

MA Fuli, JI Zhen, YANG Xiaoyan, YU Qinsi, LI Bing, LUO Wentian, TONG Jizhou. Design of the EP Satellite Data Preprocessing and Distribution Service Platform (in Chinese). *Chinese Journal of Space Science*, 2025, **45**(4): 1098–1113. CSTR:32142.14.cjss.2024-0092. DOI:10.11728/cjss2025.04.2024-0092

EP 卫星数据预处理与分发服务平台设计与实现*

马福利¹ 纪 珍¹ 杨晓艳² 于勤思¹
李 冰² 罗文甜³ 佟继周¹

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院空天信息创新研究院 北京 100190)

3(中国科学院计算机网络信息中心 北京 100083)

摘 要 随着空间探测技术发展和星地通信手段的增加, 为了能够第一时间对重要天文事件进行分析, 空间天文卫星对下行探测数据的实时性和处理效率的要求日益严格. 本文针对爱因斯坦探针 (EP) 卫星的数传 X-band、遥测 S-band、甚高频 (VHF) 数据和北斗短报文多种类型数据高时效性处理与分发需求, 开展了 EP 卫星数据预处理与分析平台关键技术研究, 设计了载荷 0 级数据处理、1 级数据处理以及基于观测组织的数据完整性判读等关键算法; 针对 VHF 数据和北斗短报文中天文警报信息的高时效性发布需求, 设计了不同优先级的数据预处理和分发流程, 最终采用分层架构设计的方法研制了 EP 卫星数据预处理与分发服务平台, 具备对 EP 卫星下行的各类数据进行自动、高效和准确预处理、管理以及分发的能力. 系统运行验证结果表明, 该平台处理的各级各类科学数据具有较高的准确性, 能够满足 EP 卫星数据预处理的时效性要求, 且具有良好的可扩展性.

关键词 北斗短报文, 空间科学卫星, 数据预处理, 数据分发

中图分类号 TP75

Design of the EP Satellite Data Preprocessing and Distribution Service Platform

MA Fuli¹ JI Zhen¹ YANG Xiaoyan² YU Qinsi¹
LI Bing² LUO Wentian³ TONG Jizhou¹

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

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

3(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083)

Abstract Along with the development of space exploration technology and the communication methods between satellite and the Earth, the requirements for the efficiency of data transmission and process-

* 中国科学院战略性先导科技专项项目资助 (XDA15040200)

2024-07-02 收到原稿, 2024-12-05 收到修定稿

©The Author(s) 2025. This is an open access article under the CC-BY 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>)

ing in astronomical missions are increasing day by day. The article introduces the design and implementation of a data preprocessing and distribution platform, which focuses on the processing and distribution requirements of various types of data in the Einstein Probe (EP) mission. The EP mission downlinks data via four different channels, including X-band, S-band, VHF channel, and Beidou channel. For each data channel, the data processing methods and efficiency requirements are different, such as the VHF and Beidou data needs to be processed in real-time, the X-band data needs to be processed comprehensively by different types of data packets. During the development process, the key technologies of the EP satellite data preprocessing and analysis platform have been studied. At the meanwhile, several key algorithms have been designed, such as the Level 0 data processing method, Level 1 data processing method and data product integrity interpretation based on observation organization. Especially, regarding the high timeliness requirements for the release of astronomical alert information in VHF data and Beidou short messages, different priority data preprocessing and distribution workflows were designed. The EP mission data preprocessing and distribution service platform adopts a hierarchical architecture design. The application results indicate that the platform has the ability to preprocess, manage and distribute various types of data in EP mission automatically, efficiently, and accurately. The operational process of the platform supports the different processing flows for different data products and simultaneous data distribution and acquisition for multiple users. The platform has good scalability and efficiency. The various levels and types of scientific data products processed by the platform have high accuracy, can satisfy the application requirements of EP scientific teams.

Key words Beidou short message, Space science satellite, Data preprocessing, Data distribution

0 引言

时域天文学作为天文学的一个重要分支,在美国天文和天体物理十年规划(2012—2021年)中被列为孕育着发现机遇的5个前沿领域之一^[1].随着空间科学探测技术的不断进步和信息处理能力的显著提高,对宇宙的认知正从静态图像转向动态过程,使得对天体的快速时变观测成为可能,这一转变不仅为探索和认知宇宙和相关物理规律提供了新的机遇,而且通过观测天体辐射的变化,例如暂现源和爆发天体等现象,揭示了天体辐射机制、历史演化等的物理信息.特别是随着引力波天文学的兴起^[2],时域天文学已经进入了一个多波段、多信使联合观测的黄金时代.

面向时域天文学的小型天文探测卫星,爱因斯坦探针(EP)卫星是中国科学院空间科学战略性先导专项(二期)规划的任务之一,法国宇航局、欧洲空间局参与合作.该卫星任务对软X射线天空进行高频率的时域巡天观测和暂现源快速后随观测,其目标为发现高能瞬态天体及其暴发现象,监测天体活动情况,对暂现源进行深度后随观测,以便更精确地定位暂现源,获取高质量X射线流量与光谱数据^[3].

EP卫星共有两台科学载荷,宽视场软X射线微

孔龙虾眼望远镜(WXT)和后随X射线望远镜(FXT).WXT载荷主要实施巡天观测,对X射线源进行触发与定位,并针对重要暂现源,触发星上自主后随观测.FXT载荷对触发源进行深度后随观测,对机遇目标进行高精度定点观测,兼具高灵敏度巡天观测. EP卫星共有巡天监测、自主后随观测、机遇目标观测和定标观测模式^[3]四种科学观测模式.在轨期间主要采用巡天观测和定点高精度后随观测相结合的工作模式,同时开展星上实时暂现源搜寻.WXT探测到X射线暂现源或爆发源时,星上实时计算源的位置坐标、流强及可能的粗略能谱.在满足触发阈值条件时,启动星上自主转向,迅速将FXT的光轴指向暂现源目标并达到姿态稳定.同时,通过中国北斗卫星导航系统的短报文系统和法国的甚高频(VHF)基站网实时上传发现的暂现源位置以及强度等基本信息至地面,用以发布暂现源天文警报,引导其他设备开展多波段后随观测.

EP任务具有巡天观测、普通机遇目标(Target of Opportunity, ToO)观测、重要ToO观测以及后随观测等观测类型,这些观测类型分别对应4个观测优先级,从高到低分别为重要ToO观测、后随观测、普通ToO观测、巡天观测,能够保证根据观测执行情

况及观测优先级快速调整观测任务计划,从而支持对暂现天体和新科学现象的发现与追踪.同时,EP任务定义的多类天文警报信息,包括WXT载荷的A,B,C三级警报信息,FXT载荷的关键数据、图像数据、能谱数据7类警报信息.

EP卫星的数据基于数传、遥测、VHF系统及北斗系统4个传输信道数据下行至地面段,下行数据都会汇集至空间科学数据中心,如图1所示.空间科学数据中心负责研制的EP卫星数据预处理与分发服务平台是EP卫星地面支撑系统的重要组成部分,是衔接卫星与科学应用系统和科学用户的关键环节,主要负责进行EP卫星下行原始数据的预处理、数据存储管理以及数据分发与发布服务.

目前,针对卫星地面数据处理与分发流程开展了广泛研究,面向多类型数据的高时效处理需求,多采用模块化的多任务处理流程,其中流式数据处理框架和科学工作流被普遍应用于卫星数据处理场景中.面向流式处理的Spark处理框架成功应用于天文图像数据处理中^[4],但是由于实际卫星下行数据存在数据延迟和乱序的场景,以及需要多次传输数据的拼接等处理需求,通过实时流的方式无法满足多次下行数据拼接以及完整性的要求^[5-10].此外,科学工作流框

架^[11,12]也被普遍应用在卫星地面系统中,支持多卫星任务、多载荷数据源的大规模数据高度复杂处理流程,但是在应对通过北斗短报文和VHF下行的天文警报数据处理的高时效性要求时有很大的局限性.

EP预处理与分发平台每天需要处理数传、遥测、VHF数据和北斗短报文4个通道下行的数据,数传信道包含卫星平台及载荷所产生的所有数据,例如事例数据、能谱数据以及工程等多种类型的原始数据,遥测信道包含实时工程数据,VHF数据和北斗短报文包含天文警报数据与部分载荷工程数据.卫星数据处理以及数据加工处理过程需要依据数据分级定义将原始数据处理成不同级的数据,数据处理过程复杂且VHF数据和北斗短报文中的警报数据的高时效性要求以及基于观测组织的数据完整性和连续性要求都对EP数据预处理与分发平台的设计提出了较高的要求.

