Increasing File Read Throughput
on Tape with TAOS

Table of Contents

Abstract

Retrieving non-consecutive files from a tape can result in large seek times between file reads due to potentially long distances between files. These large seek times between files are due to the way in which data is put on tape in a serpentine, out and back method. With an approximate tape length of 1km, creating inefficient seeks to files can become costly.

By taking in a list of read requests and determining where each read starts and ends on the tape physically, a 'Traveling Salesman' solution can be applied to reduce the overall time of reading multiple files from a tape. Spectra Logic has developed a Time-Based Access Order System (TAOS™) to create an optimal ordering of read requests that results in a four times improvement in overall time for files that are less than 100MB in size.

Objectives:

- Adds no cost to existing libraries (standard BlueScale® feature)
- Provides nearly identical SCSI interface to RAO SCSI commands implemented in IBM® TS115x drives
- Supports LTO-7 and LTO-8 (both full height and half height drives) and future LTO drives.
  - Works initially with Spectra's T950 Tape Library and TFinity® ExaScale Tape Library
  - Additional libraries may be added depending on demand
- Improves performance of restore operations
- Reduces tape media wear
- Reduces tape drive wear

The High Energy Physics (HEP) environment at CERN is one example of where this functionality has been leveraged to attain large performance improvements when moving large volumes of physics data between tape and disk and vice versa. More background on how CERN has utilized this ability to compute the file order that corresponds to the minimum access time can be found in the thesis paper: Enhancing the low-level tape layer of CERN Tape Archive software.
https://cds.cern.ch/record/2282014?ln=en

Introduction

Similar functionality can be found in IBM's TS1140, TS1150, and TS1155 enterprise tape drives. In these devices, it is referred to as Recommended Access Ordering (RAO). RAO was released with the IBM® TS1140 tape drive. The only real changes that have been made between the initial release and the latest generation are that the IBM® TS1140 could use a list of up to 3200 files and the latest IBM® TS1155 drive can use a list of up to only 2730 files. The Oracle® T10000 tape drives also had built-in Recommended Access Order functionality that was defined by an operation code and a Service action. It utilized the same Generate Recommended Access Order (GRAO) and Receive Recommended Access Order (RRAO) commands as the IBM® TS11xx tape drives. Until now, no such functionality has existed for LTO tape-drive-based systems. Spectra Logic's TAOS adds this performance improvement to Spectra enterprise tape systems with LTO drives.

Data on tape is laid out in wraps that follow a serpentine pattern, starting at the physical beginning of the tape, traversing all the way to the end of the tape, then back.

This pattern is repeated until the tape is filled, which on an LTO-7 tape is 112 wraps or on an LTO-8 tape is 208 wraps.
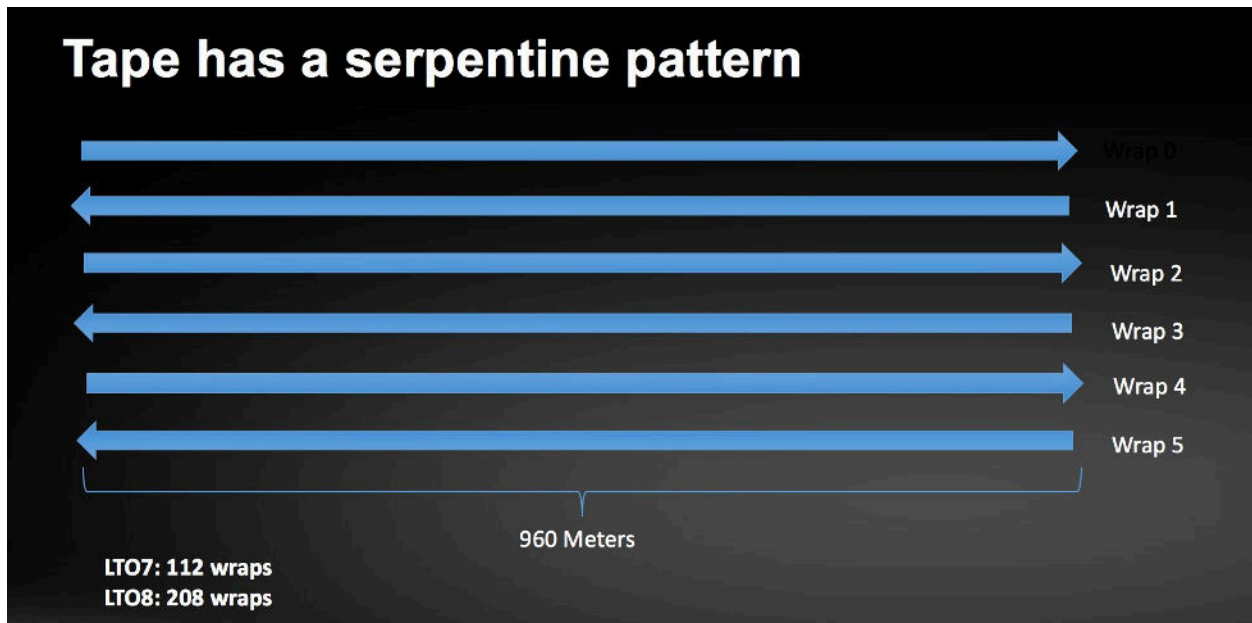
In addition, the tape head only covers about a fourth of the width of the tape so it needs to be moved to one of four positions. These four positions are known as bands. So, for LTO-8 tape, there are four bands each consisting of 52 wraps. Moving between wraps contained within a band is faster than moving to a wrap on another band. Also moving to a wrap that is in the same direction is faster than moving to one that is in the opposite direction.

