New Structure of OLG DC

Changes in Hard-, Software and Data Handling

G.Stangl[1]

Federal Office of Metrology and Surveying

1. Introduction

Switching to a new building of the Austrian Academy of Sciences (AAS) changes in the design of the EUREF LDC OLG became necessary. This does not only concern new network addresses and new hardware, the integration into the LAN of the Space Research Institute made it mandatory to adjust the security aspects too. Additionally it was felt necessary to build up a mirror DC for the EUREF RDC BKG to allow users (including the OLG itself) to have an alternative in the case of outage. At last refinements in the data checking procedures and the web services are under progress. Some important aspects for the future to make the data services more efficient and more interesting are discussed too.

2. New Structure

The OLG DC consists of some components, the data server including the data checker, the public ftp server and the administration part. The data server and the administration part are located at the intranet while the public ftp server (ftp://olggps.oeaw.ac.at) was placed in the Demilitarised Zone (DMZ) between the outer and the inner firewall (Figure 1). All processes are running automatically including a basic exception handling. The communication interface to the manual administration consists of several log files containing the messages from normal data handling, warnings and errors.

Access from the internet is possible via anonymous ftp to the **public server** (Intel Pentium III Raid system under Solaris 2.6). Some directories are opened for writing, especially those where the RINEX data (daily and hourly) should be uploaded by the station managers or operational centres. The other directories are read-only, containing RINEX data, OLG products, data handling programs and occasional data for clients with special demands (e.g. raw data). For security reasons the input directories are cleared every hour to hamper crackers to establish programs there.

Additionally to those data the **data server** (Intel Pentium II under Linux) contains a storage area which serves for medium-range online data holding, for RINEX data (30 sec) typically 1-2 years. The data checker contains intermediate directories for new incoming data, data to be forwarded for storage and bad data to be held back. The data checker is started automatically

[1] Address: Space Research Institute, Department of Satellite Geodesy; Schmiedlstrasse 6, A-8042 Graz, Austria; Tel. +43 316 4120 712; Fax +43 316 4120 790;
e-mail guenter.stangl@oeaw.ac.at

every hour for daily RINEX files and four times the hour for hourly RINEX files. Because the hardware is pretty old and small (Memory 64 MB) the number of processes had to be held very low. It has also to be considered that almost all of the services access the same partition of one of the hard disks. The daily RINEX files are available with a delay of one hour, the hourly ones within 15 minutes, therefore. The long-term storage of data on MO-disks, only daily RINEX files, will be handled separately by another PC, running under Windows NT.

The **administration part** is the link between the inner and outer data handling. It contains scripts and programs move the data in both directions and control and organize the storage. The administration is acting with root authority and has writing access to all directories. As the public ftp server of OLG is only a part of the common public server the administration is not root but a user with more rights than usual there. The administration programs act in the same rhythm as the data check programs, delayed by 5-10 minutes, however. Under normal conditions (10-50 files) this delay is sufficiently long to allow the for check programs to finish.

3. BKG mirror

While the OLG LDC has to handle about 10 EUREF-stations and about 15 non-EUREF stations, some of them providing hourly data, the OLG analysis centre needs much more stations at a regular basis (about 50). All additional data were retrieved from the BKG RDC some days before their usage which means with a delay of one week. At few times the BKG outage made these downloads impossible. To provide data at an earlier basis and to serve as a fall-back DC in case of BKG outages the BKG mirror service was started (Figure 2). Every hour RINEX files are downloaded. The first download of every day loads down the files of the former day, assuming that the bulk of data has already arrived at BKG. The following hours try to download those files which were still missing using the OLG DC storage list and the list of all possible stations. The update routines cover two days because experience shows that more than 90% of all files arrive within two days. Additionally each week one update routine is started manually looking back for very late files, four weeks approximately. The hourly files should be retrieved too, but the hardware (memory and hard disk) is presently too weak to perform all check routines to all files of OLG DC and BKG mirror. This will be resolved in the next months.

All downloaded RINEX files are checked by the same checking routines used for other OLG files. The files rejected are stored in a trash box and not sent to the public server. Thus the BKG mirror contains about 0.3% less files than the BKG. All correct files are stored for 1-2 years. If there is a need, redundant long-term storage has to be discussed.

This procedure is not yet very elegant and there may be some differences in the data holding. To enlarge the contents for e.g. GLONASS data and products seems to be no problem provided the hardware will be upgrade. The improvement with respect to the efficiency of communication between BKG and OLG and a shorter latency is more crucial. A real mirror implies the same directory structure together with a quick and sophisticated data management. There may be several solutions. One may be a frequent scanning of the server directories at BKG by the client software at OLG. Secondly a storage area at OLG may be provided for BKG via nfs. Thirdly there are possibilities for remote copies or retrieving procedures using the unix program rsync together with a secure shell like ssh. All procedures require not only technical considerations like hardware, software and network requirements, but have also security aspects.

