A Grid Enabled Visual Tool for Time Series Pattern Matching

B. Liang, M. Jessop, J. Austin

DAME team
Advanced Computer Architectures Group
Department of computer science
University of York
York, YO10 5DD
UK

Abstract

Distributed Aircraft maintenance Environment (DAME) is a major UK eScience Grid project aiming to build a Grid test-bed for Distributed Diagnostics. The main challenge of the DAME project is the design and implementation of a fault diagnosis and prognosis system over the Grid computing infrastructure. This paper examines how the Grid may effect the implementation of search engines by focusing on the Signal Data Explorer (SDE) developed within the DAME project. The tool may be used for any time series based data.

1. Introduction

Distributed Aircraft Maintenance Environment (DAME) is a major UK eScience Grid project aiming to build a Grid test-bed for Distributed Diagnostics [1]. The application demonstrator of DAME is a distributed aircraft maintenance environment for Rolls-Royce aero-engines. The main challenge of the DAME project is the design and implementation of a fault diagnosis and prognosis system over the Grid computing infrastructure. Distributed engine fault diagnostics shares many common characteristics of a similar system in many fields such as medical, transportation and manufacturing.

Engine data is stored as Engine Data Records (EDRs) consisting of vibration and performance data represented in time series. The DAME system is able to take data from a fleet of aircraft and store this data in distributed data stores for later searching and matching. A complete pattern storage and search system based on existing AURA (***Advanced Uncertain Reasoning Architecture***) search technology has been developed with the DAME project. The front end of the DAME pattern match service is an application specific, portal-based data browser and search interface called the ***Signal Data Explorer*** (SDE). This paper examines how the Grid may effect the implementation of search engines by focusing on the development of the SDE.

The rest of the paper is organized as follows. In Section 2, we state the pattern match problem in DAME and introduce the AURA search technology. In Section 3, we briefly describe the DAME Grid environment and the DAME pattern match service. In Section 4, the structure and functionality of the SDE are introduced. In Section 5, we demonstrate the application of the SDE. In Section 6, we discuss the directions of further work.

2. DAME pattern match problem

A civil airline engine is capable of generating up to terabytes of vibration and performance data per year. These data may be stored in repositories distributed across many geographic and operational boundaries. Advanced pattern matching and data mining methods which are able to operate on the large volumes of data and give a response time that meets operational demands are required. At present, these data are searched locally for the presence of features known to be associated with fault condition and for deviations from a model of known operation (novelty detection). There is no mechanism for searching a globally distributed database of past experience to find any similar novelties. Neither is there a mechanism for collecting any relevant collateral information to provide the owner of the data with decision support.

2.1 Time series data pattern match

A time series signal is a series of values of a variable at successive times. Analysis of time series can be within the time domain or frequency domain [4][5][6]. Both the engine vibration and performance data are the time series signal. The health of an operating engine can be identified by analyzing patterns and values of the monitored time series signals.

A crucial part of engine health monitoring is the search and detection of abnormal pattern and values within the vibration and performance data. The difficulties of such pattern search lies in the scale of the search, typically terabytes of data. The problem is made worse by the limited knowledge on patterns of fault conditions. The AURA technology provides solutions for high-speed pattern match from large volumes of data. A set of methods for pre- and post-processing makes it possible to handle the imprecisely defined fault patterns.

2.2 Overview of AURA

AURA (Advanced Uncertain Reasoning Architecture) is a parallel pattern matching technology developed as a set of general-purpose methods for searching large unstructured datasets. The AURA technology is based on a high-performance neural network called Correlation Matrix Memory (CMM) [2][3]. AURA defines a set of pre- and post-processing methods for different data as well as a neural network based data storage system and a set of efficient indexing techniques. AURA provides fast approximate search and match operation on large unstructured datasets.

The latest developments of AURA include the new AURA C++ library, the new hardware prototype (PERSENCE-II) and a Grid enabled AURA pattern match control service (PMC). AURA has been applied to many problems, these include a postal address matching, high-speed rule-matching systems [7], fraudulent benefit claims detection system (FEDAURA), structure--matching (e.g. 3D molecular structures) [8], human face recognition and trademark database searching [9].

