

LigoDV-web: Providing easy, secure and universal access to a large distributed scientific data store for the LIGO Scientific Collaboration

J. S. Areeda^a, J. R. Smith^a, A. P. Lundgren^b, E. Maros^c, D. M. Macleod^d, J. Zweizig^c

^a*Gravitational-Wave Physics and Astronomy Center and Department of Physics, California State University Fullerton, Fullerton, CA 92831, USA*

^b*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany*

^c*LIGO, California Institute of Technology, Pasadena, CA 91125, USA*

^d*Louisiana State University, Baton Rouge, LA 70803, USA*

Abstract

Gravitational-wave observatories around the world, including the Laser Interferometer Gravitational-wave Observatory (LIGO), record a large volume of gravitational-wave output data and auxiliary data about the instruments and their environments. These data are stored at the observatory sites and distributed to computing clusters for data analysis. LigoDV-web is a web-based data viewer that provides access to data recorded at the LIGO Hanford, LIGO Livingston and GEO600 observatories, and the 40m prototype interferometer at Caltech. The challenge addressed by this project is to provide meaningful visualizations of small data sets to anyone in the collaboration in a fast, secure and reliable manner with minimal software, hardware and training required of the end users. LigoDV-web is implemented as a Java Enterprise Application, with Shibboleth Single Sign On for authentication and authorization and a proprietary network protocol used for data access on the back end. Collaboration members with proper credentials can request data be displayed in any of several general formats from any Internet appliance that supports a modern browser with Javascript and minimal HTML5 support, including personal computers, smartphones, and tablets. To date 634 unique users have visited the LigoDV-web website in a total of 33,861 sessions and generated a total of 139,875 plots. This infrastructure has been helpful in many analyses within the collaboration including follow-up of the data surrounding the first gravitational-wave events observed by LIGO in 2015.

1. Introduction

The Advanced Laser Interferometer Gravitational-Wave Observatory (Advanced LIGO) [1] is a project to directly measure gravitational waves, ripples in space and time that were predicted by Einstein's General Relativity, from astronomical sources. In 2015, the project observed two black-hole merger events [2, 3, 4] inaugurating the field of gravitational-wave astronomy. The science related to LIGO is carried out by the LIGO Scientific Collaboration, with more than 1,000 members worldwide. Along with the two Advanced LIGO observatories in Hanford, WA, and Livingston, LA, there are also single observatories in Italy (Advanced Virgo) [5], Japan (Kagra) [6], and Germany (GEO600) [7]. In addition, prototype interferometers such as the Caltech 40m lab [8] are testing new technologies. Together, these instruments produce a vast store of scientific data. Other major astronomy and physics projects with complex and large data sets (and often unique data formats) have implemented custom web-based solutions to share and plot their data, e.g. [9, 10, 11]. In this paper we describe a software suite, LigoDV-web, that has been developed to provide secure and universal data access to members of the LIGO Scientific Collaboration.

1.1. Data sources and characteristics

Figure 1 diagrams the flow of data for the LIGO Hanford and LIGO Livingston Observatories. The primary instrument at each observatory is a large modified Michelson laser interferometer with two 4km long arms that acts as a transducer from the stretching and squeezing of space-time caused by a passing gravitational wave (called strain because the stretching is proportional to the length, $h = \delta L/L$) into light power fluctuations at the output of the detector. Each interferometer produces a calibrated time-series data stream known as $h(t)$ along with hundred of thousands of auxiliary data channels that record the behavior of the instrument and its environment. The Data Acquisition System (DAQ) collects analog and digital data with sample rates ranging from 1 to 32,000 Hz, calculates some processed channels such as detector status channels, and passes them to the frame writer for immediate storage. Post processing routines then create several other processed channels such as second and minute trends (max, min, mean, and root-mean-squared values per second and minute, respectively) as well as the calibrated $h(t)$ channel. With both Advanced LIGO observatories operating, the data acquisition rate is roughly 10MB/s or 1PB/year.

All of these data are written to local disk (cache system) in a proprietary format called Frame files [12] while

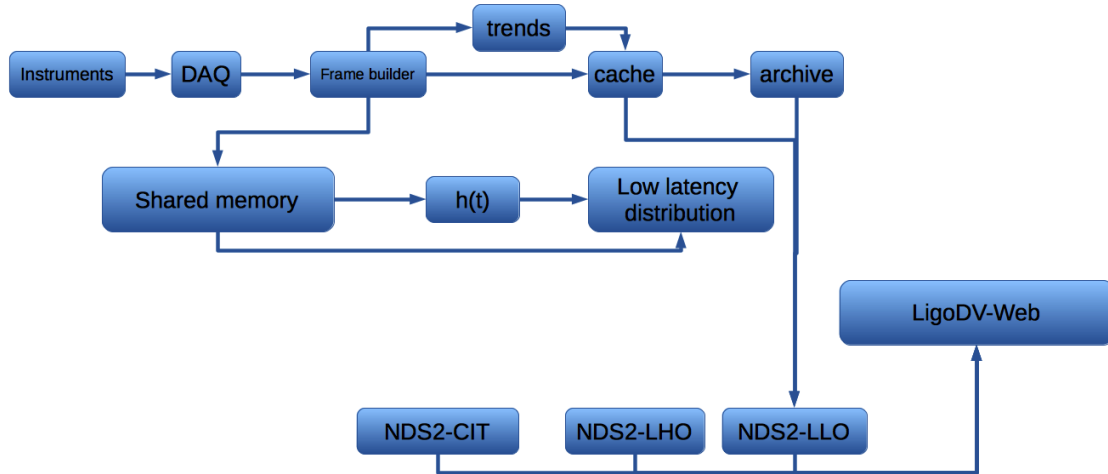


Figure 1: Data flow from instrument to the LigoDV-web application. The LIGO Hanford and Livingston Observatories collect data using local Data Acquisition Systems (DAQ), and write the data to a proprietary data storage format called Frame files. Low-latency data are written to shared memory and distributed to the local computing clusters. Frame data are cached to disk on the Tier 1 computing clusters. Both types of data are served to LigoDV-web upon request via the Network Data Servers (NDS2).

