

Investigating the effects of missing data upon VLBI correlation using the VLBI_UDP application

Simon Casey¹

The University of Manchester, Oxford Road, Manchester, UK

E-mail: scasey@jb.man.ac.uk

Ralph E. Spencer, Matthew Strong, Richard Hughes-Jones, Paul Burgess

The University of Manchester, Oxford Road, Manchester, UK

E-mail: res@jb.man.ac.uk

E-mail: mstrong@jb.man.ac.uk

E-mail: R.Hughes-Jones@manchester.ac.uk

E-mail: pb@jb.man.ac.uk

Arpad Szomoru

Joint Institute for VLBI in Europe (JIVE)

Dwingeloo, NL

E-mail: szomoru@jive.nl

Colin Greenwood

National e-Science Centre (NESC)

Edinburgh, UK

E-mail: coling@nesc.ac.uk

The work presented in this paper describes tests conducted to assess the effect that missing data has upon VLBI correlations. Results obtained show that the correlation is resilient under data losses approaching 20% and that the resulting correlation amplitude decreases as 1.2 times the packet loss rate. The results also indicate that while the Station Units are able to cope with at least 2 consecutive missing headers, they need to see at least 2 correct headers otherwise synchronisation is lost.

Lighting the Blue Touchpaper for UK e-Science - Closing Conference of ESLEA Project

The George Hotel, Edinburgh, UK

26-28 March, 2007

¹ Speaker

1. Introduction

The design of the data transport application VLBI_UDP[1], allows for portions of data to be lost before transmission. This can allow data packets to be dropped selectively in order to investigate the behaviour of the VLBI correlator. The correlator was originally designed to be fed with data from magnetic tapes and so is resilient to a certain amount of data loss[2]. The aim of the work presented here is to determine what effect varying rates of packet-loss has on the correlator output, and also to see if there is a limit at which the output is un-usable. These results may then be programmed back into VLBI_UDP such that it can selectively drop packets if congestion is detected, whilst having a minimal effect on the resulting correlated data.

2. Packet-dropping theory

This section outlines the data format used by the Mark5A data recorders, and explains the decisions behind choosing the tests that were run.

2.1 Mark5A data format

The data fed into the Mark5A data recorder comes from the Mark4 tape formatter. This has a variety of modes[3] which affect exactly how the data stream appears when written to a standard PC file. The data used in these tests were recorded at a rate of 256Mbit/s, 32 tracks with 1-bit sampling. Upon examining the recorded file, it became apparent that the VLBI data tracks were bit-interleaved as in *Figure 1*, such that 2 subsequent bits from any track would be separated by 32 bits in the PC file. Each track is framed in 2500 byte segments, and so 32 tracks give a ‘combined frame’ size of 80 000 bytes.

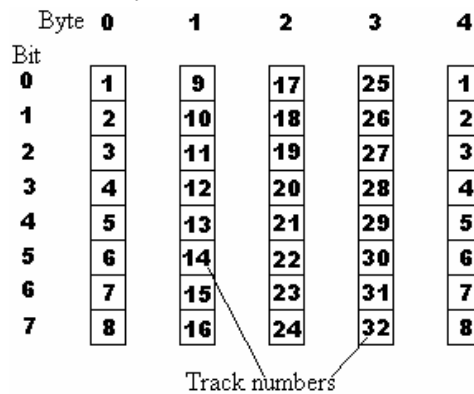


Figure 1: Visual layout of VLBI data tracks within PC file

2.2 Keeping Station Units in synchronisation

The Station Units (SUs) sit between the Mark5As and the correlator. They are a relic from the days of tape and, among other things, serve to servo the incoming data stream to keep all the streams aligned in time and check that the data has valid headers. Since the goal of the Mark5A was to emulate a tape drive, the SUs are still required. To ensure all the data streams are synchronised in time, they read the headers from each frame, perform various checks to ensure

that the headers are valid, and then read the time stamp to ensure synchronisation. Since the data frames are a constant 2500 bytes long, the SUs know exactly where in the stream the next header should be located. If the SU does not find a valid header where it expects to see one, then it is able to flywheel over that frame and check where the subsequent header should be. It is unknown how the SUs behave when there are multiple headers missing, giving rise to an investigation where only the VLBI headers are dropped. The data file used for all tests was approximately 15 minutes long, and different strategies were attempted every 2 minutes. The

<i>Minutes in file</i>	<i>Header drop pattern</i>
2-4	Drop---Keep---Keep
4-6	Drop---Keep---Drop---Keep
6-8	Drop---Drop---Keep---Keep
8-10	Drop---Drop---Keep
10-12	Drop---Drop---Drop---Keep
12-14	Drop---Drop---Drop---Drop---Keep

first 2 minutes were left intact to ensure the system was fully synchronised, then starting at 2 minutes, headers were dropped for 30 seconds followed by 90 seconds of intact data to allow for any resynchronisation if needed. The strategies chosen can be seen in *Figure 2*.