3. DAME Grid patterm match service

Within the DAME project, the AURA methods have been extended to deal with vibration based time series data and distributed search data.

3.1 DAME Grid environment

The White Rose Computational Grid (WRCG) provides a realistic and manageable Grid environment for DAME. Details of WRCG can be found elsewhere [12]. Grid technology allows the DAME system to be highly distributed and yet still appear as a single coherent system. To these ends the architecture shown in Fig. 1 has been devised to allow high performance data management and pattern matching against globally distributed engine data archives.

The DAME Pattern match services are implemented as a Globus Toolkit 3.x grid services [10], and as such are effectively web services, which are hosted inside a Jakarta Tomcat 4.1.x [11] installation.

3.2 Pattern Match Control

Since engine data arrives at airports where it is stored for access and processing. Each airport is represented by a pattern match control, PMC, service. Underneath this, there is the AURA service and services for data extraction and encoding. Data repository facilities are provided by the SDSC Storage Request Broker, SRB [13].

To perform a pattern match a client contacts a single PMC that acts as the master node for that particular search. It is responsible for fanning out the search to every other node in the system. All results are returned to the master node, which correlates them and returns them to the client.

A client is free to contact any PMC node in order to initiate a pattern match operation. This makes the distributed search engine's behaviour powerful, yet simple and transparent to clients

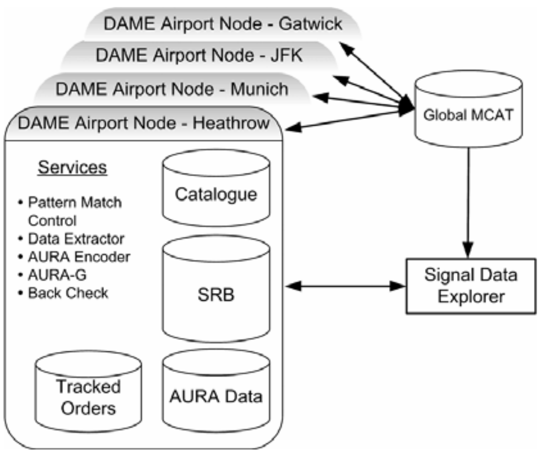

Fig. 1 Distributed Data Management Architecture

3.3 AURA-G

The AURA technology itself exists as a Grid enabled service called AURA-G. It is an extension of the AURA technology dedicated software and hardware implementations. AURA-G supports distributed search, across multiple CMM engines at different locations and scalable pattern matching. The AURA-G is OGSA compliant.

3.4 Data extractor and encoder

The Data Extractor extracts specific low-level features from raw engine data into time series data with specified format. These data are further encoded into specified binary code, which is understood by the AURA pattern search engine. The encoding is performed by an AURA Encoder service

3.5 Data Repository

The Data Repository at each node is responsible for storing all raw engine data along with any extracted and AURA encoded data. This service is provided by the SRB, a system for managing distributed storage resources from large disc arrays to tape backup systems. Files are entered into SRB and can then be referenced by logical file handles that require no knowledge of where the file physically exists. A Meta Data Catalogue is maintained which maps logical handles to physical file locations. Additional system specified meta-data could be added to each record. This means that the catalogue can be queried in a data centric way, e.g. engine serial number and flight date/number, or via some characteristic of the data, rather than in a location centric fashion.

4. DAME Signal Data Explorer

Intelligent feature extraction and data mining are necessary to provide early detection of engine faults. The DAME system provides a collection of diagnostic tools, which can be used in an automatic sequence, usually by a Maintenance Engineer, or interactively by a Domain Expert. The Signal Data Explorer (SDE) is a visual tool act as the front end of the AURA technology. It is vital to the work of the Domain Expert for engine diagnosis. The SDE bridges the gap between user and the DAME Grid pattern match services. The DAME system is able to locate the data record for the user from within the network and stream the data to the application. The distributed search engine searches all the appropriate engine data in a parallel fashion. In this way, it is possible for a user to bear the full power of large data and processing Grids without necessarily knowing anything about Grid technology itself.