本文在中国相关研究的基础上^[13,14],采用面向不同时效性以及观测优先级的分层架构设计和标准化设计思路,设计和研制了EP卫星数据预处理与分发平台,通过0级数据分离与提取处理算法、FXT载荷1级数据处理算法以及基于观测组织的数据完整性判读算法等一系列关键算法以及流程的设计,优化

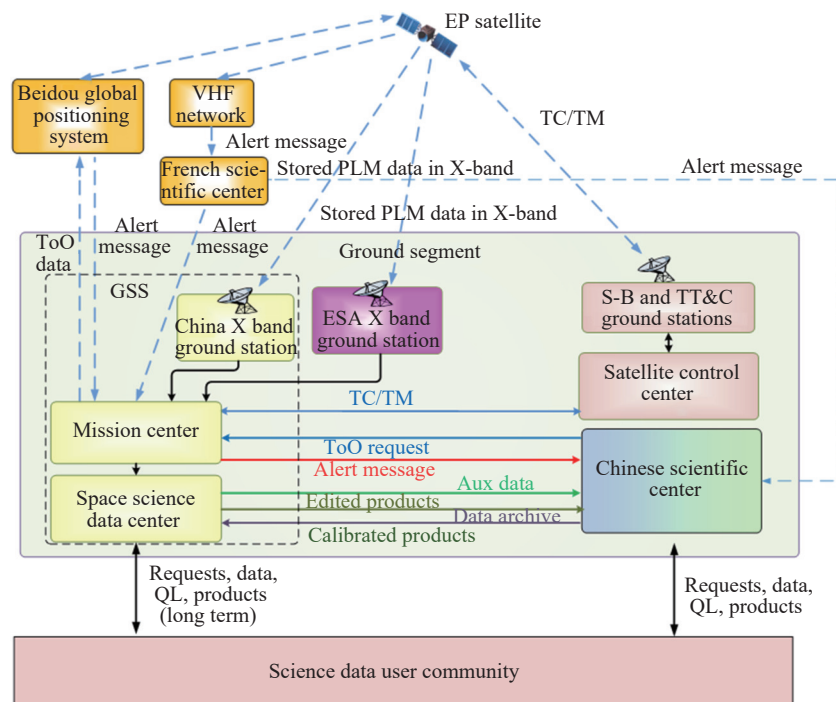


图1 EP卫星地面段运行框架

Fig. 1 EP ground segment architecture

了系统运行模式,提升了关键数据的预处理与分发效率.

1 任务特点与难点分析

EP 卫星是中国首颗同时具备数传 X-band 信道、遥测 S-band 信道、VHF 数据和北斗短报文四种信道数据下行能力的空间科学卫星任务,数据预处理与分发平台在接收到各类原始下行数据后需要基于 EP 卫星数据分级定义以及数据格式规范进行传输帧虚拟信道分离、源包提取、拼接去重以及排序等预处理操作生成各载荷 0A, 0B, 0C, 0D 和辅助数据;在 0 级数据的基础上对载荷科学数据进行科学数据解包、时间码解算和坐标系转换等处理,剔选出和科学分析无关数据,并添加该次观测相关轨道、姿态、数据筛选文件 (Make Filter File, MKF) 以及探测器状态数据等信息形成基于观测任务组织的 1 级数据.此外,该平台还需要对生产出的各类数据进行数据的连续性、完整性等质量信息进行审核,对通过审核的数据进行入库存储管理,并依据数据分发策略分发给科学应用系统用户.由于 EP 卫星四类下行通道的传输带宽以及传输时效性不同,因此每个信道下行的数据内容、数据处理级别以及数据处理与分发效率要求也不尽相同,给数据预处理与分发平台的设计带来了一定的困难.

1.1 下行数据种类多且时效性要求高

根据卫星数据下行信道不同,EP 卫星的原始数据分为数传原始数据、遥测原始数据、VHF 数据和北斗短报文四类.数传原始数据通过 X-band 下行,经 X-band 地面站接收、解调、帧同步、解扰、解码后生

成的由传输帧组成的数据文件,覆盖卫星平台和有效载荷探测的所有科学数据和工程数据.遥测原始数据通过测控信道下行,经 S-band 地面站接收、解调、帧同步、解码、解扰后生成的由传输帧组成的数据文件,主要包含稀疏化后的卫星平台和有效载荷关键工程参数数据.VHF 数据是星上自主触发的暂现源警报信息和卫星平台及载荷重要遥测信息,通过法国 VHF 天线阵实时接收,法国科学中心逐条解析为数据源包,转发至数据预处理与分发平台进行警报解析与拼接.北斗短报文是星上自主触发的暂现源警报信息及重要遥测信息,由北斗卫星转发,北斗测控站准实时接收,并转发至数据预处理与分发平台进行警报解析.数据处理与分发平台在接收到四类信道下行的原始数据后需要依据表 1 中的描述生成不同级别的数据.

此外,由于 VHF 数据和北斗短报文包含重要的暂现源警报信息,这些快速下行的警报信息在经过科学家分析后可以通过伽马射线暴协同观测网 (Gamma-ray Coordinates Network, GCN) 对外快速发布警报信息,引导国内外其他天文设备开展后随观测,因此数据处理与分发平台需要在接收到警报数据后,10 s 内完成天文警报数据处理与生成,并发送至科学应用系统,在进行平台系统设计时需考虑警报的高优先级需求.

1.2 数据完整性判读准确性要求高

面对广泛的观测空间、天文事件的突发性、有限的卫星观测时长以及前沿问题研究需求之间的隐形矛盾,科学家需要基于有限的观测资源,合理制定科学观测计划,优先保证对关键科学目标进行及时、持

表 1 EP 卫星 0 级和 1 级数据定义
Table 1 Definition of EP Level 0 and Level 1 data product

数据级别	数据级别定义
Level 0A	对卫星单次过站的数传X-band原始数据、遥测S-band原始数据、北斗短报文以及VHF数据进行解帧、源包提取、验证排序后生成的按应用过程标识符 (Application Process Identifier, APID) 拆分的源包数据
Level 0B	在0A级数据的基础上,融合多次数传接收数据,对数据源包进行拼接、去重复后形成按小时组织,按 APID拆分的数据
Level 0C	在0A/0B级科学数据的基础上,进行拼接、去重复、源包解析后生成按观测号组织的科学数据帧数据
Level 0Q	在0B/0C级科学数据的基础上,对载荷科学数据进行粗略标定后生成的光变曲线、曝光图以及能谱数据
Level 1	FXT载荷: 在0C级科学数据的基础上,进行科学数据解包、时间还原和坐标系转换,对同一观测号的数据进行再组织,剔选出和科学分析无关数据 WXT载荷: 在0B级科学数据的基础上,进行解包、科学数据帧解析、物理量转化,按观测号组织,同时整合载荷状态数据、姿轨信息等得到的数据

续、有效的观测. 因此, EP 任务中会出现低优先级观测被高优先级打断的情况, 例如重要 ToO 观测打断巡天观测, 按照时间顺序不同观测任务的数据将相互交叉. 而进行数据分析的时候需要基于不同的观测任务进行针对性研究, 科学数据需要基于观测任务进行组织, 并进行观测数据连续性和完整性判断, 但是依据观测优先级和数据顺序下行的实际执行情况, 一次下行的数据中可能包含相互交叉的多个观测任务的数据, 导致一次观测任务数据完整性的判读逻辑非常复杂, 既要区分不同观测号数据, 判读单个观测号的数据连续性, 还要判读前后观测号之间的连续性等. 以 FXT 载荷 1 级数据为例, 如图 2 所示, 包括 FXTa 和 FXTb 两个模块, 巡天观测 (obsID1) 的优先级较低, 在数据流中被重要 ToO 观测 (obsID2) 和后随观测 (obsID3) 打断, 地面顺序接收数据后需要依据观测计划以及观测优先级进行 obsID1 观测是否结束以及 obsID1 观测数据是否完整的判读.

2 EP 卫星数据预处理与分发平台设计

2.1 体系架构设计

基于面向业务分层设计思想进行 EP 卫星数据预

处理与分发平台的架构设计, 自底向上包括基础设施层、基础软件层、数据层和业务应用层四层结构, 并对平台系统研制标准规范进行设计, 例如 EP 卫星 0 级数据格式规范、1 级数据格式规范、辅助数据格式规范、平台日志管理规范以及约束各个模块间接口的接口规范等文件, 保证整个平台架构接口与数据的规范性和一致性, 具体设计如图 3 所示.

基础设施层是整个平台运行的基础, 主要包括计算机、网络安全、网络通信以及存储等各类硬件设备, 是开展数据存储、数据处理以及数据传输的物理基础, 此外还包括虚拟化软件, 通过创建虚拟机的方式实现资源的高效利用和灵活管理.

