

多重索引加时间戳校正的科学卫星源包排序法^①



马文臻^{1,2,3}, 邹自明², 黎建辉¹, 李冰⁴, 石涛⁴, 孙小涓⁴

¹(中国科学院 计算机网络信息中心, 北京 100190)

²(中国科学院 国家空间科学中心, 北京 100190)

³(中国科学院大学, 北京 100049)

⁴(中国科学院 空天信息创新研究院, 北京 100190)

通信作者: 马文臻, E-mail: mawenzhen@nssc.ac.cn

摘要: 卫星载荷数据源包的排序还原是地面数据处理流程中的重要环节, 其正确性和完备性直接影响基于数据开展的科学研究。在对国际现行 CCSDS 标准下的卫星数据组织和星地数传机制可能给地面数据处理带来的数据异常情形与源包排序挑战进行剖析的基础上, 提出了“多重索引+时间戳校正”的源包排序方法, 不仅能够解决由于空间科学卫星数据采集频率极快而导致的源包排重和拼接难题, 而且在帧计数、源包计数、时间码等关键索引信息发生异常的情况下, 仍然能够给出正确的结果, 在我国空间科学卫星任务实践中大幅提升了源包排序处理的正确率, 同时有效地提升了处理效率。

关键词: 空间科学卫星; 数据处理; 源包排序; 多重索引; 时间戳校正

引用格式: 马文臻, 邹自明, 黎建辉, 李冰, 石涛, 孙小涓. 多重索引加时间戳校正的科学卫星源包排序法. 计算机系统应用, 2022, 31(1): 37-46. <http://www.c-s-a.org.cn/1003-3254/8272.html>

Science Satellite Data Packets Sorting Based on Multiple Index and Timestamp Correction

MA Wen-Zhen^{1,2,3}, ZOU Zi-Ming², LI Jian-Hui¹, LI Bing⁴, SHI Tao⁴, SUN Xiao-Juan⁴

¹(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

²(National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China)

³(University of Chinese Academy of Sciences, Beijing 100049, China)

⁴(Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: The sorting and restoring of source packets of satellite payload data is an important part of a ground data processing flow, and its correctness and completeness directly affect data-based scientific research. This study analyzes the abnormal data and challenges to source packet sorting in ground data processing brought by the satellite data organizations and satellite-ground data transmission mechanisms under the current international standard—the consultative committee for space data systems (CCSDS) standard. On this basis, it proposes a “multiple indexes and timestamp correction” packet sorting method, which not only solves the difficulties in duplicate data elimination and data registration of ground source packets caused by the frequent acquisition of data about space science satellites but also gives correct results in face of key index abnormalities such as abnormalities in the frame serial number, packet serial number, and timestamp. This method has effectively promoted the accuracy of source packet sorting and data processing efficiency of China’s space science missions.

Key words: space science satellite; data processing; source packet sorting; multiple index; timestamp correction

① 基金项目: 中国科学院“十三五”信息化建设专项 (XXH13505-04); 北京市科技计划 (Z181100002918002)

收稿时间: 2021-03-27; 修改时间: 2021-04-26; 采用时间: 2021-05-07; csa 在线出版时间: 2021-12-17

1 引言

在卫星地面数据处理的全部流程中,对正确性和时效性要求最高的一个环节就是数据预处理,其中最关键的技术就是载荷源包排序算法.卫星载荷获取的探测或者实验数据,以载荷源包的形式存储在星上存储器中,在记录数据之后需要尽快下行到地面执行预处理,去除传输层的结构,将载荷源包还原成载荷执行探测并记录下来的原始数据内容序列,并尽最大可能保障数据的正确性和完备性,以支持后续的科学研究的.

从星上各载荷组织源包、数管将各路源包统一组成信道传输帧、数传发射机传输数据、地面接收站接收数据、地面数据处理系统进行数据解帧、源包提取与正确性校验、源包时间码处理、源包排序到数据连续性及其完备性验证^[1],每一个步骤都可能发生异常.地面系统需要识别发生的各种异常,判断是否可以通过容错处理进行地面修复,确认是否可以还原得到源包顺序及内容都正确的、连续而完备的载荷科学数据源包序列^[2].如果确认存在无法修复的异常,地面系统需要权衡缺失数据的科学重要性、星上存储是否已经覆盖、地面测控站是否有余量进行指令上行、地面接收站是否有余量进行增量接收等各种情况,尽快决定是否上行指令点播发生异常的载荷源包数据,以免在点播完成之前星上存储器中的数据被覆盖^[3].

如果数据标定处理和数据反演处理不够理想,都可以事后重新进行处理和验证.只有预处理,如果星上数据发生错误或者缺失,是无法重新获取的.关键位置的一个源包的残缺或者误码,都可能导致科学实验结果失效或者观测事例不完备.例如,在量子科学实验卫星中,科学数据源包存储了星地量子密钥,地面实验数据的分析处理中,需要对密钥进行逐 bit 的基矢比对,一个 bit 都不能错,否则该次实验就失效.所以,空间科学卫星对于载荷源包还原的正确性和完备性有着极为严苛的要求,这是不同于其它任何业务应用卫星的.其它的对地观测卫星、气象卫星等,都是允许一定的误

码率,个别的源码错误或者缺失不影响整张图片的产出或者探测量的监视.

在航天领域,通用的卫星数传误码率指标为纠错后达到 $1e-5$,这是按照 bit 位计算的,换算为帧计数的异常率平均在几分之一.在空间科学卫星数据中,误码率通常优于 $1e-6$,那么每个卫星下行原始数据文件中发生帧计数异常的概率为几分之一,也就是在每个文件中都会遇到几次帧计数错误.卫星上对源包赋予计数和时间码的时候也可能不定时发生异常.因此,排序依据的标志位发生异常情况下,如何能提高处理正确率,是非常值得研究的^[4,5].

