N94- 33812

Using Magnetic Tape Technology for Data Migration

David Therrien and Yim Ling Cheung

Epoch Systems
8 Technology Drive
Westboro MA 01581
Phone: (508)-836-4711
Fax: (508)-836-4884
dave@epoch.com

Abstract

Magnetic tape and optical disk library units (jukeboxes) are satisfying the demand for high-capacity cost-effective storage. The choice between optical disk and magnetic tape technology must take into account the cost limitations as well as the performance and reliability requirements of the user environment.

Library units require data management software in order to function in an automated and user-transparent way. The most common data management applications are backup and recovery, data migration and archiving. The medium access patterns that these applications create will be described. Since the most user visible application is data migration, a queue simulator has been developed to model its performance against a variety of library units. The major subject of this paper is the design and implementation of this simulator as well as some simulation results. The relative cost and reliability of magnetic tape versus optical disk library units is presented for completeness.

Data Management Applications

There are three main data management applications that library units are used for:

- The *Backup/Recovery* application enables data that has been lost due to magnetic disk failure or accidental user file deletion to be recovered from backup media. During backup, magnetic tape is preferred over optical disk for the following reasons:

    - Magnetic tape has a lower cost per megabyte than optical disk.
    - Magnetic tape can provide higher write data transfer rates than optical disk.
    - Backup is a sequential access process, so the random access feature of optical disk is not an advantage.

    When a large number of files must be recovered from a backup medium, optical disk could significantly speed up the recovery time. For optical disk, file to file access time is measured in milliseconds as opposed to seconds and even minutes on magnetic tape. However, recovery software that can sort the list of files to be recovered by physical location on magnetic tape has been developed, thereby minimizing search time. This sorting operation also reduces magnetic tape medium wear.

- *Migration* is a high-capacity, lower performance, user-transparent extension of a system's magnetic disk file system. A system that supports migration can provide a storage capacity that is well in excess of reasonable magnetic disk subsystems at a fraction of the cost. During the stage-out process, the migration application automatically identifies least-recently-used data on magnetic disk and moves that data to a lower cost staging medium. Since data is staged-out periodically in bulk form and written to the staging medium in sequential form, magnetic tape is as effective as optical disk. Stage-in moves data from the staging medium back to magnetic disk when requested by a user. The fast drive load/unload and seek times of optical disk make it the preferred medium over magnetic tape for stage-in. These user requests for stage-in are random and unpredictable, making software optimizations ineffective for general storage systems. Since stage-in is the most user-visible application, it was chosen as the application to model against a variety of library units using the queue simulator.

- *Archiving* moves data from magnetic disk to a lower cost archive medium when it is either not being requested by users or it needs to be replicated for increased data availability. Users expect an access time of hours or days to acquire data that has been archived. Magnetic tape provides the following advantages over optical disk for archiving.:

  - The storage density of magnetic tape is higher than optical disk.
  - The cost per megabyte of magnetic tape media is significantly lower than optical disk media.
  - Data compression minimizes the physical storage space for off-line volumes. Hardware data compression is available within most tape drives and is not found in any optical disk drives today because disks are direct access devices that create operating system dependencies.

  The advantages of optical disk over magnetic tape in an archiving application include:

  - Longer archive life. Optical disk archive life is measured in tens to hundreds of years. Magnetic tape is measured in units to tens of years.
  - Lower medium maintenance. Most magnetic tape formats require retensioning to repack the tape onto the storage reels. Magnetic tape must also be periodically cycled from the archive environment back into the active-use environment in order to monitor medium quality and expire volumes with higher bit error rates. Optical disk requires no recycling of volumes in this manner.

Data management servers today that run these applications usually employ magnetic tape for backup/recovery & archiving. Optical disk has been the preferred medium for migration. With the recent availability of cost-effective magnetic tape library units,, users are requesting that servers be configured with just tape library units, thereby eliminating the purchase of optical library units. Although this solution is attractive from a cost standpoint, there are significant performance and reliability concerns that must be addressed. The stage-in simulator has been used to quantify the performance differences between these two technologies.

Performance Comparison and the Stage-In Simulator

Motivation for Developing the Stage-In Simulator

Since stage-in is the most user-visible application of data management, the primary purpose of the stage-in simulator is to quantify the library unit service rate of various magnetic tape and optical disk library units. Optical disk provides a stage-in service time to the user of approximately twenty seconds, even in high request rate environments. Idle magnetic tape library units can service requests within minutes, but in high user request rate environments, the service time would extend to hours and possible days in extreme cases. The motivation for developing the simulator was to define the acceptable user request rate limits for a variety of library units.

Simulation Methodology

The stage-in simulator is a discrete queue simulator. The steps involved in the development of this simulator follow typical simulation methodology [3] which includes planning, modeling, verification and validation and finally running applications against it.

Simulator Planning

The statement of the problem was formed during the planning phase. Initially, the simulator was going to be designed to model all data management applications being serviced by a single library unit. This problem statement was simplified to develop a model for just the migration stage-in application. This application was chosen since it is the most user-visible application and it exhibits the most unpredictable user-access patterns.

Simulator Modeling

During the modeling phase, the following activities were undertaken:

- The model of a library unit was developed
- The data model describing input, output and simulation variables was defined
- The simulator was written based on the library unit and data modeling.
- Performance data from real devices was measured and accumulated for input to the simulator.

