

地震观测仪器的 MiniSEED 数据格式的实现

张 旻 滕云田 王喜珍 吴 琼 朱 荣

(中国北京 100081 中国地震局地球物理研究所)

摘要 为适应地震观测台网的数字化、标准化和数据网络资源共享的要求,研究实现了在地震观测仪器的 Linux 系统中直接生成 MiniSEED 数据格式,使数据的记录、网络化传输和接收处理更为快捷、准确,适用于地震观测仪器的实时数据交换与存储。

关键词 地震观测仪器;MiniSEED 数据格式;Steim2 算法

中图分类号:P315.69 **文献标识码:**A **文章编号:**1003-3246(2007)03-0110-04

引言

随着我国地震观测台网的数字化、网络化的日趋完善,如何能更方便、准确、实时地记录和传送观测数据,对于相应观测仪器的数据交换与存储技术提出了新的要求。传统的数字化地震观测仪器多采用 DOS 系统作为操作平台。由于 DOS 系统的兼容性、可移植性、稳定性等较差,使得此类仪器在使用方面出现许多问题,如数据记录丢失、经常死机等。因为在原始数据采集后使用不同的、非标准的数据格式记录,而在数据资料接收处理时还要进行统一的、标准的数据格式转换,给数据的共享和使用造成了一些困难。因此,改进仪器的操作平台和数据的记录—传输—使用过程的格式势在必行。

目前,由于嵌入式技术在地震领域的深入应用,新一代嵌入式 Linux 的地震观测仪器已经产生。它以功能强大的 Linux 系统作为操作平台,解决了以往的诸多使用问题,并完善壮大了仪器的网络功能,为观测数据的格式统一提供了仪器基础。本文采用地震行业的标准数据格式 MiniSEED,设计出在地震观测仪器中直接生成 MiniSEED 数据格式的程序,并采用 Steim2 算法对原始数据进行压缩,方便了数据处理和共享,优化了网络传输和数据资料存储。

1 原理

1.1 MiniSEED 数据格式

目前针对数字地震数据交换的国际通用格式之一是 SEED 数据格式(王秀文,2000),它是为地震学界的应用而设计的,主要用于各个机构之间交换未经处理的地面运动数据,即地震波形数据。

作者简介:张旻(1979~),男,硕士研究生,现主要从事地震地磁观测仪器的研究工作
中国地震局地球物理研究所论著编号:07AC2005
本文收到日期:2006-11-16

MiniSEED 数据格式是 SEED 数据格式的简化,对于只需了解台站特性的专业人员是一种数据传送的适宜方法。它由若干数据记录组成,每个记录包含一个固定头段部分、一个可变头段部分和一个数据部分(图 1)。记录长度常固定为 4 096 个字节。

固定头段部分	
可变头段	子块 1 ⋮

固定头段部分	
可变	子块 1
头段	⋮
部分	子块 N
数据部分	

图 1 MiniSEED 数据记录结构

固定头段部分包含非格式化的标识和状态信息,出现在每个数据记录开头的固定系列中,包括 18 个数据字段(ASCII 或二进制格式表示的一项辅助信息),共 48 个字节。

可变头段部分由一系列可选的非 ASCII 格式的子块序列构成,子块数据结构包含一个类型标识符、长度,以及一系列与特定子块类型有关的数据字段,提供有关针对通道和依赖时间的事件(如自动确定的震相到时信息或正在进行的标定)的信息。多数子块都是可选的。本设计选取只数据子块(子块[1 000])来补充记录长度和编码格式信息。

数据部分包含实际的由一个台站若干通道在一个有限时间段内连续记录下来的原始数据,即时间序列数据。数据以编码格式存放。

本设计的数据记录结构为:固定头段部分(48 字节),可变头段部分(16 字节),数据部分(4 032 字节),总长度 4 096 个字节。

1.2 Steim2 压缩算法

每个数据记录的数据都以编码格式存放, MiniSEED 有多种编码格式可选, 其中以 FDSN 台网编码格式最为广泛, 本文选取其中的 Steim2 压缩格式(陈运泰, 2000)。