在与美国航天局(NASA)和欧空局(ESA)地面数据处理机构的学术交流活动中了解到,其空间科学卫星地面系统对卫星数传原始数据进行源包排序处理的整体文件正确率平均达 92% 左右.最终发生错误导致失效的源包占总源包的比例大概在百万分之一的级别.然而,对于空间科学卫星来说,也许失效的那几个源包,恰好是一个实验或者一次观测的关键数据,会导致一次实验或者观测的失败;也许失效的那几个源包,正好错过了百年难遇的重大发现.为了支持科学发现上可能的突破,地面数据处理领域值得付出持续不断的改进努力,向着 100% 的正确率去靠近.

2 问题背景与现状分析

2.1 星地数传数据的组织结构

空间科学卫星的数据组织与通信采用国际航天领域通用的 CCSDS 标准.卫星数据有科学数据、载荷工程数据和平台工程数据 3 大类,科学数据是载荷设备通过开展观测得到的原始观测数据或者开展实验得到的原始实验数据,载荷工程数据是表征载荷运行状态的工程参数,平台工程数据是表征卫星平台设备运行状态的工程参数.卫星平台及载荷的各设备使用应用过程标识符对卫星数据采集和传输过程中不同数据源设备进行区分,构成可变长度的面向应用的数据包,称为“源包”.遥测源包的格式如表 1 所示^[6,7].

表 1 CCSDS 遥测源包格式

同步码	E-PDU 主导头							副主导头		E-PDU 数据域
	版本号	包标识			包序控制		包数据长度	时间码		
		类型	副主导头标志	应用过程标识符	分组标志	包序列计数				
16 b	3 b	1 b	1 b	11 b	2 b	14 b	16 b	64 b	字节数×8	
2 B			2 B			2 B	2 B	8 B	变长	

同步码长度为 16 bit, 一般用‘E225’h (十六进制). 每一种不同类型的源包依据应用过程标识符 (APID) 来区分. 硬 X 射线调制望远镜卫星 (HXMT) 有 35 个 APID, 暗物质粒子探测卫星 (DAMPE) 有 14 个 APID, 量子科学实验卫星 (QUESS) 有 11 个 APID, 实践十号返回式科学实验卫星 (SJ-10) 有 48 个 APID. 按照 APID 对飞行器上每个特有的源应用过程产生的源包分别进行包序列计数, 长度 14 bit, 模 16 384.

副导头标志长度为 1 bit, ‘0’b (二进制) 表示源包内没有副导头, ‘1’b 表示源包内有副导头数据. 空间科学卫星由于科学数据采集频率极高, 源包序列计数在一次下行中就会回滚很多次, 所以必须联合使用源包时间码来区分源包, 副导头标志均为‘1’b. 分组标志为长度为 2 bit, 标记该源包是独立源包还是分组源包, 用来提供源包在组内位置的信息. 副导头长度 8 字节, 自定义内容, 目的是为任何辅助数据 (如时间, 内部数据

域格式, 航天器位置/姿态等) 编入源包提供一种 CCSDS 定义的手段. 在空间科学卫星中, 副导头为时间码, 其高 2 字节为 0, 然后 4 字节为秒计数, 最后 2 字节为毫秒计数.

E-PDU 数据域是可变长度的 N 字节, 是应用过程在每个源内运行产生的真正的科学探测或科学实验信息.

卫星数管设备将众多遥测源包分成几大类, 分别组成不同的虚拟信道数据单元 (VCDU), 每个 VCDU 包括主导头和数据域, VCDU 加上同步码就构成了数据传输帧 (AOS frame). 不同的虚拟信道的 AOS 帧用虚拟信道标识符 (virtual channel identifier, VCID) 来区分. 针对工程参数区回放数据、科学数据区回放数据和实时工程参数数据等, 数据传输帧划分不同虚拟信道, 以支持不同的数据应用. 第一层的遥测源包的连续二进制码流截断填充到第二层 AOS 传输帧的数据域部分. AOS 传输帧的格式如表 2 所示^[8-10].

表 2 AOS 传输帧格式

同步码	VCDU主导头				VCDU数据域				VCDU差错控制域	编码校验区
	主信道标识符		虚拟信道标识符	虚拟信道帧计数	信号域		B_PDU位流数据			
	版本号	航天器标识符			回放标志	备用域	B_PDU导头	B_PDU数据域		
32 b	2 b	8 b	6 b	24 b	1 b	7 b	16 b	字节数×8	16 b	512 b/1024 b
4 B	2 B		3 B	1 B	2 B	自定义长度	2 B	64 B/128 B		

数据处理系统解析“ AOS 帧—遥测源包” 的双层结构的过程如图 1 所示. 首先依据虚拟信道标识符、虚拟信道帧计数、回放标志等进行信道区分, 从一个或多个传输帧的数据域中解析得到有效的源包数据, 然后解析源包结构, 依据应用过程标识符、源包计数、包长进行分路处理, 还原出每一类的遥测源包, 对其数

据内容进行有效性验证, 之后依据时间码或源包计数进行源包排序, 使其还原为星上探测的原始顺序^[9-11]. 只有排序正确、计数连续的源包序列, 其数据域的内容依次连接, 才可以形成正确的用户数据单元 (protocol data unit, PDU), 也就是正确、完备的原始科学探测或实验数据.