Library Unit Modeling

Each user request that is sent to the stage-in simulator requires that a volume be mounted in a library unit drive so that data transfer can take place. The simulator uses a two-level library unit service model where *some* requests require a robot to mount the medium into one of the available drives and *all* requests require the use of a drive to access the data from the mounted volume. A queue is created when the user requests arrive faster than the library unit can process them, because either all of the drives and/or the robot are busy servicing an outstanding request. As shown in Figure 1, the stage-in simulator takes a single stream of user requests and attempts to satisfy them based on the utilization of a single shared robotics element feeding a number of drives.

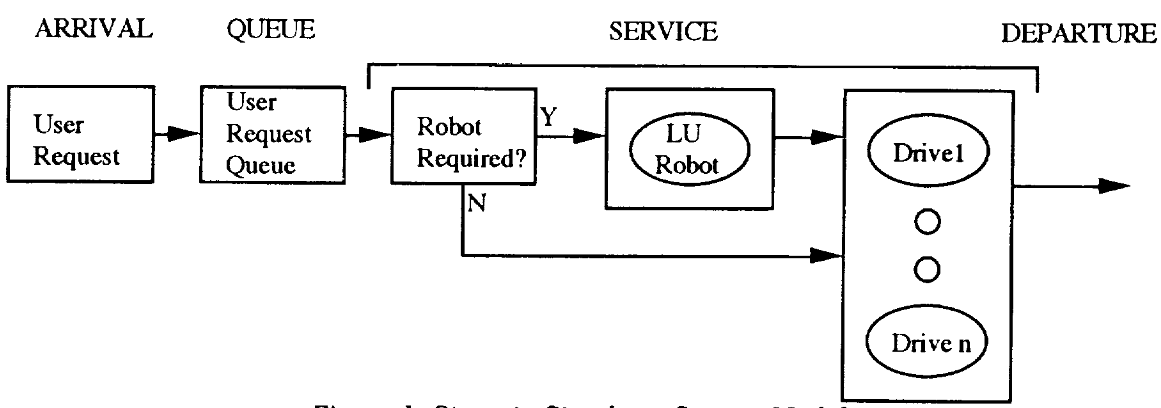

Figure 1. Stage-in Simulator Service Model

The service time for a user request involves a number of robot and drive service time components as shown in Table 1. When user requests require the use of a robot, the service time is the sum of all of the library unit and drive service time components If a user request arrives that can be satisfied by a drive that already has the right medium loaded, only the drive's access time and data transfer time are included in the service time for that request.

Table 1: Robotics and Drive Components of Service Time

Magnetic Tape (Optical Disk)	**Robot Is Required**	**No Robot Is Required**
Rewind to BOT (Spin-down) Medium	√	
Eject Medium from Drive	√	
Robotics Exchange, drive->slot, slot-> drive	√	
Drive Medium Load to BOT (Spin-up)	√	
Drive Access Time	√	√
Drive Data Transfer Time	√	√

Data Modeling

The simulator data model is comprised of input data, simulation variables and output as shown in Figure 2. The stage-in simulator accepts laboratory-measured library unit and drive performance data as input. It produces information on the percent utilization of the library unit robotics and drive(s) as well as the overall library unit service rate, average service time and maximum queue length as output. During simulation, simulation variables such as the user request rate and file size are varied to simulate different user environments.

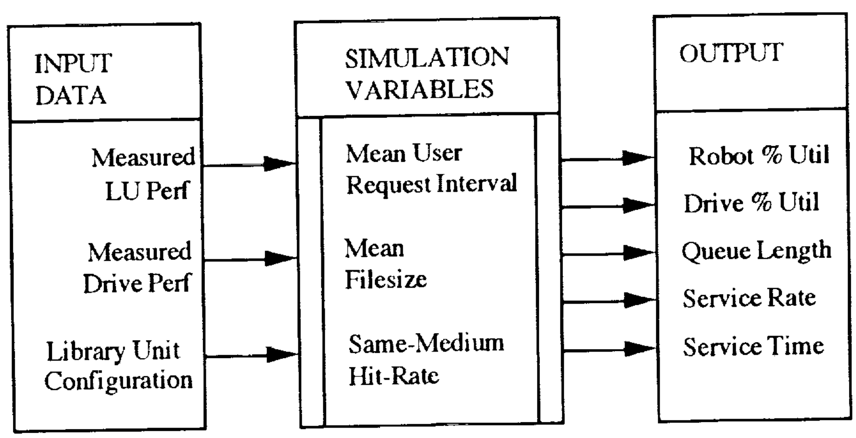

Figure 2: Data Model of the Stage-In Simulator

Simulator Output:

- Robot % Utilization - the percentage of time that the robot is busy during the simulation. Logged values near 100% indicate that the performance of the unit is limited by the robot.
- Drive % Utilization - the percentage of time that the drives in the LU are busy during the simulation Logged values near 100% indicate that the performance of the unit is limited by the drive.
- Queue Length - the size of the user request queue after servicing fifty user requests is logged to quantify the degree to which certain LU configurations fall behind in servicing simulated user request rates. For very high user request rates of very large files, the queue length of user requests to be serviced could reach into the thousands at the point in time where just the first fifty requests have been serviced.
- Service Rate - the number of user requests serviced per hour by the library unit.
- Service Time - the average service time per user request.