4 Data Checks

For further usage the DC has to provide files which can be used, e.g. by analysis centres. A basic data check is therefore necessary. Because the EPN checks the consistency of the RINEX headers with the log sheets at Brussels the checks should be confined to readability and usefulness of the contents. At the OLG DC a cascading checking procedure was introduced using some small programs (Figure 3). Only files which stand all checks are forwarded to the public server. The rest will be stored in a "trash box" and inspected manually. It is intended to implement an automatic information via e-mail to the data providers about those files.

The first check considers whether the uncompressed file can be unix compressed or if the already compressed files can be uncompressed. The first case occurs if the provider omits the necessary extension (usually Z) delivering a compressed file which is assumed to be uncompressed. The second case may come from a broken transmission. Those files are not usable at all.

The second check consists of compressing and decompressing the file with the Hatanaka algorithm. Because all DCs use the Hatanaka compression to save disk space and transfer time the compression is mandatory. Compression stops if there is no time information in the header or there are some additional control characters in the file. Unfortunately it stops also at lines which are correct and may come from file splicing or changes in the observation types. A file which cannot decompressed or only partially because of transmission or generating errors are claimed as "broken". Notwithstanding that a part may be readable it contains a crucial error.

The third check is to test if the naming convention of the file is correct with respect to its contents. The marker name should be equal to the first four characters of the file name. The data within the file have to correspond to the day of the year in the file name. Some downloading programs seem to store old data into new files in cases of bad date or gps week retrieval. In every case the file is not usable because there is only a guess whether the outer or the inner information might be correct.

The last step of checking consists of the program teqc by UNAVCO. In addition to formal errors teqc detects epochs which are not ordered correctly. It seems that sometimes downloading programs or RINEX converters tend to duplicate epochs or to fill gaps by mixing epochs for minutes or hours. Most of the analysis programs cannot handle those data correctly, therefore it is better to manage those errors within the DCs because all users should have a right on data which are formally correct.

5. Further Improvements

There exists also a web server (http://gps.tu-graz.ac.at) which provides information and manages orders for RINEX data of some Austrian permanent stations. It is intended to provide information of the whole data base via an http port. It is not yet clear if and how to unite the existing Windows web server (using MS IIS) with the storage at unix work stations. This server should also provide some services like automatic decompression and data reduction.
Participation in the new initiative within EUREF providing real time via internet is intended. For users which want to have access by phone the technical equipment together with the

communication software has to be established. All projects do not only have technical but financial and security aspects too.