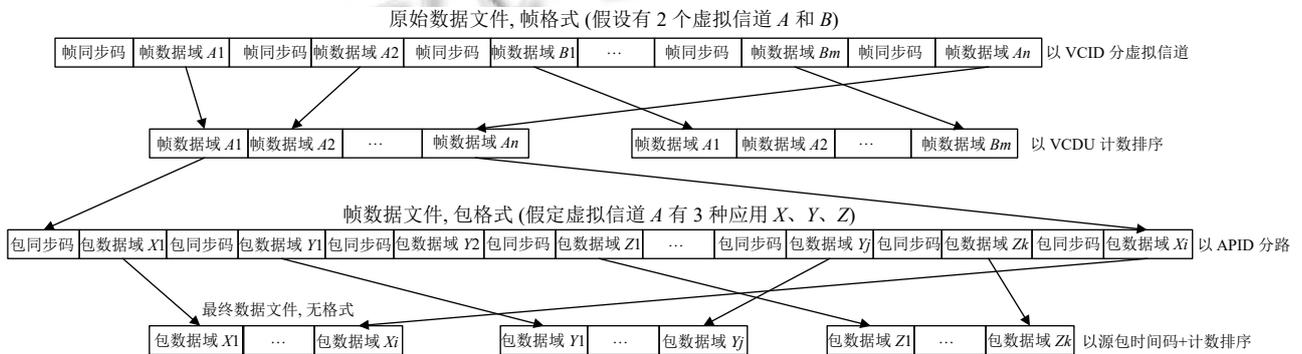


图 1 传输帧与源包组织结构的帧解析过程

2.2 传统的源包排序处理算法及问题

在遥感卫星、气象卫星等应用卫星的地面数据预

处理工作中, 源包排序算法是基于“虚拟信道帧计数+源包计数”进行的. 其基本排序策略为先按照源包

计数进行排序,如果源包计数相同,则按照 VCDU 计数进行排序.这种排序策略并未考虑到计数溢出而重置的情况,也不支持同一轨接收过程中进行历史数据点播的操作^[12-14].

在空间科学卫星数据处理领域,由于两层计数循环速度太快,重置频率高,单轨数据中常出现源包计数和 VCDU 计数都相同的情况,所以“虚拟信道帧计数+源包计数”的双重计数排序法不满足空间科学卫星数据的处理需求.所以空间科学卫星在源包结构的副导头中都增加了“时间码”字段,扩展为“源包时间码+源包计数+虚拟信道帧计数”的源包排序处理策略,时间码相同的源包按照源包计数进行局部排序,虚拟信道帧计数和源包计数仅仅是作为局部变量来使用的.虚拟信道计数主要用来拼接跨两个虚拟信道的源包数据,使其按照下行顺序首尾衔接.由于存储空间有限,CCSDS 为避免大量冗余,源包的时间码一般精确到秒,最多精确到毫秒,而很多观测数据的采集频率极高,最高情况下 1 s 可以产生上万个源包,源包计数的深度为 14 bit,计满 16384 条数据就会循环重新开始一个计数周期,所以源包计数主要用来区别时间码相同的源包的真正采集顺序^[15-17].

这种“源包时间码+源包计数+虚拟信道帧计数”的通行方法对于星上数据和时间码都正常的情况下是没有问题的,但是如果帧计数、源包计数和时间码等关键索引信息由于各种异常发生错误而不能代表源包的真正采集顺序呢?

不同的空间科学卫星,由于载荷对数据采集记录时间的精度要求的不同,给星上时间系统的设计带来了不同的要求和实现难度.源包中的时间码,一般有两种来源,有的来源于 GPS 设备,有的来源于自身时钟累计的相对秒数和毫秒数.如果载荷组源包的过程发生通讯或者软件异常,没有取到时间码,那么源包中时间码字段会变成默认值^[17].如果时钟的秒和毫秒计数不完全同步,会导致时间错齿跳变.如果受到单粒子时间打翻存储器的某一位或者某几位的记录,那么时间码会发生无规则的跳变.源包计数和虚拟信道帧计数也都是如此,当接收设备故障、雨雪天气、突发信号干扰等原因造成的误码率超出可纠错范围时,当星上数传设备出现故障时,遭遇单粒子打翻某些 bit 位时,当点播历史数据时,都可能出现重复、错误、乱序等异常情况,而且有时候多种异常情形会联合发生.这时

候,“源包时间码+源包计数+虚拟信道帧计数”的算法就会发生排序错误,而且很难有一种万全的源包排序算法,能够保证在所有异常情况下都能百分百正确地实施源包排序.对于发生排序错误的源包,有的可以通过耗时漫长(可能几天到几个月)的人工分析去修复,有的没有得到彻底修复,造成个别科学数据源包失效.

纵观国内外的主流航天总体机构,都针对性地开展了卫星数据源包提取与排序技术的研究,尽量提高排序正确率,降低因各种时间码异常情况而导致的大量人工干预工作和对科学研究的负面影响.

3 多重索引加时间戳校正的源包排序算法

经过对我国空间科学先导专项一期发射的 DAMPE、QUESS、HXMT 和 SJ-10 四颗卫星的星上数据组织结构与地面数据预处理过程进行研究,本文设计了一种把诸多异常情形综合考虑在内的“多重索引+时间戳校正”的源包排序方法.

3.1 多重索引构建

AOS 帧解析、源包提取与多重索引构建的整个处理过程如图 2 所示.

首先,构建信道级帧索引.获取数传原始数据文件后进行文件解析与校验,之后进行 AOS 帧的完整性和有效性校验,然后进行虚拟信道区分、直传与回放数据的区分,按照虚拟信道的不同,形成每个信道的帧文件,分别进行帧的排序与连续性校验,提取帧索引表所需的信息,构建信道级的帧索引,如图 3 所示.

帧索引信息不仅包含传输帧的虚拟信道标识符(VCID,十六进制)、虚拟信道帧计数(VCDU Count,十六进制),还包括直传回放标志(Flag,二进制 0 或 1)、帧 B-PDU 数据域在原始数据文件中的偏移位置(VCDU Offset,十六进制字节数)、下行轨道圈次(Track No,十进制)等信息^[18].

其次,构建应用过程级源包索引.对每个虚拟信道的帧文件中的帧进行数据的连续性校验,在校验正确的情况下从帧中提取源包数据域,按照不同的 APID 拼接形成完成的源包(其中,一个虚拟信道标识符 VCID 对应一个或者多个应用过程标识符 APID),并进行源包的有效性校验,通过校验的源包,将其信息解析后提取源包索引表所需的信息,构建应用过程级的源包索引,如图 4 所示.