a subset of it is also distributed to a low-latency network. Great care is needed to ensure data integrity and prevention of loss so the systems involved in data acquisition are tightly controlled and processes involved are limited to critical tasks. After the data are written to the local cache system several processes come into play. Current data are kept in the cache system for 4 to 6 weeks, depending on available disk space, to allow rapid access for investigations and analysis. It is then moved to an archive file system which is hierarchical in nature with a large tape archive and limited disk space.

The LIGO Data Replicator [13], is responsible for distributing data to the many data centers and computing clusters involved in data analysis. The Tier 1 [14] data clusters are located at the Hanford and Livingston Observatories and LIGO Laboratory at Caltech. They provide data for online data collection and analysis. The observatories store data acquired on-site and some data from the other observatory used for analysis. Caltech maintains all data with latencies varying from a few seconds for the low latency network to an hour or two for some trends. The LSC has other computer clusters designated Tier 2-4 that maintain storage of data pertinent to current investigations.

Multiple Network Data Servers (NDS2) are available that provide access to data stored in frame files and in the low latency network at Tier 1 sites. NDS2 provides critical back end services to the enterprise application (LigoDV-web). NDS2 analyzes available frame files to produce channel lists and expected availability. It responds to requests for channel lists, data availability and arbitrary intervals of data from specific channels. NDS2 uses a proprietary network protocol to communicate with clients over the Internet. Users are authenticated using Kerberos against the LIGO online directory built on the Lightweight Directory Access Protocol (LDAP) [15], which contains all collabora-

tion members and approved client software packages such as LigoDV-web. Each NDS2 server provides access to frame files on disk and in the tape archive at one location. A single location may have multiple NDS2 servers for increased throughput and reliability.

1.2. LigoDV-web overview

LigoDV-web is a web-based data viewer that provides access to data recorded at the LIGO Hanford, LIGO Livingston and GEO600 observatories, and the 40m prototype at Caltech. It grew out of the LigoDV project (which itself grew out of DV6, developed for GEO600 by Martin Hewitson) which is a mature MATLAB program for viewing LIGO data with general analysis features. LigoDV is supported on Macintosh, Windows and several Linux platforms but requires a relatively expensive Matlab license and installation of the NDS2 client. These burdens, while not unreasonable, are enough to discourage casual investigation into the usefulness of LigoDV. LigoDV-web was designed to greatly simplify access for all collaboration members by eliminating local installation and providing an interface that is compatible with standard internet devices and web browsers while providing many of the features of LigoDV. LigoDV is maintained as the more appropriate tool for in depth analysis. LigoDV-web was thus designed to move all proprietary and commercial software to the server with the only requirements of the client being a modern web browser and Internet connection. The browser requirements are common HTML5 features and Javascript support. A side benefit is the availability of LIGO data on most Internet appliances including smartphones, tablets, Internet TV with browser capabilities as well as personal computers and workstations.

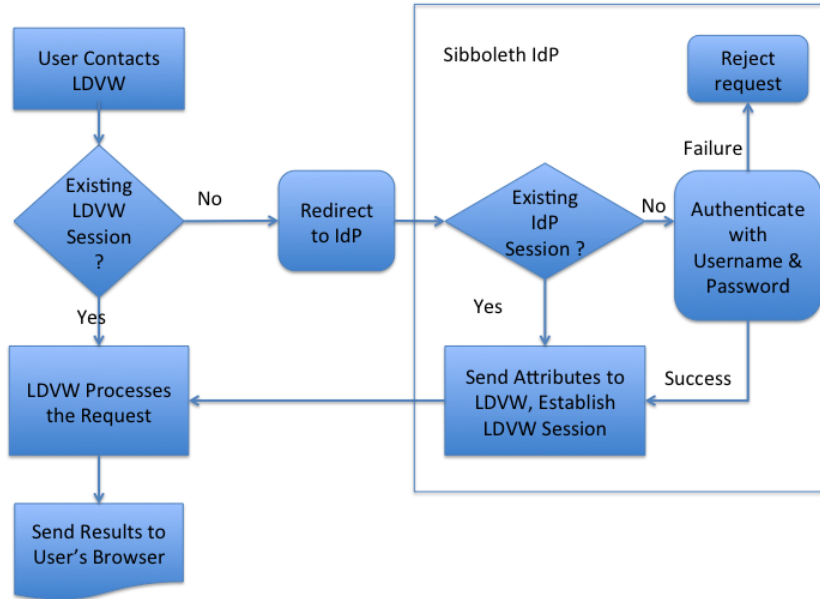


Figure 2: Flow of the Shibboleth Authentication Process. Shibboleth authenticates new sessions with a central Identity Provider to establish a new session. If the user has an existing session with any service provider there is no need to reenter credentials. If they do not have an existing session they are authenticated with username and password.

2. Methods

2.1. Authentication and authorization

LigoDV-web is technically implemented as a Java Enterprise Application running in a Tomcat 7 container [16]. It uses Apache as a reverse proxy to leverage support for Shibboleth Single Sign On services allowing members to log into one supported service and have access to all other supported services for which they are authorized. The back end relational database uses MySQL to store local indexes of all available channels, a cache of recently transferred data, all plots created, and several log files and performance statistics.

