Saltzer

To: Distribution

From: T. H. Van Vleck

Date: March 7, 1976

Subject: New Storage System Long Range Plans (revised)

This document reports changes since MTR-112.

OVERVIEW

The following table shows the major phases of the implementation of the new storage system.

Phase		Date
I	Command Level One user at command level	May 75
II	Prototype Running Several users	June 75
III	Design Review Error recovery, backup, mount/demount	Oct 75
IV	Installable System Run mini-service at CISL	Nov 75
V	Initial Installation at MIT No mount/demount	Feb 76
VI	Follow-up Installation at MIT Operational enhancements	March 76
VII	MR 4.0 Installed at MIT With mount/demount	April 76
VIII	Release MR 4.0	June 76
IX	Further Enhancements Administrative improvements	-

CHANGES SINCE LAST REPORT

Since MTR-112 was published, the following events have occurred:

1. The new storage system was installed on the MIT service as system 28.0 on February 15, 1976. This represented a slip of four weeks. The additional time was spent in curing several serious performance problems, waiting for an additonal disk drive to be delivered, and in finding and fixing several bugs.

2. The scheaule for mount and demount has been rethought. Because key people for this phase were unavailable, a simpler interim version of mount and demount will be done for release 4.0. This plan is described in MTB-260.

WORK COMPLETED

1. Statement of Problem.

MTB-017, November 1973. The Storage Problem

2. Preliminary Design.

MTB-055, April 1974. Overview
MTB-060, May 1974. Paging Analysis
MTB-065, April 1974. New Bound on Performance
MTB-095, June 1974. Removable Disk Packs
MTB-110, August 1974. Implementation of New Storage System
MTB-167, February 1975. Disk Usage
MTB-203, June 1975. Attributes of Backup System
MTB-206, June 1975. SAVE and RESTOR
MTB-213, July 1975. Disk Definition
MTB-220, September 1975. Salvager
MTB-221, September 1975. Salvager Implementation
MTB-229, October 1975. Demountable Logical Volumes
MTB-233, November 1975. Backup
MTB-237, November 1975. Overview
MTB-238, November 1975. Initial MIT Installation
MTB-239, November 1975. Error Recovery
MTB-243, December 1975. User Ring Changes
MTB-246, December 1975. List Command
MTB-260, February 1976. Interim Mount
MTB-261, March 1976. Master Directory Control

3. Preliminary Task Schedules.

MTR-068, October 1974. Implementation Plans
MTR-081, March 1975. Long Range Plans
MTR-084, April 1975. Task List Phases I and II
MTR-095, September 1975. Long Range Plans
MTR-112, December 1975. Long Range Plans

17. Interim Backup.

This task modifies the current backup programs to dump and reload the new directory quota cell and the logical volume ID for a directory. This change allows the current incremental/catchup/complete dumper to be used for backup until a new version is designed and built.

Finished.

29. Improved Directory Format.

This task redesigned the directory to be more easily verified for correctness. All storage system modules which reference the directory must be recompiled with the new declaration. The various redundancy checks are not inserted by this task, though. See MTB-221 and MTB-220 for details.

Finished.

31. New Directory Salvager

Rewrite salvager to operate on a new expanded directory structure, without reference to the VTOC entry.

Finished.

33. Phase IV: Make System Installable.

The first version of the new storage system installed at MIT did not have all the functional improvements which will be provided with release 4.0. The ability for a user to request the mounting of a logical volume is not present in this version of the system, and the interim backup is used. What is provided is the reformatting of disk storage and directories and the consequent improvements in reliability.

MTB-238 describes the contents of this installation.

Target date: January 1976.
Finished January 1976.

34. Formalities of Submission.

This step covers filling out submission forms, auditing of all programs, running final performance runs, fixing last-minute problems, etc.

Finished February 1976.

35. Phase V: First Installation at MIT.

Target date: January 18, 1976.
Actual date: February 15, 1976.

CURRENT TASKS

52. Continuing Performance Improvement

This task includes the investigation and metering of the system's performance as installed at MIT, and the identification and removal of performance bottlenecks. For example, the locking and I/O strategies used by the supervisor will be studied.