源包索引信息不仅包括最重要的应用过程标识符

(APID, 十六进制)、源包时间码 (packet time, 转化为统一的时间格式“年月日 T 时分秒.毫秒”)、源包计数 (packet count, 十六进制), 还包括源包长度 (packet length, 十六进制字节数) 和源包 E-PDU 数据域在原始数据文件中的偏移位置 (packet offset, 十六进制字节数) 等信息, 还包括一个格外重要的字段, 就是校正时间戳 (corrected time stamp, 格式与源包时间码相同), 这个字段存放的是经过时间戳校正之后的源包时间码, 本文提出的算法将主要依据这个校正时间戳进行源包

排序, 此外, 还有异常标记 (anomaly flag, 十进制), 代表不同的时间戳校正策略。

帧索引和源包索引中的信息按照一定规则组合起来, 就构成了多重索引. 多重索引与数据帧及源包中的相关字段一一对应, 于是把对数据本身的处理转化为对索引的操作, 而索引的数据量相对于数据本身是非常小的, 所以可以全部在内存中处理, 避免了对数据本身进行处理而产生的大量内外存数据交换, 提高数据源包排序处理的效率。

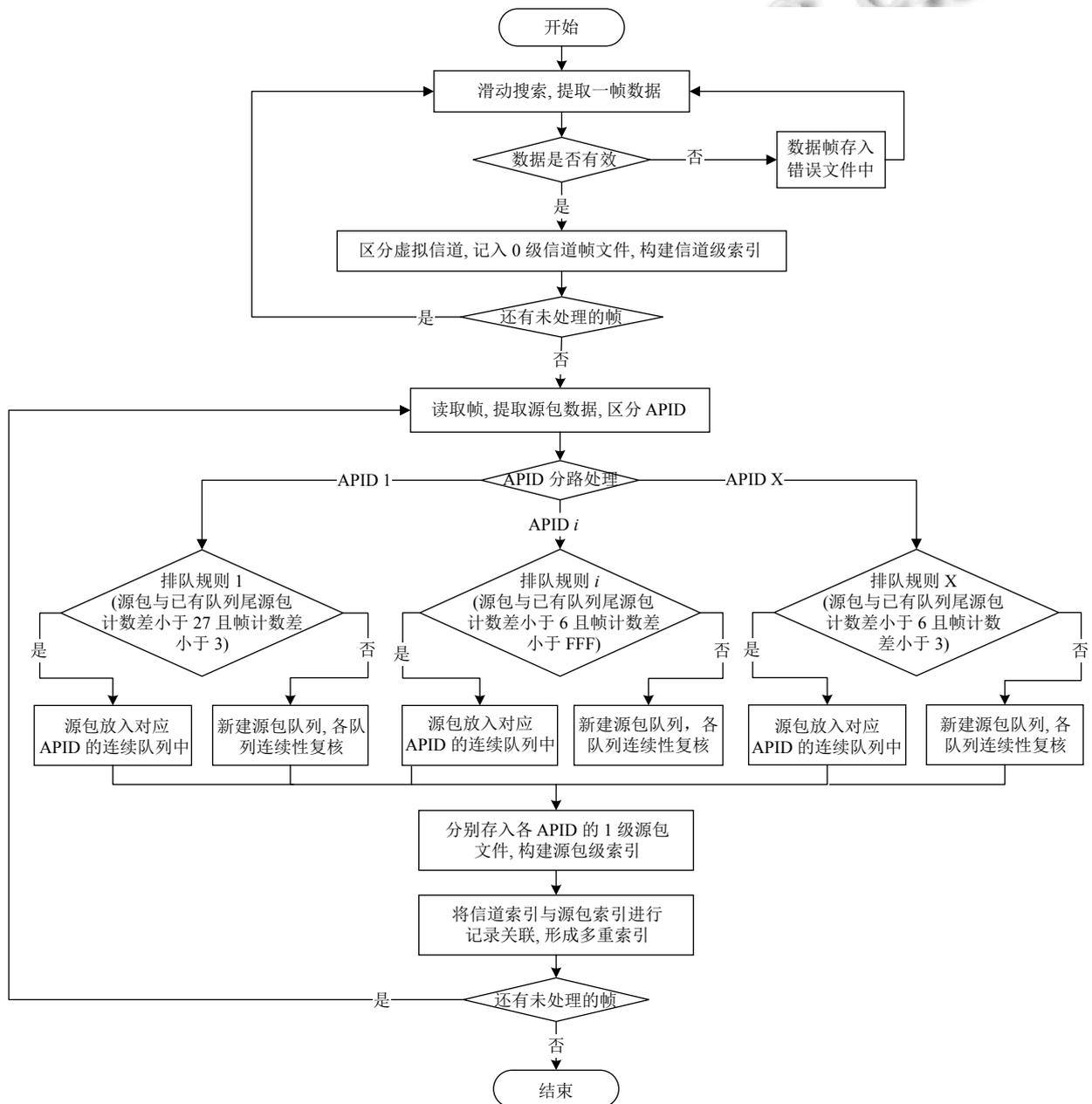


图2 多重索引构建

VCID	虚拟信道帧计数	回放标志	帧偏移	下行圈次
A1	BF589A	0	A7F000	9001
A1	BF589B	0	A7F400	9001
24	100010	0	B1200	9001
3F	000000	1	0	9001
3F	000001	1	400	9001