Security is an important consideration for any web site. The LSC deploys many web services with different access rules depending on the type of data presented. LigoDV-web is only accessible to members of organizations in the collaboration that have a data sharing agreement with LIGO.

LIGO uses multiple technologies to implement secure Single Sign On (SSO) authentication and authorization: Shibboleth, Grouper, Kerberos, LDAP, and the encrypted web protocol, https. When combined with a custom program called myligo, this infrastructure provides a convenient, secure and powerful system to restrict access to the assets of the collaboration (for example, gravitational-wave data) without undue burden on the authorized users or the administrators.

A new person joining a group within the collaboration creates a ligo.org account on myligo, adding their contact information and group affiliation. The principal investigator is then notified via email and prompted to log in to myligo to accept this person into their group. Once

they have been accepted, they become an official group member, are added to the LSC roster.

Shibboleth [17] is an open-source, standards-based system for SSO system authentication and authorization. Shibboleth resources are broadly divided into two categories: Identity Providers (IdP) and Service Providers (SP). A trust relationship is established between them through signed certificate based communication and sharing of certain metadata. LigoDV-web is a Service Provider in this context. As such it knows nothing about its users but it trusts the LIGO Identity Providers.

Fig 2 shows the flow of the Shibboleth authentication process. When a user contacts a Service Provider, if they do not have an active session, they are redirected to the Identity Provider. If they have a session with the Identity Provider they are redirected back to the Service Provider and the Identity Provider includes an appropriate set of attributes used to establish a user session with the Service Provider. If no Identity Provider session is active, the user is prompted for a username and password, which is validated against their LDAP record. Once a session is established with the Identity Provider, the user may establish a new session with any participating Service Provider without reconfirming their identity. This provides the convenient Single Sign On process. Subsequent requests to the Service Provider are processed immediately without further involvement of the Identity Provider.

Shibboleth also supports the use of multiple Identity Providers through a process known as Federation. In that case each institution manages the identities of its members. During the authentication process the user is redirected to the appropriate Identity Provider. LigoDV-web does not

use this feature but may in the future if other gravitational wave research groups, such as KAGRA in Japan, enter data sharing agreements with LIGO.

The establishment of an active session gives the Service Provider a trusted identity of the user and additional attributes including a name, email address and group memberships. A user's group membership authorizes their access to certain parts of the LigoDV-web service.

Grouper [18] is a distributed access management system. Privileges are granted or denied by membership in groups. When members enter or leave the collaboration their identity is assigned to the appropriate groups. In the case of LigoDV-web the groups that are allowed access to the instrument data are determined by data sharing agreements. The membership of those groups is managed by their Principal Investigators (PIs). An additional group was created for LigoDV-web, the Service Provider, and subgroups were created by the developers to authorize certain users (developers and testers) for access to administrative functions and unreleased functions for testing.

Kerberos [19] is an authentication system that is used in program-to-program communications by LigoDV-web but is not used directly by people accessing the web site. The LigoDV-web application uses Kerberos to authenticate access to the Network Data Servers for data retrieval and programs that use the RESTful interfaces to LigoDV-web may use Kerberos to initiate their Shibboleth sessions without user interaction. REST (REpresentational State Transfer) [20] is a technology that allows remote programs or scripts run in a browser to access services provided by LigoDV-web (e.g., to achieve automated production of plots for the time of a gravitational-wave event candidate).

2.2. Network data services

Frame files are designed to efficiently record large amounts of time series data from many channels in real time. A typical file contains over 100,000 channels that have been directly acquired from the instrument or derived from acquired data. The duration of a given frame file ranges between 1 second for data distributed for low latency analysis and 1 hour for minute trend data. Trend data are generated from the raw time series with minute and second sample rates. They record the minimum, mean, maximum, root-mean-squared and number of samples during each sample and are used to monitor the detectors' behavior over long periods. Most frames containing raw data are 32 seconds long. A channel is described by metadata that contains a name, sample rate, data type (integer, long, float, double, etc.) and may contain calibration factors and physical units. Channels are sampled at rates that range between once per minute and 65,536 Hz.

The efficiency of the frame file design is weighted heavily toward recording data at the expense of retrieval and analysis. Recording the data is the real time process, while most retrieval and analysis are done without the same time constraints. This storage scheme provides challenges for a service such as LigoDV-web, which presents data for a

small number of channels over a long period of time as opposed to a large number of channels over a short period. The Network Data Server (NDS) was designed to ease such data access.

There are two versions of the Network Data Server in operation, one is oriented toward real time data and is used in the control room of the observatories, known as NDS or NDS1. NDS2 is mainly concerned with archived data although it does provide access to the low latency distribution network. There is one NDS2 server at each observatory and one at the central archive maintained at Caltech. A fourth NDS2 server is used by the 40 m prototype interferometer at Caltech. The NDS1 servers are unauthenticated and are only available on the local network of the observatories. NDS2 servers use Kerberos backed by the collaboration's LDAP for secure access.

The NDS2 server and client software provides a more natural interface to the data in frame files stored on the local computer cluster. Requests are made for a list of channels or data from a list of channels during a specific time interval. The server maintains a list of channels, tracking changes over time. When a request is made the appropriate frames are located and data for the requested channels is sent to the client.

2.3. LigoDV-web

LigoDV-web can be seen as another layer of abstraction on the NDS2 model; its primary purpose is to give users access to the large store of data in a natural, convenient manner without the need to understand how or where the data are stored. A user need only know which channel and what time to access the data. Thus it is not necessary for them to know how and where the data are stored. LigoDV-web is typically used to help answer research questions about an interesting period of data, such as one containing a gravitational-wave candidate or poor quality data [21]. The purpose may be to identify the channels that are relevant to the question or to investigate the features present in those channels. Data are accessed by channel and time regardless of what type of frame contains the data and which servers host the frame files. Data are presented in graphical form or as a downloadable file for use in local analyses. The design criteria for LigoDV-web are:

1. It presents a simple, easy to use interface to the user.
2. It requires only a modern browser (with common HTML5 components and javascript). It cannot require plugins or any special components to be loaded on the client side.
3. It provides robust, secure and fast access to as much gravitational wave data as possible.