基础软件层是在基础设施层的基础上为上层业务应用提供基础软件运行支撑环境, 包括操作系统、数据库以及 MQ 和 MESOS 等中间件软件. 为了支撑 EP 卫星任务, 本平台采用开源 OpenEuler Linux 操作系统, 该操作系统支持多种硬件架构, 具备良好的兼容性和可扩展性; 采用 MYSQL 数据库进行数据存储业务的应用, 另外基于 Elasticsearch 数据库实时查询和分析的高性能表现, 采用 Elasticsearch 数据库进行缓存处理数据的存储与查询分析; 采用 Kafka 消息队列中间件用于进行业务系统内外部消息接口的传输; 在整体计算资源调度方面, 采用 MESOS

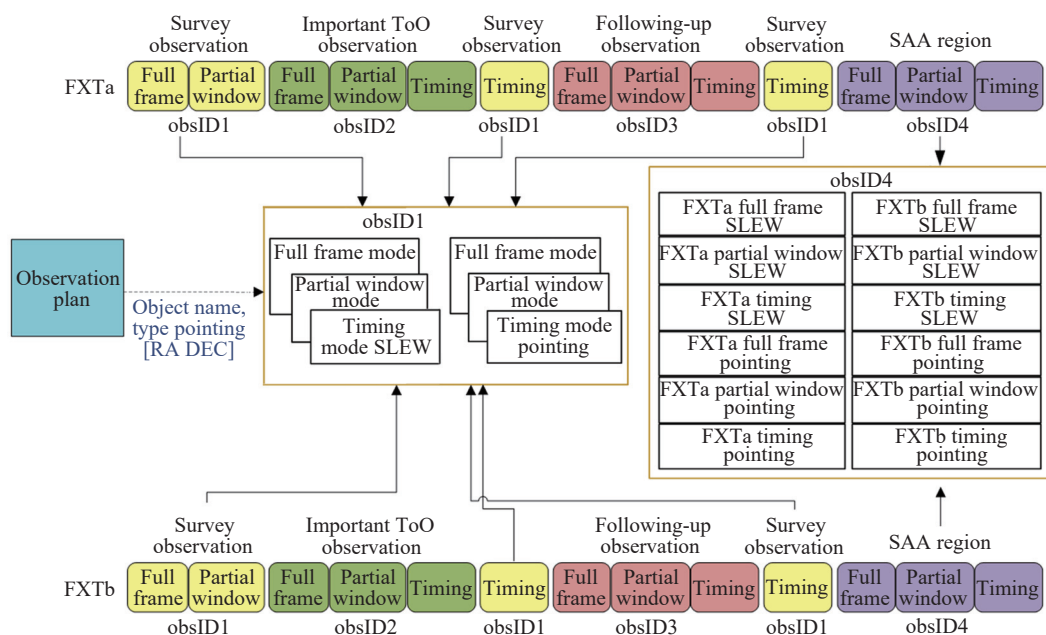


图 2 FXT 载荷观测数据组织结构

Fig. 2 FXT observation data structure

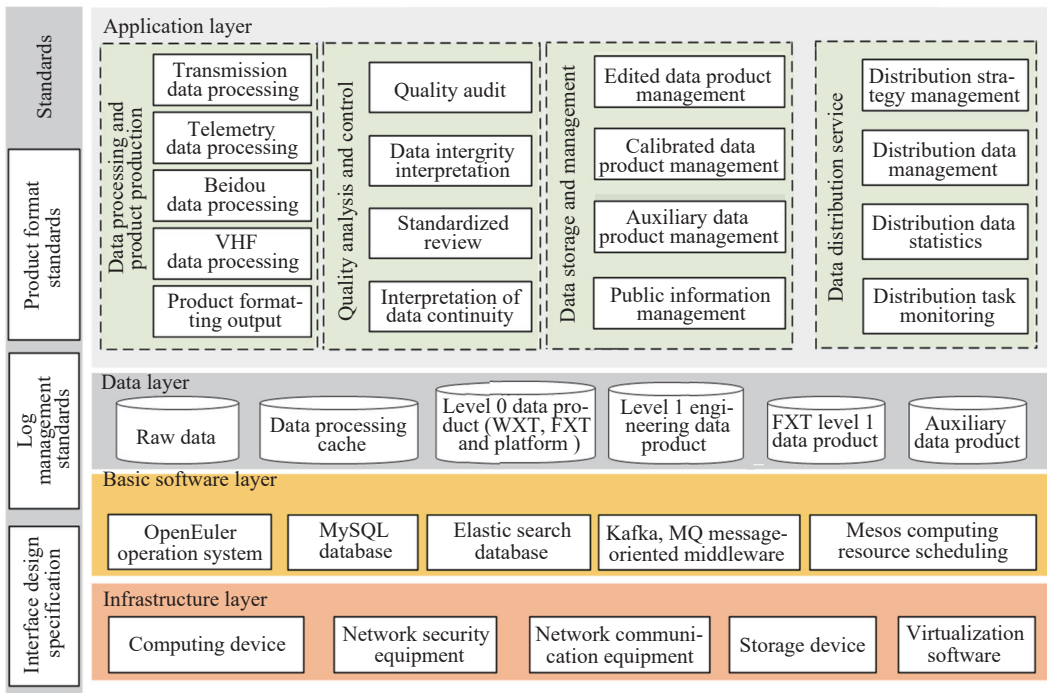


图 3 EP 数据处理与分发平台分层架构

Fig. 3 EP data processing and distribution platform architecture

资源管理软件建立统一的计算资源池, 依据计算任务的类型以及优先级统一进行各类计算资源的调度。

数据层负责管理和维护数据处理与分发平台所有业务数据的存储、组织和访问, 这一层主要建立在 MySQL 数据库和 ElasticSearch 数据库之上, 按照 EP 卫星数据分级定义、数据组织模型、数据结构与依赖关系来组织数据资源, 确保数据的逻辑性和一致性, 包括卫星下行的遥测、数传、VHF 数据和北斗短报文等原始数据、经过处理的 0 级、1 级、辅助和数据以及数据处理业务过程中的缓存数据。通过精心设计的数据层, 可以有效提升数据处理效率, 降低数据管理复杂性, 为整个信息系统成功实施提供坚实的数据基础。

业务应用层面向数据处理与数据生产、质量分析与控制、数据存储与管理以及数据分发服务 4 个方面的业务需求, 结合 EP 卫星任务特点开展具体业务软件研制, 其中数据处理与生产主要负责依据处理流程设计, 进行各类各级别卫星数据处理以及标准格式化数据生产; 质量分析与控制负责进行数据连续性、数据完整性以及规范性的审核; 数据存储与管理、数据分发服务负责 EP 卫星各类数据的存储管理以及面向科学应用系统的数据分发。

2.2 平台流程设计

2.2.1 数据预处理与分发平台运行流程设计

设计的基于数据驱动的 EP 卫星数据预处理与分发流程如图 4 所示, 为了兼顾北斗短报文以及 VHF 数据中暂现源警报信息处理的高时效性需求, 对这两种数据涉及的数据处理任务进行了高优先级的设置, 保证在计算资源调度服务中优先获取计算资源, 实时开始数据处理作业任务。平台在接收到卫星下行的遥测数据、数传数据、北斗短报文以及 VHF 数据后, 依据数据类型驱动不同的数据处理作业任务。

(1) 针对北斗短报文以及 VHF 数据, 完成数据处理之后直接进行数据分发, 通过数据获取客户端发送至科学应用系统, 同时进行数据的存储与管理。

(2) 针对遥测数据, 在完成遥测数据生产后首先进行数据规范性审核, 确保数据格式正确, 之后将通过审核的数据分发至科学应用系统, 并完成数据存储与管理。

(3) 针对数传数据, 由于全部科学数据和工程数据都会在该信道下行, 因此该信道数据内容最全、数据体量最大, 在完成数据预处理之后, 对数据分别进行数据质量分析和基于观测组织的数据完整性判读, 确保数据连续性和完整性, 之后将判读完整的数据进

