Network Backup at the University of Kiel Using HYPERtape and a UNICOS Server

Uwe H. Olias, Computer Center of Christian Albrechts University at Kiel, Kiel, Germany

ABSTRACT: *The University of Kiel has been running HYPERtape for network backup since September 1993 with a SIEMENS-Mainframe as a server platform. In Spring 1996 we changed the server platform to UNICOS and run a CRAY Y-MP EL as the server. One of the outstanding results is the increase of backup performance. This presentation will discuss the actual backup system at Kiel and the use of HYPERtape in a real user system.*

1 Past History of Backup at UniKiel

On the occasion of a considerable reorganization in our computer center in 1992, I have been appointed to the post of the head of our systems department. I am now leading and coordinating all kind of system maintenance in our house, beginning with workstations up to mainframes and CRAYs. Before that appointment, I have maintained several mainframe operating systems and have been concerned with mass storage and hierarchical storage management. And now, leading this systems department, I suddenly recognized how careless everybody was treating the data of the workstations. Owners, responsibles and administrators of workstations likewise. They have treated their workstations like their PC at home. But nobody realized that a workstation kept a lot of data, and along with that the results of partially expensive and valuable work. I therefore gave the order to find a suitable tool for network backup which was dedicated to replace the arbitrary backup behaviour. This tool should provide for a controlled backup without operator intervention, without exception and oblivion. It was just the way we were accustomed to from mainframes and vector processors. Selecting such tools, we had to consider several marginal conditions:

- Using a system as a server that was already available in the computer center

- Backup media using the available StorageTek ACS4400

- Automatic operation without operator intervention but interception possible from central point of administration

- Backup jobs at distinct scheduled date and time

- Backing all central platforms (except those whith a direct access to STK ACS 4400)

- Capable to back almost every platform at the campus

- Control the backup organisation and procedure with an available platform

- Extend the backup support step by step throughout the entire campus

- The strongest restriction seemed to be: Network backup had to be performed by one single backup tool suitable for almost every UNIX-Platform, and we expected it should be for long term disposition.

Now we should have a glance at the near surroundings of our StorageTek ACS 4400 with one ExtendedStore. At the University of Kiel, the StorageTek ACS is used as a background storage with native capacity of 3 TB. At that instant, it was applied to

- offline data management, that is archiving, for some special groups within the university

- backing of central systems with direct connect to the STK ACS

- data migration by DMF under UNICOS

- At that time, we had a STK-Connection to the following systems:

  - a CRAY Y-MP M92; a supercomputer.
  - STK-Controller was of FIPS-60 channel type
  - a CRAY Y-MP EL; used as compute server.
  - STK-Controller was of SCSI type
  - a SIEMENS computer with OS BS2000; a mainframe as a general purpose computer. STK-Controller was of FIPS-60 channel type

Obviously, a supercomputer is rather too expensive for network backup, which is backing of workstations! But the backup server should provide for a reasonable I/O capacity

without annoying the interactive end-user. Our CRAY Y-MP EL was dedicated to be a compute server and should not spend compute power for network backup. Network backup is part of the work of a fileserver. But we had a SIEMENS mainframe with some spare performance and I/O capacity.

Unfortunately we missed a fileserver which was predestined to act as a backup server. So we decided to act in two steps:

- Step-1 at that time would be to use BS2000 as a network backup server.

- Step-2 later on: We intended to have a future CRAY Y-MP EL as a fileserver and then acting as backup server, too.

There existed only one network backup tool at the date of procurement that matched our requirements: HYPERtape from MultiStream Systems. Yet we have to confess HYPERtape suffered from some disadvantages while using it with BS2000 as a server. We new that from discussions with some friends from SAVE (an organisation of users of SIEMENS IT, like CUG of CRI users). So we made a grand study of specifications to abandon these disadvantages. It resulted in the product HYPERtape/Media-BS2000. Today, this is a model for all HYPERtape/Media implementations on every platform. According to our experience, HYPERtape and HYPERtape/Media are continuously developped in close contact to Ht customers.

2   HYPERtape Usage, Step II

We skip the first phase of HYPERtape usage at the University of Kiel, which was accompanied by a very close discussion of new Ht-Features and Enhancements. I must confess that it has been useful to exchange experiences and knowledge with other users of Ht and BS2000. As I've been very engaged with BS2000 and SAVE for some twenty years, I invoked a first meeting of Ht & BS2000 users in Fall 1995, and we'll have - encouraged by the participants - our next HYPERtape-Meeting this October on the eve of a BS2000-Meeting.

