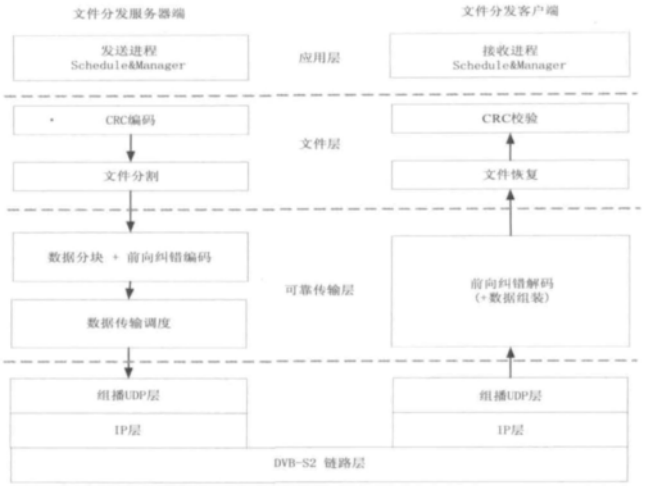

图1 基于可靠组播的文件分发系统解决方案流程图

2 LT Code

LT Code 是一种喷泉码（Fountain Code），于 1998 年首次由 Michael Luby 等人提出 [6]，并于 2002 年发表 [7, 8]。LT Code 是一种无比率码（Rateless Code），该码是由原始数据生成且译码符号位数可以无限长。它具有相对简单的可扩展的非规则图结构和相对简单的设计，每个编码信息块是独立生成的，编码信息块的数目没有限制。通常情况下，输入信息单元的长度可以为任意长 (从 1 bit 到通常的 L bits)。如果输入数据包含 k 个信息单元，则在编码过程中，每个编码信息单元都可以通过进行$O(\ln(k/\delta))$次异或运算产生，该过程与其他编码信息单元的产生相互独立。在解码过程中接收端收到的 k 个信息单元可以在概率为 1~δ 的情况下通过进行$O(k\ln(k/\delta))$次异或运算恢复，参与上述过程的编码信息单元数量为$k+O(k^{1/2}\ln(k/\delta))$。

2.1 原理简介

在介绍 LT Code 原理之前，先做一个有趣的试验。

假设将球随机地丢到K个桶里，K=1000 或 K=10000。那么，为了使每个桶中都至少有一个球，那么至少要丢多少个球呢？

根据概率论的相关定理，K 个桶中丢了 N 个球后，其中某一个桶为空的概率是 :

$$(1-\frac{1}{K})^N \approx e^{-N/k} \quad \cdots\cdots (1)$$

从而得出，为了使空桶的概率为 δ，即每个桶都至少有一个球的概率为 1-δ，至少要丢$N > K\ln(\frac{K}{\delta})$个球。

2.2 度数分布函数：鲁棒性孤波分布

假设有 K 个信号 (K 个桶) 需要传递，从丢球试验中，可以推得 ：至少需要$N > K\ln(\frac{K}{\delta})$个球，即 LT Code 编码度数的总和$N > K\ln(\frac{K}{\delta})$，才能使误差为 1~$\delta$。对此，Luby 提出了一种适用度数分布 $\rho(d)$ 设计方法 [2] 满足上述要求 ：

$$\rho(d)\begin{cases}\frac{1}{K} & d=1\\ \frac{1}{d(d-1)} & d=2,3,...,K\end{cases} \quad \cdots\cdots (2)$$

$$S = c\ln(\frac{K}{\sigma})\sqrt{K} \quad \cdots\cdots (3)$$

$$\tau(d)=\begin{cases}\frac{s}{K} & d<\frac{K}{S}\\ \frac{s}{K}\ln(\frac{S}{\sigma}) & d=\frac{K}{S}\\ 0 & d>\frac{K}{S}\end{cases} \quad \cdots\cdots (4)$$

$$\mu(d)=\frac{\rho(d)+\tau(d)}{Z} \quad \cdots\cdots (5)$$

其中，$Z=\sum_d \rho(d)+\tau(d)$，$\mu(d)$ 为修正后的 LT Code 的度数分布函数。

2.3 LT Code编码

假设输入信号为 $u=(u_1, \cdots, u_k)$，编码后输出为 $c=(c_1, c_2, \cdots)$，算法描述为 ：

1: 计算 LT Code 的分布函数 $\mu(d)$

2: repeat

3: 从 $\mu(d)$ 中随机产生某一节点的度数为 d

4: 从输入信号 u 中抽取 d 个数据块$m(i_1), m(i_2), ...m(i_d)$

5: 计算第 i 个输出信号$c_i = m(i_1)\otimes m(i_2)\otimes ...m(i_d)$，并发送给接收者

6: until（从 C 中恢复出输入信号 u）

2.4 LT Code解码

假设接收序列为$c' = (c'_1, c'_2, ...)$，每个接收信号的度值为$d = (d_1, d_2, ...)$。对于每次收到的新信号，都进行 LT Code 的解码。相关的 LT Code 的解码算法描述为 [3] ：

1 ：repeat

2 ：接收新节点c'_k，度为 d_k

3 ：if $d_k > 1$

4: for all m(j) in，已解码队列 R :d_k includes m(j) do

5: $d_k = d_k \otimes m(j)$

6: end for

7 ：if $d_k > 1$

$B \leftarrow c'_k$

8 ：else

goto 13

9: end if

10 ：else

11 ：goto 13

12 ：end if

13: while 在候选队列 B 中不存在 d=1 节点

14: B = received packet-known blocks.