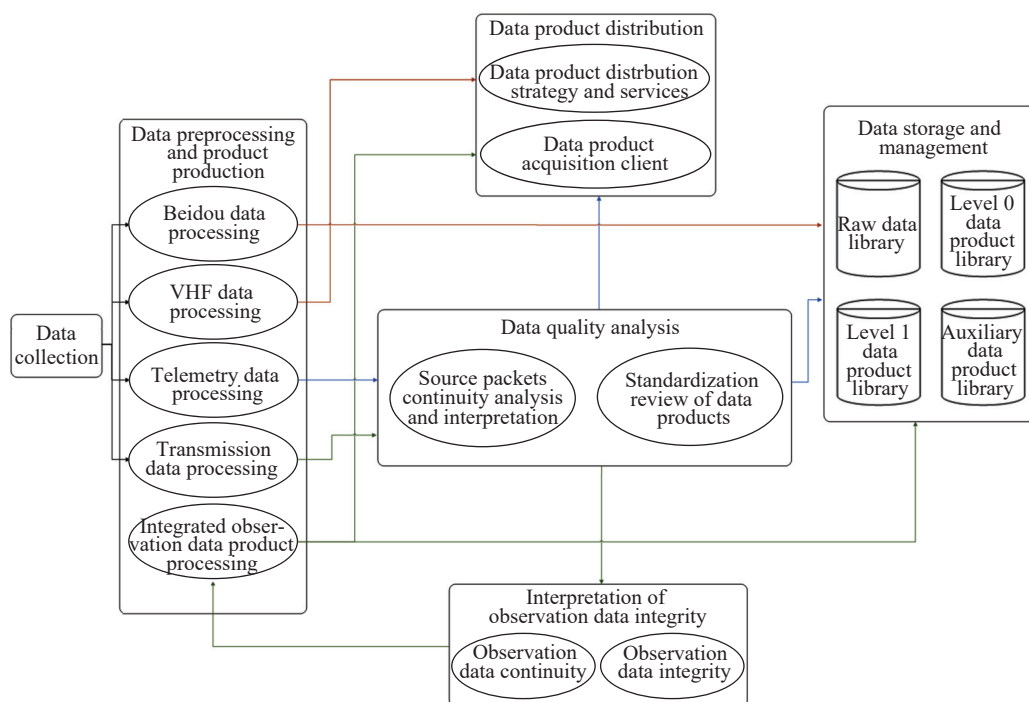


图4 EP卫星数据预处理与分发运行流程

Fig. 4 EP data processing and distribution operation flow chart

行完整观测数据的处理, 最后将完整的观测数据进行数据分发、存储与管理。

2.2.2 基于数据类型和处理依赖关系的数据

预处理流程设计

数据处理与分发平台从卫星运控系统、中国科学中心以及法国科学中心分别接收卫星下行的各类原始数据, 其中数传(X-band)、遥测(S-band)、北斗信道的数据由地面卫星运控系统(见图1中的Mission Center)获取, VHF数据由法国科学中心(图1中的French Science Center)获取处理, 后随观测信息由中国科学中心(见图1中的Chinese Science Center)生成, 最后上述所有数据汇集至空间科学数据中心进行数据生产。不同信道中包含不同的数据, 其处理方法及流程不同, 具体如下。

(1) 中、欧方数传信道下行数传数据包含的数据类型最多, 所有载荷和卫星平台产生的数据均在该信道中下行, 其处理流程最为复杂, 不仅需要按照不同的虚拟信道、数据类型、数据级别开展各载荷0级和1级数据处理, 同时在处理过程中各类数据处理间存在相关依赖。例如, FXT载荷1级数据处理过程中, 不仅需要载荷科学数据的输入, 还需要抽取卫星平台数据中必要的工程参数信息以及观测信息库中的观

测任务计划信息才能组织生成完整的科学数据。

(2) S-band的遥测数据处理仅包括载荷和卫星平台产生的工程参数, 需进行0级和1级工程参数的数据处理, 无须额外处理。

(3) 北斗短报文以及VHF数据中主要为天文警报信息和部分载荷工程参数, 内容包含暂现源的位置、能量以及卫星轨道姿态等其他辅助信息, 对天文警报信息依据观测计划进行拼接, 完成0级以及1级警报数据和工程数据处理。

(4) 当从外部科学应用系统收到后随观测任务的基本信息后还要进行相应1级观测数据的更新处理。

设计基于数据级别与处理依赖关系的EP卫星预处理流程, 其中数传信道的处理流程如图5所示。依据不同类型的数据特征以及EP卫星数据需求, 分别进行独立数据处理与生成, 图5中每一条横线都是一条单独类型的数据预处理流程设计, 且依据数据的输入来源和数据依赖建立不同的流程之间的关联关系, 在具体的计算业务调度过程中, 需要依据关联关系进行处理任务的输入状态判读, 只有当流程所需的所有输入数据全部齐全时才能启动处理作业任务。

2.2.3 数据快速分发与获取流程设计

针对EP卫星数据分发时效性要求高、分发对象

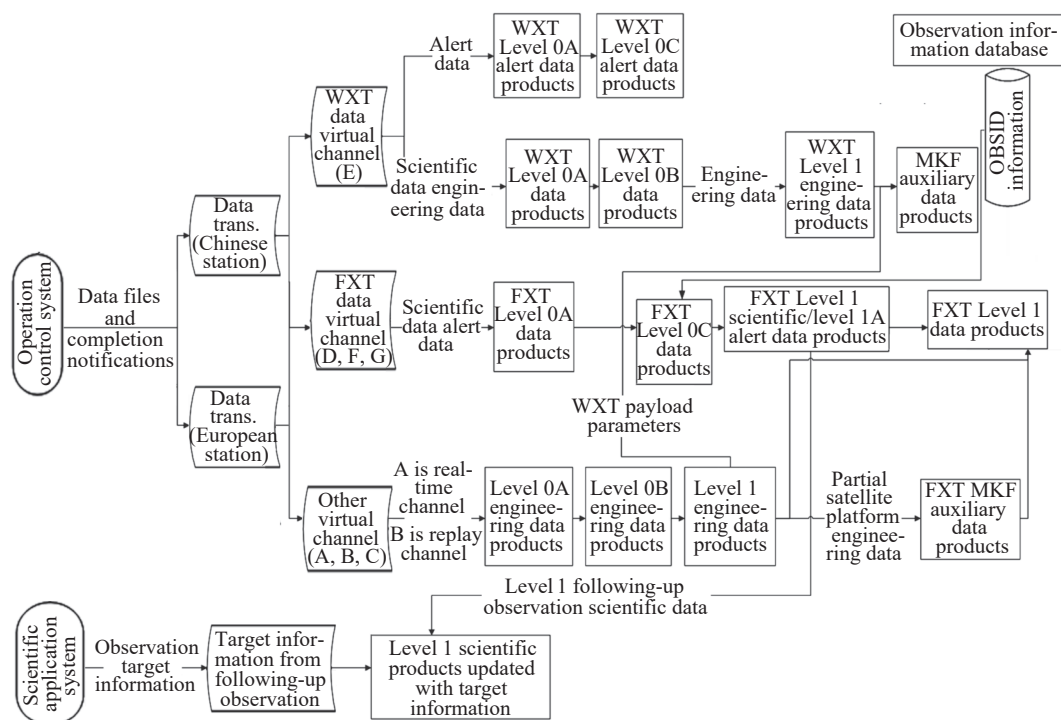


图5 EP 卫星数传信道数据预处理流程设计

Fig. 5 EP X-band data processing flow chart

多、分发数据类型复杂, 分发依赖强等需求, 设计了数据快速分发与获取流程, 具体流程如图6所示。数据分发与获取通过数据分发服务端和数据获取客户端服务实现, 针对EP任务分发需求, 还设置了不同的数据分发策略, 例如针对不同用户设置可分发的数据类型集合(载荷、数据级别、数据类型等信息)以及各类数据的数据分发路径规则, 科学数据与工程数据匹配分发等, 为后续科学分析流程提供正确输入。

数据分发基于数据处理与分发服务平台的数据预处理以及质量分析与控制得到数据分发需求, 依据分发需求中的数据级别以及类型信息, 基于数据分发关联需求进行分发策略匹配, 分发服务端依据分发策略生成面向不同用户的数据分发任务, 并将分发任务推送至数据获取客户端。

数据获取客户端部署用户系统运行环境中, 负责接收分发服务端的数据分发通知, 收到通知后根据配置的用户分发根目录和各载荷各级数据分发路径规则, 将分发通知中包含的数据按照配置规则从分发服务端的存储中复制到本地系统存储中, 完成数据的复制后, 生成该批次的数据分发清单列表文件, 并将数据的获取状态上报给分发服务端, 完成整个数据分发流程的闭环。

3 关键处理算法

3.1 数传 X-band 信道 0 级数据分离与提取处理算法设计

0 级数据分离与提取处理算法是处理数传 X-band 数据的关键第一步, 由于数传 X-band 信道中所包含的科学数据和工程数据类型最全、数据体量较大, 算法首先按照虚拟信道标识符对数传传输帧进行识别与分离, 并对传输帧关键域进行校验, 关键域校验过程中, 如有误码则将传输帧加入到错误数据中, 保存为 error 文件。虚拟信道分离后对传输帧进行连续性校验(首先根据传输帧数据格式定义分别检查版本号、航天器标识符、虚拟信道标识符、信令域、B_PDU 导头规范性, 其次检查 RS 编码是否能够通过校验)。由于传输帧的数据域以数据流的方式存储载荷源包数据, 因此算法约束在提取数据时要求相邻的两个传输帧计数必须连续, 并且设置传输帧缓存区域用于存储历史传输帧数据。