Steim2 压缩算法是一种整形数据的压缩算法。在 32 B(4 个字节)有符号整数(二进制补码格式)的时间序列中,每个整数代表正常地震背景情况下地震仪输出的时间序列数据。由于正常地震背景的频率与数据采样率相比一般都相当低,所以连续数据样本之间的差值与满量程 32 Bit 数字值相比一般很小。

设一个原始时间序列样本为 $X_{-1}, X_0, X_1, \dots, X_n$, 其中每个 X_i 是一个 32 Bit (或小于 32 Bit) 有符号整数。设 d_0, d_1, \dots, d_n 为 32 Bit 宽的一阶差值时间序列, 即 $d_0 = x_0 - x_{-1}$, $d_1 = x_1 - x_0, \dots, d_n = x_n - x_{n-1}$ 。根据差值 d_i 的数值大小, 可以用 4, 5, 6, 8, 15, 30 Bit 表示这一差值。通过对连续 d_i 划分差值类型, 可以由几个连续 d_i 组成一个 32 Bit 字 W_K 。因此, 用差值字 W_K 的记录格式来存储取代原始时间序列, 能在不丢失任何信息的情况下显著地压缩数据, 节省存储空间。

在 MiniSEED 数据记录中,用 Steim2 压缩的数据部分(4 032 字节)分成 63 个 64 字节帧,每一帧包含 1 个差值类型标记(W_0)和 15 个 32 Bit 差值字($W_1 \sim W_{15}$),在数据帧 0 中, $W_1 = X_0, W_2 = X_1$ (图 2)。

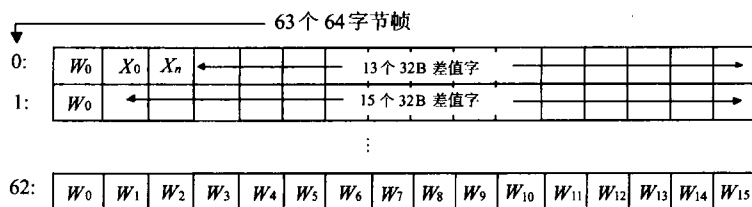


图 2 Steim2 压缩数据格式

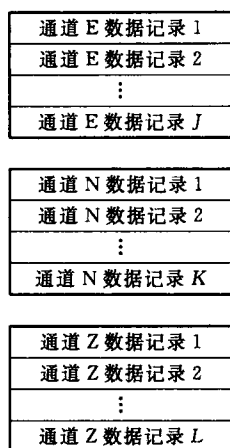


图 3 MiniSEED 文件结构

2 实现

2.1 文件结构

原始数据为现场台站仪器经由一组特定滤波器产生的若干通道的时间序列的数字记录。它们在几乎同一时刻被采样,并逐一存储在一个固定系列中。地震观测仪器的原始数据一般为三通道数据,即三分量记录 E(东西)、N(南北)、Z(垂直)。由于各通道的相关性不同,所以不应把几个地震数据通道多道组合进单一记录中。因此,MiniSEED 的一个数据记录中只应含有一个通道的数据。地震观测仪器的三通道 MiniSEED 文件应以一定顺序将各通道的数据记录按时间序列分别存放。由于采用 Steim2 压缩格式,故相同时间序列的三通道数据记录总数不一定相等,文件结构见图 3。

2.2 程序设计

地震观测仪器系统主板的主频为 70 MHz,内存为 32 MB,系统平台为 Linux 2.4.5(内核版本号),系统程序语言为 C 语言。为节省系统消耗,保证整个系统的正常运行,本程序共使用内存约 90 kB。实现将 32 Bit 有符号整数(二进制补码格式)表示的三通道原始数据,连续进行 MiniSEED 交换与存储(邹思轶,2002)。

首先进行变量、参数初始化;根据台站和仪器的情况输入标识和状态信息,如:台站通道标识符、原始数据采集的起始时间、仪器采样率等;顺序读取定量(固定内存)的单通道数据;然后计算出数据的差值序列,用 Steim2 算法压缩存储数据;满一个数据记录长度后,通过压缩的样本数、原始数据采集的起始时间和采样率计算记录起始时间,输入记录起始时间、记录序号、通道标识符等记录信息;存储数据记录,直至单通道数据读取结束;依次进行各通道数据处理,生成相应的 MiniSEED 文件,循环处理下一原始数据(任永铮,2000)(图 4)。