15: end while

16: $m(j)$ = 从候选队列 B 中新发现的 d=1 的节点 j

17: for all C' in B : C' includes $m(j)$ do

$$c' = c' \otimes m(j)$$

18: end for

19: until 所有节点解码成功

3 LT Code 编码实现

如上所述，LT Code 编码对节点的操作只有异或操作，因此，理论上对输入节点的大小没有要求，一个节点既可以是 1 bit，也可以是 x bits（x>1）。对于源节点个数 n 的选择，我们对 LT Code 进行 1000 次的测试。试验结果如表 1 所示。

表1 基于鲁棒孤波分布的LT编码性能参数统计表

源节点个数 n	平均需要节点个数 T	标准差 Std(T)	冗余度 p (%)
100	134.6	26.9	34.6
1000	1179	42	17.9
10000	10904	91	9.04
20000	21406	86	7.03

通过实验数据我们可以看到，节点个数为 100 的时候解码所需冗余最大，节点个数为 20000 的时候冗余最小。事实上，在以鲁棒孤波分布（Robust Soliton Distribution）为随机分布的 LT Code 中，随着源节点个数 n 的增大，成功解码所需冗余度减小[8-10]。但是，LT Code 的编码效率是 O(kn)，解码效率是 O(knlog n)，其中 k 为源节点 bit 数，n 为源节点个数。

因此，当 k=6000×8 bit 时，在 CPU 为 Intel(R) Core(TM)2 Duo 2.33GHz，内存为 1G 的计算机上，对于不同 n 编解码所需要的时间如表 2 所示。

表2 不同节点个数n编解码所需时间统计表

源节点个数 n	编码时间 (s)	解码时间 (s)
100	0.06	0.04
1000	0.61	0.69
10000	6.42	7.33
20000	11.66	16.32

考虑到多任务并发的可能性，选取节点个数 n=10000。显然，不同的冗余对于解码成功率会产生影响。当 n=10000 时，解码成功所需节点数统计表如表 3 所示。

表3 对n=10000试验1000次所需总解码节点区间分布表

所需节点区间	解码成功案例个数	成功累计百分比 (%)
10000～11000	808	80.8
11000～12000	176	98.4
12000～12500	14	99.8
12500～13000	1	99.9
>13000	1	100

从以上表格统计数据可以看出，冗余数据越大，区间分布越稀。从信道冗余性和文件传输成功解码性能平衡考虑，选取冗余为 25% 的时候，成功解码率达到 99.8%。

4 结束语

LT Code 作为一种新型编码喷泉码，通过以上研究，我们看到它满足以下四个特性：

1）任何输出都是由随机个源节点生成；2）对任意接收到的 m 个节点，都有 p 个可能性恢复出原来 k 个节点，其中，m ≥ k，p = 1-1/kc，m/k-1 就是这种喷泉码的冗余；3）节点大小可以为任意长度；4）对于较低的冗余可以有较高的解码成功率。

这四个特性让 LT Code 特别适合基于无回传的链路中，在软件层实现。本文通过在实际链路中对 LT Code 进行的多个试验，通过对编码源节点个数 n、冗余 p 和编解码时间等参数进行设置，在 LT Code 解码成功率和编码冗余率方面进行了平衡。

但是，通过对试验程序的跟踪，我们也发现，LT Code 在解码的过程中，具有一定的概率出现极个别节点滞留无法解码成功的现象。正是这种现象，导致 LT Code 解码成功在某些小概率情况下需要大冗余数。而这种情况出现的原因正是因为上述的 LT 编码的第一个特点：对于任何输出都是由随机个源节点生成。当某个节点解码条件没有被满足时，由于编码仍然是随机产生的，而不能为该节点的成功解码进行编码。因此，我们下一步的工作是找到另外一种编码随机分布取代鲁棒孤波分布，或者将 LT Code 同另外一种编解码结合使用，消除这种解码不成功的情况。（责编 杨晨）

参考文献:

[1] ETSI EN 302 307(V I.1.1 Draft). Digital Video Broadcasting(DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, June,2004.

[2] 姚春光，郭秀英，张健，王建新 . 自主标准卫星数据广播分发系统关键技术研究 [J]. 电信科学，2007，(04)：22-25.

[3] 金晓成，吴耀军，陈敏 . ST Turbo TC 与 LT Code 相结合的性能研究 [J]. 计算机与数字工程，2008，(02)：125-128.

[4] 张有志，李子木，郝英川 . 基于卫星的可靠组播协议研究 [J]. 无线电通信技术，2008，(01)：6-8.

[5] 张有志 . 卫星组播通信中的分组级 FEC 技术研究 . 南京：解放军理工大学，2004.

[6] J.W.Byers, M.Luby, M.Mitzenmacher, A.Rege. .A Digital Fountain Approach to Reliable Distribution of Bulk Data., SIGCOMM. 1998, pp. 56-67.

[7] M.Luby and M.Mitzenmacher,A. Shokrollahi,D.Spielman,.Ef_cient Erasure Correcting Codes.,IEEE Transactions on Information Theory, vol. 47, issue 2, pp. 569-584, February 2001.

[8] M.Luby, .LT Codes., 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002.

[9] David J.C. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge University Press, 2004.

[10] Patrick Farrell and Jorge Castinera Moreira, Essentials of Error-Control Coding, John Wiley and Sons, 2006.