在获取连续的传输帧数据后提取传输帧的数据域, 提取传输帧的头和尾部字段, 将连续传输帧的数据域进行拼接成源包数据流, 并依据源包同步字搜索

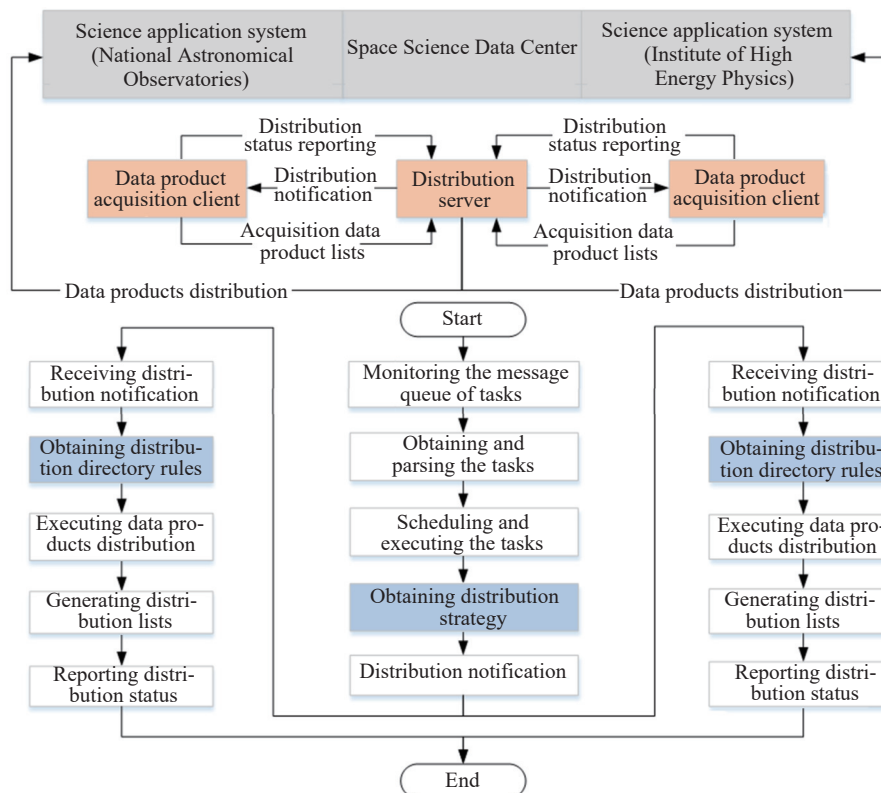


图6 数据快速分发与获取流程

Fig. 6 EP data distribution flow chart

定位和提取完整长度的源包数据,并对源包数据的关键域进行校验(检查版本号、类型、副导头标志、应用过程标识符、分组标志、包序列计数是否符合规范),将校验错误的数据所在传输帧以及源包分别加入 error 文件中,最终经过排序、去重以及数据格式化形成 0 级数据. 0 级数据分离与提取算法如图 7 所示.

3.2 FXT 载荷 1 级数据处理算法设计

FXT 载荷具备全帧、开窗、时变和节能等工作模式,在相同工作模式下又可分为定点、巡天、定标等工作状态,其 1 级数据处理过程中需要将依据载荷工作模式以及工作状态等信息对科学事例数据帧进行区分,同时需要单独提取南大西洋异常区、地影区等特殊区域的载荷数据.因此,设计了基于卫星观测计划、载荷事例类型、载荷工作模式、载荷滤光轮角度以及卫星姿态状态等参数标识的载荷科学事例数据帧识别、分离和提取算法,如图 8 所示.

由于科学用户需要依据每次观测任务开展科学数据分析,因此首先生成基于观测号组织的 0 级观测数据,基于该数据进行 FXT 载荷 1 级处理,具体处理算法如下.

(1) 对基于观测号组织的 0 级载荷源包进行处理,包括提取科学源包、科学数据帧、科学事例,获取其中探测器异常的数据,生成 Abnormal 数据.

(2) 按照科学数据帧格式,解析每一帧包含的 GPS 事例、UTC 事例,结合科学数据帧的本地时间,进行时间解算,生成每一帧的时间.

(3) 根据工程参数对每一帧数据的地球遮挡区 (Earth Occultation, EOC) 状态和南大西洋异常区 (South Atlantic Anomaly, SAA) 状态判断.

(4) 对科学数据帧中的物理事例、offset 事例、强制触发事例、背景物理事例、背景强制触发事例,按照载荷类型、工作模式、工作状态、滤光轮角度、姿态状态进行拆分后,去除其中状态为 EOC 和 SAA 的科学数据帧.根据时间、帧计数、坐标转换生成数据块,根据帧计数连续性生成时间数据块,融合后随观测信息后,从观测计划和后随观测信息中获取目标信息,计算实际指向后生成各类型事例数据.

(5) 对其中的 GPS 事例、UTC 事例根据载荷类型、姿态状态进行拆分,从观测计划和后随观测信息中获取目标信息,计算实际指向,生成时间数据.

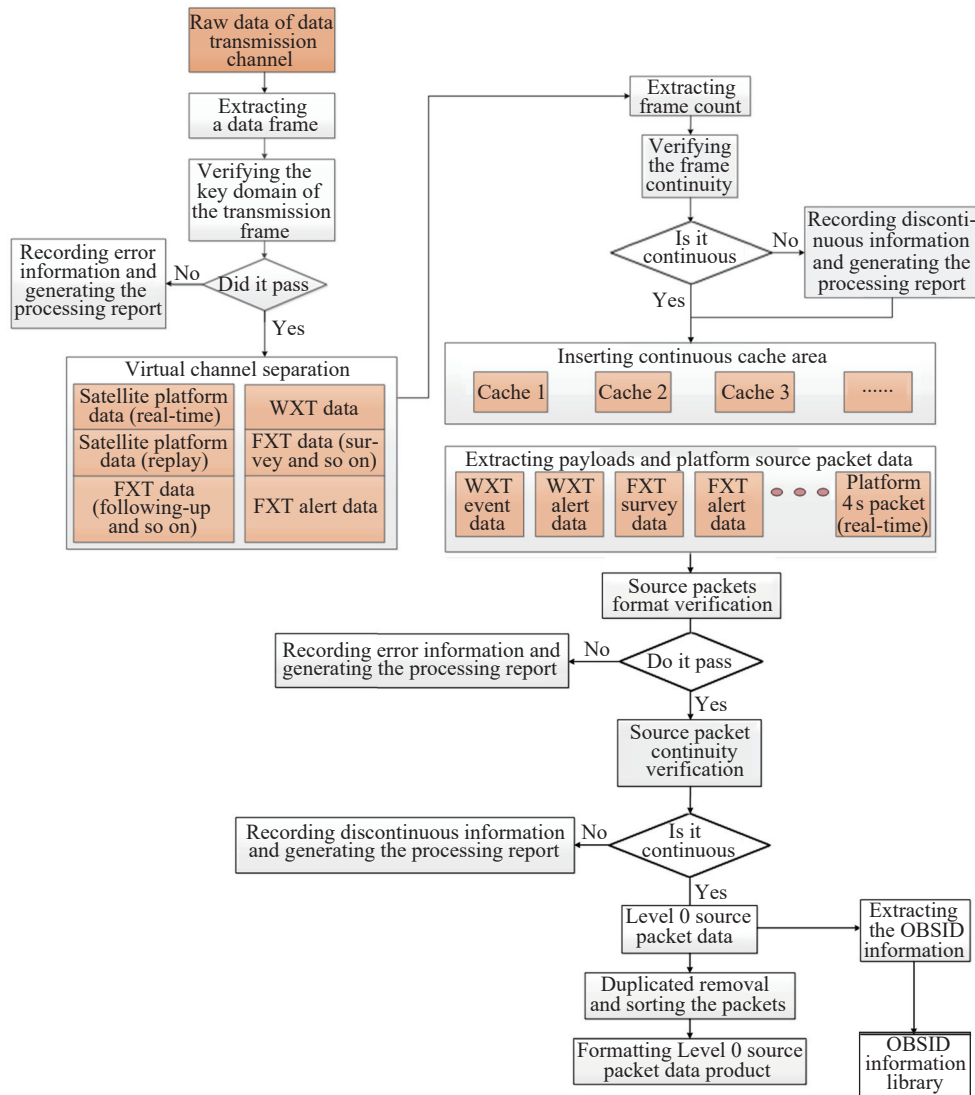


图 7 0 级数据分离与提取处理算法流程

Fig. 7 Level 0 data processing algorithm

(6) 对物理事例、offset 事例根据载荷类型、姿态状态、工作模式进行拆分, 获取状态为 EOC 和 SAA 的科学数据帧, 根据时间、帧计数、坐标转换生成数据块, 从观测计划和后随观测信息中获取目标信息, 计算实际指向后, 生成 SAA 数据和 EOC 数据。

3.3 基于观测任务组织的数据完整性判读算法设计

基于观测组织的数据完整性主要有两方面基本要求: 一是该次观测的数据内容与前、后观测任务的数据连续, 两次观测之间没有数据丢失; 二是该次观测时间段内组织的数据没有数据缺失, 保证基于观测组织的数据中间以及首尾都没有数据缺失。但是由于 EP 卫星设计了巡天观测、后随观测、机遇目标观测、

