

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HAYSTACK OBSERVATORY

WESTFORD, MASSACHUSETTS 01886

Telephone: 978-692-4764

Fax: 781-981-0590

16th December 2004

TO: Distribution
FROM: D.Lapsley & J.SooHoo
SUBJECT: e-VLBI Production Process and Configuration Files for K5 to Mark5 Experiments (DRAFT)

Introduction

This document describes the production process and tools used for the e-VLBI transfer of data recorded using K5 format and stored at Japan to the US where it is converted and stored on Mark5 disc pack.

All e-VLBI K5 to Mark5 Experiments must follow the procedure outlined in this memo. Any changes in the procedure must be reflected in this document. Any feedback is welcome. Additional production e-VLBI experiments require a separate document and procedure to describe them (of course, they may borrow from this document as much as is useful).

Hardware Platform

Experiments will be conducted using the e-VLBI Production Facility located at MIT Haystack Observatory. This includes:

- Two high speed servers for the transfer and conversion of data:
 - Turtle (1.266 GHz Intel Pentium III Dell PowerEdge 2500)
 - Enterprise (a dual 2.4 GHz Xeon Dell PowerEdge 2600)
- Two 1.0 TB Lacie Bigger Extreme Firewire 800 external hard drives for the temporary storage of data
- A Mark5 (Mark554) for the transfer of data from system disc to Mark5 disc pack.
- 2 Dell Powerconnect 5224 Managed Ethernet Switches

Software Platform

All servers involved in these e-VLBI experiments will run Linux RedHat Enterprise 3.

Server Network Configuration

All servers will be configured to allow passwordless ssh sessions between them.

The following servers will be involved in the transfer of data:

Server Name	Server IP Address	Server Location	Server Description
vlbi2.jp.apan.net	203.181.194.50	Kashima	File server for K5 files.
copc0053	210.229.104.69	Tsukuba	File server for K5 files.
Turtle	140.173.174.200	Haystack	File storage, e-VLBI process server and conversion processing.
Enterprise	140.173.174.201	Haystack	File storage, e-VLBI process server and conversion processing.
Mark5-54	140.173.174.203	Haystack	Transfer of data to Mark5 disc pack.

Transfer Process

The following diagram shows the steps involved in the transfer, conversion and storage of data.