Simulator Variables:

- Mean User Request Interval - This variable represents the rate that user requests arrive for stage-in at the server. During simulation, mean user request intervals of 512, 256, 128, 64, 32, 16, 8, 4, and 2 seconds per request were run. This range was selected because it showed the region of user request rate that created drive and robot bound conditions for both magnetic tape and optical disk library units. During simulation, a Poisson distribution was applied to this mean user request interval to induce variability in arrival time. This distribution has been widely used to model arrival distributions and other seemingly random events [3].

- Mean File Size - mean file sizes of 10KB, 100KB, 1MB, and 10MB were selected for simulation. A Poisson distribution was applied to this mean file size to induce variability in user request file size. The drive's measured data transfer rate was multiplied by the file size during simulation to create the data transfer service time component of the total user request service time.

- Same-Medium-Hit-Rate (SMHR) - This variable allowed the simulator to model the behavior of servicing user requests that either exhibit a high degree of same-medium locality (SMHR = 100%) or a low degree of same-medium locality (SMHR = 0%). Each user request that arrives is tagged with a flag that indicates whether or not it requires the use of the robot based on the SMHR % value. Any SMHR percentage can be modeled. When the SMHR is 100%, the service time only includes a drive access time and a drive data transfer component. When the SMHR is 0%, the service time is the sum of all possible drive and robot times as shown in the "Robot Required" column of Table 1.

Simulator Input:

The first real application of the simulator was to model the stage-in performance of a number of magnetic tape and optical disk library unit configurations. For these devices, the following data was collected as input to the simulator:

- Library Unit (LU) Performance - Each real library unit that was modeled had its robotics exchange time measured to be used directly by the simulator. The exchange time includes the time to move a medium from a drive to a storage slot plus the time to move another medium from a storage slot into a drive. For the purpose of this simulation, some conceptual library units were created. Their exchange time was set to exchange times of similar commercially available library units.

- Library Unit (LU) Configuration - The number of media and drives associated with commercially available as well as conceptual library units.

- Drive Performance - the following drive parameters were measured for input to the simulator:

  - Drive Load Time - the time it takes a drive to load and spin up an optical disk or to load and get a magnetic tape to its BOT point.

  - Drive Unload Time - the time it takes a drive to spin-down and eject an optical disk or to eject a tape that was already rewound and at BOT.

  - Drive Data Transfer Rate - the rate at which the drive transfers data to/from the host computer. This rate was measured while servicing stage-in requests for all simulated drive devices. The measured data rate is generally lower than the manufacturer's published data transfer rate, due to drive and host latencies. For this reason, it was important to provide this measured data to the simulator.

- Drive Access Time - For optical disk drives, access time is the sum of seek time plus the rotational delay and is usually well under one second. The access time for magnetic tape drives is its search time which can be measured in minutes. Since magnetic tape drive search time is a major service time component for random stage-in requests, it was important to accurately model search characteristics for magnetic tape. The method of capturing this data involved first writing to the entire medium with a fixed file size and then performing random file reads on that volume while recording the time for each access. Six-hundred random access time samples were taken for a number of storage technologies. Table 2 shows the calculated mean and standard deviation of these six-hundred random access times.

Table 2: Measured Mean and Standard Deviation for Various Device Random Access Times

Medium Type	**Tape Length (Opt. Disk Diam.)**	**Median (seconds)**	**Standard Deviation (seconds)**
Eraseable Optical Disk	(5.25")	0.044	0.011
WORM Disk	(12")	0.429	0.199
8mm Tape	54m	31	15
4mm Tape	90m	47	25
8mm Tape	112m	53	31
DLT Tape	1100'	54	31
VHS Tape	T120	67	19

Random access times could have been generated for the simulator using the mean and standard deviation values in Table 2, but these two values alone did not capture the inherent skew visible in some of the distribution histograms (see Figures 3 and 4). When a random access service time component was required, one of the six-hundred random access time data points was selected.

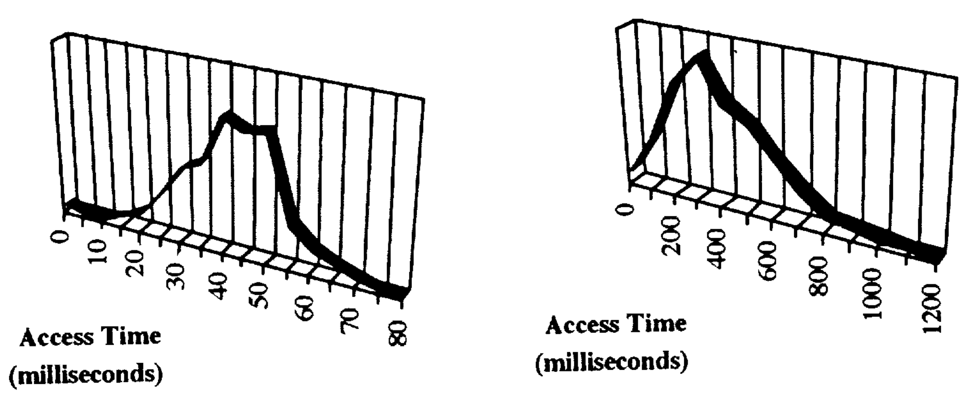

Figure 3: 5.25" Eraseable Optical Disk and 12" WORM Disk Random Access Time Distribution