ToOMM 观测等不同级别的观测任务, 存在依据观测优先级的高低对当前观测进行打断或恢复的情景以及高优先级观测任务数据优先下行, 因此地面系统在进行数据完整性判读时如果发现载荷科学数据帧不连续时, 有两种数据不连续情况为正常场景: 一个是高优先级数据先于低优先级数据下行, 低优先级数据出现不连续; 另一个是执行中的观测任务被高优先级观测任务打断, 在高优先级观测任务结束后, 再继续执行。但是同时也有可能由于星地传输等原因而使得数据丢失, 这给地面系统数据完整性判读带来了挑战。本文设计了基于观测优先级的数据完整性判读算法, 如图 9 所示。

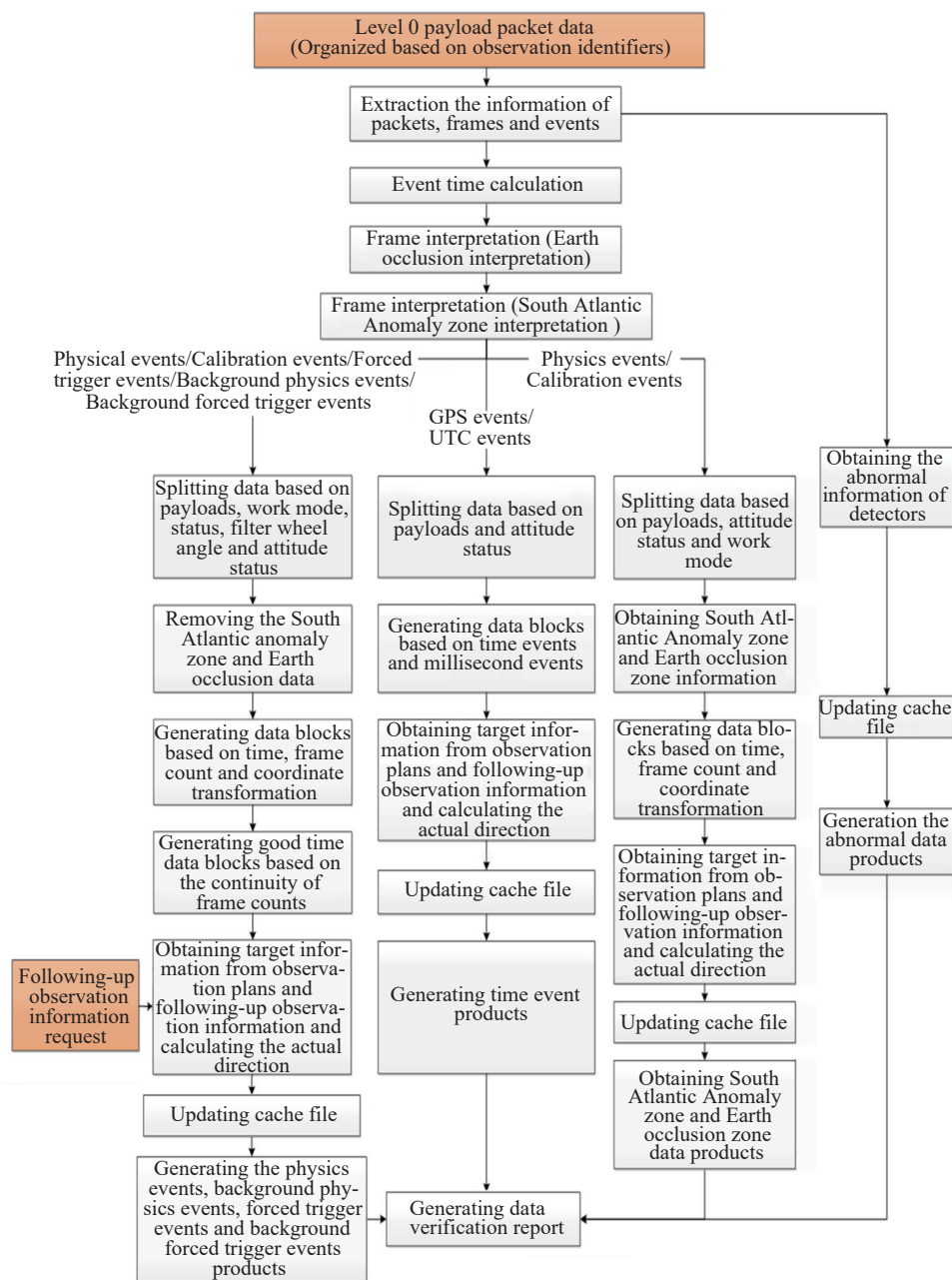


图 8 FXT 载荷 1 级数据处理算法流程

Fig. 8 FXT level 1 data processing algorithm

主要算法步骤如下.

步骤 1 判读传输连续性. 获取同一观测号的所有数据, 依据每次卫星数传数据处理报告, 对高级在轨系统 (Advanced Orbiting Systems, AOS) 帧计数和空间数据系统咨询委员会 (Consultative Committee for Space Data Systems, CCSDS) 包计数进行连续性判读, 当 AOS 帧或 CCSDS 包出现计数不连续时, 判读为观测数据不完整, 结束判读; AOS 帧和

CCSDS 包均连续时, 执行步骤 2.

步骤 2 判读科学数据包连续性. 获取连续 CCSDS 包中的科学数据包, 对其时间和包计数进行排序. 当科学数据包不连续时, 则需判读当前时间段内是否存在高优先级的其他观测数据, 执行步骤 4; 当科学数据包连续时, 执行第 3 步骤.

步骤 3 判读与前后观测号的连续性. 基于数据时间向前找到时间最近的前一个观测号, 判读时间与

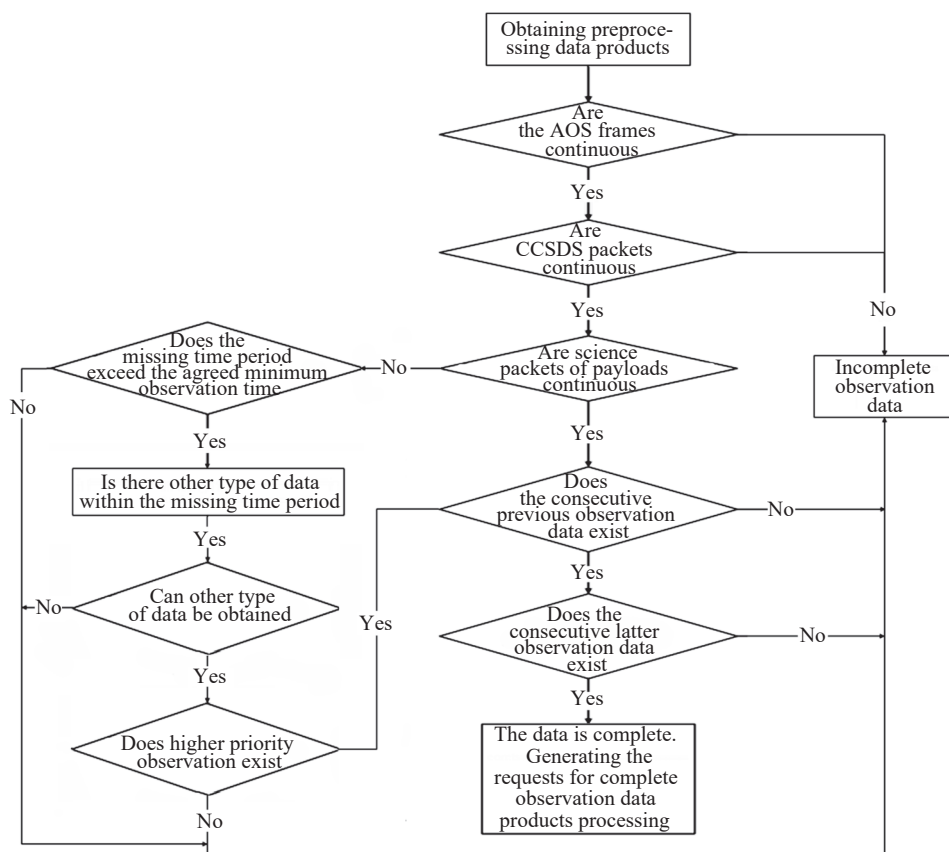


图9 观测数据的完整性判读算法流程

Fig. 9 Observation data completeness analysis algorithm

科学数据包计数是否连续。如果连续, 则向后寻找时间最近的一个观测号, 判读时间与科学数据包计数是否连续。如果与前后观测号均连续则判读观测数据完整, 触发数据生产请求, 判读结束。如果不连续, 判读为观测数据不完整, 结束判读。