LigoDV-web is implemented as Java Servlets (small programs that are run on the server to respond to user requests) that run inside the Tomcat containers. The servlets handle the user interface, database interactions

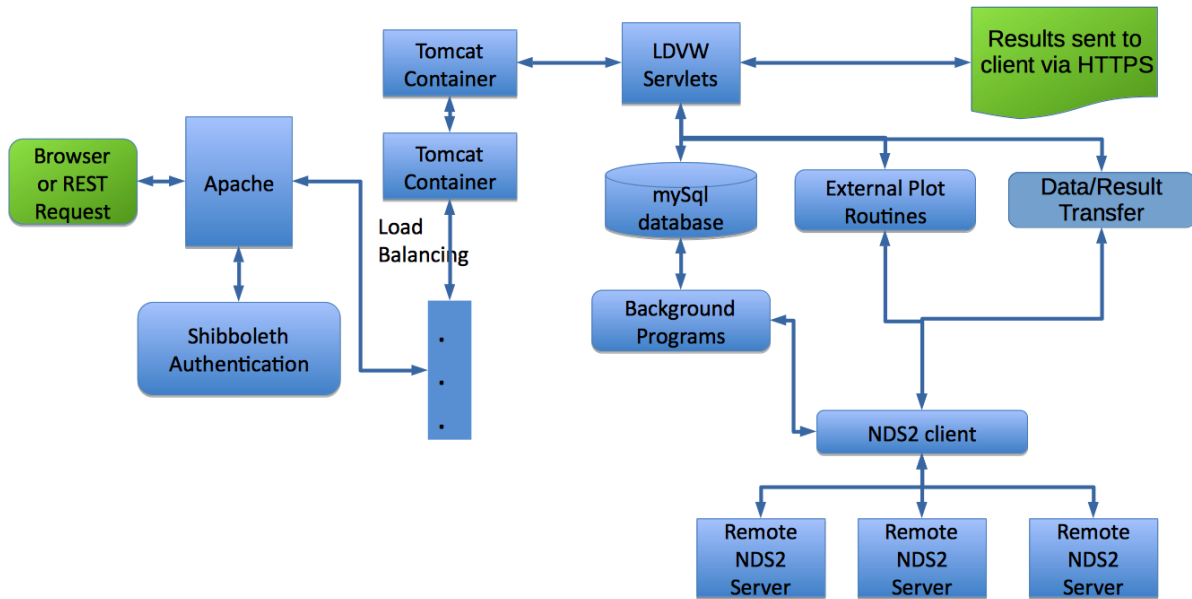


Figure 3: Overview of LigoDV-web Components. Apache is used to provide authentication and load balancing among multiple tomcat containers. The LigoDV-web servlets verify users are authorized, then handle the user requests by accessing the database and interfacing with external routines that produce plots or format data for transfers. The NDS clients retrieve the gravitational-wave data stored remotely.

and some of the simpler plot products. They also call external programs for the more complicated plot products. There are also background tasks that do not interact directly with the user but facilitate timely interactions such as maintaining the channel database tables.

Figure 3 is a high-level block diagram of the LigoDV-web application. Users make requests via a browser to a single Apache web server. Apache handles the Shibboleth negotiations and only requests from authenticated users are passed to a Tomcat container which provide the Java Enterprise Application Environment for the LigoDV-web servlets. The load-balancing feature of Apache allows easy expansion of capacity by adding more Tomcat containers to meet growing demand. One Apache server is expected to handle as many Tomcat instances as needed for any foreseeable demand.

2.4. Channels, database, and background processes

LigoDV-web uses a mySql database to control most of the user interactions. A major function of the database tables is to provide quick intuitive access to the data available via the NDS2 servers.

The NDS2 servers provide lists of all available channels and are also capable of listing only those channels that have data available during a specific time period. Considering channels available from multiple servers, the current list contains over 14 million channels. The complete list takes on the order of 10 minutes to transfer over university grade Internet connections. However, NDS2 servers provide a hash of each channel type at each server, thus we only need to download channel lists when a change is detected.

To simplify the channel list, LigoDV-web coalesces each channel into what is called a “base channel.” This represents the channel’s raw data, 5 associated minute trends, 5 second trends, any reduced sample rate data and copies at multiple locations. This consolidation reduces the effective channel list in LigoDV-web to about 811,000 base channels. This greatly increases the speed of user searches for specific channels and makes it much easier to select the channels of interest needed to answer a specific research question.

LigoDV-web also keeps a semi-permanent database of all plots and analyses created. It is less than permanent in the sense that the creator of a plot or an administrator can delete individual images but this is rarely done. Each image has a permanent link that can be shared with others or used to include analyses in other web pages without downloading and copying the images themselves. Images can also be grouped together under a user specified name (such as “my favorite seismic spectrograms”) and permanent links to these groups can also be shared or linked from other web resources. Links to images and image groups are a very popular means to share plots within the collaboration.

The other background programs that assist this service perform functions such as monitoring the status of the web server and NDS2 servers, verifying the integrity of the database, and automatic backups of the tables in the database that cannot be recreated from the NDS2 channel lists and near real-time plot generation (described below).

LIGO LigoDV-Web – v0.3.40
Welcome Joshua Smith (admin)

Home History Status Help Admin

Use the optional fields below to narrow down the list of base channels, then hit the Retrieve Channel List button.

