Geophysical Research Abstracts
Vol. 12, EGU2010-5629, 2010
EGU General Assembly 2010

Streaming data at the IRIS DMC: collecting, consolidating and re-distribution

Chad Trabant, Sandy Stromme, Rick Benson, and Tim Ahern
IRIS Data Management Center

The IRIS Data Management Center (DMC) began collecting near real-time streams in 2002, within a very short time real-time streams became the largest source of new data for the DMC. The DMC collects data in real-time primarily to streamline the archiving process and allow collection of data not available otherwise. Additionally, data is rapidly available for users and can be redistributed with minimal latency. Currently the DMC is collecting data from over 2.100 stations delivering more than 18.000 channels. At a rate of 33 gigabytes per day the DMC archives 12 terabytes of real-time data per year. Even though this rate of data flow is significant for seismic recordings the volume is not the most challenging aspect of our data collection effort. The real-time data system at the DMC collects data from approximately 100 network operators using 8 different streaming systems utilizing open (e.g. SeedLink, Earthworm, ISI) and proprietary (e.g. Antelope, Reftek, SCREAM!) protocols. The variety and number of connections result in a complex system. All collected data are converted to Mini-SEED if necessary and organized into a Buffer of Uniform Data (BUD). Quality control measurements are performed while data are in the BUD with the results publicly available through a web interface. After settling in the BUD for an average 18 hours data are transferred to the archive in a fully automated process. All open data in the BUD are publicly available using the open SeedLink protocol with minimal additional latency. The DMC's SeedLink service offers users a uniform interface and format for all DMC export data streams. Over 270 unique client connections download 65 gigabytes per day using this service. To handle the scale and dynamic nature of our SeedLink service the DMC has developed high performance server software dedicated to re-distribution of data streams. The DMC is making this software available to promote data exchange and increase availability. This software allows any data center writing data as 512-byte Mini-SEED records in near real-time to serve data streams using the SeedLink protocol with minimum resource and configuration requirements. With a multi-threaded design the server software scales very well with any level of usage and works under Linux, Solaris and Mac OSX operating systems.