步骤4 判读中断时长情况。当科学数据包不连续时, 如果间隔时间小于设计的最短观测时长, 则判读为观测数据不完整。如果间隔时间超过最短观测时长, 则存在该观测号数据未下行或该观测被打断的可能性, 执行步骤5进行判读。

步骤5 判读中断时间内是否存在其他观测数据。根据观测号的时间范围, 检索获取该时间段内的所有观测号数据, 按照时间排序, 逐段判读连续性。如果时间范围内数据连续, 则执行步骤6; 如果获取不到中断时间内的其他观测号数据, 或者该段时间不连续, 则判读观测数据不完整, 结束判读。

步骤6 观测优先级判读。当获取中断时间内的其他观测数据, 解析观测号中的观测类型, 与当前观

测号的观测类型比较, 判读是否为更高优先级的观测类型, 如果为同级或更高优先级观测类型, 则跳至步骤3, 完成与前后观测号连续性的判读。如果为低优先级类型, 则判读为数据不完整, 结束判读。

4 系统设计与实现

EP 卫星数据预处理与分发平台由数据处理与快视处理子系统、数据质量分析与控制子系统、数据存储与管理子系统、数据分发服务子系统和系统业务支撑子系统共5个子系统组成, 如图10所示。

数据预处理与快视处理子系统负责处理EP卫星下行的数传原始数据、遥测原始数据、北斗短报文以及VHF数据, 完成单站接收数据的处理、多站接收数据的排重/拼接、数据跳变的统计/容错、工程参数的解算/转换等处理, 依据EP卫星数据分级定义和数据格式说明生成0级、1级科学和工程数据; 进行轨道根数、星地时差、卫星星历等辅助数据的生产; 对载荷科学数据进行快视处理, 以期快速展现科学探

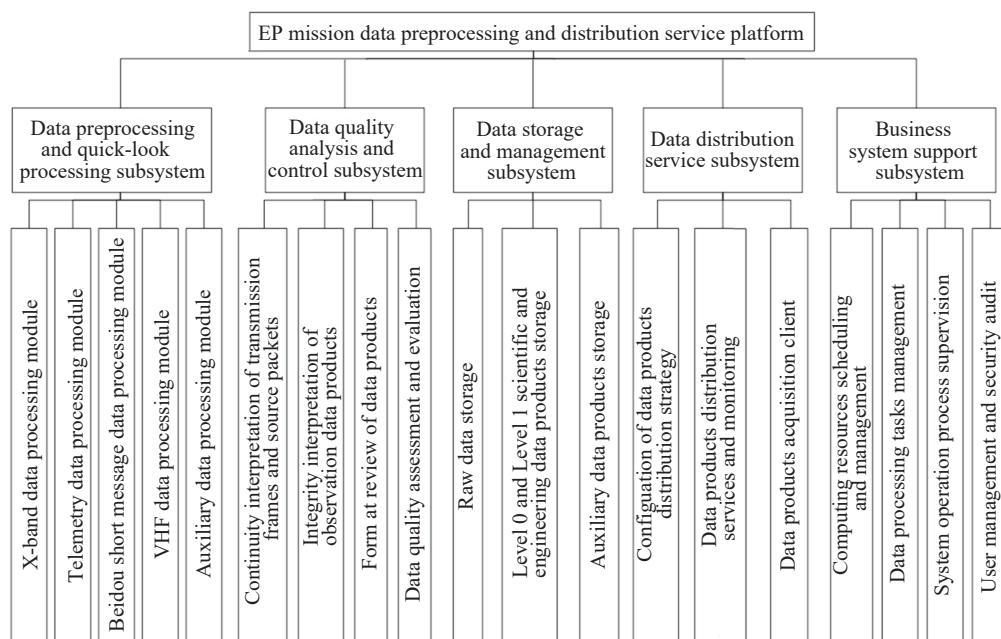


图 10 EP 卫星数据预处理与分发服务平台组成

Fig. 10 EP data processing and distribution platform structure

测对象及其数据结果。

数据质量分析与控制子系统包括数据质量分析和数据质量控制两部分。其中,数据质量分析部分负责对 EP 卫星数传原始数据、0 级和 1 级数据内容的连续性、观测号数据完整性等进行分析;数据质量控制部分负责对 EP 卫星数据格式的符合性、规范性等进行检查,同时对数据质量信息进行综合评价,生成数据质量评估报告,确保数据的正确性、完整性和可用性。

数据存储与管理子系统负责建立卫星数据库、公共信息库,对 EP 卫星工程产生的遥测原始数据、数传原始数据、0 级、1 级的科学数据和工程数据、辅助数据等数据进行集中统一管理。

数据分发服务子系统负责向科学应用系统、卫星数据用户和载荷团队用户分发 0 级、1 级科学数据、工程数据以及辅助数据,并提供数据获取客户端,支持用户通过客户端的方式获取各类、各级以及自定义时间段内的数据。

系统业务支撑子系统负责收集卫星遥测、数传原始数据、北斗短报文以及 VHF 数据,为卫星数据预处理提供数据输入,并进行作业任务的调度,统一调度平台内各类计算资源;对分系统的软硬件状态、业务运行状态进行全方位监视和综合显示;负责建立通

用的授权操作框架,进行系统内、外部用户的统一管理,并提供统一的用户注册和统一认证服务。

5 系统运行验证

EP 卫星数据预处理与分发服务平台部署在中国科学院国家空间科学数据中心的业务集群运行环境中,采用 MESOS 计算资源调度软件进行计算资源的调度分配,该业务集群共计包括 8 台物理机计算节点,合计 544 个 CPU, 2 TB 内存,单机最大内存 250 GB。由于该集群同时还支持怀柔一号 (Gravitational wave high-energy Electromagnetic Counterpart All-sky Monitor, GECAM) 卫星、悟空号 (Dark Matter Particle Explorer, DAMPE) 卫星、夸父一号 (Advanced Space-based Solar Observatory, ASO-S) 卫星以及慧眼 (Hard X-ray Modulation Telescope, HXMT) 卫星等多颗空间科学卫星任务,在计算资源分配策略上设计采用“共用资源池+专用资源”的方式,其中专用资源是固定分配给某一颗卫星任务的计算任务,资源由该卫星的计算任务独享,共用资源池中的计算资源根据每颗卫星任务的计算任务需求进行动态分配。目前分配给 EP 卫星任务 2 台专用 OpenEuler 服务器,合计逻辑 CPU 总数为 256 个,内

存大小为 512 GByte, 单机最大内存约为 250 GByte.

EP 卫星于 2024 年 1 月 9 日成功发射, 每天下行约 10 轨数传数据, 每次下行原始数据量约 5 GByte, 每次下行数据包括卫星平台数据、载荷工程数据和科学数据, 并根据触发情况, 不定期产生警报数据. 此外, 还有每天约 4 轨遥测数据, 1500 余条 VHF 信息以及 300 余条北斗短报文数据. 在接收到卫星下行原始数据后, 该平台每天可生成 0A 级、0B 级、0C 级、1 级和辅助数据 200 余种, 总体量可达 230 GByte.

一般情况下, 北斗和 VHF 信道每次下行的数据约 KByte 量级, 遥测信道每次下行的数据约 MByte 量级, 数传信道每次下行的数据约 GByte 量级, 如表 2 所示. 基于不同下行通道数据处理复杂度以及时效性需求, 平台将北斗及 VHF 数据处理的资源调用优先级设为最高级, 遥测信道数据次之, 数传信道数据为最低. 由于数据处理与分发延迟与原始数

据体量以及处理资源分配及时性相关, 北斗以及 VHF 信道数据量远小于数传信道数据量, 且能够及时获得数据处理资源, 其处理及分发延迟远低于数传数据. 表 2 列出了常规情况下数传、VHF、北斗及遥测信道 0 级数据处理时效性和分发时效性.

卫星发射后该平台持续对接收的原始数据和处理生成的 0 级和 1 级数据进行存储管理, 截至 2024 年 5 月, 对超过 200 余种卫星数据进行统一存储管理, 该平台管理的数据详情见表 3.