Using the SDE a Domain Expert can view raw engine data, extract and view features from the raw data, select a pattern or Region Of Interest (ROI) and search for similar patterns in the terabytes of data stored from previous and recent flights. The SDE allows the viewing and searching of time series data stored locally or remotely. Insight can be gained from viewing and searching such data for the presence of features known to be associated with fault conditions and for deviations from a model of normal operation.

The SDE provides an interactive graphic user interface (GUI) for the viewing and analysis of data from individual EDRs. It is equipped with the necessary signal processing tools to extract low level features and parameters from the input vibration data and performance data. It permits large scale searching for patterns in an engine data set (many EDRs). A built-in pattern template library allows the user to manage his “interested pattern templates” for engine diagnosis or feature detection.

In addition to the provision of an integrated environment for interactive pattern matching, feature extraction and detection, the SDE also provides programmable facilities. A Domain Expert can use the SDE to perform a programmed set of parallel and sequential operations to search for multiple patterns in the course of fault detection. This is a very powerful feature of the tool.

4.1Architecture

Fig-2 shows the architecture of the SDE and the relevant Grid services. All the Grid services are general purpose (i.e. can be used in other applications). An outline of the operation of the SDE is as follows. Raw data is input from the Data Store and is displayed in the appropriate sub-window in the GUI. The user can then select a feature to view – this feature is then extracted by the SDE from the raw data and displayed. The user can “play” the data at various speeds for rapid observation. A fragment of the feature (ROI) can be selected by the user and used as a search query within the AURA search engine (this represents terabytes of engine data). The fragment has to be encoded prior to submission to the search engine. The results of the search and the scores are displayed – individual results (matched patterns) can be displayed graphically in a popup window.

To allow the SDE be used separately without the support of the network, some services such as the extractor, encoder and search engine are also made available locally. The work mode requires that the data is available locally.

Referring to the SDE, the main components on Fig.2 are described below.