图3 帧索引表

3.2 时间码提取与时间戳初标

针对不同卫星的不同源包, 依据时间码来源与异常情况的不同, 其时间码提取、时间戳初次标记、异常标记的策略也有所不同, 可以配置不同的处理流程和参数。

以HXMT卫星的几个典型APID为例进行说明,

源包时间码提取与时间戳初标的过程如图5所示。高能望远镜(HE)的科学和工程参数源包时间码来自于GPS设备, 因此需要判断GPS设备是否处于非定位状态来确定时间码的正确性, 当其处于正常的定位状态时, 时间戳初始标记为GPS的时间码, 在处于非定位状态时, 时间戳初始标记为0, 同时进行异常标记。中能望远镜(ME)的源包时间来自于UTC事例, 而且多组FPGA源包与UTC事例源包轮流排列, 因此需要判断FPGA分组编号的连续性并查找对应的UTC事例, 时间戳初始标记为UTC事例中的时间码, 存在异常时时间戳初始标记为0。平台工程数据由一系列不同类别的源包首尾相接组成, 需要检查源包组的完整性, 如果恰好有不完整的源包组, 则与之前20轨数传下行数据的不完整缓存进行拼接, 内容完整的源包从GPS定位数据包或者时间数据包中获取时间码。

源包索引表 (APID=0282)						
APID	源包时间码	源包长度	源包计数	偏移位置	校正时间戳	异常标记
0282	20180101T010101.958	56	0000	A7F000	20180101T010101.958	
0282	20180101T010101.000	56	0001	A7F056	20180101T010102.000	2
0282	20180101T010102.016	56	0002	A7F0AC	20180101T010102.016	
源包索引表 (APID=039A)						
APID	源包时间码	源包长度	源包计数	偏移位置	校正时间戳	异常标记
039A	20180101T010101.185	188	01CB	A7F102	20180101T010101.185	
源包索引表 (APID=0696)						
APID	源包时间码	源包长度	源包计数	偏移位置	校正时间戳	异常标记
0696	20180101T010101.185	220	1A11	A7F2F8	20180101T010101.185	
源包索引表 (APID=02FE)						
APID	源包时间码	源包长度	源包计数	偏移位置	校正时间戳	异常标记
02FE	20180101T010101.185	256	001C	B1200	20180101T010101.185	
02FE	20120101T010101.008	256	001D	B1456	20180101T010101.185	1
源包索引表 (APID=049B)						
APID	源包时间码	源包长度	源包计数	偏移位置	校正时间戳	异常标记
049B	20180101T010101.185	446	0155	0	20180101T010101.185	

图4 源包索引表

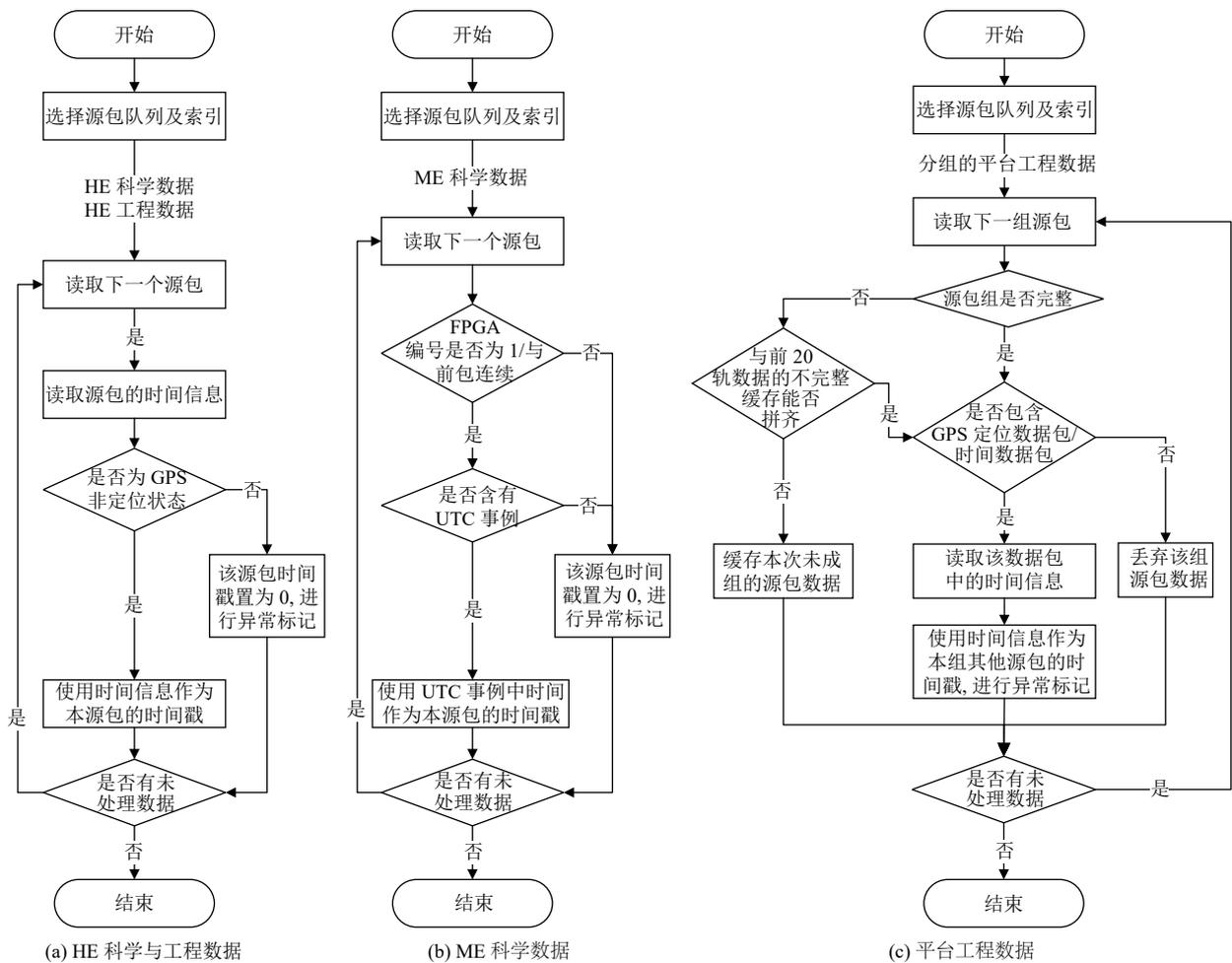


图 5 源包时间码提取与时间戳初标过程

关键参数之一是队列连续性判别参数. 根据虚拟信道与应用过程标识符的对应关系、源包的分组特性, 对各 APID 的源包队列连续性判别参数进行分别设置. 举例如下:

