

# 录波数据文件的有损压缩传送及文件传输协议的选择研究

赵渊

(重庆大学电气工程学院电力系, 重庆 400044)

**摘要:**针对采用公用电话网进行电力系统故障信息远程传输的录波测距装置,对其信息传输过程的两个方面即信源编码和信道编码分别进行了分析,提出了一种应用于信源编码中的能定量控制误差的有损数据压缩算法,并对信道编码中的文件传输协议(FTP)在传输效率上进行了比较,通过对故障录波数据文件的有效压缩并选择较好的文件传输协议,能实现录波数据文件的远程快速传送。

**关键词:**信源编码; 信道编码; 数据压缩; 文件传输协议

**中图分类号:** TM73 **文献标识码:** A **文章编号:** 1003-4897(2002)10-0033-04

## 1 引言

电力系统故障录波测距装置是记录故障后系统参数变化的重要设备,是调度人员进行故障分析的重要信息源,实现故障录波数据的快速远传具有重要意义:(1)可使调度中心及时、全面、准确地掌握故障情况,从而快速准确地处理系统故障;(2)便于录波测距装置运行工况的集中监测和装置的集中管理;(3)可实现双端故障测距,从而提高测距定位精度,大大减少寻线工作量及故障恢复时间<sup>[1]</sup>。由于尚未普遍实现录波测距装置联网,同时由于调度中心往往在系统发生故障后才需要从录波测距装置索取故障信息,通信双方对信道的占用是断续性的,因此采用覆盖社会各个角落的公用电话网来传送故障数据成为经济合理的方式。但电话网传输带宽有限,同时由于录波数据文件很大,传送时间将很长,往往需要十多分钟<sup>[3]</sup>。解决这一问题的有效途径是在信源编码中采用高效压缩技术,并在信道编码中采用合理的文件传输协议,通过以上处理,传输一个200 k的录波文件,仅仅需要10 s时间。

## 2 故障录波数据文件传输系统模型

图1中的信源编码和信源解码,统称为信源编码;而信道编码和信道解码统称为信道编码。信源编码:主要解决有效性问题,通过对录波数据文件的压缩、扰乱、加密等一系列处理,力求用最少的数码传递最大的信息量,使信号更适应传输。因此,从信息论的角度看,信源编码的一个最主要目的,就是要解决数据的压缩问题,它构成了数据压缩的理论基础。信道编码:主要解决可靠性问题。即将通过信源编码处理后的信号按通信规约进行组装发送,以

尽量使处理过的信号在传输过程中实现无差错地传输,即使出错,在信道解码时也能自动检错和尽量纠错,由于加入了控制信息,这将不可避免地增加要传递的信息的冗余度。由图1可见,录波文件的整个传送过程涉及信源编码和信道编码两个过程,因此这两个过程的编码效率将影响录波文件的整个传送时间。

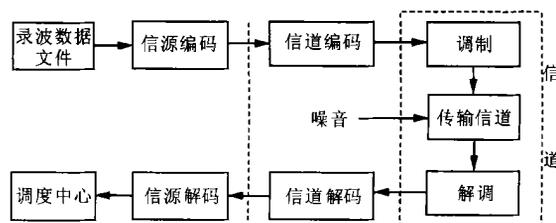


图1 传输系统模型

Fig. 1 The model of transmission system

## 3 故障录波数据文件的信源编码——数据压缩

作者曾经对故障录波数据文件用通用压缩算法进行压缩对比测试,所应用的算法有 Huffman 编码、自适应算术编码、LZ77 编码、LZ78 编码,在压缩效果和压缩时间上都比较好的是:2 阶自适应算术编码。它的压缩率基本上达到了 50%,但压缩效果仍然不能令人满意,这主要是因为录波数据文件的数据记录格式是按照 IEEE 标准 COMTRADE 格式来记录的,文件中记录的不仅有大量的 AD 采样数据,同时也记录了其它一些信息。而算术编码能有效压缩数据的一个条件就在于精确统计各信源符号的概率,因此利用算术编码对整个文件进行压缩时,由于其它信息的影响,导致采样值的统计概率将小于其实际的概率分布,因此其编码的位数将增大,而文件中