All of that co-operation has been a great benefit for Multi-Stream Systems, UoK and other customers, and it has been a successful preparation for our Network Backup Phase II. Since the end of 1995 we implemented the following testbed:

- ßUse a CRAY Y-MP EL as a fileserver, and additionally as a HYPERtape BackupNode

- ßWe used a beta-version of Solaris ControlNode, now replaced by the official release

- ßWe had a remarkable benefit from network improvement both in direct CRAY environment and our central workstations

Today, we follow our final HYPERtape usage concept, already planned in 1993. We support the following types of clients, totally numbered 50 when writing this paper:

HW Platform	SW Platform
DEC Alpha	Digital UNIX
SUN	Solaris & SunOS
IBM RS6000	AIX
SGI	IRIX
SNI RM400	SINIX
VAX , Single &Cluster	VMS
Tapeless CRAY Y-MP EL	UNICOS

3   Organization and Backup Technique

We use one single SUN with Solaris as a ControlNode, but we have an option to use an additional ControlNode, if necessary. A ControlNode is the scheduler, dispatcher or just the controller of a backup complex. Our one and only BackupNode is a CRAY Y-MP EL. The clients, called ServiceNodes, run their client-local backup-utility and put the files to be backed directly into the virtual device ftp-put, for example. Thus, they send their SAVESETs over the network to the BackupNode CRAY Y-MP EL, where the SAVESETs are primarily stored on a special filesystem on disk. Comparatively, a SAVESET is a container which exactly contains the group of files that have been saved within one backup job. From the point of view of the UNICOS-BackupNode (the server), this container appears as one file. From the point of view of the ServiceNode (the client), this container is a virtual device consisting of a group of backed files.

We have said the SAVESETs are stored on the BackupNode's disk. We chose a special filesystem for these SAVESETs, and this filesystem is under control of UNICOS DMF. If the high watermark is reached, the SAVESETs are migrated to the dedicated part of cartridges within the STK ACS 4400. In case of a recall, a single file (or a list of files) will be restored from the container it is included in. That is, the according SAVESET is either on the BackupNode's Disk, or a dual state file, or it will be recalled from migration level to foreground storage, that's the disk. Our organizational model is to group all objects to be saved up to estimated 2GB. So we prevent SAVESETs from becoming too large. It helps preventing the BackupNode to handle too large SAVESETs, because

- a save-job may die due to network or workstation problems, and then we have to repeat the whole job, or

- in case of restore, we probably have to recall the SAVESET onto BackupNode's disk and then transfer it to the ServiceNode, were the wanted file is extracted from the virtual device ftp-get by local restore. Yes, you are right, in case of a restore the whole container is xferred over the network towards the client, and then the wanted file is extracted from the virtual device ftp-get unless you have HYPERtape/Media for UNICOS: There, the SAVESET will be recalled onto the BackupNode's disk; only the desired file is extracted and put over the network to the client. So imagine the great advantage

with HYPERtape/Media for your BackupNode (Ht/Media is working together with both, cpio and HYPERtapeBackup, which is very close to cpio). We suffer from HYPERtape/Media for UNICOS is not yet available, but for other UNIXes for example.

At the current state, we run up to two parallel backups during night time (single jobs normally, sometimes two in parallel). In case of increasing numbers of Ht-Clients, we intend to have more parallel backing jobs. The technique putting the SAVESET on disk and then migrating it on cartridges is a great advantage in performance and reliability.

For example using direct write from network on cartridges and having fast devices, you will have the cartridge device always running with start and stop. That results in low usage of capacity and even less perfomance of the backup itself. Faster devices will strengthen this effect!

4   Increase in Performance

We have achieved five major advantages by changing our BackupNode platform from BS2000 to UNICOS, and by a side effect where the central workstations are connected with the BackupNode via FDDI instead of Ethernet connection:

- The BS2000 general purpose computer has no spare performance available for more than three parallel runs. If you have these three parallel runs, you have rather bad response times for interactive BS2000-Jobs up to "connection lost", for this computer keeps busy with I/O and interrupt handling. And switching from Ethernet to FDDI, the calamity increases. So HYPERtape backup with SNI-BS2000 BackupNode was strongly restricted to night-windows three parallel runs.

- Switching from old BackupNode to UNICOS-BackupNode, we had an increase of performance at once for workstations that were able to deliver more input (e.g. Alphas and UltraSparcs). The performance didn't speed up for slow workstations, e.g. some slow SUNs.

- Switching from Ethernet to FDDI caused two improvements, that is direct FDDI connected workstations had a better performance (if they were fast enough), and the other workstations coming from Ethernet to the central systems had a faster throughput. This is due to the fact that the central parts of the network are no longer a bottleneck.

