

DEVELOPMENT OF A DMT MONITOR FOR STATISTICAL TRACKING OF GRAVITATIONAL-WAVE BURST TRIGGERS GENERATED FROM THE OMEGA PIPELINE

JUNWEI LI

*Research Institute of Information Technology, Tsinghua University
Beijing 100084, P. R. China*

JUNWEI CAO

*Research Institute of Information Technology
Tsinghua National Laboratory for Information Science and Technology
Tsinghua University, Beijing 100084, P. R. China*

One challenge in large-scale scientific data analysis is to monitor data in real-time in a distributed environment. For the LIGO (Laser Interferometer Gravitational-wave Observatory) project, a dedicated suit of data monitoring tools (DMT) has been developed, yielding good extensibility to new data type and high flexibility to a distributed environment. Several services are provided, including visualization of data information in various forms and file output of monitoring results.

In this work, a DMT monitor, OmegaMon, is developed for tracking statistics of gravitational-wave (GW) burst triggers that are generated from a specific GW burst data analysis pipeline, the Omega Pipeline. Such results can provide diagnostic information as reference of trigger post-processing and interferometer maintenance.

1. Introduction

Many scientific applications require manipulation of large amount of data in real-time, e.g. astronomical and astrophysical experiments. The Very Long Baseline Array (VLBA), a system of ten radio telescopes, has a wideband, high-density recording system which generates data at a sustained rate of 128Mb/s, and peak rate of up to 512Mb/s¹. The Sloan Digital Sky Survey (SDSS)'s motivation is to digitally map about a quarter of the whole sky. The data rate from its 120 million pixels camera is 8 Mb/s².

Another common feature for most of scientific applications is a distributed system and environment. For example, in VLBA there is an array of ten 25-m-diameter high-performance antennas scattered across the US. Another example is the High Energy Stereoscopic System (HESS) which consists of four imaging Cherenkov telescopes³.

In order to maintain the integrity of the whole system and to obtain data quality information as the reference for subsequent data analysis, an efficient set of tools which can monitor critical system activities and important data statistics is required. Due to large amount of data and the distributed environment, demands for monitoring become greater.

LIGO (Laser Interferometer Gravitational-wave Observatory) initiated the Data Monitor Tool (DMT) project, aiming at defining tools and environments necessary to support continuous data monitoring of LIGO interferometers⁴. In this work, the DMT is applied for gravitational-wave burst (GWB) data analysis. In particular, a DMT monitor, OmegaMon, is developed for tracking statistics of gravitational-wave (GW) burst triggers that are generated from a specific GW burst data analysis pipeline, the Omega Pipeline.

The paper is organized as follows: the overview of LIGO and DMT is given in Sections 2 and 3, respectively. We introduce OmegaMon implementation in Section 4. Finally, we summarize our conclusions and future works in Section 5.

2. Overview of LIGO

Albert Einstein predicted the existence of gravitational waves in 1916 as part of the theory of general relativity in which he described space and time as different aspects of reality in which matter and energy are ultimately the same⁵. And the existence of gravitational wave is the most important prediction in general relativity theory.

Gravitational wave is a fluctuation in the curvature of space-time which propagates as wave. Until now, gravitational wave hasn't been directly detected, but observations of the orbital decay of the first binary pulsar PSR B1913 + 16 have provided significant indirect evidence for their existence since the late eighties⁶.

To directly detect gravitational wave, LIGO was built by California Institute of Technology (Caltech) and the Massachusetts Institute of Technology (MIT). LIGO uses Michelson laser interferometers to detect gravitational waves by measuring the interference of two laser beams whose length is changed when the gravitational wave passes by⁷. In other words, the laser interferometer is similar as a microphone that converts gravitational waves into electrical signals. Currently LIGO has two observatory sites in the US with a total of three laser interferometers. The LIGO Hanford Observatory (LHO) in Washington State has two laser interferometers called H1 and H2. The LIGO Livingston Observatory (LLO) in Louisiana has a single interferometer called L1⁸. In the fifth science

run of LIGO, the total data acquisition rate of three interferometers exceeded 16Mb/s⁹.