No channels are selected.

Interferometer: L1
Subsystem: any
Sample Frequency: 64
Channel name filter: PEM CS ACC HAM2 PR | LSC PRC
 show only currently acquired

Retrieve Channel List

If you prefer the old way, click here.

Select one or more sub channels, then click on the Continue button.

8 channels match query. Displaying channels 1 through 8,

Page: 1 of 1 Select raw Select m-trend Clear all Select more Continue

Name	Raw rate(s)	RDS rate(s)	Info links
L1:LSC-PRCL_CTRL_256_DQ	256		
<input type="checkbox"/> Raw. Minute-trend: none Second-trend: none			
L1:LSC-PRCL_IN1_DQ	16384		
<input checked="" type="checkbox"/> Raw. <input type="checkbox"/> Online. Minute-trend: none Second-trend: none			
L1:LSC-PRCL_OUT_DQ	16384		
<input type="checkbox"/> Raw. <input type="checkbox"/> Online. Minute-trend: none Second-trend: none			
L1:LSC-PRC_CTRL	16384	4096	
<input type="checkbox"/> Raw. <input type="checkbox"/> RDS. Minute-trend: none			
L1:PEM-CS_ACC_HAM2_PRM_Y_DQ	2048		
<input checked="" type="checkbox"/> Raw. <input type="checkbox"/> Online. Minute-trend: none Second-trend: none			
L1:PEM-CS_ACC_HAM2_PRM_Y_OUT_DQ	2048		
<input type="checkbox"/> Raw. Minute-trend: none Second-trend: none			
L1:PEM-CS_ACC_HAM2_PRM_Z_DQ	2048-16384		
<input type="checkbox"/> Raw. <input type="checkbox"/> Online. Minute-trend: none Second-trend: none			
L1:PEM-CS_ACC_HAM2_PRM_Z_OUT_DQ	2048		
<input type="checkbox"/> Raw. Minute-trend: none Second-trend: none			

1. Select time(s)

Start time: 2016-09-06 22:10:00 (gps or YYYY-MM-DD HH:MM)
Duration: 20 (day HH:MM:SS or sec)
Group by: time

2. Select one or more plot types

Time Series (GWpy):
Spectrum (GWpy):
 Generate Spectrum

Set appropriate parameters below:
Seconds per fft: Default = 1.0
Overlap: Overlap fraction [0-1], default = 0.5
High pass filter: Cut off frequency, default = no filter
 Linear F-axis Default = log
F-minimum: Default is auto
F-maximum: Default is auto
 Linear Y-axis Default is log
Y-minimum: Default is auto
Y-maximum: Default is auto

Spectrogram:
Spectrogram (GWpy):
Coherence (GWpy):
Coherence-spectrogram (GWpy):
Omega scan (dmt_wplot):
BLRMS plot:
ODC Plot:

To proceed, click on plot or download

Plot Download format: LigoDV export Download

Spectrum: L1:LSC-PRCL_IN1_DQ
2016-09-06 22:10:36.000 - 1157235017 (20s), Fs=(16384.0 Hz), (2048.0 Hz), scppff=1.0 (bw=1.000), overlap=0.50

ASD (Counts/√Hz)

Frequency (Hz)

Legend: L1:LSC-PRCL_IN1_DQ (blue), L1:PEM-CS_ACC_HAM2_PRM_Y_DQ (red)

Figure 4: Example flow through the LigoDV-web user interface to generate a plot. 1) Users can enter one or more parameters to limit the search for channels. 2) Individual channels are then selected from the search results. 3) One or more time intervals are selected for the analysis as well as the way to group multiple channels and/or time intervals. 4) One or more plot products are chosen and the optional parameters are set. 5) Users can produce the plots specified or download raw data. 6) The resulting plots are presented as images that can be linked to or downloaded.

2.5. The User Interface

The design requirements of the User Interface (UI) are that it be as easy to use as possible and require no software on the client side other than a modern browser that supports minimal HTML5 features and have JavaScript enabled. This allows most Internet appliances to use LigoDV-web including smartphones, tablets as well as laptops and desktop workstations.

The UI, shown in Figure 4, provides the following major functions to non-administrative users:

- Search the Channel database and select a subset of channels
- Request one or more plot products (see Section 2.6) for one or more channels at one or more times
- Automatically upload images to the database and optionally view or group them
- Review previous plots and plot groups made by any user
- Check status of the NDS2 servers or channel data base
- Provide links to information about specific channels and their availability

2.6. Plot products

The purpose of LigoDV-web is to provide quick easy access to gravitational-wave data not to provide comprehensive plotting and analysis tools. Therefore we provide an expanding set of basic tools with the development focused on ease of operation and robustness. The reliability, ease of use and quality of plot products are of primary importance to our users.

Plotting is requested in one of three ways: a simple plot of one or more specific times, a recurring plot, or automatic near real-time displays. Recurring plots use on-line data with a latency of one or two seconds. Near real-time plots use a background job to create a self-updating web page that show status of the instrument a few minutes old, typically 5 minutes depending on the latency of raw frames being inserted into the NDS2 servers available data at Caltech. The main advantage of near-real time plots over recurring plots is they are served as static web pages, thus many people can view the same content simultaneously without overloading the server.

The current list of plot products available on LigoDV-web are listed below and shown in Figure 5. The majority of these call on external programs to produce the plots. The use of external programs has increased robustness because they perform one function then exit. This means memory and other resource leaks are less damaging. Primary among these is the GWpy software package [22], an inter-connected software ecosystem that has significantly

advanced GW detector characterization by enabling collaborators in the LSC to perform rigorous, complex analyses that were not previously possible, with minimal programming overhead. This package is at the heart of the LIGO Summary Pages [23] that have become the first port of call for a large fraction of the collaboration to access myriad statistics and figures or merit regarding operational sensitivity and performance of the LIGO and GEO600 detectors. Because the plots produced by this package are attractive and have benefited from feedback from many users, LigoDV-web has integrated them whenever possible. Additionally, LigoDV-web calls an external program to create “Omega” plots, time-frequency representations of data based on Sine-Gaussian wavelets [24], which are also very widely used in the collaboration.