Figure 2: Strategies for dropping headers

2.3 General packet-loss

These tests are to examine the effect that loss of any part of the data stream has on the resulting correlation. As has been discussed, the SUs are able to cope with 1 missing header, but any more than 1 consecutive missing header and synchronisation may be lost resulting in a temporary loss of correlation.

If, therefore, it is assumed that the SUs will stay in synchronisation when every other header is present, then the following can be used to determine a theoretical maximum packet-loss rate, below which synchronisation will be maintained. The 32 recorded data frames, when interleaved together, create a large 80 000 byte ‘frame’. This can then be broken down into 79 360 bytes of VLBI data sandwiched between a 384 byte header and 256 byte footer. VLBI_UDP allows for 1444 bytes of user data in each UDP packet, hence one 80 000 byte frame will need 55.4 UDP packets. Working on the case where losses are statically distributed, if no more than 1 in every 57 packets (1.75%) is lost, then 2 consecutive headers will never be lost. With this in mind, loss rates of 0.5%, 1%, 1.5%, 1.75%, 2% and 2.25% were chosen for the initial series of tests. Based upon the results of these tests in September 2006, follow up tests were performed in December 2006 at rates of 1%, 2%, 5%, 7.5%, 10%, 15% and 20%.

3. Experimental details

The correlation tests used data provided by JIVE, taken during an NME experiment from 3 antennae: Jodrell Bank, UK; Westerbork, NL; Effelsberg, DE. Only the Jodrell Bank data were modified, giving a total of 2 modified baselines and 1 unmodified.

The Jodrell Bank data were taken from a Mark5A disk pack and written to a standard PC file using the Mark5A software suite. The resulting file was ~28GB and transferred to a PC (Huygens) at JIVE which had a RAID array large enough to accommodate all the subsequent data sets. VLBI_UDP was run in a stand-alone mode, where it simply reads 1444 byte size chunks from one file, filters them through a packet dropping function, and then writes the resulting data out to another file.

For each of the loss rates listed previously in the September tests, two output files were created; the first where single packets were dropped at a constant rate; the second where packets were dropped in bunches. For the December tests, 4 files were created for each loss rate. Two files created as before, with a further 2 created by replacing the missing data with a fill pattern 0x11223344 instead of random data. The fill pattern is processed by the Mark5A's I/O board, and when the pattern is observed, the interface board signifies to the SUs that the data is invalid. The output files were placed back onto Mark5A disk packs, ready for correlation.

4. Results

4.1 September 2006

Unfortunately due to a miscalculation on the position of the headers within the file, the header dropping test was not a success. The varying loss rates however, produced good results. *Figure 3* shows a typical plot produced from one of the runs. It can clearly be seen at 32 minutes (the file starts at 30 minutes) the amplitude differences shoot up as the packets are being dropped. It became apparent that even at rates of 2.25% the SUs had no problem with the packet losses, thus giving rise to the higher loss rates used in the December 2006 tests.

By averaging over all the amplitude differences for each run, it is possible to create a linear plot of packet loss rate against amplitude decrease; *Figure 4* is an example of such a plot.

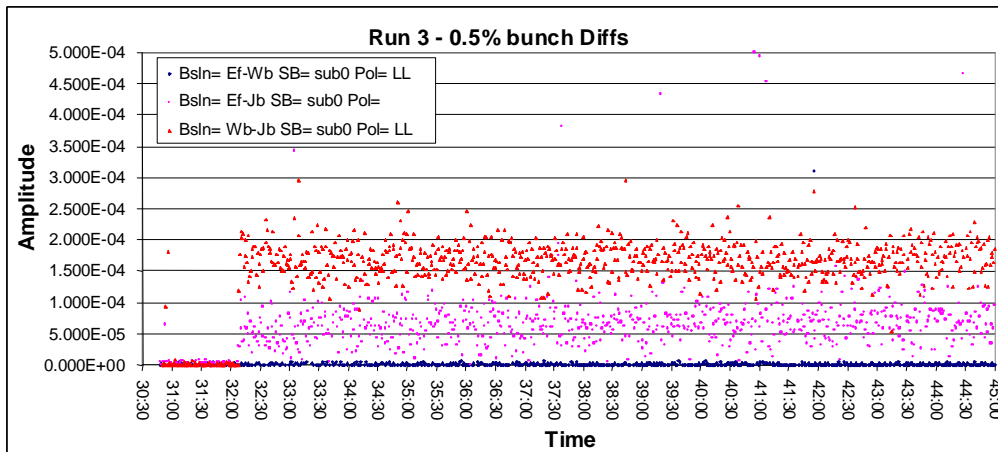


Figure 3: Plot of modified correlation amplitude subtracted from unmodified amplitude

4.2 December 2006

Most of these runs produced good results, but produced spurious amplitudes, thus badly skewing the average plot. Upon closer investigation, it was apparent that the correlation weighting values for the spurious points were close to zero whereas they should be close to 1 for good data. Recalculating the averages using a weighted average based on the correlation weight suppressed the spurious points and brought the average plot back in to line. The set that had this problem was at 20% loss rate, and this would have been caused by the SUs being unable to cope with the amount of lost data and so feeding invalid data to the correlator. It would therefore appear that loss rates of 15% and below can be handled adequately by the SUs and correlator.