In addition to the two observatory sites in the US, LIGO also collaborates with the British-German GEO¹⁰ 600m detector located near Hannover, Germany; and shares data with the Italian/French VIRGO¹¹ detector. VIRGO's data acquisition rate is at about 20Mb/s, which means about 690GB per day written on disk¹². GEO's data acquisition system is required to be capable of handling a data rate of up to 1Mb/s¹³.

Summing up the above, these gravitational wave detecting projects all have large amount of real-time data streams to monitor. LIGO's DMT project commendably performs the monitoring on large volumes of data.

3. Overview of DMT

3.1. DMT architecture

DMT consists of three main components:

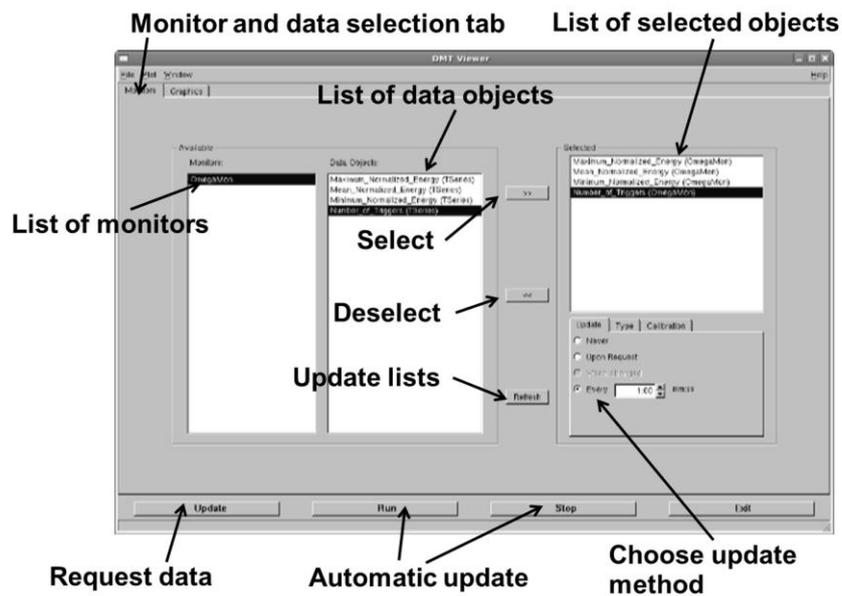


Figure 1. The DMT viewer GUI which mainly consists of list textboxes of available monitors, available and selected data objects.

- The monitors in various types. Each monitor has its own one or several specific data types to monitor. A data type could be a shared memory buffer or simple text file on disk. For each data type, a monitor could generate several classes of monitoring results. A class of corresponding monitoring result is called a class of data object.
- The name server. It is responsible for handling monitor registration and monitor information query.
- The DMT viewer. It is a tool for visualizing monitors' monitoring results, as shown in Figure 1.

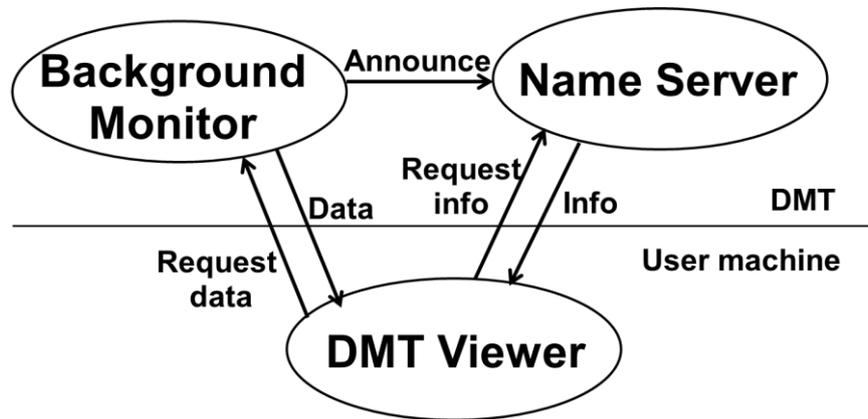


Figure 2. Three main components of DMT: monitors, the name server and DMT viewer.