The currently available plot products are as follows.

- Time series (GWpy): Displays the time series on a linear-linear axis. Options include configurable x - and y -axis limits, logarithmic x - and y -axis scaling, and configurable high-pass filtering.
- Spectrum (GWpy): Displays an amplitude spectral density on a log-log axis. Options include seconds per Fast Fourier Transform (FFT), overlap fraction, high-pass filter and its cutoff frequency, x - and y -axis limits and linear x - and y -axis scaling.
- Spectrogram: Displays a spectrogram (time-frequency representation with color as the z - (amplitude/power) axis) on a linear-log axis. Options include windowing function type, scaling (amplitude, power, amplitude spectral density, power spectral density), seconds per FFT, overlap fraction, normalization, smoothing and interpolation, colormaps, x -, y - and z -axis limits and linear x - and y -axis scaling.
- Coherence (GWpy): Displays the coherence versus frequency of two or more channels on a log-linear axis. Options include choice of reference channel (which channel the coherence of other channels will be computed with respect to), seconds per FFT, overlap fraction, high-pass filter and its cutoff, x - and y -axis limits and linear x - and y -axis scaling.
- Coherence Spectrogram (GWpy): Displays a time-frequency representation with color representing coherence on the z -axis. Options are the same as for coherence, with the additional options of normalization and z -axis scaling.
- Omega plot (DMT Omega): Displays a whitened spectrogram centered at a user-specified time on a linear-log plot with color representing energy. Additional plots generated by the Omega algorithm can be optionally produced such as eventgrams and time-series with no filtering, whitening, or autoscaling. Options include plot time ranges, downsampling frequency, search time, frequency and Q range, and y - and z -axis limits.

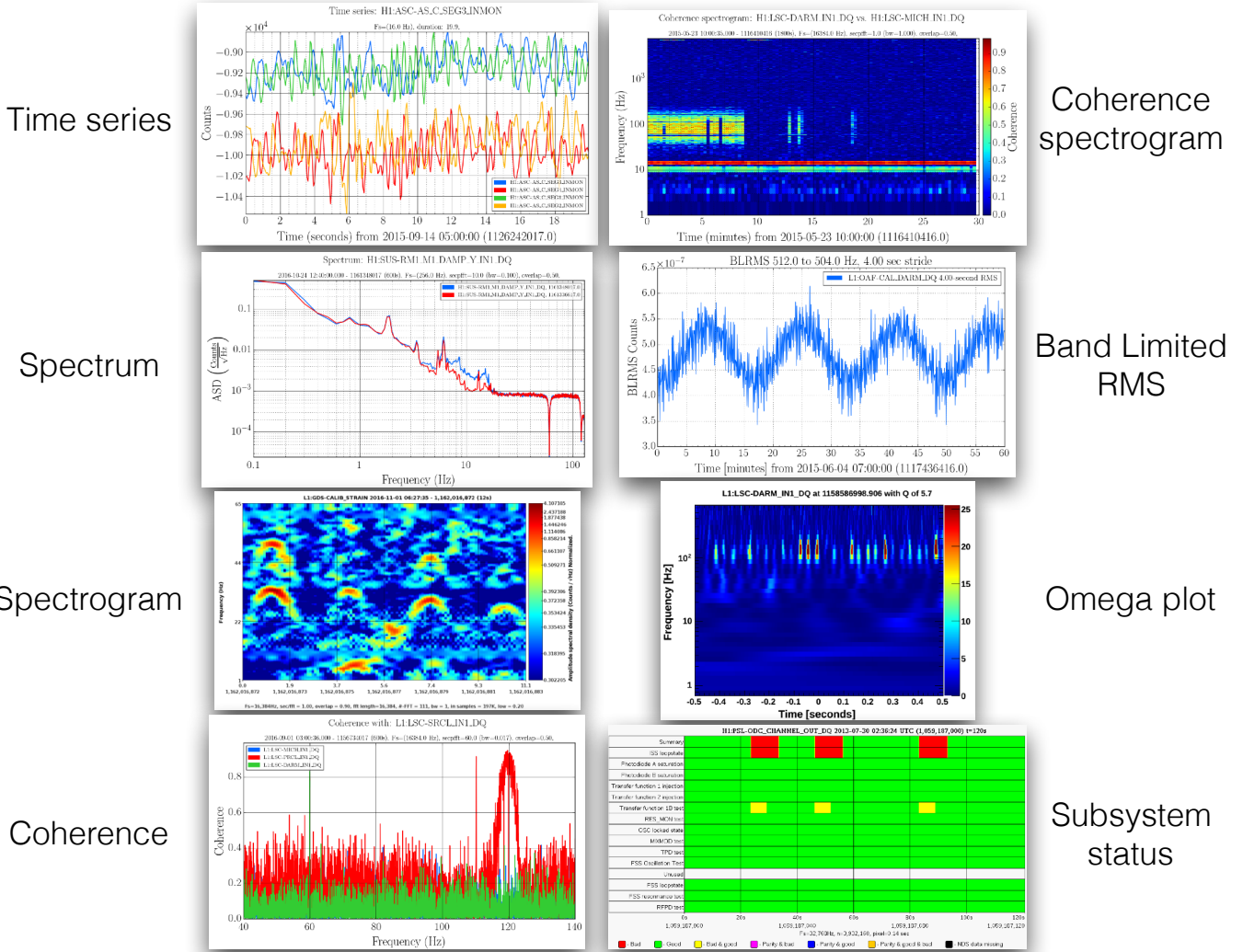


Figure 5: One example each, chosen from the database of user-made images, of the eight plot types available through LigoDV-web.