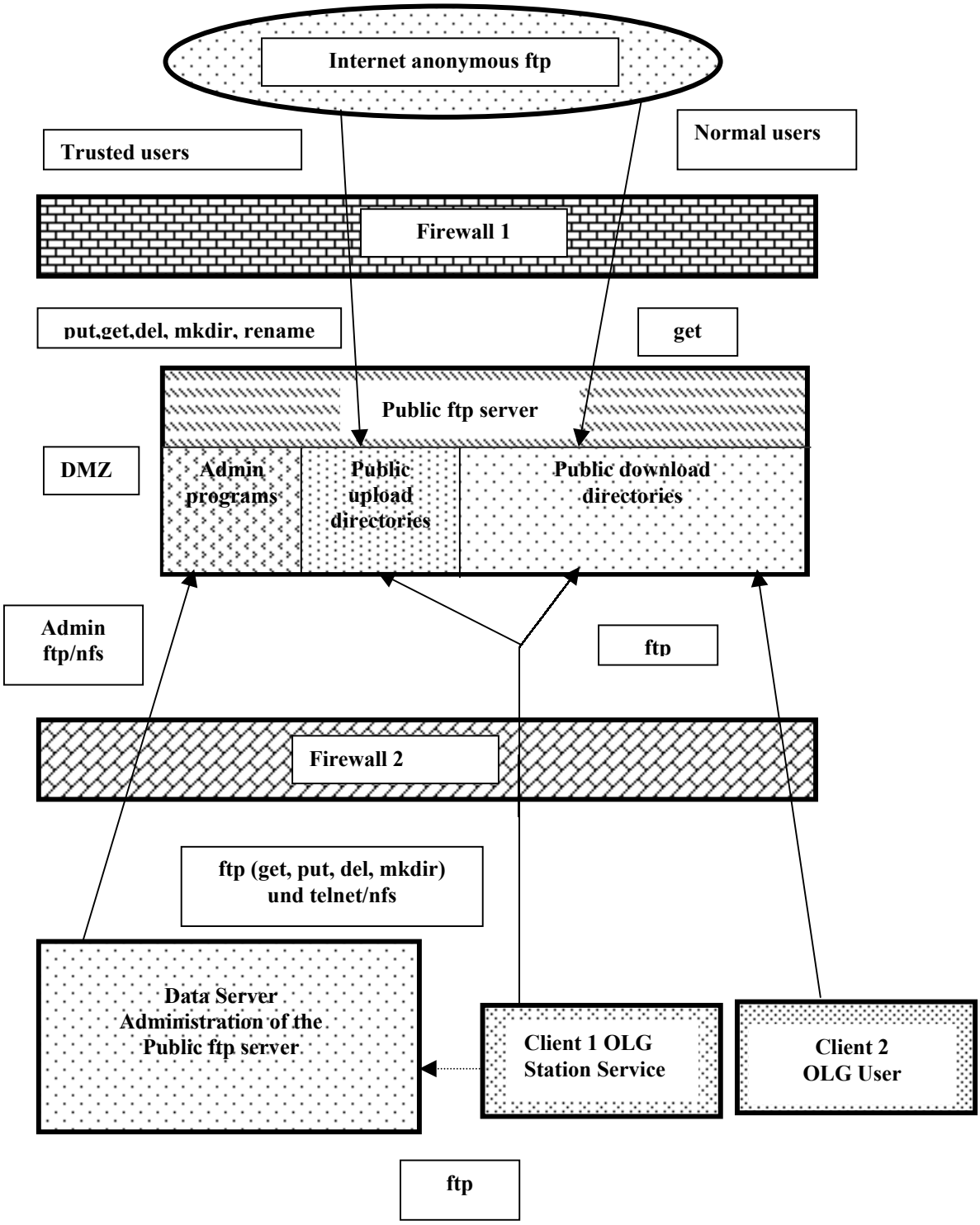

Figure 1 Concept of Data Centre OLG - ftp server

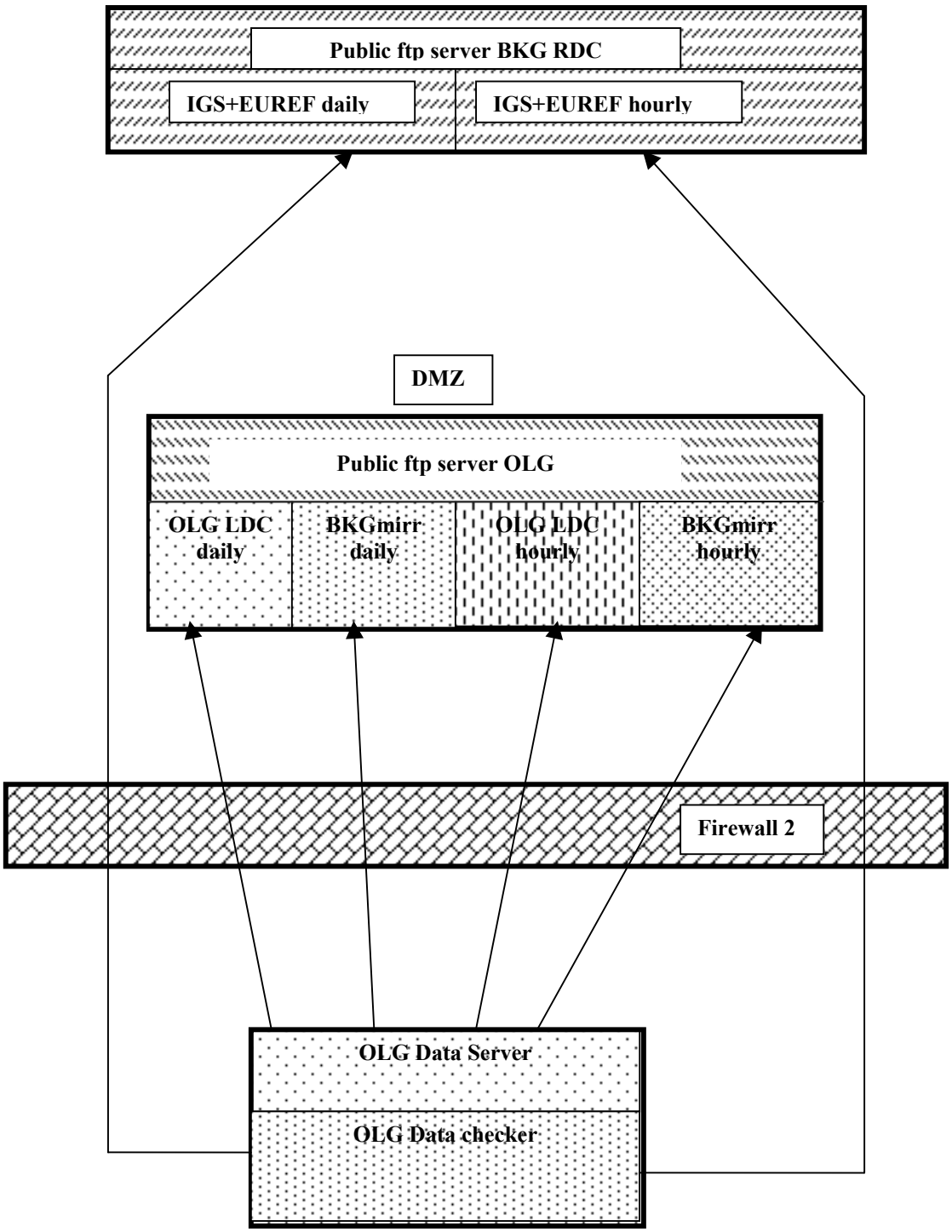