Figure 2¹⁴ shows the overview of the DMT architecture. In general, monitors run at background without displaying any graphics by itself, but serving data to the DMT viewer. The procedure is as follows. Firstly, a monitor is launched, and at the same time, monitor registers its name, socket information and monitored data types to the name server. The name server receives monitors' registration information and starts to keep track of the status of monitors. Then users can launch a DMT viewer. The DMT viewer will connect to the name server and requests for the list of monitors currently running with corresponding information. Then the DMT viewer will visualize these monitors' names and data objects. Users can select monitors in the DMT viewer, and the DMT viewer then requests a list of data objects' names directly from the selected monitors. Users select one or more data objects in the DMT viewer, the data objects are transferred from the selected monitor to the DMT viewer for displaying.

3.2. DMT features

To adapt large data volume situation and distributed environments, DMT is designed to have the following features:

- **Extensibility:** DMT is written in C++ using the object-oriented method. DMT has developed several sub-packages which consist of some basic classes. Take the MonServer class as an example. Users can easily develop their own monitors by inheriting the MonServer class without considering how their monitors keep running, but focusing on which data type they want to monitor and which data object they want the monitor to generate and send to the DMT viewer. And in DMT the frequently monitored data types have already been written in C++ classes, users can expediently invoke them. Even if there is a new data type to monitor, it only needs write a new C++ class and add it once in DMT. So DMT provides good extensibility for developing new monitors. Besides, DMT has defined multiple standard output file formats, such as html, xml. The external programs can easily access the monitors' output file in seamless way, which also improves DMT's extensibility.
- **Flexibility and scalability:** The name server's port number is set in advance and this port number is defined in basic classes related to monitors and the DMT viewer. So all the monitors and the DMT viewer can always communicate with the name server via this port number. It can be inferred that all the monitors, the name server and the DMT viewer can be at totally different locations. They can communicate with each other via the TCP/IP protocol. Users scattered all over the world can freely start or shutdown their own DMT viewers in their local hosts at any time without influence to each other. And DMT supports multiple users in different places to see a same monitor at the same time. This greatly improves the user experience in a distributed environment.
- **Light-weight and fine-grained:** Since DMT is required to support continuous data monitoring, its overhead on monitoring must be low enough to follow large data streams. The low overhead is achieved by light-weight monitor design and fine-grained division of monitoring jobs. Light-weight design means only some simple statistic calculations can be kept in the monitor. Fine-grained division of monitoring jobs is to assign as few data types as possible to each monitor. The optimized assignment is that each monitor has only one data type to monitor, and then the load of each monitor is the lowest.

4. OmegaMon for monitoring burst triggers

In this section, we introduce in details a DMT monitor, OmegaMon, which monitors statistical outputs of a gravitational-wave burst search pipeline, namely Omega Pipeline¹⁵.

4.1. Overview of Omega Pipeline

Gravitational wave bursts (GWBs) are one of the most interesting classes of signals being sought by GW detectors¹⁶. The core collapse of supernovae, the merger of binaries, the gamma ray bursts and other relativistic systems could be the sources of GWBs¹⁷. Figuratively speaking, a burst is just like a pulse signal, the duration of which is typically shorter than 1 second.

In LIGO, a search pipeline called Omega Pipeline is developed for GWB detection. For each interferometer of LIGO, it performs time-frequency decomposition by filtering data against bisquare-enveloped sine waves, in what amounts to an over-sampled wavelet transform¹⁸. The filtering procedure will generate a value for each tile of a time-frequency plane. This value is called the normalized energy Z ¹⁹. A tile is called a trigger if its Z energy exceeds a predefined threshold. The Omega Pipeline writes every trigger's attributes as an item into a plain text file, called trigger files. Currently each trigger is characterized with 5 attributes: the central time, the central frequency, the duration, the bandwidth and the normalized energy Z . All trigger files of different interferometers are sent to the central location of LIGO. Omega Pipeline on the central location will perform post-processing to triggers in these trigger files.

Since in a single interferometer Omega Pipeline performs a time-frequency decomposition for each 64 seconds interferometer data block, and every two adjacent 64 seconds interferometer data blocks has a 32 seconds overlap. So the relationship between each trigger file is shown in Figure 3.

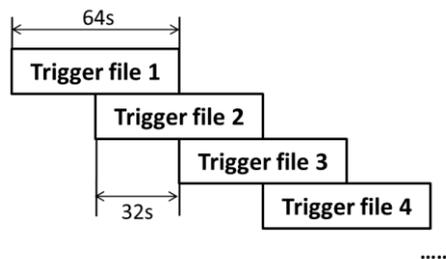


Figure 3. The relationship between trigger files.