(1) 对于 HXMT 卫星的低能望远镜 LE 科学数据, 判断队列连续性时使用源包计数差小于 6 且帧计数差小于 FFF 进行判定. 因为 LE 科学数据包含 3 个独立的 APIDI, 分别为 LE 科学 1、LE 科学 2、LE 科学 3, 它们共用一个虚拟信道, 而每一帧数据存储 2 个独立的源包, 各源包最大计数为 $0x3FFF$, 因此每类源包一个计数循环最多占用 $1FFF$ 个帧计数. 为保证源包排序, 使用半个源包计数周期占用的帧计数 FFF. 通过对发现问题当轨数据比对, 两个源包间帧计数差未超过 FF, 因此使用 FFF 作为参数可以满足要求.

(2) 对于 HXMT 卫星的高能望远镜 HE 科学数据、中能望远镜 ME 科学数据, 判断队列连续性时使

用源包计数差小于 6 且帧计数差小于 3 进行判定. 存储 HE 和 ME 科学数据的每个虚拟信道仅存储一类源包, HE 科学数据虚拟信道每帧数据存储 1 个源包, ME 科学虚拟信道每帧数据存储 2 个源包, 当对 3 帧数据容错时, 可能丢失源包数应不大于 6.

(3) 对于 HXMT 的载荷工程数据, 判断队列连续性时使用源包计数差小于 27 且帧计数差小于 3 进行判定. 各载荷的工程数据虚拟信道仅存储该载荷的工程数据源包, 每帧数据存储 9 个源包, 当对 3 帧数据容错时, 可能丢失源包数应不大于 27.

关键参数之二是成组源包的不完整数据回溯文件数参数. 对于分组源包, 一组连续源包共同组成一条探测记录的情况, 占用多个传输帧进行传输, 因此需要判断该组源包的完整性. 举例如下: 对于某些类型的源包组不完整的数据, 回溯 20 轨数据尝试进行源包组的补充.

3.3 时间戳校正

从卫星原始数据中提取出来的各种索引字段,没有一个随着源包的真正顺序而永远递增的,帧计数、源包计数都会周期性回滚归零重新累计,源包时间码可能发生各种错误和乱序而没有依据源包真实的先后顺序而递增,因此作为排序依据使用起来有一定困难.而想要对乱序下行的多份重复源包进行正确排序,必须建立一个真正随着源包的先后顺序一直递增的标志,基于这种情况,本文设置了“校正的时间戳”.源包时间戳校正环节的主要流程如图6所示.

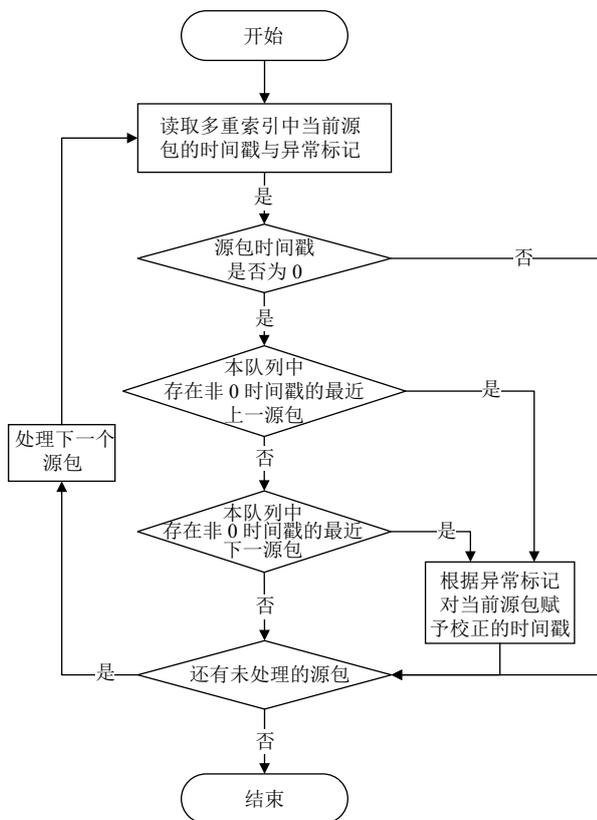


图6 源包时间戳校正

根据不同的异常标记采取不同的时间戳校正策略.

(1) 当因无法获取 GPS 时间导致源包时间码变成默认的填充码时,将其时间戳校正为该源包之前的最近一个正常的源包时间码;

(2) 当载荷异常开关机导致累积秒归零,造成时间码复位到初始位置,重新累积,将发生复位后的源包的时间戳校正为当前累积秒叠加复位之前累积秒的总秒数经换算得到的时间码;

(3) 当时间的秒部和毫秒部的进位不完全同步造

成“跳秒”现象,也就是判断当前源包的时间码小于上一个源包,并且很接近(设置为2s之内),那么将当前源包的时间戳校正为增加1s,并判断校正后的时间码大于上一个源包的时间码,且不大于下一个源包的时间码,如果不满足则把时间戳校正为上一个源包的时间码;

(4) 当源包时间码因单粒子事件等造成无规则异常时,检索当前源包时间码,如果小于上一个正常源包的时间码或大于下一个相邻源包的时间码,将其时间戳校正为上一个正常源包的时间码;

(5) 如果本队列中在当前源包之前不存在非0时间戳的正常源包,那么依据其之后相邻最近的非0时间戳的正常源包来进行时间戳校正,如果本队列时间戳均为0,即时间码有误,那么回溯上一轨数据继续寻找时间戳正常的源包作为校正基准.

