Efficient Downloading of NSO Service Mode Data

The NSO Service Mode data is presently made available via anonymous FTP from the NSO/NISP servers located at: **ftp://vso.nso.edu/DST/DST_Service_2/**

The data are organized in subdirectories according to both date (`/DST/DST_Service_2/date/`) and program identifier (`/DST/DST_Service_2/programs/`)
The directory hierarchy under the program organization actually contains links to directories or files under the date hierarchy (to avoid having two copies of the data).

In order to achieve the best data transfer rates, it is generally necessary to execute several file transfers simultaneously. This overcomes the effects of throttling on each single connection (imposed not by the server, but by intervening routers) that often limits the transfer rate for a single file to less than 250 kbytes per second.

At the same time, the data are organized in a hierarchy of subdirectories for different dates, instruments, channels, and observation type (instrument dependent), and it is important to maintain this directory structure when copying the data.

We have identified a program that is able to achieve both of these goals, transferring 10 or more files simultaneously and recreating the directory structure on the local machine. The program is called **aria2c**, and is available at:

http://aria2.sourceforge.net/

as source code and as an executable for a variety of operating systems. The program is also available as a package from the standard repositories for different Linux systems.

The program can be provided with an input text file containing a list of FTP URI's to be transferred and the program will process them sequentially, recreating the necessary directories as it proceeds. Each program directory on the FTP server contains the required lists of all files that can be directly fed to the **aria2c** program. For example, for the P490 program, the list of files would be available here:

```
/DST/DST_Service_2/programs/p490/uri.list.490.txt
```

The file listings for the data related to the desired observing program can be transferred quickly with any normal ftp client (or with **aria2c** itself), and then provided as input to the program:

```
aria2c --split 10 -j 10 -x 10 -R --file-allocation=none -i uri.list.p490.txt
```

This will start up to 10 simultaneous transfers of the data files listed in the provided input file (this number can be adjusted depending on overall network capacity). There are many more command-line options available, which are explained in the online manual:

http://aria2.sourceforge.net/manual/en/html/aria2c.html

Using this program, we have achieved sustained transfer rates greater than 10 MB/sec, which allows 500 GB to be transferred in about 14 hours. The file will print out status lines (which can be suppressed) that indicate the achieved aggregate transfer speeds at start of the line in text of the form: `[DL:676KiB UL:0B]` where "DL" indicates the

download speed in kilo- o mega-bytes per second. We would be interested in hearing the typical transfer rates you are able to achieve with this approach at kreardon@nso.edu.

Please note that the use multiple simultaneous transfers has the potential to saturate the network capacity on your local network, reducing the available bandwidth for other users. Monitor your usage to avoid raising the ire of your colleagues and system administrator!

<div align="right">

kevin reardon
19 March, 2014

</div>