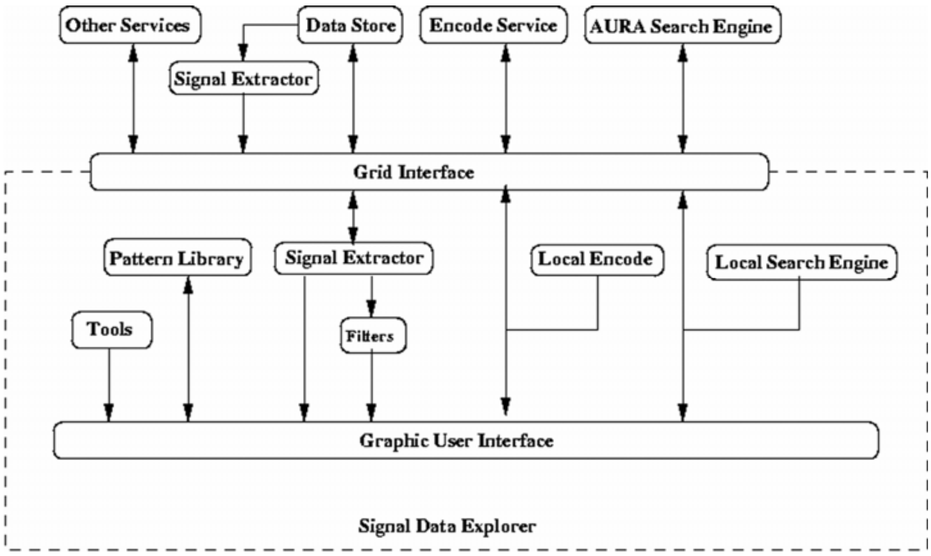

Fig. 2 Architecture of the signal data explorer

- **The Data Store** is a repository for the engine data records and provides the SDE with the EDR data for viewing and processing. Data stores are distributed over the Grid and available for all other services.
- **The Graphical User Interface** integrates the input and output to and from all local and remote services with the purpose of providing facilities to explore the data.
- **The Signal Extractor** extracts the user selected low level features from the unstructured raw EDR. Such low level features can be directly displayed or fragments of these can be selected for uses as a search query with the AURA search engine. The SDE can use either it's own local version of the extractor or the more sophisticated external extractor Grid service.
- **The AURA Search Engine**. The SDE can use either it's own local search engine or the more sophisticated external search engine implemented as a Grid service. The user uses the provided tools to generate a query pattern (fragment). This pattern is firstly encoded into a query vector and then fed to the search engine. The searching results feedbacks to the SDE and displays to the user. The training data of the search engine can be of pre-built or built in time depending on applications.
- **The Encoder** encodes the time series data into a binary vector according to the consequent pattern match.
- **Filter** (programmable) module provides a means of signal preprocessing such as noise suppression. Signals can be preprocessed before the search activity.
- **Pattern Library** contains pattern templates. These can be generated using built-in tools and used as the query pattern for the searching. The sources of pattern templates can be ROI of current displayed low level feature, a library pattern or a user constructed one. The built-in pattern template library stores and manages the pattern templates of interest, which might be "examples of fault" or "examples of unknown event". These templates can be used separately or organized as a set of features to represent a particular fault.
- **The Tools** module provides probing and scaling facilities. The probing tools allow the user to point to areas of the data display and view information about the area. The information can be frequency spectra or data source information as required. The scaling facility allows the user to apply scales to the various display axes.
- **Other Services.** The SDE also integrates seamlessly with other DAME diagnostics tools and services. For example, the results from the AURA search engine may be "sent" directly from the SDE to the DAME Case Based

Reasoning tools for further processing in the diagnostic sequence.

4.2 Operation and Functionality.

The SDE has the flexibility to be used as a separate tool or as a front end GUI for the DAME search services. It can be used as a viewer to display the input time series in at various speeds, switch between different extracted data sets, measure parameters of the variable at the given time instance and probe the data hidden in the raw data. The low level features are extracted at real time while the data are playing. The input data can be local or from the Grid.

The search engine supports correlation similarity measure, Euclidean and City-block distance measures. Depending on the applications, one or more measures can be applied to the AURA search engine. Based on the defined measure, a searching task can be "find me all examples of vibration pattern similar to the input template" or "find me the first N best matches of the query pattern from the data store".

A set of tools for extraction of data values from the raw engine data record and manipulation of the feature patterns is built into the data explorer. When searching for a similar pattern, the user can choose the query pattern from the template library or from highlighted ROI of the current time series pattern displayed on the window. Once the search is invoked, the Grid pattern match controller dispatches the task to the relevant distributed search engine. The search engine will look for similar patterns from the data store or testing data and locate the position of the best match. Results are displayed on the results window. The user can activate each result to exam the actual "shape" of the pattern and call for further operation over the Grid. Results and pattern displayed on the window can be printed or save to file.

The main scenario of the use of the SDE is as follows:

1. Display the vibration ZMOD data pattern and the time series of other parameters extracted from the data record as movie sequence. Locate and highlight positions of the parameters from the main window.
2. Extract and display instantaneous parameters of the engine data record.
3. Search for the best match of the query pattern from each record in the data store over the network.
4. Search for the next similar pattern from the current data set -- an equivalent of find word from a text file.
5. Manage the "interested pattern templates" data store for fault detection. The user can add, delete and edit the pattern stored in the library.
6. Probe for particular parameters instance hidden in the data record from the ROI of the pattern view window.
7. Synchronize and display correlations of multiple features.

5. Application examples

The following screenshots show particular examples of the operation of the Signal Data Explorer.

1. ***Search for similar pattern to the input template from large volume of history data store (Fig. 3).***

   The *region of interest* is highlighted from the feature window (the bottom view) as the search template. The best four matching patterns are displayed. Search results are listed in the results window.

2. ***Find similar pattern to the template from the current test data record (Fig. 4).***

   A query pattern is imported from the template library (shown on the popup window). A similar pattern in the current test data record is displayed in the feature window. The input ZMOD data view (the main window) is synchronized to the same time instance.

3. ***Correlations of multiple patterns (Fig.5).***

   Multiple patterns are displayed and synchronized in the main window.