There are many environment and equipment sources that could be related with trigger statistics, for example, an earthquake happens near an interferometer or the laser intensity fluctuates due to the unstable power of the laser generator in the interferometer. Data during the time interval in which the earthquake happens or the laser intensity fluctuates will be corrupted or contaminated, which may lead to changes of some trigger statistics. For example, if the average trigger rate in the past one hour is 500 per minute, which is two times larger than the normal trigger rate, it implies that the interferometer is highly possible not working well. This reminds operators or scientists to check the interferometer immediately and to discard the triggers in the past one hour in the post-processing of triggers. Based on the above consideration, OmegaMon is developed to monitor the triggers generated from the Omega Pipeline.

4.2. *Implementation of OmegaMon*

OmegaMon is a typical DMT monitor, which inherits the MonServer class, and has the following four typical functions²⁰.

- **OmegaMon Constructor:** This function initializes OmegaMon. It reads command line arguments and initialize the monitor process, such as serving OmegaMon's name and data objects' names to the name server, etc.
- **OmegaMon Destructor:** This function cleans up after OmegaMon is shut down. It typically flushes the pending output.
- **Process Data:** This is the primary function of OmegaMon for calculating trigger statistics. It is called repeatedly over a predefined period. In each cycle, firstly it reads trigger files in a certain time interval, the length of which is defined in the command line arguments. And then it calculates the statistics of triggers in the time interval. At last, the time interval is updated for next cycle to use.
- **Attention:** This function handles interrupt requests, e.g. the quit request.

4.3. *Key issues in OmegaMon*

During the development of OmegaMon, several key issues are considered and addressed:

- **Multi-resolution monitoring:** Monitoring requirement on the time dimension ranges from the long term trend to the short term fluctuation. Take the trigger rate as an example. We concern about the number of triggers per hour, and we also concern about the number of triggers per minute since it represents latest status. In the multi-resolution monitoring case, OmegaMon uses the maximum resolution value to set the time interval

in the process data function, so that OmegaMon can get all data objects' values for all resolutions. In this trigger rate example, the time interval is set to 1 minute.

- Name conflicts in the name server:** When there are multiple same monitors running simultaneously, only the first registered monitor's information is kept in the name server, the others are ignored due to name conflicts. In a distributed environment since users don't know each other at all; several users could all start a same monitor. But since the monitor name is set in OmegaMon's constructor function, we can solve name conflicts by adding suffix to monitor names. User can set their preferred suffix in the command line arguments of OmegaMon. When user starts a new OmegaMon, this OmegaMon will tell the name server its name with suffix. It is almost impossible to have the same suffix for different users, so it effectively avoids name conflicts in the name server.

4.4. Monitoring results of OmegaMon in the DMT viewer

Figure 4 shows four OmegaMon data objects illustrated in the DMT viewer. They are the maximum normalized energy (top left corner), the minimum normalized energy (top right corner), the mean normalized energy (lower left corner) and the number of triggers (lower right corner).

