ATHARVA- An Ocean Acoustic Data Management Framework

Parvathy Ramakrishnan P, Dr.Sahely Bhadra, Sudhina Dinesh, Dr.Satyajit Das
Department of Data Science, Indian Institute of Technology, Palakkad

Introduction

In today's data-driven world, effective data management is paramount due to the exponential growth in data volumes. This challenge is particularly pronounced in the domain of ocean observation systems [1], where the continuous advancement of ocean observation platforms, sensors, and sonar technologies has resulted in the generation of vast datasets. These data collected from sound navigation and ranging (sonar) applications hold immense potential for modeling the underwater acoustics environment and informing scientific analysis and policy decisions. Effectively managing sonar data necessitates the integration of data from various sources, each with its own data format, packet rate, and size. The complexity is compounded by the fact that sonar data is received in binary packets, which require considerable time for decryption and interpretation. This demands both manual effort and specialized domain knowledge for effective data analysis. Furthermore, manual search through file systems for data retrieval is inherently inefficient. Recognizing these challenges, there is an evident need to modernize data management systems to improve data accessibility and utility while reducing the burden of manual intervention. Our work aims to address this need for an efficient and streamlined process to work with sonar data by offering efficient data processing, integration, retrieval, and visualization capabilities.

Background

Sonar array consists of hundreds of sensors that pick up the acoustic signals, convert them to electrical signals, and capture them into binary files. During data acquisition, several oceanographic and transmission-specific parameters are also captured along with the acoustic signals. These non-acoustic parameters including temperature, depth, latitude, longitude, frequency, GPS Speed, and range, collectively referred to as **metadata**, are critical for understanding and retrieving the data effectively. The generic digital data recorder (**GDR**) receives the data packets without any particular order. Acoustic data packets contain high-frequency components, contributing to a significant data volume, while non-acoustic packets typically have lower update frequencies, often less than 1Hz. As shown in Fig.1, GDR appends a **meta-header** with each packet received. Meta-header includes the time-stamp of packet arrival, IP Address, port, sequence number, and packet size. These details are pivotal for packet filtering and the replay of acoustic signals.

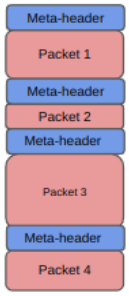

Fig 1. GDR File Structure

Proposed Framework:

In this paper, we propose a framework for ocean data management, Acoustic daTa Handling ARchival Visualization and Analytics (ATHARVA) which includes 2 components (i) An *indexed metadata database* for fast and efficient data discovery with minimum overhead and (ii) A GUI-based software tool for interactive information retrieval and visual analysis of data.

The central concept of our framework is to create an *indexed metadata database* from datafiles, enabling easy discovery of acoustic data based on metadata queries. Additionally, the byte offsets and timestamps are also captured into the *indexed metadata database* which facilitates pinpointing relevant data within GDR files. This will eliminate the need to read entire files during retrieval. To optimize performance, we employ indexing and partitioning techniques within the database. These strategies reduce the data volume involved in queries, resulting in faster query execution [2].

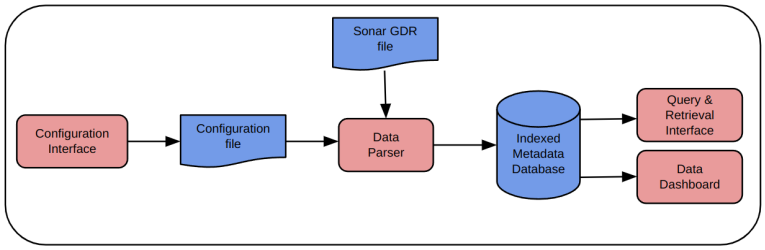

Fig 2. The Data Flow Schema

The software tool is designed to perform the following tasks. (i) Process binary-encoded sonar GDR files (ii) Load the metadata extracted from these files into the database (iii) Facilitate easy data integration (iv) Enable users to search and find the sonar data files for specific information (v) Provide a visual representation of the metadata and acoustic data.

Fig.2 shows the data flow in the proposed framework. The Users can effortlessly define configurations for GDR files through the configuration interface. The tool generates a JSON file that can accommodate variations in data formats, acoustic data, and metadata packet structure. The data parser module in the framework is responsible for processing and extracting the metadata from the sonar data files stored in the file system. The metadata is then stored in the *indexed metadata database*. This database is used to search and discover relevant data. The query &retrieval interface and the data dashboard allow users to search relevant data, analyze the data, understand trends and anomalies, and download relevant acoustic data or metadata for further analysis. The metadata operations access the database only while the acoustic data operations will query the matching files and the byte offsets within the file from the database and then read the GDR file for the relevant data.

Fig 3. Data Visualizations (a) Data Dashboard with query widget, time series comparisons of different metadata parameters and acoustic sensors and acoustic sensor heatmap plots (b) Region wise data distribution over map layout (c) Power spectral density overlay plot of different sensors (d) Stem plots for sensor amplitude mean value

Results/Discussion

ATHARVA has achieved high-speed data parsing and metadata capture, with the capability to process data at a remarkable speed of 6GB per minute. Furthermore, our database design achieves fast data discovery within the file system with a minimal overhead of 1GB of database over 500 GB of GDR files. Table 1 shows the average response time for the major tasks. The metadata operations are in order of milliseconds and acoustic data operations are in seconds. This is because metadata is fetched directly from the *indexed metadata database* while acoustic data needs to read the GDR file also. Also, the volume of data handled is very huge in the case of acoustic data hence it is processed and visualized in blocks of 60 seconds. Our tool provides efficient data visualization capabilities, as showcased in Fig. 3, through a data dashboard featuring a query window and various plots for exploring and analyzing both metadata and acoustic data comprehensively.

Task	Response Time (sec)
Parse 1GB GDR file	10.0
List files matching the metadata query	0.35
Retrieve and Visualize metadata	0.59
Download metadata into CSV	0.18
Retrieve acoustic data	3.5
Visualize Acoustic data	0.81
Download acoustic data	8.0

Table 1. The average response time for major tasks

Conclusion

In this work, we introduced an integrated framework, ATHARVA for handling ocean acoustic data ensuring data accessibility and its potential for reuse. The framework is capable of integrating data from different sources having diverse data formats, variable packet rates, and sizes. It has minimized manual intervention in data analysis and modernized the data management system. We proposed an *indexed metadata database* that enhances data discovery within file systems through tagging and indexing metadata, ensuring reduced data retrieval times. The experimental results demonstrate that the framework performs data discovery, analysis, retrieval, and visualization efficiently.

Acknowledgments

The authors extend their gratitude to the team at the Naval Physical and Oceanographic Laboratory, Kochi, India, for their invaluable guidance and support throughout this research project. This work was made possible through funding provided by the Naval Research Board.

References

[1] Subramanian, Ramasundaram, Sundar Ranganathan, and Venkatesan Ramasamy. "Gateway to Ocean Data Management: Ocean Best Practices, Information and Visualisation for moored ocean observation network." OCEANS 2022-Chennai. IEEE, 2022.

[2] Sim, Hyogi, et al. "Tagit: an integrated indexing and search service for file systems." Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2017.

[3] Hilburn, Ben, et al. "Sigmf: the signal metadata format." Proceedings of the GNU Radio Conference. Vol. 3. No. 1. 2018.

[4]Harwani, Bintu M. Qt5 Python GUI Programming Cookbook: Building responsive and powerful cross-platform applications with PyQt. Packt Publishing Ltd, 2018.