3.4 源包排序

源包排序流程如图7所示.首先依据校正后的源包时间戳对数据进行排序,进行时间大小比较时,当时间戳无法区分时,借助源包计数、帧计数、源包偏移量、轨道圈次等索引信息进行联合判断.对于时间差在一定时间内的数据(这个时间是可以对不同的卫星和载荷进行区分设置的),作为时间相等的数据进行处理.例如在HXMT卫星中,依据卫星系统的设计,卫星的GPS非定位时间状态不会超过20min就会复位正常,因此设定时间差在20min作为判断依据.

4 方法应用与验证

本文使用HXMT、QUESS、DAMPE三个卫星任务2018年度的全部数传原始数据进行了实验验证,数据资源来自国家科技资源共享服务平台-国家空间科学数据中心.

三星共5347个数传原始数据文件,总体量22.78TB,使用传统的“源包时间码+源包计数+虚拟信道帧计数”源包排序算法处理,源包排序完全正确的文件占比平均为93.55%.处理错误的文件中,平均约15%是由多种异常联合发生所引起.使用本文提出的“多重索引+时间戳校正”源包排序处理算法处理,源包排序完全正确的文件占比平均为99.84%,其中排序错误的几个文件均是由于星上本身设备故障引起的数据错误.可见,本文提出的算法,大大提升了源包排序处理的正确率,具体实验结果如表3所示.运行环境为12个处理节点

的服务器集群 (4 个 CPU Intel Xeon E7-4820v2, 主频 2 GHz, 内存 64 GB), 由中国科学院国家空间科学中心公共技术服务中心提供.

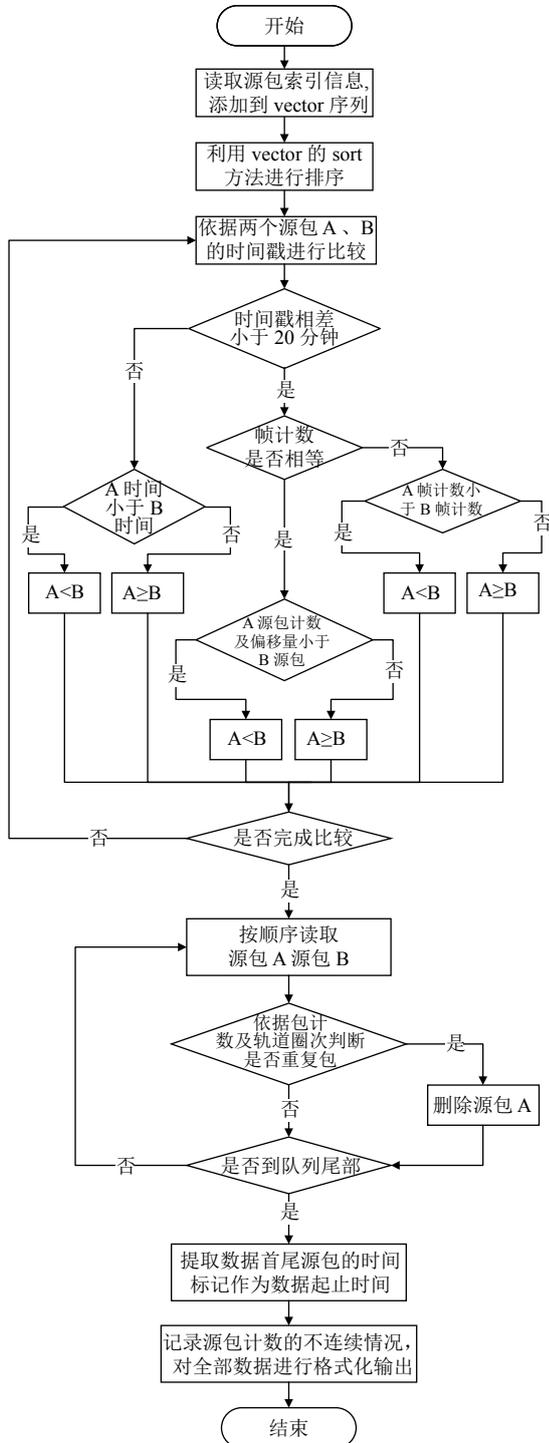


图 7 源包排序流程

在同一个 HXMT 卫星数传原始数据文件中分别截取源包计数无循环、源包计数有 3 次不完整循环的

大小为 1 GB 的数据片段, 帧长为 1 024 字节, 源包长度为 488 字节. 使用本文提出的处理方法和传统处理方法进行源包提取与还原排序, 分别记录两种方法的处理时长及最大占用内存, 实验结果如表 4 所示. 为增加运行时长而使性能指标对比明显, 选择低配置的运行硬件环境, 在一台普通台式计算机 (Intel Xeon CPU E5649 @2.53 GHz 16 GB RAM) 上进行实验. 结果说明“多重索引+时间戳校正”源包排序算法能够实现源包计数有不完整多次循环和无循环情况下的正确源包提取与排序还原, 同时相较于传统的“源包时间码+源包计数+虚拟信道帧计数”的源包排序算法, 所需的处理时间缩短 (约 3%), 内存的占用量减小 (约 28%). 这是由于传统处理算法是对帧数据本身进行操作的, 本文所提算法是对源包的多重索引进行操作, 而源包多重索引的长度明显小于帧长, 所以在处理时间和内存占用量方面有性能优势.