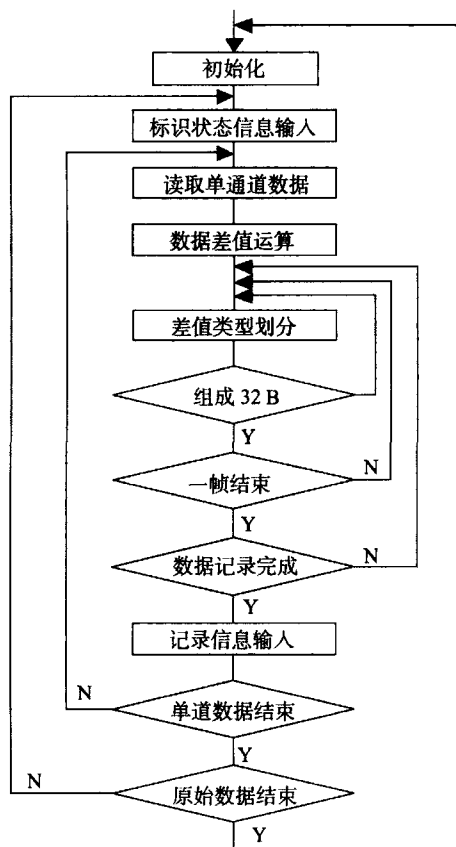


图 4 程序流程

地震观测仪器采样率范围为 62.5~2 000 样/s。为检验程序在其采样率范围内的执行情况,在正常地震背景下,选取 10 min 的不同采样率的 4 组原始数据进行 MiniSEED 处理,计算出执行时间和压缩率,结果见表 1。

从表 1 中数据文件压缩率看,相同的正常地震背景下的数据压缩率与原始数据大小无关,

3 实验结果

符合 Steim2 算法压缩原理。并且压缩率在 3 倍以上,减轻了地震观测仪器因数据量大而造成
的数据存储负担,进而优化了数据资
料的网络传输。由于程序消耗固定的
较小的内存资源,故表 1 中的执行时
间与原始数据大小成正比关系。虽然
相同时间的原始数据采样率大的执行
时间也长,但是一般情况下,地震观测
仪器的采样率都在 500 样/s 以下,因
此,对于一般的原始数据,程序能在较短的时间内生成相应的 MiniSEED 文件。

表 1 MiniSEED 程序执行结果

仪器采样率 /样·s ⁻¹	10 min 原始 数据大小/MB	生成 MiniSEED 文件大小/MB	数据文件 压缩率	程序执行 时间/s
62.5	0.43	0.12	3.58	0.21
250	1.72	0.47	3.66	0.78
500	3.44	0.95	3.62	1.65
2 000	13.76	3.75	3.67	6.26

4 结论

本文实现了在地震观测仪器的 Linux 系统中直接生成 MiniSEED 数据格式。使数据的记
录—传输—使用过程达到格式统一,方便了数据的处理和共享;采用 Steim2 压缩算法,优化了
数据的存储和网络传输,适用于地震观测仪器的实时数据交换与存储。

参考文献

陈运泰,吴忠良,王培德等. 2000. 数字地震学[M]. 北京:地震出版社,7~72
任永锋. 2000. Linux C 程序员指南[M]. 北京:国防工业出版社,28~268
王秀文,姚立平,赖德伦等. 1994. 地震数据交换标准[J]. 地震地磁观测与研究, 15(2): 1~38
邹恩秩. 2002. 嵌入式 Linux 设计与应用[M]. 北京:清华大学出版社,18~126

Application of the MiniSEED data format of the seismic observation in-
strument

Zhang Yang, Teng Yuntian, Wang Xizhen, Wu Qiong and Zhu Rong

(Institute of Geophysics, China Earthquake Administration, Beijing 100081, China)

Abstract

For the demand of the digitalization, standardization and data share of seismic observation
network, the data production in MiniSEED format on Linux system of the seismic observa-
tion instruments is achieved. So the recording, transporting and receiving of data by network
becomes quicker and more precise. The method can be applied in the real-time exchange and
storage of the seismic data.

Key words: seismic observation instrument, MiniSEED data format, Steim2 algorithm