主要存储的是采样值,故文件的压缩效果会变差。因此为了获得较好的压缩效果,必须在压缩前对故障录波文件的记录格式进行逻辑分析,然后在逻辑分析的结果上改进压缩算法。

3.1 COMTRADE 二进制记录格式及故障动态记录准则

为实现各种数字式装置(故障录波装置和数字保护及微机监测装置)中的数据交换(共享)问题,IEEE于1991年提出了电力系统暂态数据通用格式标准<sup>[4]</sup>。按此标准,二进制格式的记录文件格式如图2所示。该文件按采样点的方式进行记录,每一采样点有(n+2)个数据,其中n=(模拟量采样路数+状态量采样路数/16)。采样点数目的大小取决于记录时间的长度,因而影响整个文件的大小,而每一采样点中采样数据数目则取决于录波系统,它同时也影响整个文件的大小。每一采样点的第1个数据是采样点序列号,从1开始向后递增;第2个数据是以微秒为单位的从记录开始算起的时间标记;第3个以后的采样数据是记录那些电压、电流和开关量的数据<sup>[4]</sup>。

$n \quad t \quad A_1 \quad A_2 \dots A_k \quad D_1 \quad D_2 \dots D_m$

n为采样序列号,用4字节表示;t为用微秒表示的采样时间,用4个字节表示;A<sub>1</sub>...A<sub>k</sub>为模拟采样数据,每一信道用2字节表示;D<sub>1</sub>...D<sub>m</sub>为状态值,每一个数据用2字节表示,可以代表16个状态量;用双字节表示的量,先存低字节,再存高字节。

图2 数据文件的记录格式

Fig. 2 The record format of the data file

电力系统动态记录过程的主要任务,是记录系统发生大扰动(如短路故障、系统震荡、频率崩溃、电压崩溃等)后的有关电参量的变化过程及继电保护与安全自动装置的动作行为。依据这一技术准则的数据记录特点则是分时段记录,以适应分析数据的要求,满足运行部门故障分析和系统分析的需要,并尽可能只记录和输出满足实际需要的数据<sup>[5]</sup>。为此,我国部颁220~500kV电力系统故障动态记录

技术准则确定了ABCDE分段模拟量采样方式,如图3所示<sup>[6]</sup>。这里,技术准则的数据记录有两个主要特点,一是分段记录,二是记录的数据可以保留不同的输出值,既有按某一采样率的采样数据,又有按一定间隔储存输出的有效值。

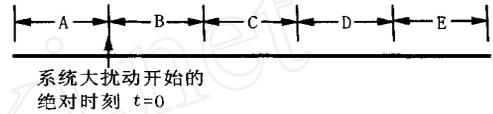


图3 分段模拟量采样方式

Fig. 3 The sampling mode based on different segment

3.2 文件压缩算法

通过分析COMTRADE标准和部颁技术准则,可以知道数据文件中存在一些明显的冗余信息:文件的每一采样点都有一个采样点序列号,而文件的接受方只需要采样点的总数目,即可生成各采样点的序列号。文件的每一采样点都有一个时间标记,用以表示此采样点的采样时间,由于A、B、C段的采样率是一样的,其相邻采样点的时标以等间隔(若每周波采样20点,则为1000μs)增加,对于D、E段,则分别以其数据的保留周期100000μs和1000000μs为间隔向后递增。这些大量的时间标记信息量并不大,因为数据接受方只需要数据文件中有几种采样率(或数据保留时间间隔)及其大小,以及该采样率下的最大采样点序列号,就可算出各采样点的采样时间。故障录波装置中往往接入了大量的开关量信号,按COMTRADE标准,不管状态量是否发生变位,每一采样点也要重复记录它们。而数据接受方仅需要各个发生变位的状态量,及该状态量下的最大采样点序列号,就可完全恢复出数据文件中的全部状态量数据<sup>[7]</sup>。

为获得较大的压缩率,应在压缩前对数据文件进行预处理,以消除这些冗余信息。通过预处理,数据文件明显地减少了,而文件信息量未变,其效果如表1所示。

表1 数据文件预处理效果

Tab. 1 The effect of the data file disposal before compression

原始文件大小(Byte)	322 400	269 200	291 280	286 000	287 280	256 000	256 000	256 000	256 000	256 000
预处理后临时文件(Byte)	257 963	215 395	233 067	228 843	229 867	204 839	204 839	204 839	204 839	204 839
用自适应2阶算术编码压缩(Byte)	83 434	71 187	77 875	77 027	75 974	73 049	71 644	71 057	73 274	73 375
压缩率(%)	75	74	74	74	74	72	73	73	72	72