- Band Limited RMS (BLRMS) (GWpy): Displays a timeseries of BLRMS on a linear-linear axis. Options include upper and lower cutoff frequencies for the bandpass filter used, and the duration of each RMS calculation.
- Detector or Subsystem Status: Displays a colored bar bitmap timeseries of the status of the detector or a given detector subsystem, with colors indicating a good, bad or unknown status at a given time.

Timeseries and spectrum data are downloadable as comma separated values files or as MATLAB structures. Additionally, timeseries data can be converted into audio files and downloaded. Expanding data download capabilities to more of the above plot types is currently a high priority.

3. Status and future

LigoDV-web is a successful project that has provided easy, secure and universal access to a large distributed scientific data store for the LIGO Scientific Collaboration. Since being made available to the LSC in 2012, 634 LSC members have logged into LigoDV-web in 33,861 sessions and have produced a total of 139,875 plots, with the ten most heavy users having generated more than 2,000 plots each, as shown in Figure 6. A snapshot of the statistics for when people typically use LigoDV-web is shown in Figure 7. The usage varies greatly depending on how interesting recent data are. Around the discovery of GW150914 the website was visited by 40-90 unique users per day, while the mean number of unique users per day is around 10. These usage patterns imply acceptance by the community.

To build upon this, future development of LigoDV-web will be focused on expanding the number of plot products and making their options and presentation more consistent by leveraging GWpy and to make data downloadable for all plot products. Development will be done to add user preferences, such as favorite channel lists and settings. Additionally, load balancing will be expanded to handle the growing number of users and in anticipation of future public access to the LIGO data store.

Acknowledgments

We thank our colleagues in the LIGO Scientific Collaboration and Virgo Collaboration for providing many fruitful ideas, bug reports, and feature requests for LigoDV-web and for thoughtful input on this manuscript. Thanks to TJ Massinger for careful review of the manuscript leading to many improvements. This work was supported by National Science Foundation awards PHY-1104371 and PHY-0600953.

References

- [1] J Aasi et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, 2015.
- [2] B. P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016.
- [3] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Gw151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Physical Review Letters*, 116(24):241103, 2016.
- [4] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Binary black hole mergers in the first advanced ligo observing run. *arXiv preprint arXiv:1606.04856*, 2016.
- [5] F Acernese et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, 2014.
- [6] Yoichi Aso et al. Interferometer design of the KAGRA gravitational wave detector. 88(4):043007, 2013.
- [7] C Affeldt et al. Advanced techniques in geo 600. 31:224002, 2014.
- [8] Keisuke Goda, Osamu Miyakawa, Eugeniy E Mikhailov, Shailendhar Saraf, Rana Adhikari, Kirk McKenzie, Robert Ward, Steve Vass, Alan J Weinstein, and Nergis Mavalvala. A quantum-enhanced prototype gravitational-wave detector. *Nature Physics*, 4(6):472–476, 2008.
- [9] C Tsarouchas, S Schlenker, G Dimitrov, and G Jahn. Dcs data viewer, an application that accesses atlas dcs historical data. *Journal of Physics: Conference Series*, 513(3):032097, 2014.
- [10] R. L. Akeson, X. Chen, D. Ciardi, M. Crane, J. Good, M. Harbut, E. Jackson, S. R. Kane, A. C. Laity, S. Leifer, M. Lynn, D. L. McElroy, M. Papin, P. Plavchan, S. V. Ramrez, R. Rey, K. von Braun, M. Wittman, M. Abajian, B. Ali, C. Beichman, A. Beekey, G. B. Berriman, S. Berukoff, G. Bryden, B. Chan, S. Groom, C. Lau, A. N. Payne, M. Regelson, M. Saucedo, M. Schmitz, J. Stauffer, P. Wyatt, and A. Zhang. The nasa exoplanet archive: Data and tools for exoplanet research. *Publications of the Astronomical Society of the Pacific*, 125(930):989, 2013.
- [11] MAST. <https://archive.stsci.edu/>. Accessed November 2, 2016.
- [12] Specification of a common data frame format for interferometric gravitational wave detectors. Technical report, LIGO Data and Computing Group and Virgo Data Acquisition Group, 2009. LIGO-T970130-v1, VIR-067A-08, <https://dcc.ligo.org/LIGO-T970130/public>.
- [13] Jeffery Kline. ldrv1: an update of the ligo data replicator. Technical report, LIGO Scientific Collaboration, 2012. LIGO-T1200481-v1, <https://dcc.ligo.org/LIGO-T1200481>.
- [14] Duncan Brown, Stephen Fairhurst, and Abe Singer. Policies and procedures for ligo data grid tier n centers. Technical report, LIGO Scientific Collaboration, 2012. LIGO-M0900325-v13, <https://dcc.ligo.org/LIGO-M0900325>.
- [15] LDAP. <http://www.openldap.org/pub/>. Accessed October 4, 2016.
- [16] Apache Tomcat. <https://tomcat.apache.org/>. Accessed: October 3, 2016.
- [17] RL Morgan, Scott Cantor, Steven Carmody, Walter Hoehn, and Ken Klingenstein. Federated security: The shibboleth approach. *Educause Quarterly*, 27(4):12–17, 2004.
- [18] Grouper Web Services, <https://spaces.internet2.edu/display/Grouper/Grouper+Web+Services>.
- [19] JT Kohl, BC Neuman, and TY Tso. The evolution of the kerberos authentication system. distributed open systems (pp. 78-94), 1994.
- [20] The Java EE 6 Tutorial. <http://docs.oracle.com/javaee/6/tutorial/doc/gijqy.html>. Accessed: October 3, 2016.
- [21] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, M Adamo, C Adams, T Adams, P Addesso, et al. Characterization of transient noise in advanced ligo relevant to