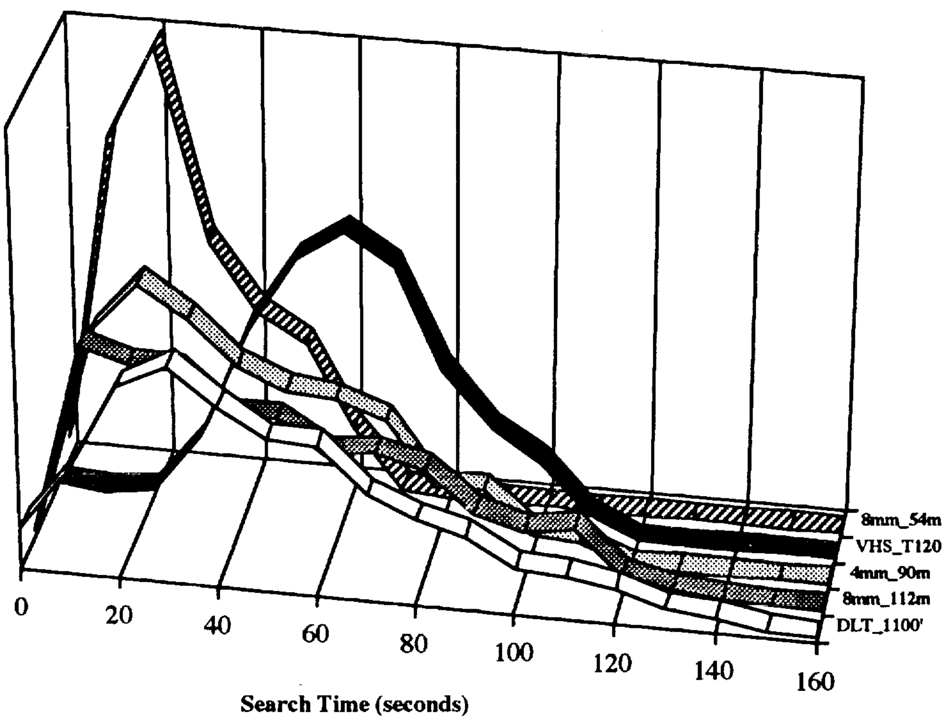

Figure 4: Magnetic Tape Drive Random Access Time Distributions

Simulation Verification and Validation

During the verification and validation phase, the program produced a significant amount of logged data to allow the servicing of each arrival to be studied. This data was helpful in identifying functional bugs in the early implementations of the simulator. Special simulation runs were executed that modeled the operating extremes of a device so the simulated results could be compared against calculated results for validation purposes. The simulator was executed over the same input data and simulation variables repeatedly to ensure the results produced were within a reasonable deviation from all other simulation runs. Also, by varying simulation variables and simulating different library unit configurations, sanity checks of the change in the output data revealed that the simulator was functioning properly.

During this phase of simulator development, it was important to identify the number of departures that had to be produced to provide consistent output data. Simulation runs of 25, 100 and 500 departures were executed with similar output results. For this application, the simulator was run for each user request rate, file size and SMHR value until 50 departures were completed.

Simulator Application

The simulator has the capability of modeling the performance of commercially available library units as well as those that are only conceptual. For this application, a total library unit capacity of 300GB was selected as a product normalizing criterion. Also, each library unit had a configuration of four drives. The library unit and media configurations are shown in Table 3:

Table 3: 300GB Library unit configurations used during simulation

Medium Type/Size	Media/LU	Media Cap. (GB)	LU Cap. (GB)	Real /Conceptual LU
EO_5.25"	215	1.3	279	Real (DISC w/HP1.3GB)
WORM_12"	47	6.5	307	Real (Sony WDA-930)
4mm_90m	150	2.0	300	Conceptual
8mm_54m	116	2.5	300	Real (Exabyte EXB120)
8mm_112m	60	5.0	300	Real (Exabyte EXB120)
VHS_T120	20	14.5	290	Conceptual
DLT_1100'	50	6.0	300	Conceptual

For this application of the simulator, the three SMHR percentages shown in Table 4 were run.

Table 4: Effect of SMHR on Service Time

SMHR	Effect on Service Time
0%	All user requests require a robotics exchange, drive load and unload
50%	Half of the requests do not require robotics exchange and drive load/unload
100%	Robot only used to load each drive once

The theoretical maximum library unit service rate (requests per hour) is bounded by the user request rate as shown in Table 5. This is a units conversion from seconds per user request to library unit service rate expressed in requests per hour. For example, a user request every two seconds generates a theoretical maximum library unit service rate of 1800 requests per hour.

Table 5: Maximum LU Service Rate based on the User Request Rate

User Request Rate(Sec/Req)	64	32	16	8	4	2
Maximum LU Rate (Req/Hr)	56	112	225	450	900	1800

After running the simulator across many library unit models while varying the mean file size, mean user request rate and SMHR, it was observed that the library unit service rate was file size insensitive (from 10KB to 10MB) for lower SMHR percentages (0%, 50%). When SMHR approached 100%, file sizes at 10K, 100K and 1MB had similar service rate performance and 10MB files had measurably lower service rate performance, due to the significant service time component associated with data transfer. For this reason, the simulator output data was condensed to the four cases shown in Table 6.