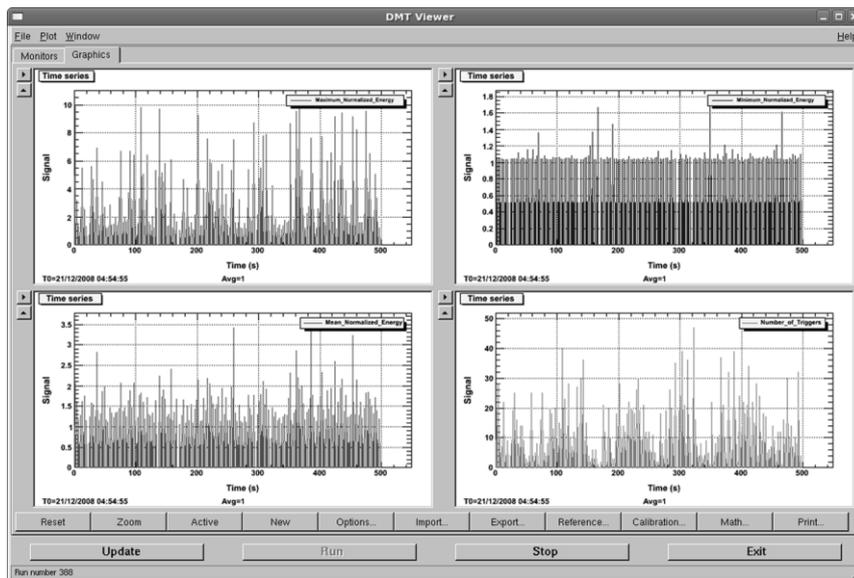


Figure 4. Monitoring results of OmegaMon in the DMT viewer.

For example, we can see from Figure 4 that the maximum normalized energy in one second interval varies acutely in the sub graph of the top left corner. Some values are obviously too high. It can be due to large transients of gravitational waves or some non-gravitational wave origins, such as loud noises. So in the post-processing, triggers in intervals of which the maximum normalized energy is too high should be discarded or tagged for further detailed analysis.

5. Conclusions and future work

We have introduced a C++ based real-time data monitoring toolkit dedicated for LIGO. It has the advantages of extensibility, flexibility and scalability. The architecture of the toolkit is introduced, and it is very suitable for distributed usage. An example DMT monitor, OmegaMon, is also introduced in details to show statistical tracking of GWB triggers generated by the Omega Pipeline.

More data objects are being developed in OmegaMon so as to provide more diagnostic information for the reference of trigger post-processing and interferometer maintenance. The OmegaMon is also being integrated into a real-time LIGO burst data computing infrastructure currently being developed at Tsinghua University.

Acknowledgments

This work is supported by National Science Foundation of China (grant No. 60803017), Ministry of Science and Technology of China under the national 863 high-tech R&D program (grants No. 2008AA01Z118 and No. 2008BAH32B03).

The authors would like to show their gratitude to Erik Katsvounidis at LIGO Laboratory of Massachusetts Institute of Technology and John Zweizig at LIGO Laboratory of California Institute of Technology for their extensive supports on this work.

References

1. P. J. Napier, D. S. Bagri et al, *Proc. IEEE*. **VOL. 82**, NO. 5 (1994).
2. A. S. Szalay et al, *ACM SIGMOD Rec.* **VOL. 29**, Issue 2 (2000).
3. J. A. Hinton, *New Astron. Rev.* **48**, 331 (2004).
4. Data Monitor Tool Project. <http://www.ligo.caltech.edu/~jzweizig/DMT-Project.html>
5. Overview of LIGO. <http://www.ligo-la.caltech.edu/LLO/overviewsci.htm>
6. B. Abbott et al, *Phys. Rev.* **D77**, 062002 (2008).
7. A. Abramovici et al, *Science*. **VOL. 256**, No. 5055 (1992).

8. B. Abbott et al, *Phys. Rev.* **D69**, 122001 (2004).
9. G. Mendell, <http://www.ligo.caltech.edu/docs/G/G070127-00/G070127-00.ppt>
10. B. Willke, P. Ajith, *Class. Quantum Grav.* **23**. S207-S214 (2006).
11. VIRGO project. <http://www.virgo.infn.it/>
12. F. Acernese et al, *IEEE Trans. Nucl. Sci.* **VOL. 55**, No. 1 (2008).
13. K. Kotter et al, *Class. Quantum Grav.* **19**, 1399-1407 (2002).
14. DMT system diagram. <http://www.ldas-sw.ligo.caltech.edu/cgi-bin/cvsweb.cgi/~checkout~/gds/doc/quickstart/MonitorAPI/monserv1.jpg?rev=1.1;content-type=image%2Fjpeg;cvroot=GDS>
15. Omega Pipeline. <https://geco.phys.columbia.edu/omega>
16. P. J. Sutton et al, *arXiv.org*. 0908.3665 (2009).
17. C. Cutler and K. S. Thorne, *arXiv.org*. gr-qc/0204090 (2002).
18. B. P. Abbott et al, *arXiv.org*. 0905.0020 (2009).
19. B. P. Abbott et al, *arXiv.org*. 0904.4910 (2009).
20. hMon status update.
http://emvogil-3.mit.edu/~shourov/fan/dmt/update_20080822.html