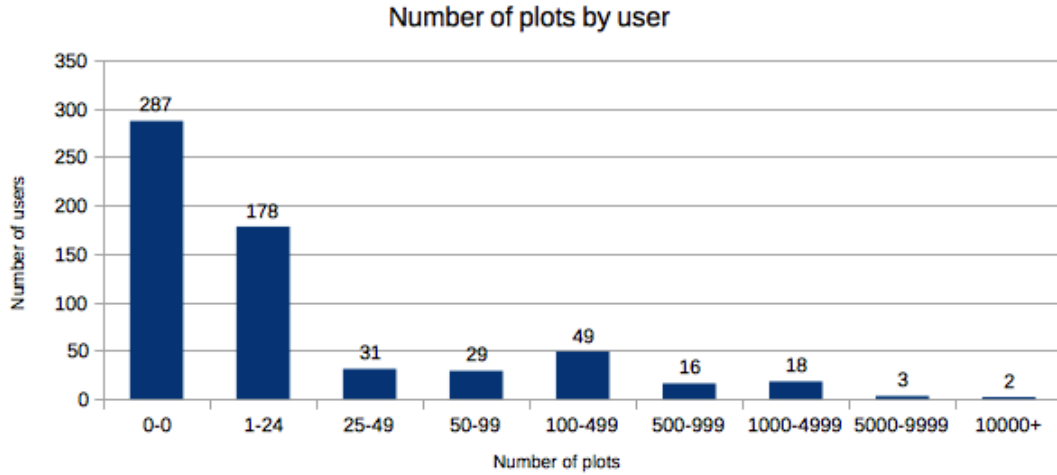


Figure 6: The number of people (unique users) that have made a given number of plots on LigoDV-web. Of the more than 1000 LSC members 613 members have used LigoDV-web. Of those 287 have not created plots of LIGO data but only visited the site likely to view links to plots made by others. Of the 326 users that have made plots about half have made less than 25 plots. Many people use the site regularly: 88 people have produced more than 100 plots each and the top users have created more than 10,000 plots each.

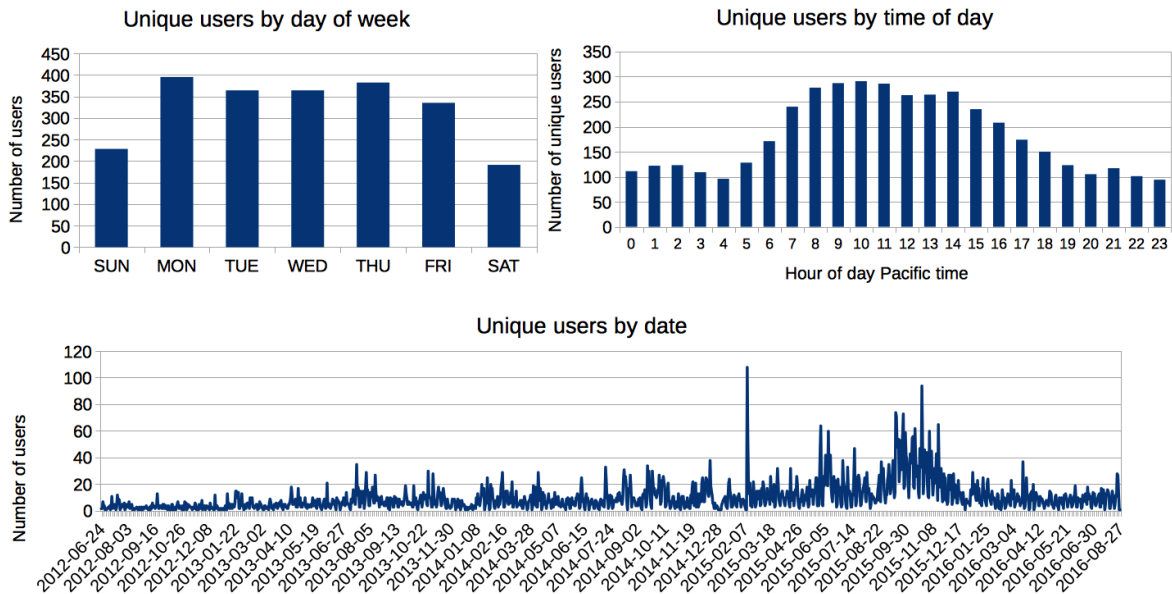


Figure 7: Usage patterns of LigoDV-web. During a typical visit to the website a user may generate a plot, check channel availability, download data, or view plots made by others. All graphs show the number of people (unique users) who visit the site for that period. *Upper left*: by day of the week; *Upper right*: by hour of the day; *Lower*: day by day since the launch of the website. The broad bump in daily users in September 2015 coincides with the first direct detection of gravitational waves. The LSC has more than 1000 members, so these plots give an indication of the acceptance of LigoDV-web.

gravitational wave signal gw150914. *Classical and Quantum Gravity*, 33(13):134001, 2016.

- [22] GWpy. <http://gwpy.github.io>. Accessed: August 25, 2016.
- [23] GWsumm. <https://github.com/gwpy/gwsumm/>. Accessed: August 25, 2016.
- [24] S. Chatterji et al. Multiresolution techniques for the detection of gravitational-wave bursts. *Classical and Quantum Gravity*, 21:S1809–S1818, 2004.