Table 6: Effect of File size on Service Rate for various values of SMHR

SMHR	File sizes (Bytes)	Service Rate Computation
0%	10K, 100K, 1M, 10M	average of service rate for 10KB, 100KB, 1MB and 10MB files
50%	10K, 100K, 1M, 10M	average of service rate for 10KB, 100KB, 1MB and 10MB files
100%	10K, 100K, 1M	average of service rate for 10K,100K and 1MB files
100%	10M	service rate for 10MB files

Tables 7 through 10 display the simulated service rate of a number of library units expressed in requests per hour. This data represents the capability of each library units to service user requests that arrive at various request rates and file sizes.

The values in Tables 7 through 10 are coded with an indication of whether the unit was *drive bound* (shown in italics) or **robot bound** (shown in boldface). In either case, user requests were being placed in a queue for service and the overall library unit service rate was limited. Drive-bound service rates indicate that the library unit could not service requests at the required user request rate because the drive access time and data transfer characteristics were the limiting factor. Robot-bound service rates indicate that the unit was dominated by robotics exchanges and drive load/unload/search/rewind times.

The queue size data in the rightmost column of Tables 7 through 10 indicates the number of user requests that were waiting in the queue at the point in time when 50 requests were serviced and when the user request rate was at 2 seconds per request which is the worst case user-request rate condition.

Table 7: LU Service Rate - SMHR = 0% - all file sizes

User Req Rate(Sec/Req)	64	32	16	8	4	2	**Queue**
Max. LU Rate (Req/Hr)	56	112	225	450	900	1800	**Size**
EO_5.25" 215c4d	57	112	220	**230**	**232**	**232**	330
WORM_12" 47c4d	71	112	210	440	**470**	**480**	160
4mm_90m_150c4d	**45**	**45**	**47**	**48**	**47**	**47**	1900
8mm_54m_116c4d	45	**48**	**48**	**48**	**48**	**48**	1800
8mm_112m_60c4d	**36**	**37**	**37**	**37**	**37**	**38**	2400
DLT_1100'_50c4d	**30**	**30**	**30**	**30**	**30**	**30**	3000
VHS_T120_20c4d	**39**	**39**	**40**	**41**	**40**	**40**	2200

Table 7 Observations:

- All magnetic tape library units were able to service user requests at a rate of 128 seconds per request (this user rate was simulated, but not shown in the table). Magnetic tape library units are limited to servicing only 30 to 50 requests per hour for SMHR = 0%.
- 12" WORM disk outperformed 5.25" eraseable optical disk in this model primarily because the 12" library unit robotics exchange time was faster. Optical disk technology can service user requests in the 8-16 second per request range.
- All library units became **robot bound** as the user request rate increased.
- After servicing only 50 user requests, very significant request queues were created for magnetic tape. With the average service time per user request at ~100 seconds for magnetic tape, the last user requests in the queue of ~2000 entries would not be serviced for 2.3 days. The first 50 requests to magnetic tape library units were serviced in approximately one hour.

Table 8: LU Service Rate - SMHR = 50% - all file sizes

User Req Rate(Sec/Req)	64	32	16	8	4	2	**Queue**
Max. LU Rate (Req/Hr)	56	112	225	450	900	1800	**Size**
EO_5.25"_215c4d	56	110	210	400	**500**	**420**	150
WORM_12"_47c4d	56	112	236	472	700	**1050**	55
4mm_90m_150c4d	58	**78**	**84**	**97**	**95**	**88**	1000
8mm_54m_116c4d	55	**82**	**103**	**90**	**110**	**105**	700
8mm_112m_60c4d	56	**63**	**65**	**70**	**70**	**70**	1300
DLT_1100'_50c4d	56	**62**	**60**	**62**	**60**	**60**	1600
VHS_T120_20c4d	50	**70**	**70**	**90**	**80**	**80**	1500

Table 8 Observations:

- All magnetic tape library units were able to service user requests at a rate of 64 seconds per request.
- 12" WORM disk outperformed 5.25" eraseable optical disk in this model primarily because the 12" library unit robotics exchange time was faster. Either of these technologies is capable of servicing user requests at a rate of 8 seconds per request.
- All library units became **robot bound** (as shown in boldface) as the user request rate increased.
- Magnetic tape library units are limited to servicing only 60 to 100 requests per hour.
- Using shorter 54m tapes instead of the longer 112m 8mm tapes improved the LU service rate from ~90 request per hour to ~105 requests per hour.
- After servicing only 50 user requests, very significant request queues were created for magnetic tape. With the average service time per user request at ~60 seconds for magnetic tape, the last user requests in the queue of ~1500 entries would not be serviced for ~1 day. The first 50 requests to magnetic tape library units were serviced in approximately one hour.

Table 9: LU Service Rate - SMHR = 100% - file size <= 1MB

User Req Rate(Sec/Req)	64	32	16	8	4	2	**Queue**
Max. LU Rate (Req/Hr)	56	112	225	450	900	1800	**Size**
EO_5.25"_215c4d	56	110	230	450	860	1900	0
WORM_12"_47c4d	55	110	200	470	880	1670	2
4mm_90m_150c4d	54	110	212	*285*	*285*	*285*	230
8mm_54m_116c4d	56	114	190	370	*450*	*440*	160
8mm_112m_60c4d	75	110	*175*	*260*	*290*	*270*	270
DLT_1100'_50c4d	56	104	200	*270*	*270*	*263*	290
VHS_T120_20c4d	56	112	*190*	*200*	*200*	*200*	400

Table 9 Observations

- Magnetic tape library units were able to service user requests at a rate of ~16-32 seconds per request.