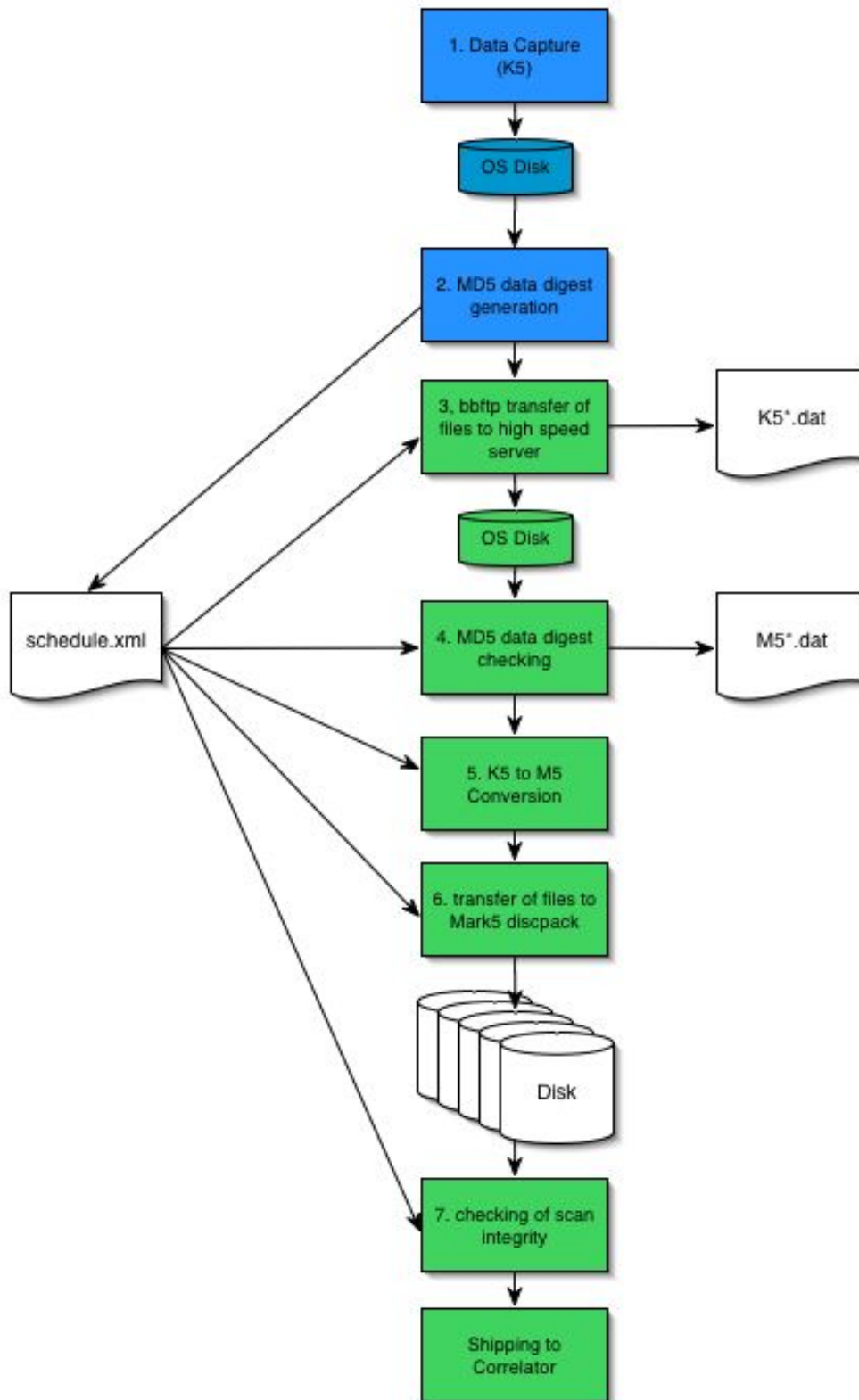



Figure 1 K5 to Mark5 e-VLBI Production Process

The following provides a high level view of the steps involved in a production e-VLBI K5 to Mark5 experiment:

1. Data is first captured to K5 file using the K5 Data Acquisition System and is stored to OS Disk on a server in Japan
2. The data has MD5 checksums run against it and the checksums stored to check data integrity after the transfer phase.
3. K5 Data is transferred from the Japanese server to a server at Haystack where it is stored on system disk.
4. K5 Data integrity is then checked by calculating MD5 checksums for each file and comparing against those generated in step 2.
5. K5 Data is then converted to Mark5 format and stored on system disk.
6. Mark5 Data is transferred from system disk to Mark5 disc pack.
7. Mark5 disc pack is then shipped to correlator for processing.

Steps that are shaded in blue are carried out in Japan. Steps shaded in green are carried out at Haystack. The e-VLBI process is controlled by a single XML configuration file generated prior to the e-VLBI experiment.

Directory Structure

All of the software used for the transfer and processing of e-VLBI data will be stored on  Enterprise in the /scripts directory.

Each experiment will have two directories located on Enterprise dedicated to the storage of e-VLBI data and configuration files. The directory structure will be as follows:

```
/scripts/  
/raw/<experiment name>/etc  
/raw/<experiment name>/var  
/raw/<experiment name>/data  
/processed/<experiment name>/data
```

The "raw" and "processed" partitions are stored on separate Lacie hard drives.

Each experiment has its own subdirectory in the "/raw" directory. Within each experiment directory, there are three subdirectories. The "etc" directory contains an XML configuration file that describes the entire e-VLBI experiment: which files are to be downloaded, conversion parameters, etc. Refer to section FIXME for a complete description of the configuration file. The "var" directory contains any log files generated as well as the "checkpoint.log" file. This file is used when recovering from an interruption to an experiment. The "data" directory holds all of the unconverted K5 data that is downloaded from the servers in Japan.

Each experiment also has its own subdirectory in the "/processed" directory. The "data" subdirectory contains all of the data that has been converted to Mark5 format and is waiting to be written to Mark5

disc pack.

Configuration File Format

As discussed previously, the entire e-VLBI process is controlled by a single XML configuration file. In this section, we describe the format of the configuration file.

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- New document created at Tue Dec 07 16:12:42 EST 2004 -->

