Gaps in Archived Data from the E3 and E4 Engineering Runs

Peter Shawhan

LIGO / Caltech

LIGO Scientific Collaboration Meeting
August 15, 2001

Archived Data

The HPSS archive at CACR contains three datasets:

- "Full" raw frame files   (1 second long)
  - Assembled into large files (200 sec) using tar
  - Written to tape, sent to Caltech, read into HPSS
- Second-trend frame files   (1 minute long)
  - Copied over network
  - Assembled into large files (6 hours for LLO, 1 hour for LHO) using tar
  - Put into HPSS
- Minute-trend frame files   (1 hour long)
  - Copied over network
  - Assembled into large files (24 hours) using tar
  - Put into HPSS

All of these datasets are retrievable using getFrames

Some second-trend files were intially omitted, but have been restored

Gaps From DAQ Restarts

Each time the DAQ is restarted, some "full" frames are missing

If the downtime is longer than ~1 minute, then one or more second-trend frames are missing as well.

When DAQ starts up, it writes full-length (1 minute / 1 hour) trend frames with zeros for the times prior to the DAQ restart

This results in a loss of minute-trend data since the previous hour boundary — which typically includes a period of good running before the DAQ restart

	E3 LLO	E3 LHO	E4 LLO	E4 LHO
Length of run	270443	269204	270000	269092
DAQ downtimes	4	8	0	15
Missing raw frames	439	2804	0	2647
Missing second-trend data	451	3088	0	3247
Missing minute-trend (sec)	5940	11580	0	17040

Other Causes of Data Loss

Some archived minute-trend files were found to have been truncated

Only for data from LHO

Affected 2 hours of E3 data, 7 hours of E4 data

Found to be a synchonization problem in the copying process — fixed

During E4, one second of LHO second-trend data was found to be zero (at least for the few channels I was spot-checking) — no known reason

Notes

LIGO

This was a tedious process —
had to get data out of the archive to check it

Discovered the zeros in the trend data by reading frames into Matlab

The list of valid time intervals (i.e. data exists and is not zero) for each dataset is on the web

Hopefully, the DAQ will be more stable in the long run