- All magnetic tape library units became *drive bound* (shown in italics in the table) due to long drive search time as the user request rate increased.

- 5.25" eraseable had a performance advantage over 12" WORM, primarily due to the faster seek time of the smaller 5.25" medium (see Figure 3). It should be noted that a 5.25" medium contains only one-fifth the data of a 12" WORM medium. Either of these technologies is capable of servicing user requests at a rate of 2 seconds per request.

- Magnetic tape library units are limited to servicing only 200 to 400 requests per hour for this SMHR and mean file size range of 10K-1MB.

- Using shorter 54m tapes instead of the longer 112m 8mm tapes improved the LU service rate from ~270 request per hour to ~440 requests per hour.

- After servicing only 50 user requests, significant request queues were created for magnetic tape. With the average service time per user request at ~15 seconds for magnetic tape (because 4 user requests are being serviced simultaneously), the last user requests in the queue of ~275 entries would not be serviced for ~1 hour.

Table 10: LU Service Rate - SMHR = 100% - file size = 10MB

User Req Rate(Sec/Req)	64	32	16	8	4	2	**Queue**
Max. LU Rate (Req/Hr)	56	112	225	450	900	1800	**Size**
EO_5.25"_215c4d	56	101	218	417	843	*1130*	29
WORM_12"_47c4d	54	97	236	*396*	*582*	*504*	111
4mm_90m_150c4d	59	103	172	*201*	*219*	*202*	232
8mm_54m_116c4d	60	92	219	*236*	*278*	*267*	259
8mm_112m_60c4d	52	113	*184*	*169*	*191*	*184*	442
DLT_1100'_50c4d	60	121	*158*	*228*	*203*	*196*	381
VHS_T120_20c4d	58	85	*164*	*194*	*180*	*196*	993

Table 10 Observations:

- All magnetic tape library units were able to service all requests at a rate of ~16-32 seconds per request.

- All magnetic tape library units became *drive bound* (shown in italics in Table 10) due to search rate and low data transfer rate as the user request rate increased. The 5.25" eraseable and 12" WORM library units became *drive bound* because of their relatively low read data transfer rate.

- 5.25" eraseable optical and 12" WORM are capable of servicing user requests at a rate of 16 seconds per request. This simulation set of parameters produced lower performance than that from Table 9, indicating the increased contribution of data transfer rate to the overall service time and the low data transfer rate characteristics of optical disk drives.

- Magnetic tape library units are limited to servicing only 200 requests per hour for this SMHR and file size.

- Using shorter 54m tapes instead of the longer 112m 8mm tapes improved the LU service rate from ~184 request per hour to ~267 requests per hour.

- After servicing only 50 user requests, significant request queues were created for magnetic tape. With the average service time per user request at ~18 seconds for magnetic tape (because 4 user requests are being serviced simultaneously), the last user requests in the queue of ~350 entries would not be serviced for ~2 hours.

Summary of Simulation Application

Tables 7 through 10 indicate that any library unit can be driven to either being drive or robot bound under various user request load characteristics. If a user can determine a mean file size for the environment and estimate a user request rate, the SMHR percentage can be varied from 0% to 100% across a number of library unit models to determine the best technology fit for that environment.

Cost Comparison

Today, most systems that support data management applications employ optical disk library units for migration and magnetic tape library units for backup/recovery. From the overall system cost perspective, there is a strong motivation to have all data management applications running on a single magnetic tape library unit to eliminate the cost of the optical disk library unit altogether.

When comparing various library unit options, the total cost of the library unit, its drives and its media must be considered. Magnetic tape library units with their media and drives are two to five times more cost effective than optical disk library units of a similar capacity.

The cost of library unit drives becomes a major factor in deciding on a storage technology for data management applications. Random stage-in requests from users can be serviced more effectively when more drives are available to service requests simultaneously. Middle and high-end magnetic tape drives (VHS, 3480, D2) and larger optical disk drives (12", 14") can be from three times to hundreds of times more expensive than smaller form-factor drives (3.5", 5.25"). For servicing high-volume stage-in requests, the preferred library unit configuration would house many low-cost drives as opposed to a few large drives. This assumes that the outstanding requests are serviced by as many different media as there are drives.

The cost per megabyte of optical disk media can be from three times to twenty times more expensive than magnetic tape media, depending on the two specific media types being compared. The cost of having to replace worn magnetic tape should be factored into the comparative media cost calculation. Media cost comparisons become important for environments where a significant amount of data will be archived off-line outside of the library unit.

The simulation data presented in Tables 7-10 represented the service rate performance of a variety of 300GB library units, each having four drives. The range of service performance that a single library unit can exhibit can be plotted against the estimated cost of the sum of the library unit, its drives and media to create a stage-in performance versus cost chart as shown in Figure 5. The service rate minimum and maximum values were taken from the 2 seconds per request column of Tables 7-10.