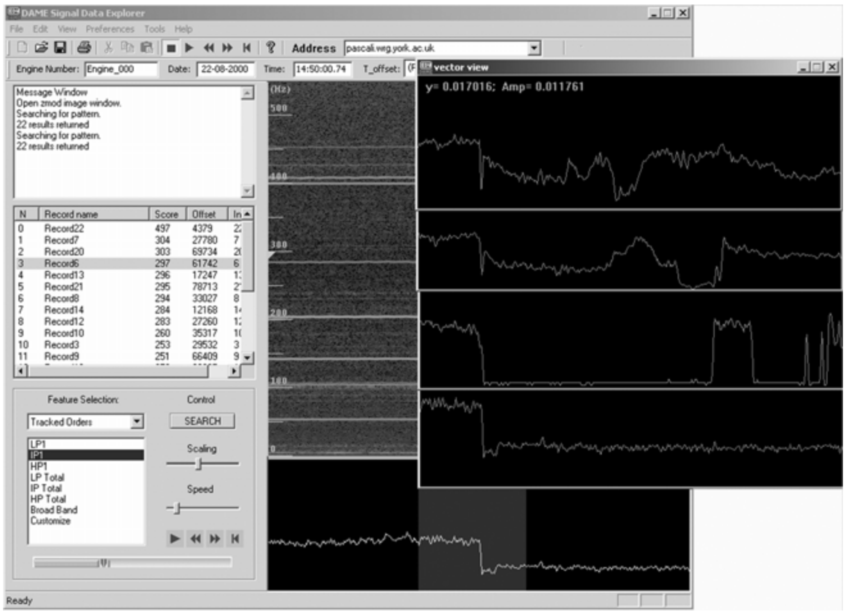

Fig. 3 Pattern display as a result of a search

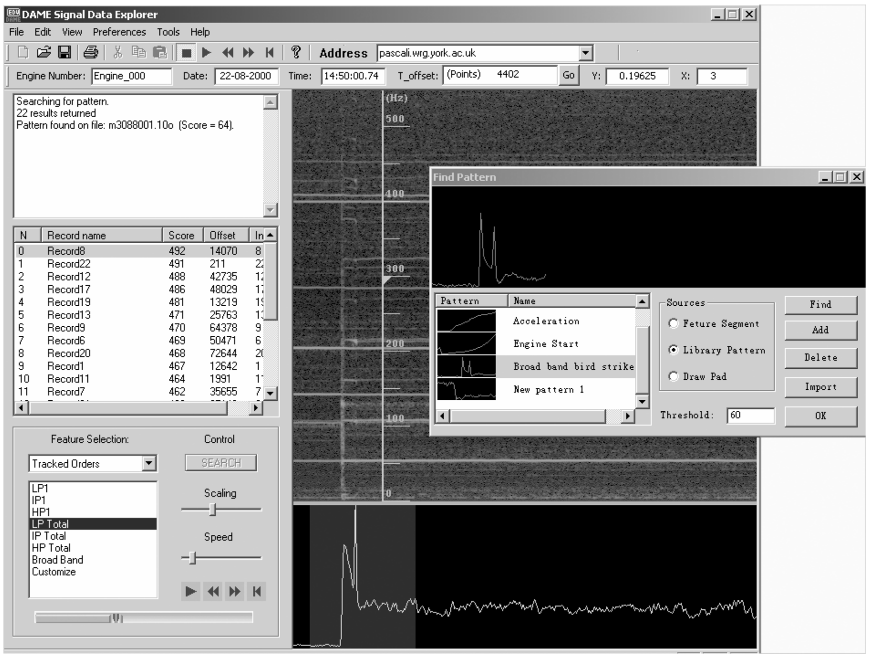

Fig. 4 Pattern found from current feature set matching a template

6. Further work

At the moment, the SDE only supports single query pattern match in separate search tasks. For more complicated pattern match applications, it is necessary to fulfil multiple pattern searching and provide more sophisticated pre- and post-processing methods. In the future, the search engine and the searching task can be configured in time over the Grid by using a task planner equipped with the SDE. A user can choose and combine together different modules and the relevant parameters in a search to suit his use.