用自适应2阶算术编码,对表1中的原始录波数据文件经过预处理后再进行压缩,压缩率可达到

70%左右。但是要想再提高压缩率,除了增加压缩阶数(即采用3阶以上的算术编码)使压缩率有微小

提高外,再难有较大改进。这是因为波形数据文件所含的信息量是一定的,其信息熵也是一定的,而熵代表了离散无记忆信源无失真编码的理论极限,故无论采用什么方法,只能使压缩编码的平均码长逼近或等于熵的大小,而不可能使平均码长小于熵的大小,要想将文件压缩的更小,只有在压缩编码时,允许损失一定的信息量,即降低熵的大小,这在图像和数字化语音的压缩中被普遍采用。另外,由“率失真”理论可知,当允许失真时,数据压缩的极限数码率随着失真度的增大而降低,因此失真越大,压缩效果越好。

综上所述,对于故障录波数据文件来说,要获得较大的压缩率,可以采用有损压缩。但由于故障录波的数据文件除用于直观地显示故障时的电压、电流波形外,还是故障测距、谐波分析等功能的数据来源,如果压缩引起过大失真,则计算出的结果将有较大误差。所以在压缩时,对文件中用于故障测距和谐波分析的原始故障周波数据不能改变其数值,同时必须做到将失真控制在较小范围的基础上力求获得较大的压缩率。

由统计模型可知,设有熵函数  $H_m(p_1, p_2, \dots, p_m)$ ,若其中有两个事件,不妨设为  $a_1$  和  $a_2$ ,若将它们合并为一个事件,且  $p_1 + p_2 > 0$ ,则新的信源熵为<sup>[2]</sup>

$$H_{m-1}(p_1 + p_2, p_3, \dots, p_m) = (p_1 + p_2) \log(p_1 + p_2) - \sum_{j=3}^m p_j \log p_j = H_m(p_1, \dots, p_m) - (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) < H_m(p_1, p_2, \dots, p_m)$$

按此原理,可将文件中采样值的范围(例如在 8191 ~ -8192),划分为由多个子区间的组合,如  $[-8192, 8191] = \sum_{j=1}^m [a_j, b_j]$ ,子区间  $[a_j, b_j]$  的区间距离  $L_j = b_j - a_j$ ,在编码采样值  $C$  时,首先判断其所在子区间  $[a_j, b_j]$ ,将其按数值  $(a_j + L_j/2)$  进行编码,换句话说,就是将此子区间中的  $(L_j + 1)$  个事件合并为了一个事件,因此波形文件的信息熵自然降低,但也将引起一定的编码误差。在子范围  $[a_j, b_j]$  中,绝对误差最大的是采样值  $a_j$  和  $b_j$ ,分别  $EA(a_j) = L_j/2$ ,  $EA(b_j) = -L_j/2$ ,相对误差最大的是采样值  $a_j$ ,为  $ER(a_j) = L_j/2 a_j \times 100\%$ ,最大相对误差由子区间  $[a_j, b_j]$  的区间距离  $L_j$  和  $a_j$  决定。为了能定量地将相对误差控制在允许范围内,可作如下处理:将

采样值集合划分成各个子区间时,在采样值比较小时,采用较小的区间距离;而在采样值较大时采用较大的区间距离。选取 10 个文件进行预处理后再用改进有损压缩的自适应算术编码算法压缩,其结果如表 2 所示。

表 2 改进有损压缩算法的压缩效果

压缩前文件大小(Byte)	压缩后文件大小(Byte)	压缩率(%)
322 400	48 598	85
269 200	43 081	84
291 280	45 757	85
286 000	46 432	84
287 280	44 280	85
256 000	44 864	83
256 000	43 386	84
256 000	42 278	84
256 000	46 166	82
256 000	45 722	83

#### 4 故障录波数据文件的信道编码——文件传输协议的选择

故障录波测距装置利用 MODEM 拨号方式通过电话网传输录波数据文件,为了保证数据传送的准确性,MODEM 本身提供了差错控制协议:V.42 & MNP4,利用它们能克服电话线路上的噪声、异常突变,提供无差错的数据链路。尽管 V.42 & MNP4 能在两个 MODEM 间建立可靠的联结,但在文件传输时还是需要文件传输协议。当 MODEM 发送或接收一个文件时,计算机必须知道这个文件的细节:文件名、文件尺寸、文件属性等。这些细节都必须用文件传输协议来处理。以下对常用的二进制文件传输协议进行介绍。

(1) XModem。传输数据单位为 128 个字符的信息包,包含一个标题开始字符 SOH,一个单字节包序号,一个包序号的补码,128 个字节数据和单字节的校验和。XModem 是一种发送等待协议,每发送一个小包都要检查是否正确,有错重发,具有流量控制功能,传送可靠率能达到 96%。其优点是简单通用,但传送时间很长,只能以 8 个数据位的格式传送文件,而且在流控状态时易造成系统异常。