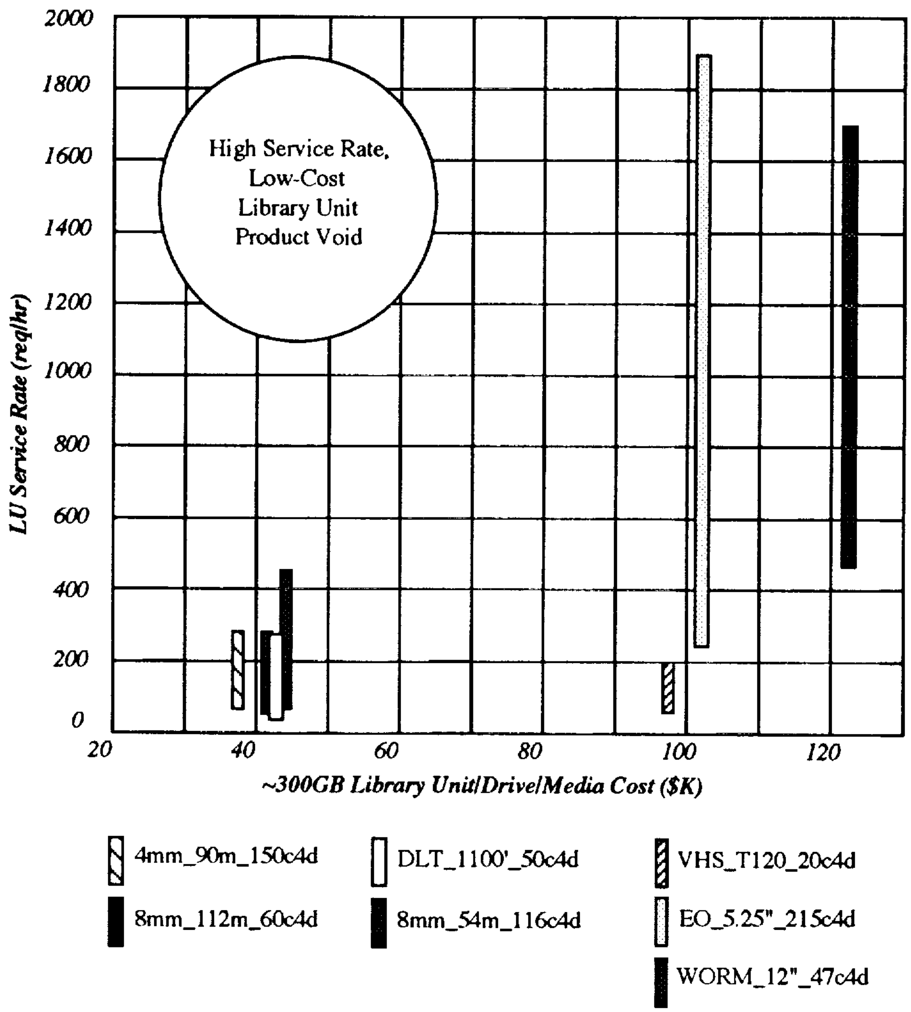

Figure 5: Cost vs. LU Service Rate Performance of Simulated 300GB Library Units

From the data presented in Figure 5, the following cost & performance observations can be made:

- Of the devices simulated, only optical disk library units provide service rate capability over 500 requests per hour.
- EO_5.25" is faster than WORM_12" for high SMHR values because its access time and data transfer rates are greater than WORM_12". EO_5.25" can also have a lower service rate than 12" WORM in very low SMHR environments, since WORM_12" has faster robotics exchange, drive load and unload times.
- Most magnetic tape technologies are clustered in the low service rate, low cost corner of the chart, with the exception of VHS. VHS tape drives are expensive, but they can transfer data faster than any other tape drive that was simulated.
- VHS produced the narrowest range of stage-in performance. This can be explained by the shifted random search distribution for VHS as compared to 4mm, 8mm, and DLT (see Figure 4). Although VHS did not perform well for stage-in, it would most likely outperform all other tape technologies when used with backup/recovery and stage-out data management applications.

- The 8mm configuration that used the shorter 54m tape had a high-end service rate that was significantly higher than the same library unit with fewer cartridges and longer 112m tapes. This is primarily due to the reduced search/rewind times of shorter tapes as shown in Figure 4. This may be an option for customers who are willing to significantly reduce the library unit capacity for an increase in overall stage-in performance.

- 4mm tape library units can provide service rate performance similar to DLT and 8mm library units at a reduced cost. This is primarily due to the lower cost of the 4mm drives.

- There is a high-service rate, low cost library unit product void that has not yet been filled by new library units as shown in Figure 5.

Reliability and Data Availability Comparison

There are many optical disk and magnetic tape library units available that provide high reliability and high availability of user data. The critical reliability features of a library unit include:

- Robotics MEBF - the mean exchanges between failure of the robotics mechanism. A mean of one million exchanges has become the standard that most library units are expected to perform to.

- Drive MIBF - the mean insertions of media into the drive before drive failure occurs. For optical disk drives, MIBF is usually greater than 400,000. Magnetic tape drive MIBF values are usually much lower.

- Adaptive robotics system that can compensate for robotics wear or mechanical alignment drift over time.

- Robust robotics retry mechanisms to compensate for marginal physical alignment. Some tape library units exceed optical disk library units in their ability to recover from soft robot-movement errors.

The critical data availability features of a library unit include:

- Safe operator access to media and drives when the robotics fails. This allows an operator to "play the robot" while spare robotics parts are in transit for replacement. Most optical disk library units do not provide user access to media and drives while many magnetic tape library units do.