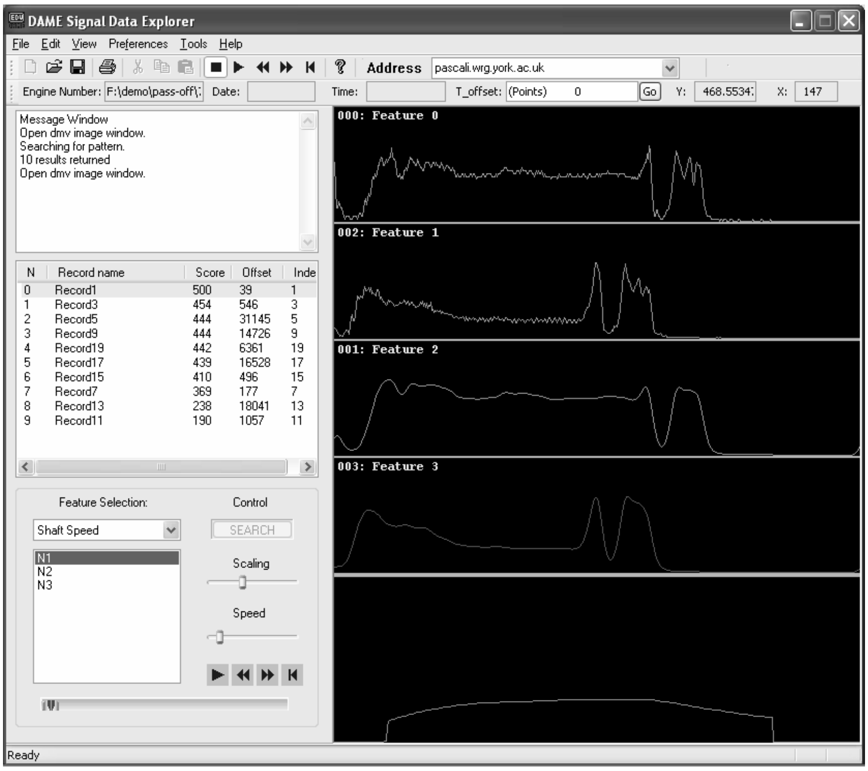

Fig. 5 Correlation of multiple patterns

7. Acknowledgements

This project was supported by a UK eScience Pilot grant from EPSRC (GR/R67668/01), as well as support from our industrial collaborators Rolls-Royce, Data Systems and Solutions and Cybula Ltd.

References

[1] J. Austin, T Jackson, M Flecher, M Jessop, P Cowley, P Lobner, "Predictive Maintenance: Distributed Aircraft Engine Diagnostics", Ch 5, p 70-78, The Grid, Second Edition, I Foster, C Kesselman, 2004

[2] J. Austin, "Distributed associative memories for high speed symbolic reasoning." In, IJCAI working Notes of workshop on connectionist-symbolic Integration, From Unified to Hybrid Approaches. (1995) pp. 87-93.

[3] J. Austin, J. Kennedy, and K. Lees, " The advanced uncertain reasoning architecture". In weightless neural network Workshop, 1995.

[4] Priestley, M. B., "Spectral Analysis and Time Series", Vols. 1 and 2, Academic Press, New York. 1981.

[5] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C., "Time Series Analysis, Forecasting and Control". 3rd ed. Prentice Hall, Englewood Clifs, NJ. 1994.

[6] Brockwell, Peter J., and Davis, Richard A., "Time Series: Theory and Methods", Springer-Verlang 1987.

[7] Austin, J., Kennedy, J., and Lees, K., "A neural architecture for fast rule matching", In Artificial Neural Networks and Expert Systems Conference (ANNES 95), Dunedin, New Zealand, December 1995.

[8] Turner, A., and Austin, J., "Performance evaluation of a fast chemical structure matching method using distributed neural relaxation", In fourth International Conference on Knowledge-based Intelligent Engineering Systems, August 2000.

[9] Alwis, S., and Austin, J., "A novel architecture for trade-mark image retrieval systems", In Electronic Workshops in Computing. Springer, 1998.

[10] The Globus Alliance, http://www.globus.org

[11] The Jakarta Project, http://jakarta.apache.org/tomcat/index.html

[12] The White Rose University Consortium. http://www.whiterose.ac.uk/Home.aspx

[13] Storage Request Broker at http://www.npaci.edu/DICE/SRB/