(2) YModem。这是 XModem 的改良版, YModem 的“包”有 1024 个字节,因此非常适合传送篇幅很长的大块文件。在线路状况良好的情况下,将会有效地缩短传送时间。如果线路状况不佳,最好不要使

用这个协议,因为频繁地重传那些被查出错误的大“包”将得不偿失。

(3) YModem - G。是一种本身不提供检错或纠错能力的流式协议,但需有 MNP(Microsoft Networking Protocol) 硬件协议为其提供检错和纠错能力。YModem - G 的每“包”容量也是 1024 个字节,但每一个“包”在传送后并不理会是否有接收端 ACK 的响应,只是不停地把所有的“包”都发出去,检错和纠错工作完全由 MNP 系统完成。因此它是一种连续 ARQ 协议。与 YModem - G 合作的最佳硬件协议是 MNP4。就传送速度而言,YModem - G 要比前面介绍的几种协议的速度快得多。但如果文件在传输过程中发生了问题,那么将导致整个文件传输失败。

(4) Kermit。Kermit 在传送文件时也是将数据打“包”并使用“校验和”。前面介绍的几种协议,只要一经选定,“包”的大小就会固定下来。Kermit“包”的大小则是随数据的整体结构和线路质量的变化而改变的。在理想的线路状态时“包”最大,大小为 1 kB(1024 字节);线路质量下降时“包”就要随之变小,或是 512 字节、或是 256 字节、或是 128 字节。当线路出现瞬时中断时,Kermit 能迅速恢复数据同步传输。大多数协议都是用 8 数据位方式传送字符,Kermit 则可在需要时将 8 数据位转换为 7 数据位传送,这样的转换使 ASCII 字符集中的 28 个控制字节在传输中增加了可靠性。Kermit 还有批处理的功能,可以用通配符“\*”代替文件名或扩展名传送数个文件。

(5) ZModem。ZModem 采用了串流式(streaming)传输方式,不再以“包”来分割发送数据,而是按连续的数据流进行处理,并把错误检验码遍布文件的自始至终。接收方对整个文件进行校验,如果有错,它只将出错部分重发,因此效率很高,而且还具有自动改变区段大小和断点续传、快速错误侦测等功能,能在传送意外中断后,迅速重新恢复数据的传送。这是目前最流行的文件传输协议。ZModem 协议能够在 7 位的信道上用编码的控制字符和其它特殊字符的办法传输 8 位数据。支持一次传多个文件,还能使用通配符“.”和“?”号,支持断点续传。在常规情况下,ZModem 的传送可靠率能达到 99%。

由上面的介绍可见,由于各种文件传输协议原理不同,特点各异,因此它们的传输效率必然相差很大,现选取 10 个录波数据文件用它们进行传送,串口波特率为 115 200 Byte/s,其结果如表 3 所示。

表 3 不同文件传输协议的传输效果

Table 3 The transmission efficiency of different file transfer protocol

文件大小 /Byte	传送时间 t/s				
	Kermit	XModem	YModem	YModem-G	ZModem
256 240	199	91	32	23	23
448 000	336	159	55	40	40
334 080	251	118	41	36	29
274 080	213	96	33	24	24
281 360	217	99	34	24	24

由表 3 可见,ZModem 传送时间最短,并且具有短点续传功能,因此应是首选的文件传输协议。

## 5 结语

用 MODEM 拨号方式,通过公用电话网传输录波数据文件,传输时间较长。为解决这个问题,本文综合考虑了文件传输的信源编码和信道编码过程,提出了在信源编码中对录波数据文件进行预处理,然后运用改进的可定量控制误差的无损压缩算法进行压缩,而在信道编码中采用具有短点续传功能的高效 ZModem 文件传输协议,将会大大缩短传送时间。传送一个 200 k 的录波数据文件仅需 10 s 时间。

## 参考文献:

- [1] 孟照勇,梁军,李欣堂,等.故障录波数据远传及综合数据处理系统[J].电力系统自动化,1997,21(8):72-75.
- [2] 吴乐南.数据压缩的原理与应用[M].北京:电子工业出版社,1995.
- [3] 苗世洪,孙扬声,吴小辰.基于电力系统故障信息远程通信的高效数据压缩与解压技术研究[J].电力系统自动化,1996,20(5):53-55.
- [4] IEEE Standard for Common Format for Transient Data Exchange (COMTRADE) for Power Systems IEEE 1999;Softcover;1999;ISBN 0-7381-1666-1;Product No.: SH94743-TBR;IEEE Standard No.: C37.111-1999.
- [5] 刘建飞,夏瑞华,秦立军,等.故障录波数据交换及故障再现研究[J].继电器,1998,26(3):25-28.
- [6] 220~500kV 电力系统故障动态记录技术准则[Z].中华人民共和国电力工业部,1994,11-7.
- [7] 赵渊.故障录波数据文件压缩及远动通信规约转换研究[D].重庆大学,1999.6.

