Postprint: A Data Preprocessing Method for Astronomical Satellite Data

Authors: Zhao Haisheng, Ge Mingyu, Li Zhengheng, Jianyin Nie, Song Liming

Date: 2017-09-26T00:00:00+00:00

Abstract

Astronomical satellites acquire data that must undergo a series of operations to generate data products releasable to scientific users, among which data preprocessing constitutes a crucial step and forms the foundation of data release. Taking the Medium Energy X-ray Telescope of the Hard X-ray Modulation Telescope satellite and the Gamma-ray Burst Polarimeter onboard Tiangong-2 as examples, this work elaborates on the detailed procedures and methodologies of data preprocessing. A comprehensive set of data preprocessing steps is proposed, accompanied by strategies for data parsing and time calculation methods, which holds reference value for the development of data processing systems for Chinese astronomical satellites.

Full Text

A Method for Astronomical Satellite Data Preprocessing

Authors: Zhao Haisheng, Ge Mingyu, Li Zhengheng, Nie Jianyin, Song Liming
Affiliations:
1. Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
2. Key Laboratory of Particle Astrophysics, Chinese Academy of Sciences, Beijing 100049, China

Abstract

Data acquired by astronomical satellites must undergo a series of operations to generate data products that can be distributed to scientific users. Data preprocessing is a crucial step in this pipeline, serving as the foundation for data release. Using the Medium Energy X-ray Telescope (ME) on the Hard X-ray Modulation Telescope (HXMT) satellite and the Gamma-ray Burst Polarimeter (POLAR) on Tiangong-2 as examples, this paper describes in detail the preprocessing steps and methods for astronomical satellite data. We propose a standard set of preprocessing procedures and present strategies for data parsing and time calculation, which can serve as a reference for the development of data processing systems for other Chinese astronomical satellites.

Keywords: Data preprocessing; Satellite data; Packet decoding; Time calculation

1. Introduction

Scientific data from satellite payloads are packetized according to predefined formats by the data management system and transmitted to the satellite platform. These data are then downlinked to ground stations via telemetry channels. After simple decoding, the data conform to the Consultative Committee for Space Data Systems (CCSDS) standard [?]. The preprocessing system performs operations such as file splitting and merging to generate data units suitable for further processing. Although different satellite payloads vary in processing details, most require such preprocessing steps [?].

Prior to data analysis, it is necessary to unpack the data, calculate accurate timestamps, and establish a standard processing pipeline. This paper focuses on the Medium Energy X-ray Telescope (ME) [?] on the HXMT satellite and the Gamma-ray Burst Polarimeter (POLAR) [?] on Tiangong-2 to illustrate the preprocessing workflow and methods.

2. Level 1A Processing

Level 1A processing primarily involves unpacking all data and performing initial processing on scientific data. The input consists of minimally processed satellite telemetry data, while the output comprises complete data files ready for further processing. The workflow is illustrated in [Figure 1: see original paper], where Level 0 and Level 1 represent the input telemetry data and output data (releasable products), respectively.

2.1 Data Parsing

Data parsing converts binary data into decimal values according to predefined formats. A single data packet may contain multiple data structures. For example, ME science packets contain 54 identical event structures, while POLAR packets contain 5 different science data structures, 1 trigger packet type, and 1 data packet type. Parsing involves extracting all these structures from each packet. Additionally, certain keywords in packet headers and data structures can be treated as special data structures. These structure definitions are typically designed by the payload manufacturers and implemented by software developers.

We employ a configuration file-based parsing method. The configuration file format is shown in . The fields include: ID (identifier number), Title (name of data to be parsed), FB/EB (start/end byte positions), F2B/E2B (start/end bit positions for multi-byte fields), Repeat/Step (array/step parameters), and Bit/Byte (bit or byte mode). The parsing modes include: FB-EB (parse specified region), F2B-E2B (parse specified region as low/high bytes), Repeat (parse specified region then jump to next region), and Repeat-Step (parse with variable array dimensions). In all modes, parsing follows big-endian order (most significant byte first), with a maximum supported width of 64 bits.