<schedule experiment_name="T2000" experiment_type="K52M5_NRT" version="0.1">
  <source ip="192.168.1.1" type="server" user="vlbi">
    <bbftpd>/home/vlbi/bin/bbftpd</bbftpd>
  </source>
  <destination ip="192.168.1.2" port="2620" type="mark5" user="oper"></destination>
  <scan format="k5" mjd="139" name="132173000">
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173000a.dat
    </file>
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173000b.dat
    </file>
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173000c.dat
    </file>
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173000d.dat
    </file>
  </scan>
  <scan format="k5" mjd="139" name="132173020">
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173020a.dat
    </file>
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173020b.dat
    </file>
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173020c.dat
    </file>
    <file checksum="123456789abcdef" checktype="md5" format="unix" length="80000000">
      /k51a/ad6/t2029/O132173020d.dat
    </file>
  </scan>
</schedule>
```

Figure 2 Sample XML Configuration File

Figure 2 shows a sample configuration file. The table below describes each of the configuration tags:

Tag	Description	Attributes
<schedule>	Delimits the configuration of a single e-VLBI Mark5 to K5 production transfer.	" version ": specifies the version of the configuration file. This will help to prevent configuration file/software version mismatch in the future. " experiment_name ": specifies the experiment that the schedule is associated with. " experiment_type ": specifies the type of experiment. Currently, only one type of e-VLBI experiment is supported: K52M5_NRT. It is expected that in the future the number of e-VLBI experiment types supported will increase.
<source_host>	Delimits the information for a particular host.	" type ": can be either server or mark5. Specifies the type of server. " user ": user account on server. " ip ": IP address of server. " port ": for mark5 servers this is the port to which commands should be sent.
<bbftpd>	Delimits the location of the bbftpd server on the source_host.	None
<destination_host>	Delimits host information for the destination host.	Has the same attributes as <source_host>
<scan>	Delimits a single scan object.	" format ": k5 or mark5. Describes the format of the scan. " mjd ": contains the modified julian day for this scan. " scan_name ": the name of the scan. For K52M5_NRT, exactly four file objects will be contained within one scan object.
<file>	Delimits information describing a single scan file. The full path to the file on the source host is delimited by the file tags.	" format ": unix or mk5. Defines the current format of the file (either a UNIX file or a mk5 scan). " md5sum ": md5sum of scan as calculated prior to transfer of data. " length ": length of scan in bytes.

The XML file in Figure 2 is described by the following Document Type Definition:

```
<!ELEMENT schedule ( source | destination | scan )*>
<!ATTLIST schedule
experiment_name CDATA #REQUIRED
experiment_type CDATA #REQUIRED
version CDATA #REQUIRED
>
<!ELEMENT source ( bbftpd )>
<!ATTLIST source
ip CDATA #REQUIRED
type CDATA #REQUIRED
user CDATA #REQUIRED
>
<!ELEMENT bbftpd ( #PCDATA )>
<!ELEMENT destination EMPTY>
<!ATTLIST destination
ip CDATA #REQUIRED
port CDATA #REQUIRED
type CDATA #REQUIRED
user CDATA #REQUIRED
>
<!ELEMENT scan ( file )*>
<!ATTLIST scan
format CDATA #REQUIRED
mjd CDATA #REQUIRED
name CDATA #REQUIRED
>
<!ELEMENT file ( #PCDATA )>
<!ATTLIST file
checksum CDATA #REQUIRED
checktype CDATA #REQUIRED
format CDATA #REQUIRED
length CDATA #REQUIRED
end_byte CDATA #IMPLIED
start_byte CDATA #IMPLIED
>
```

Figure 3 e-VLBI XML DTD

Running an Experiment

1. Pre-experiment checks:

a. Network Reachability

- i. Check that all servers are reachable from Haystack using "ping"
- ii. Check that the path from Haystack to Japan (in both directions) traverses the high speed network using "traceroute".
- iii. Check that it is possible to "ssh" from turtle and enterprise into vlbi2 and copc0053 without a password using the following command: "ssh user@server -x "ls"". You should see a directory listing of the "users" home directory on the remote server.

b. Storage Space

- i. Ensure that there is sufficient storage space on Turtle and Enterprise
- ii. Ensure that a freshly SSease'd disc pack is inserted in Mark554

- c. Software
 - i. Ensure that Mark5A is running on Mark554.
 - ii. Ensure that bbftp is installed on enterprise, turtle, vlbi2 and copc0053.
- d. Configuration
 - i. Ensure that you have the experiment XML configuration file
 - ii. Validate the XML configuration file using the validation tool
 - iii. Run the "init.sh" script to create the experiment directories on Enterprise in the /raw and /processed directories.
 - iv. Store the XML configuration file in the /raw/<experiment name>/etc/ directory
- 2. Experiment execution
 - a. Log in to enterprise as "evlbi" user.
 - b. Execute the following commands:
 - i. "cd /raw/<experiment name>/etc"
 - ii. nohup /scripts/run.sh --config=<experiment name>.xml --mode=init
 - iii. tail -f nohup.out
- 3. Experiment verification
 - a. Verify data transfer integrity by comparing Japanese and Haystack md5sums
 - b. Verify scan integrity by checking scan_check.res report
- 4. Recovery from interruption:
 - a. Log in to enterprise as "evlbi" user
 - b. Execute the following commands:
 - i. "cd /raw/<experiment name>/etc"
 - ii. nohup /scripts/run.sh --config=<experiment name>.xml --mode=recover
 - iii. tail -f nohup.out
 - c. The experiment should recover transfer/conversion at the point from which it left off.