收稿日期: 2001-12-21; 修回日期: 2002-07-17

作者简介:

赵渊(1974-),男,博士研究生,研究方向为微机在电力系统中的在线应用、电力系统规划及可靠性评估。

(下转第 54 页)

## 7 结论

(1) 注入法是一种解决低压树状配电网故障定位问题的行之有效的方法。与传统定位方法相比,它具有明显的优势,能广泛适用于树状配电网,且定位准确、通用性好、接线简单、不受分布电容的影响,大大提高了故障定位效率;

(2) 恒流信号电源控制器设计中采用闭环直流反馈 PID 控制并引入非线性环节使信号电源在负荷大范围变化的条件下仍能保持恒流恒频输出;

(3) 信号探测装置中切比雪夫滤波器和故障探测装置中编码机制的使用增强了装置的抗噪能力和灵敏度。探测器体积小重量轻,方便携带;

(4) 故障检测装置既能在停电的方式下运行也可以在带电的情况下进行检测定位,能避免停电检测带来的不便。

### 参考文献:

[1] 韩凤玲,许承斌,柳焯. 树型分支电网短路点特征识别

和定位[J]. 中国电机工程学报,1997,17(6):25-31.

[2] 韩凤玲,许承斌,柳焯. 树型分支电网单相对地短路真伪故障辨析[J]. 电力系统自动化,1998,22(8):42-47.

[3] 王慧,范正林,桑在中.“S注入法”与选线定位[J]. 电力自动化设备,1999,19(3):35-39.

[4] 桑在中,张慧芬,等. 用注入法实现小电流接地系统单相接地选线保护[J]. 电力系统自动化,1996,20(2):11-13.

[5] 薛定宇. 控制系统计算机辅助设计——Matlab 语言及应用[M]. 清华大学出版社,1996.

收稿日期: 2002-04-11

作者简介:

马佳(1976-),女,硕士研究生,主要研究方向为电力系统控制及 FACTS 应用;

余文辉(1968-),男,硕士研究生,工程师,主要从事电力系统自动化方面的研究和技术管理工作;

程时杰(1945-),男,博士,教授,博士生导师,研究方向为电力系统分析与控制、人工智能在电力系统中的应用。

## A novel fault localizer of tree form distribution networks based on an improved S injection method

MA Jia<sup>1</sup>, YU Wen-hui<sup>2</sup>, CHE Wei-yang<sup>2</sup>, XU Ken<sup>1</sup>, CHENG Shi-jie<sup>1</sup>

(1. Dept. of Electrical Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;

2. Gaozhou Electrical Company of Guangdong Province, Gaozhou 525200, China)

**Abstract:** Based on an improved S injection method, a novel fault localizer for tree form distribution networks is developed. The localizer is composed of two parts: a constant current constant-frequency signal source and a signal detector. The signal source with a close loop DC feedback and a PID controller is able to keep a constant current output under different load conditions which enables the developed system suitable for a very wide range of grounding impedance. Coding and high rank filtering technologies are introduced into the designing of both the signal source and the signal detector. The localizer is high sensitive, portable, and has a strong anti-noise ability and adaptability to various distribution networks. The effectiveness of the developed system has been verified by practical uses.

**Key words:** injection method; tree form distribution network; fault localization

(上接第 36 页)

## Research on transmission of fault recorder file based on lossy compression method and the selection of file transfer protocol

ZHAO Yuan

(The Electrical Engineering Department of ChongQing University, ChongQing 400044, China)

**Abstract:** Through public telephone network the transmission of fault information in power system which is produced by Fault Recorder is researched. The message sink coding and channel coding in the transmission process is analyzed. Based on this analysis, a kind of compression algorithm which has loss and can control the distortion quantitatively is brought forward in message sink coding. Moreover, the transmission efficiency of some kinds of file transfer protocol are compared in channel coding. Through the compression and the selection of file transfer protocol, the time in transmission of data file produced by fault recorder is shorten greatly.

**Key words:** message sink coding; channel coding; data compression; FTP