- Standard drives that can be installed in the library unit without drive modification. Because of the complicated medium loading mechanism of certain tape drives, some tape library units require that the standard drive be modified before installation into a library unit.

- Customer replaceable drives with foolproof drive alignment during drive replacement. Most optical disk library units are not designed with customer replaceable drives, but some tape library units do have this feature.

- No required periodic maintenance for drives, media and robotics.

Periodic maintenance is required on many magnetic tape drives and optical disk drives. Magnetic tape drive heads wear as the medium is passed over them. Helical scan drives like 8mm, 4mm, D2, and VHS have low head life ratings between 1,000 and 5,000 hours [1] while non-helical scan drives like QIC, DLT, and 3480 tape technology have head life ratings between 5,000 and 10,000 hours after which drive heads have to be replaced. Certain optical disk drives require periodic maintenance in the form of an adjustment to the laser "head" that is responsible for writing and reading data. In either the magnetic tape drive case or the optical disk drive case, the cost of adjusting or repairing a worn head is usually a significant cost-of-ownership for lower-volume larger form-factor drives.

Overall media reliability can be segmented into archive reliability and active-use reliability. The archive life of most magnetic tape media is between 10 to 30 years and is significantly affected by temperature and humidity conditions in the archive environment. Many tape medium formats require retensioning in order to repack the tape onto the cartridge reel to eliminate stresses or to separate tape that is beginning to adhere to adjacent layers. For example, Exabyte suggests rewinding 8mm tape once every three years if kept in an archive environment of 20°C, and once every three months if kept in an archive environment of 30°C [2]. Optical disk media can provide stable archive storage from 25 to 100 years.

Active-use magnetic tape media reliability is mostly affected by the amount of wear between the drive head and media. Helical scan technologies like 4mm, 8mm, VHS, and D2 specify the number of passes against the head at ~1500 [1], where a pass is any forward or backward movement that creates contact with the head. Non-helical scan technologies like QIC, 3480, and DLT specify the number of passes of media at 5,000 to 20,000.

The limited medium pass count for helical scan tape media has not been a significant problem for use in a backup/recovery application, since backup is sequential and recovery is infrequent. When data is staged-out, it creates sequential access to magnetic tape which minimizes tape wear. **Stage-in requests, on the other hand, are random and unordered, and will impose a high number of passes over a tape during routine stage-in activity. Most tape technology cannot withstand this random-access activity.** To compensate for this lack of medium durability, data management software must be developed that provides improved media quality monitoring, data replication, and volume expiration features. From a hardware reliability and data integrity standpoint, the medium with the highest number of head to medium passes is preferred for the stage-in application.

Summary

Magnetic tape library units are more cost-effective than optical disk library units. Unfortunately, magnetic tape drives and media are less durable and reliable than optical disk drives and media. Magnétic tape library units should only be used with user-request rates that don't cause the library unit to be drive or robot bound as shown in Tables 7-10.

The stage-in simulator has been used during system planning exercises to estimate the overall performance of very high capacity system configurations. It has been effective in quantifying the weakness of sequential devices that are perceived to be "high performance" but have been designed for high data transfer rate, not fast random-access to data.

Improved data migration software needs to be developed as the use of magnetic tape as a migration device becomes more widespread. Because of the relatively low magnetic tape medium and head reliability and durability, data management software must perform more media defect management and historical soft error logging to find the "best" point in time to expire a volume. In terms of performance, improved data placement algorithms must be developed that provide a high degree of data locality during stage-in.

Future Simulation Activity

The simulator has been used for a number of other applications since its development. It has been effective in assisting library unit vendors in planning their next generation library units. The stage-in simulator can model the effect of changing the number of drives, cartridges and robotics elements within the library unit. The simulator can also assist in migration data management research by modeling a variety of stage-out data placement algorithms against real library unit devices. The goal of this researrh is to increase the locality of stage-in data.

A number of simulator enhancements are planned. These enhancements include:

- Adapting the current simulator to model library units with more than one non-conflicting robotics element to increase low SMHR performance.
- Producing a UNIX version of the program and providing it to customers for what-if analysis. It is currently written in ThinkC for an Apple Macintosh.
- A graphical output of the simulator progress as well as direct program charting of simulation results.
- Continued data acquisition of performance parameters for newer devices.
- Consideration for drives like VHS that allow the medium to be ejected without rewinding.

Also, the following applications are planned:

- Perform simulation of many library units in the 50-100GB range and the 1-10TB range to compare against the results of the 300GB simulation presented in this paper.
- Assist library unit vendors in planning their next generation library units. For instance, it is simple to create library unit configurations that show the results of changing the number of drives in the library unit from 1 to n drives to arrive at an optimal number of drives to robotics elements.
- Model the user-perceived effect of modifying library unit service rate components and configurations. For instance, if the drive load time could be cut in half from the present time, what effect would that have on the user-perceived service rate.
- A number of papers have been written on the subject of data placement on media during stage-out in order to optimize stage-in performance in the future [5-10]. Using the simulator, various data placement algorithms could be modeled against a variety of library units, user request rates and mean file sizes to quantify the effectiveness these algorithms. For example, a simulation could be run that quantifies the stage-in performance when data is staged-out *across* all magnetic tape volumes within a library unit instead of filling each volume to end-of-tape before starting the next volume. This scheme would be effective for library units that have fast robotics exchange times and magnetic tape drives that have fast load/unload/rewind times but relatively slow search times.