- Additionally, we save direct mount-rates for cartridges for every backing job. That counts, especially in case of incremental backups (-> connect hours).

- The EL-BackupNode turnaround with no parallel backup jobs equals the SNI-BackupNode turnaround with three parallel jobs. If network load and library capacity allows, we'll extend our service to three parallel EL-BackupNode streams per night and one per day. For more details, have a look at conclusion-2 within section *"V. Statistical Evaluation"*.

Now let's have a glance on our current performance table (throughput in KB/sec):

Client-Platform	BS2000-Server Ethernet only	UNICOS-Server Ethernet/FDDI
DEC/OSF	190	1,100 (max 2,300)
SUN	200	700 (max 1,500)
IBM/AIX	180	420
SGI	200	800
SNI/SINIX	210	700
VAX/VMS	140-200	300
CRAY Y-MP EL	60	760

For further details, we should have a later look at the statistical evaluation. We use cpio as the client-local backup-utility, except for VAX/VMS it's BACKUP. Some remarks should be made about performance:

- We have tested DEC/OSF with FDDI-Connection to BS2000-BackupNode and found it was significant slower than with EL-Server. Thus, the Alpha-I/O system has spare performance going to EL. Above all, a speedy Alpha 8000 delivers 2,300 KB/sec with EL-Server, an Alpha 3000 delivers up to 2,100 KB/sec peak each.

- SUN-Clients have nearly the same performance with both, BS2000-BackupNode and EL-Server (improvement about 10%) for ordinary "slow" SUN-Clients. But the performance speeds up to 1,500 with new UltraSparcs.

- So we have recognized an I/O-Limitation due to slow SUNs up to 700 KB/sec, but much more power available with some Alphas (2,300 KB/sec with Alpha 8000) and UltraSparcs (1,500 KB/sec).

- The EL-Client is of low performance, depending on her I/O-Architecture or OS-/IOS-Release. The CE is very engaged with this problem.

5   Statistical Evaluation

During the 1st HYPERtape-Meeting in Germany, attendees asked for frequent statistical overviews about performance, backup amount, reliability and some other stuff. As MultiStream Systems asked for detailed specifications, and UoK was rather interested in statistical evaluation too, we collected and prepared those specifications in a paper and installed a small pre-beta-version of an evaluation-package. A complete tool will be delivered by MultiStream Systems matching these requirements within a short time. I don't want to discuss this statistical tool now, but let's have a look upon some statistical data from a late summer month:

	EL-BN	BS2000-BN
monthly # of Ht-Client-Jobs	7,000	6,000
monthly backup-amount (MB)	176,000	160,000
monthly job-connect (hours)	240	> 700
number of clients	50	42
total sum of backup-objects	300	220
avarage size of a full-backup (MB)	350	350
avarage incremental-size (MB)	10	14
% from full	3	4

[1] (BN stands for BackupNode)

Conclusion-1: BS2000-BackupNode has less I/O-Performance and requires additional tape-mount for every backup-job, even in case of a tiny incremental. Thus, the connect-factor with BS2000-BackupNode has been three against UNICOS-BackupNode.

Conclusion-2: Today, we have a clients connect time of 240 hours/month and 50 clients. That is a daily eight hours service window for non-parallel HYPERtape jobs. And we utilize about 60% of HYPERtape's quota from our storage library capacity. If we extend our service to 100 clients we'll use two jobs in parallel, but then the HYPERtape's quota may be overdrawn. Next step will be to have 36-track cartridge drives, short tape. At that instant, we can extend our service window to three jobs in parallel for eight hours per night (not eleven hours as calculated, for at that time we have completed FDDI network installation all over the campus). Then we'll service some 200 to 250 clients with our small storage library capacity.

This is a small part only from a statistical evaluation routine by UniKiel. We are a medium HYPERtape-Installation, of course. I suppose the largest one at this time is SAP/Walldorf (Germany) with some 2TB/day of avarage backup amount.

6   Future Plans

Some plans depend on storage library capacities, some other on software availability or budget:

- expand the HYPERtape-Service throughout the entire campus (library capacity and budget)

- use the official statistical evaluation utility (software availability)

- have HYPERtape/Media for UNICOS (software availability and budget) for

   -filtered restore

   -password encryption from ServiceNode to BackupNode (not   yet implemented for at least one Ht/M-Release)

- include support of online backup for both, INFORMIX- and INGRES-DB (offline available for INFORMIX, online planned for both, INFORMIX and INGRES; budget-problems)

So we look forward to offer an increasing service for network backup with a unique and reliable utility, that is HYPERtape from MultiStream Systems. And might be we meet some more customers at CUG Conferences discussing HYPERtape-Usage.