截至 2024 年 5 月, EP 卫星已经执行 6000 余次观测任务, 数据预处理与分发平台基于每次观测任务进行 1 级数据的处理与数据生成, 并对基于观测组织的数据进行完整性判读, 数据完整性判读准确性为 100%. 表 4 给出了 5 月 15—18 日的观测数据信息. 例如观测号为 13600006198 的巡天观测任务 (2024-05-17 21:44:03 UTC 至 2024-05-19 18:38:47 UTC)

表 2 EP 卫星各类数据处理与分发时效性
Table 2 EP data processing and distribution performance

数据来源	数据类型	数据级别	数据体量	数据个数	处理延迟	分发延迟
数传数据	科学	0B	1484.8 MByte	5	45 min	23 min
		1级	3686.4 MByte	19	69 min	1.6 min
	工程	0B	146 MByte	185	45 min	24 min
		1级	902 MByte	193	67 min	18 min
	警报	1A	4.74 MByte	11	14 min	11 min
VHF数据	警报	0A	9 KByte	1	21 s	1 s
	工程	1级	864 Byte	1	7 s	1 s
北斗数据	警报	0C	8.43 KByte	1	1 s	1 s
	工程	1级	1 KByte	1	12 s	1 s
遥测数据	工程	1级	2.44 MByte	36	17 s	4 s

表 3 EP 卫星数据管理详情
Table 3 EP satellite data management

序号	数据级别	种类数量	文件个数	体量
1	原始	4	42666	4.4 TByte
2	0A级	87	273044	1.99 TByte
3	0B级	39	480540	1.78 TByte
4	0C级	7	15125	1.24 TByte
5	1级	94	824349	9.22 TByte
6	辅助	4	1427	102.16 GByte

表 4 EP 卫星观测数据完整性判读结果
Table 4 Observation data integrity interpretation results

序号	观测号	载荷	观测类型	观测开始 时间(UTC)	观测结束 时间(UTC)	更新时间 (UTC)	观测是否 结束	工程数据 完整性	科学数据 完整性
1	08500000099	FXTA	重要ToO	2024-05-18 15:00:01	2024-05-18 18:33:44	2024-05-19 10:03:07	结束	完整	完整
2	13600006198	FXTB	巡天观测	2024-05-17 21:44:03	2024-05-19 18:38:47	2024-05-20 05:48:44	结束	完整	完整
3	13600006197	FXTB	巡天观测	2024-05-17 20:07:44	2024-05-17 21:44:03	2024-05-18 14:06:07	结束	完整	完整
4	08500000098	FXTA	重要ToO	2024-05-17 13:35:01	2024-05-17 20:07:44	2024-05-19 09:46:00	结束	完整	完整
5	13600006196	FXTB	巡天观测	2024-05-16 10:24:59	2024-05-17 13:35:01	2024-05-18 10:51:00	结束	完整	完整
6	08500000097	FXTA	重要ToO	2024-05-16 02:23:18	2024-05-16 10:24:59	2024-05-18 08:57:13	结束	完整	完整

执行期间是被观测号为 08500000099 的重要 ToO 观测任务 (2024-05-18 15:00:01 UTC 至 2024-05-18 18:33:44 UTC) 打断, 08500000099 观测任务结束后载荷又重新回到 13600006198 观测任务, 地面系统接收到 13600006198 观测数据中间缺少 2024-05-18 15:00:01 UTC 至 2024-05-18 18:33:44 UTC 时间段的数据, 本文设计的数据完整性判读算法能够准确地将两次观测的数据判读均为完整.

6 结语

随着时域天文学研究的深入, 对探测数据的实时性和处理效率提出了更高的要求, EP 卫星数据预处理与分发服务平台是针对时域天文学领域的发展需求, 特别是为了满足 EP 卫星数据的快速处理与分发而建立的, 本文详细阐述了该平台的设计与实现. 平台的设计采用了分层架构, 包括基础设施层、基础软件层、数据层和业务应用层, 确保了数据处理的高效性和灵活性. 平台针对不同数据类型和观测任务设计了专门的处理流程, 能够快速响应并处理包括数传 X-band、遥测 S-band、VHF 数据和北斗短报文在内的多种数据源. 此外, 平台还实现了基于观测优先级的数据处理与数据组织, 确保了数据的完整性和连续性. 在关键算法方面, 重点介绍了 0 级数据的分离与提取算法、FXT 载荷 1 级数据处理算法及基于观测任务组织的数据完整性判读算法, 这些算法的设计充分考虑了数据的准确性和处理的高效率. 平台还具备数据快速分发的能力, 能够根据不同用户的分发策略

和权限, 实现数据的自动分发.

系统运行验证表明, EP 卫星数据预处理与分发服务平台能够满足 EP 卫星的数据处理与分发时效性要求, 处理延迟和分发延迟均在可接受范围内. 此外, 基于观测组织的数据完整性判读准确性达到 100%, 显示出平台在数据管理和质量控制方面的高效能力. 该平台的成功设计与实现不仅为科学家开展 EP 卫星天文数据处理分析以及科学研究提供了重要的支撑, 也为类似科学任务的数据处理系统设计提供了宝贵的经验和技术参考.

参考文献

[1] National Research Council. New Worlds, New Horizons in Astronomy and Astrophysics[M]. Washington: The National Academies Press, 2010

[2] ABBOTT B P, ABBOTT R, ABBOTT T D, *et al.* GW151226: observation of gravitational waves from a 22-solar-mass binary black hole coalescence[J]. *Physical Review Letters*, 2016, **116**(24): 241103

[3] YUAN Weimin, ZHANG Chen, CHEN Yong, *et al.* Einstein probe: exploring the ever-changing X-ray universe[J]. *Scientia Sinica Physica, Mechanica & Astronomica*, 2018, **48**(3): 039502 (袁为民, 张臣, 陈勇, 等. 爱因斯坦探针: 探索变幻多姿的 X 射线宇宙 [J]. 中国科学: 物理学 力学 天文学, 2018, **48**(3): 039502)

[4] ZHANG Z, BARBARY K, NOTHAFT F A, *et al.* Kira: processing astronomy imagery using big data technology[J]. *IEEE Transactions on Big Data*, 2020, **6**(2): 369-381

[5] GAO Lijing, CHEN Zhimin, WANG Chunmei, *et al.* Real-time processing method of space science satellite downlink

- data based on streaming computing[J]. *Computer Simulation*, 2023, **40**(7): 26-31,290 (高立京, 陈志敏, 王春梅, 等. 基于 Flink 的空间科学卫星数据实时处理方法 [J]. *计算机仿真*, 2023, **40**(7): 26-31,290)
- [6] WANG Jing, WANG Chunmei, ZHI Jia, *et al.* Technology of data processing for high speed data stream of payload[J]. *Computer Engineering and Design*, 2017, **38**(4): 941-945 (王静, 王春梅, 智佳, 等. 面向有效载荷高速数据流的数据处理方法 [J]. *计算机工程与设计*, 2017, **38**(4): 941-945)
- [7] ZHANG Wenbin, WANG Chunmei, WANG Jing, *et al.* Processing method of parameter analysis of payload based on spark[J]. *Computer Engineering and Design*, 2020, **41**(2): 587-591 (张文彬, 王春梅, 王静, 等. 基于 Spark 的有效载荷参数解析处理方法 [J]. *计算机工程与设计*, 2020, **41**(2): 587-591)
- [8] QI Hongyu. Survey of distributed data flow processing framework[J]. *Informatization Research*, 2019, **45**(6): 1-8 (戚红雨. 流式处理框架发展综述 [J]. *信息化研究*, 2019, **45**(6): 1-8)
- [9] KARAKAYA Z, YAZICI A, ALAYYOUB M. A comparison of stream processing frameworks[C]//Proceedings of 2017 International Conference on Computer and Applications (ICCA). Doha, Qatar, 2017: 1-12
- [10] ALMEIDA A, BRÁS S, SARGENTO S, *et al.* Time series big data: a survey on data stream frameworks, analysis and algorithms[J]. *Journal of Big Data*, 2023, **10**(1): 83
- [11] ALKHANAK E N, LEE S P. A hyper-heuristic cost optimisation approach for Scientific Workflow Scheduling in cloud computing[J]. *Future Generation Computer Systems*, 2018, **86**: 480-506
- [12] FENG Yang, ZHANG Guoqiang. Design and implementation of earth observation satellite data receiving scheduling system based on workflow technology[J]. *Radio Engineering*, 2018, **48**(11): 1004-1008 (冯阳, 张国强. 基于工作流技术的遥感卫星数据接收调度系统的设计与实现 [J]. *无线电工程*, 2018, **48**(11): 1004-1008)
- [13] MA Fuli, CHEN Ling, LI Bing, *et al.* Design and implementation of GECAM preprocessing pipeline[J]. *Astronomical Research & Technology*, 2022, **19**(3): 274-282 (马福利, 陈玲, 李冰, 等. GECAM 卫星快速预处理流程设计与实现 [J]. *天文研究与技术*, 2022, **19**(3): 274-282)
- [14] CHEN Yajie, WANG Feng, DEGN Hui, *et al.* The application of elastic search in the massive astronomical data retrieval[J]. *Acta Astronomica Sinica*, 2016, **57**(2): 241-251 (陈亚杰, 王锋, 邓辉, 等. ElasticSearch 分布式搜索引擎在天文大数据检索中的应用研究 [J]. *天文学报*, 2016, **57**(2): 241-251)

作者简介



马福利 男, 1986 年 10 月出生, 中国科学院国家空间科学中心高级工程师, 主要从事卫星地面数据处理技术和大数据计算与存储技术的研究等。
E-mail: mafuli007@nssc.ac.cn