The parsing strategy proceeds as follows: Since ME and POLAR data are organized in packets, we parse one packet at a time. First, we check packet sequence numbers for continuity and perform Cyclic Redundancy Check (CRC) validation. When the program starts or after packet loss occurs, we search for keywords to determine data structure boundaries. For POLAR, a data structure may span two packets, requiring packet assembly to form complete structures. Only structures passing CRC validation are retained; incomplete or invalid structures are discarded.

2.2 Data Storage

Science events may be stored temporarily in memory or written directly to files depending on subsequent processing requirements. Engineering and platform data are typically written directly to files. Storage design must facilitate later file merging and splitting. For ME, which has 54 independent detector units each packaged separately, we recommend storing data from each unit in separate regions. This can be implemented as independent files or as separate data regions within a ROOT format file [**?**], such as 54 tree structures in one file.

2.3 Time Calculation for ME and Event Matching for POLAR

The primary challenges are event loss and packet loss. Both instruments use special events as reference points.

ME Time Calculation: ME uses internal crystal oscillators to represent time. Each event records only the lower bits of the time; higher bits are recorded via carry events. The reference for time calculation is adjacent carry events. The method scans lower time bits and records overflow counts. If the overflow count matches adjacent carry events, intermediate carry times can be accurately determined. With only one carry event, time must be inferred forward or backward. After packet loss, subsequent events cannot obtain accurate times.

POLAR Event Matching: This involves matching trigger packets with corresponding data packets. Matched trigger packets represent physical events. The method uses baseline trigger and data packets as references, calculating relative times for other packets. If the time difference between a trigger packet and the baseline matches that of a data packet and its baseline, and the trigger flag indicates the correct module, the data packet belongs to that trigger. The baseline is updated when new matches are found to avoid electronics rollover effects.

shows a sample matching result, where the first record serves as the baseline and Flag indicates triggered modules.

For engineering and platform data, Level 1A processing mainly involves unpacking without special handling, storing data sequentially.

3. Level 1B Processing

Level 1B processing calculates GPS time for event files and converts parsed quantities to physical values.

3.1 GPS Time Calculation

This requires establishing a correspondence between local time and GPS time. We define "correspondence points" (individual time pairs) and "correspondence segments" (time periods). The crystal oscillator period is needed, though nominal values may differ from actual values due to temperature variations and aging.

For ME, the local time of 1PPS (one-pulse-per-second) events has been calculated in Level 1A, providing a direct local-to-GPS correspondence. POLAR writes this relationship to housekeeping files, updated every minute. Both instruments' correspondence is affected by crystal oscillator instability and GPS jitter. Correction involves accumulating correspondence points: given N points (T^i_{GPS}, T^i_{Local}), we calculate time differences ΔT_{GPS} and ΔT_{Local}, then compute the average ratio $G = \Delta T_{GPS}/\Delta T_{Local}$ to obtain the crystal period. The required number of points depends on temperature variation and required time precision.

Event GPS time is calculated as:

$$T_{evt} = T^0_{GPS} + G \times (N_{evt} - N^0_{Local})$$

where T^0_{GPS} and N^0_{Local} are the reference GPS time and local count, and N_{evt} is the event' s local count.

ME' s carry events accumulate continuously, so one correspondence point could theoretically apply to all data. However, ME may restart, resetting the crystal time. POLAR' s central trigger module resets its counter periodically. The solution: In Level 1A, record the start/end times of each local time period (using packet timestamps from platform or GPS time) and mark which period each event belongs to. In Level 1B, read these periods and establish correspondence within each period. ME uses accumulated carry events, while POLAR uses 1PPS events.

GPS time error is random with poor short-term stability but good long-term stability, while local time has good short-term stability but long-term drift. Accumulating sufficient correspondence points yields the crystal period and characterizes long-term drift, with error approaching zero.

3.2 EHK File Generation

Satellite-downlinked orbit and attitude data alone cannot fully characterize the space environment. Extended Housekeeping (EHK) files describe the payload' s operational environment, identifying high-background regions, South Atlantic Anomaly (SAA) passages, field-of-view occultations, and pointing instability. EHK files record at regular intervals: (1) satellite position and attitude, (2) geomagnetic cutoff rigidity, (3) SAA relationship, (4) Sun/Moon pointing relationships, and (5) pointing jitter. Definitions reference standard data products [?]. Note that satellite pointing represents averaged instantaneous attitudes. EHK generation depends on Earth boundary calculations, geomagnetic cutoff rigidity files, SAA files, and leap second files.