表 3 原始数据文件源包排序处理的正确率

卫星	DAMPE	QUESS	HXMT
原始数传数据文件数 (个)	2 107	862	2 378
包含多种异常联合发生情况的文件数 (个)	22	4	39
传统排序方法的文件处理正确率 (%)	92.45	98.72	89.49
本文排序方法的文件处理正确率 (%)	99.85	99.88	99.79

表 4 两种源包排序处理算法的性能比较

数据中有源包计数循环	本文方法			传统方法		
	处理时间 (s)	占用内存 (MB)	排序是否完全正确	处理时间 (s)	占用内存 (MB)	排序是否完全正确
有 3 次循环	9.84	1.13	是	10.18	1.61	否
连续无循环	10.03	1.19	是	10.36	1.66	是

使用本文提出的算法对 HXMT 卫星的不同大小的同类型数据进行源包提取与排序处理 (运行环境同上), 测试数据大小分别截取或者拼接成 1 GB 到 12 GB 的大小, 记录处理时长和最大内存占用量等性能指标, 测试结果如图 8 所示, 虚线条代表处理所用的时长与数据量之间的关系, 实线条代表算法的最大内存占用量与数据量之间的关系.

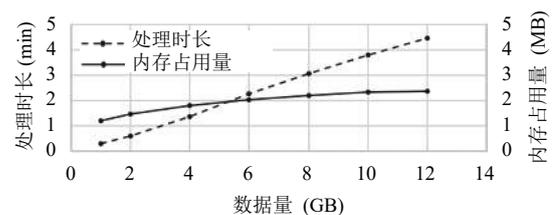


图 8 源包排序处理的时间与内存占用量

结果表明,处理时长随着数据量增加而有明显增长,因为源包排序所占用的时间随着数据量的增加而成倍增加,程序的内存占用量随着数据量增加变化不明显,因为而由于采用索引进行处理,而非数据本身,索引本身占用内存不大,主要是程序运行占用的部分,随着数据量的增加变化不大。

5 总结

“多重索引+时间戳校正”源包排序处理算法,可以解决源包时间码秒部与毫秒部进位不一致导致的“跳秒”问题、两次地面站接收数据之首尾源包不完整导致的丢包问题、源包计数深度不足导致的源包计数循环问题、GPS非定位状态导致源包时间码缺失的问题、载荷异常导致累积时间码异常归零的问题等多种排序困难,是一种相对完备的自动解决方案。经过对HXMT、QUESS、DAMPE三个卫星的大量数据进行实验验证,结果表明该算法可以将卫星原始数据文件的源包排序整体正确率明显提升,同时有效提升处理效率。此外,该算法从原理上来说适用于所有卫星的源包排序处理,具有通用性和推广价值。

参考文献

- 1 马苗,朱岩.基于CCSDS标准的卫星数据处理软件设计.电子设计工程,2015,23(1):16-20.[doi:10.14022/j.cnki.dzsjgc.2015.01.006]
- 2 杜冬梅.航天器CCSDS遥测数据包处理方法研究.电讯技术,2007,47(2):176-180.[doi:10.3969/j.issn.1001-893X.2007.02.042]
- 3 张利萍.CCSDS在我国航天领域的应用展望.飞行器测控学报,2011,30(S1):1-4.
- 4 梁磊,易予生,陈春亮.CCSDS分包遥测在航天器通信中的设计及应用.电子科技,2018,31(4):77-80.[doi:10.16180/j.cnki.issn1007-7820.2018.04.020]
- 5 杨奕飞,丛波,李强.CCSDS标准在新一代航天测控系统中的应用研究.电讯技术,2007,47(5):166-168.[doi:10.3969/j.issn.1001-893X.2007.05.043]
- 6 吕良庆,周玉霞.欧洲航天标准化合作组织的包应用标准模型化设计思想.中国航天,2020,(9):25-28.[doi:10.3969/j.issn.1002-7742.2020.09.005]
- 7 杜国明,陈逢田.CCSDS遥测包处理分析.航天器工程,2007,16(5):110-114.[doi:10.3969/j.issn.1673-8748.2007.05.020]
- 8 龙吟.一种源包和虚拟信道相结合的AOS调度算法.航天器工程,2013,22(4):77-81.[doi:10.3969/j.issn.1673-8748.2013.04.014]
- 9 孙小涓,石涛,胡玉新,等.基于流式计算的空间科学卫星数据实时处理.计算机应用,2019,39(6):1563-1568.[doi:10.11772/j.issn.1001-9081.2018122602]
- 10 刘秀秀.基于CCSDS标准的卫星数据地面实时测试系统的设计[硕士学位论文].北京:中国科学院大学(中国科学院国家空间科学中心),2017.
- 11 贾铂奇,姚焯,李悦.复接帧与载荷包联合处理架构的设计与实现.信息技术,2016,(4):48-51,55.[doi:10.13274/j.cnki.hdzt.2016.04.012]
- 12 权爽,张静,马千里,等.嫦娥四号任务中继通信链路地面测试验证及数据处理.航天器工程,2019,28(4):109-115.[doi:10.3969/j.issn.1673-8748.2019.04.017]
- 13 肖鹏,李景山.遥感卫星数据预处理功能扩展模型设计与实现.遥感信息,2020,35(6):76-84.[doi:10.3969/j.issn.1000-3177.2020.06.012]
- 14 侯仰拴,石涛,胡玉新.HJ-1C卫星实时软件帧同步算法研究与实现.雷达学报,2014,3(3):326-331.[doi:10.3724/SP.J.1300.2014.13053]
- 15 赵海升,葛明玉,李正恒,等.一种天文卫星数据预处理方法.天文研究与技术,2017,14(3):376-381.[doi:10.14005/j.cnki.issn1672-7673.2017.03.002]
- 16 赵海升,李承奎,贾淑梅,等.天文卫星下行数据处理系统的规划.天文研究与技术,2019,16(3):366-371.[doi:10.14005/j.cnki.issn1672-7673.20180713.001]
- 17 元勇,王文平,潘腾.HXMT卫星时间同步系统设计.计算机测量与控制,2018,26(10):160-164,215.[doi:10.16526/j.cnki.11-4762/tp.2018.10.035]
- 18 孙小涓,石涛,李冰,等.空间科学卫星数据快速处理方法.计算机工程与科学,2018,40(8):1351-1357.[doi:10.3969/j.issn.1007-130X.2018.08.003]