37. Master Directory Operations.

This task adds ring-1 support for operations on master directories. User calls are create, delete and list. The create_dir and delete_dir commands are modified for this

case.

The ring-1 programs use the volume registration data base, access control segments, and the Master Directory Control Segments (MDCSs). Administrative commands to manage these data bases are provided. For the volume librarian, we need register, unregister, modify, and list.

38. Ring 1 Volume Mount Module (interim).

When a logical volume is to be mounted the operator must type several commands to tell the system that the volumes are online. These commands check the volume label against the registration and when the volume is completely mounted call the hardcore to cause the logical volume to be accepted for paging.

39. User Request to Connect Logical Volume (interim).

User requests to connect to a logical volume will be passed through ring 1. If the user process is permitted to connect to the logical volume, and the logical volume is already mounted by the operator, the request is passed to the hardcore.

40. Hardcore Check on Volume Connection.

The hardcore is changed by this task to require the connection call from ring 1 before allowing a process to initiate a segment on a demountable volume. (The root logical volume is never demountable and other "public" volumes can be declared not demountable.) This insures that ring 1 is not bypassed, and makes sure that all programs using segments on removable volumes execute independently of whether some other process has caused a pack to be mounted. The list of demountable physical volumes which the process is connected to will be stored in the KST.

41. Phase VI: Follow-up Installation at MIT.

Operational experience will lead us to make many improvements to the interface and behavior of the storage system. Performance measurements under actual load may also show use where to concentrate our programming effort in order to speed the system up; if these improvements are possible we will install them soon.

Target date: March 1976.

42. Command System Changes.

These changes are the ones specified in paragraph 27. In addition to the changes to handle new error and state conditions, the create_dir command must accept and check the

new parameter which specifes the logical volume in which storage will reside, and the status command must be modified to show this attribute.

43. Phase VII: Install MR 4.0 at MIT.

    Target date: April 1976.

44. Phase VIII: Release MR 4.0

    Target Date: June 1976.

FURTHER ENHANCEMENTS

24. Implementation of Hardcore Primitives for Backup.

The hardcore primitives to support the new backup system must be able to maintain the list of modified segments on each physical volume for the use of incremental dumping; and to activate and dump or reload a segment by volume ID and VTOC index without referencing the branch.

25. Implementation of Backup Dumping Programs.

The new complete and incremental dumping programs can be much simpler than the current dump programs, since all hierarchy walking and access forcing code is eliminated. The hardcore primitives do most of the work. These programs are easy given the format of the output records to be produced.

26. Implementation of New Reload and Retrieve.

The reloading and retrieval programs will use the output of the dumping programs to reconstruct volumes and to recover the contents of single segments.

32. Directory Control Checking.

This task adds to directory control new code for maintenance of the various redundancy fields added to the directory structure, and appropriate in-line checks and repair operations. MTB-220 describes the details of this change.

36. Backup Integration.

This task integrates the new backup mechanisms into the system and ties backup in with salvaging.

Although most of the parts of the new backup system will be available by the time release 4.0 is frozen, the new backup may not have had sufficient testing and operational experience to allow us to depend upon it.

51. Change Volume Mounting to use RCP.

Modify the mount and demount commands to use RCP for mounting and demounting volumes.

When a logical volume is to be mounted, the LVRF must be consulted to find the list of physical volumes to be mounted. Calls must then be made to RCP to mount each of the physical volumes, the volume labels must be checked, and the hardcore must be called to tell it that the volumes are accepted.

46. Keep Duplicate Copies of Selected Volumes.

Once this task is completed, crucial volumes in the system can be maintained in duplicate; all modified pages will be written out to both devices. In a configuration which places the secondary copy on a different disk subsystem from the primary copy, the cost of maintaining two copies will be very low.

47. Automatic Use of Secondary Volume on Error.

Once the duplicate copy facility is available, the system can be modified so that when a disk record is unreadable, the system automatically switches to the use of the secondary copy.

49. Calls to Initializer Process During Connection.

This step causes RCP to pass all connection requests through the system control process, so that charging can be done, mount messages can be routed, and so that operator commands affecting the request can be issued.

50. Billing.

Modifications must be made to the administrative and billing package to enhance the administrator's ability to manage the system resources. Some of these improvements cannot be specified until we have obtained some operational experience.