Figure 2 Concept of Data Centre OLG - BKG mirror

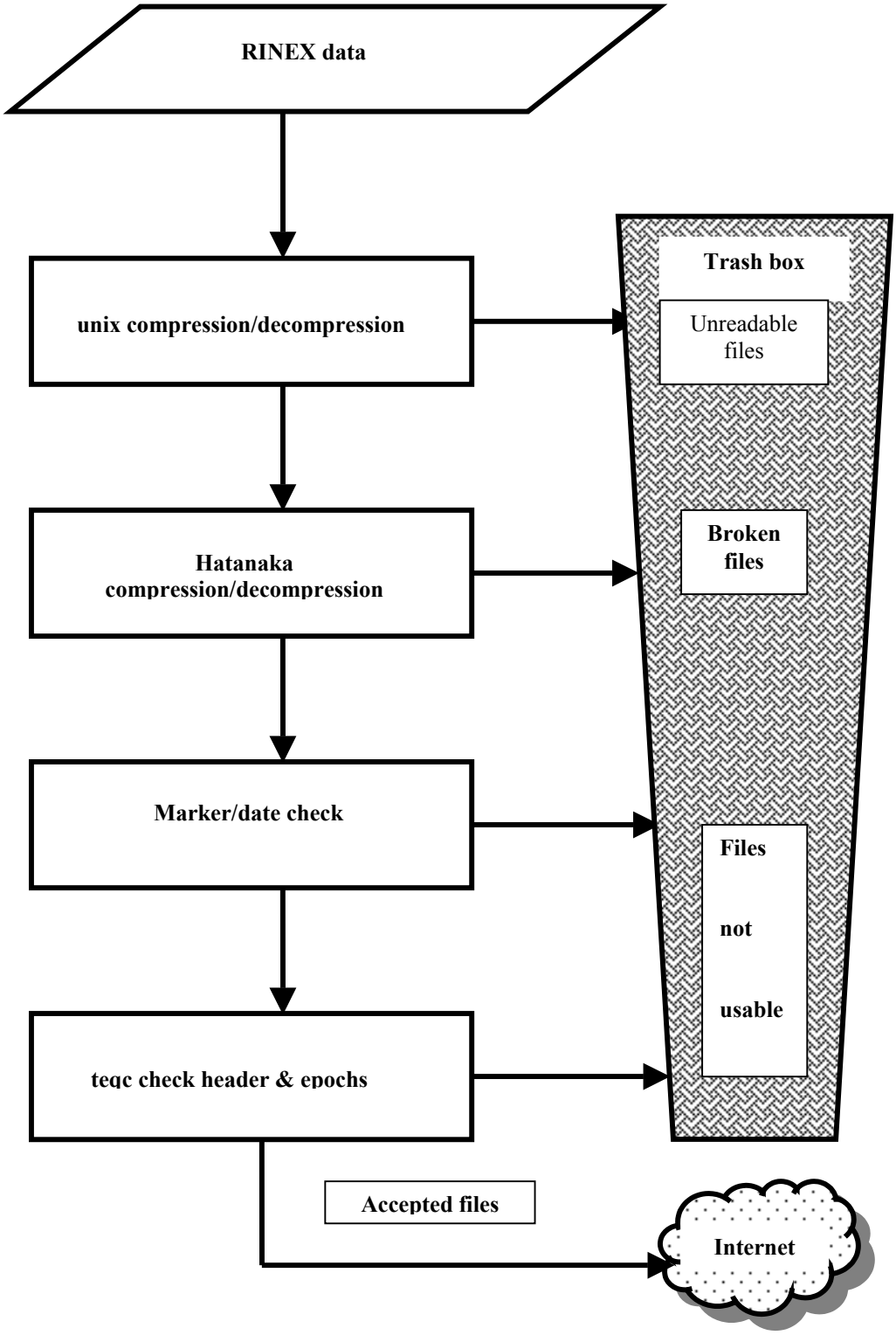

Figure 3 Concept Data Centre OLG – Data check