4. Level 1C Processing

Level 1C merges and splits data files and performs additional processing such as good time interval selection and bad detector rejection.

ME' s Level 1A events are stored by detector unit and must be time-merged. A sorting algorithm merges events from all units. POLAR' s data are organized by trigger packets without requiring sorting. Good time intervals exclude payload shutdown, high-voltage reduction, and SAA passage periods. Data from malfunctioning detectors must be rejected.

File retrieval uses a MySQL database [?]. Level 1A/1B file names and attributes (start/end times) are stored. Given a data type and time range, the database queries all matching files with temporal overlap. Events are then filtered, sorted, and processed. This stage may include format conversion to FITS, with key information (start time, observation position) written as keywords [?].

5. Summary

Data preprocessing is a critical step in astronomical satellite data processing, forming the basis for calibration and scientific analysis. While payload details vary, most follow similar steps. Scientific data undergo unpacking, processing, time calculation, and file merging/splitting. Engineering and platform data require unpacking and calculation. Some satellites generate EHK files later due to correlation considerations. Throughout, physical quantity conversion and inter-step dependencies must be considered.

Implementation details include: configuration-file-based parsing (bridging hardware and software design), processing between special events as units (improving correctness at some computational cost), and data storage strategies (split by unit initially, merge during Level 1C). These methods leverage all available data information and accommodate relationships between data dependencies.

Acknowledgments

We thank the HXMT ground system and POLAR team at the Institute of High Energy Physics for their support.

References

1. Space Communications Protocol Specification-File Protocol. Recommendation for Space Data System Standards [R/OL]. 2016-12-06. https://www.researchgate.net/publication/2593511_{{{Recommendation}}*{{For}}}*{{{Space}}*{{Data*
2. Suzaku Term. The Suzaku data reduction guide [OL]. 2013-08-01/2016-12-06. http://heasarc.gsfc.nasa.gov/docs/suzaku/analysis/abc/.
3. RXTE Term. The ABC of RXTE [OL]. 1999-09-14/2016-12-06. http://heasarc.gsfc.nasa.gov/docs/xte/abc/contents.html.
4. Capalbi M, Perri M, Saija B, Wells D C, Greisen E W, et al. The swift XRT data reduction guide [OL]. 2005-04. https://swift.gsfc.nasa.gov/analysis/xrt_{{{swguide}}_{{v1}}}2.pd
5. Harten R H, Griesen E W. FITS: A flexible image transport system. Astronomy & Astrophysics Supplement, 1988, 44: 363-370.
6. NASA. A primer on the FITS data format [OL]. 2014-10-28/2016-12-06. http://fits.gsfc.nasa.gov/fits_{primer}.html.
7. Li Tipei, Wu Mei. The hard X-ray modulation telescope mission. Nuclear Physics B, 2016: 131-139.
8. Li Tipei. HXMT: A Chinese high-energy astrophysics mission. Nuclear Physics B, 2016: 648-651.
9. Produit N, Barao F, Deluit S, et al. POLAR: A compact detector for gamma-ray burst polarization measurements. Nuclear Instruments and Methods in Physics Research A, 2005, 550: 616-625.
10. Lin Yuting, Zhao Dongming, Gao Weiguang, et al. GPS time system and its application to time comparison. Geospatial Information, 2006, 4: 30-32.
11. Han Jiawei, Zhang Hongqun. Standard remote sensing satellite downlink data error correction technology. Chinese Space Science and Technology, 2014, 5: 112-116.
12. The ROOT Term. ROOT user' s guide [OL]. 2014-05-01/2016-12-06. https://root.cern.ch/root/htmldoc/guides/users-guide/ROOTUsersGuide.html.
13. Zhong Shoubo, Han Bo, Zhang Yanxia, et al. Design and implementation of a software package for managing massive astronomical data. Astronomical Research & Technology, 2016, 13: 510-517.
14. Mei Ying, Liu Donghao, Wang Feng, et al. A study of the FITS-IDI format for the Chinese Spectral Radio Heliograph. Astronomical Research & Technology, 2016, 13: 388-395.