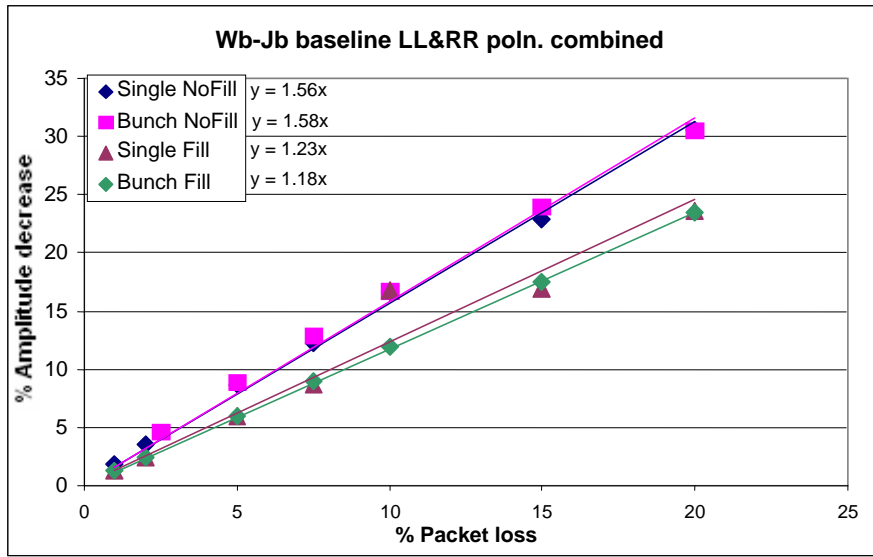
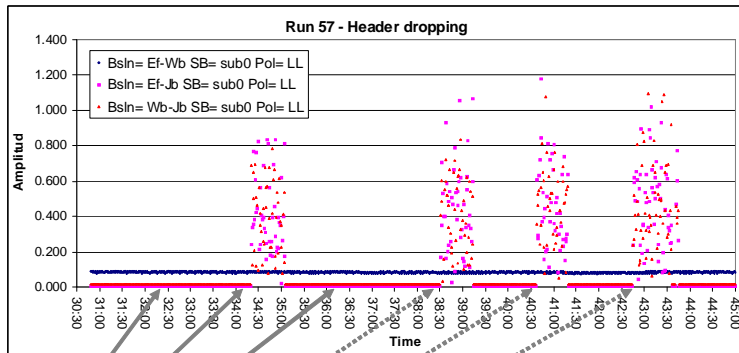


Figure 4: Amplitude decrease with packet loss for all 4 dropping algorithms, with/without fill pattern, single loss/ bunch loss, Left-Left and Right-Right polarisations averaged together

From figure 4, it is clear that by replacing missing data with the Fill Pattern instead of random data, the correlator is better able to compensate, the slopes decreasing from 1.5 without the Fill Pattern to 1.2 with. Losing packets singly or in small bunches appears to have a negligible affect on the amplitude.



Minutes in file	Header drop pattern	Result
2-4	Drop---Keep---Keep	✓
4-6	Drop---Keep---Drop---Keep	✗
6-8	Drop---Drop---Keep---Keep	✓
8-10	Drop---Drop---Keep	✗
10-12	Drop---Drop---Drop---Keep	✗
12-14	Drop---Drop---Drop---Drop---Keep	✗

Figure 5: Results of header dropping tests

5. Conclusion

The work presented above shows that the Station Units and correlator are able to cope with packet loss rates of up to 20% without causing a serious effect on the correlated result. If the packet dropping shown here is implemented during a real VLBI transfer, then it should be

possible to drop perhaps 10% of a 1024Mbit/s VLBI stream and transmit the remainder over a Gigabit Ethernet link.

6. Acknowledgements

I'd like to thank JIVE for giving me access to the correlator and their help in creating the datasets and processing the correlator output, ESLEA for overseeing the UKLight links, and Jodrell Bank Observatory for giving me access to the initial Mark5 recordings.

References

- [1] S. Casey et al., *Investigating the effects of missing data upon VLBI correlation using the VLBI_UDP application*. In proceedings of Lighting the Blue Touchpaper for UKe-Science – Closing Conference of ESLEA Project, POS(ESLEA)038, 2007
- [2] R.E. Spencer et al., *Packet Loss in High Data Rate Internet Data Transfer for eVLBI* in proceedings of 7th EVN symposium. Toledo 12-15 October 2004
- [3] W. Aldrich, *Mark5 memo #013: Mark5A Operating Modes, Mark 5 Memo Series*, viewed 26 April 2007, <http://www.haystack.edu/tech/vlbi/mark5/memo.html>