Due to the out and back way files are written to tape, retrieving files in their logical sequential order is one of the least efficient methods of reading files.

This is a result of locating up and down the tape unnecessarily with a maximum penalty of traversing 970 meters to get to the next file.

In contrast, if many files are needed from the same tape, the order in which the files are retrieved can be reordered such that shortcuts can be made on the tape reducing the time required to retrieve all of the files. These shortcuts reduce the amount of tape needed to locate over between reading all the files.
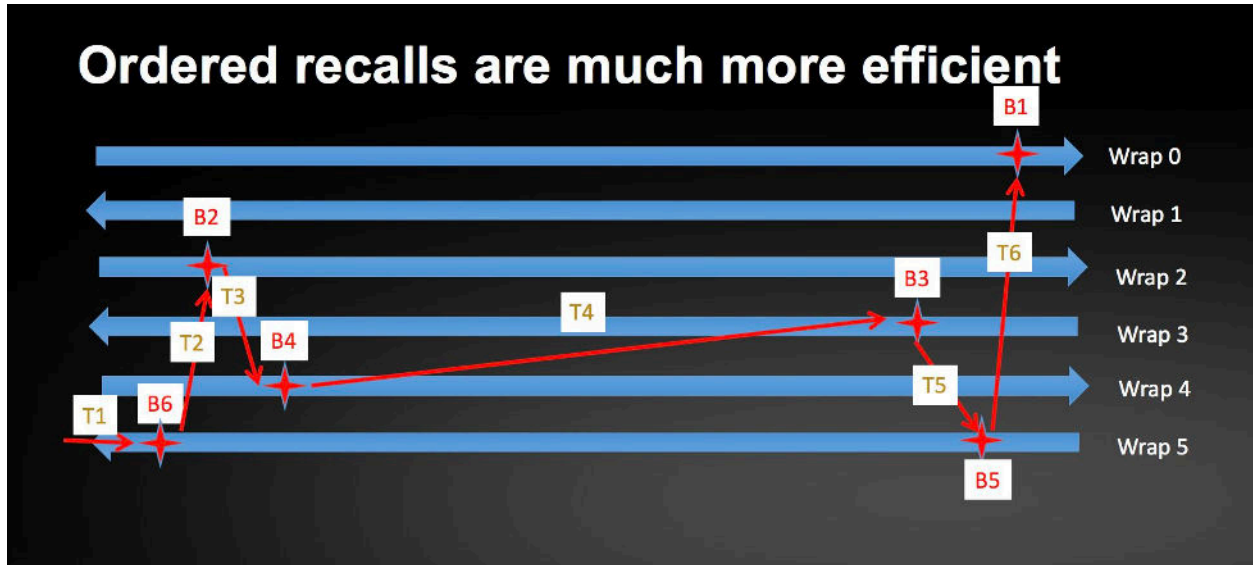

Theory

If a list of multiple files to recall are provided, these files can be mapped out to a physical dimension on the tape. Once each file's start and end position is translated to a physical position on the tape, an estimate can be calculated for how long it will take in milliseconds between each file's end position and every other file's beginning position. With this information, a directed weighted graph can be created that enables an optimal ordering of files to recall (aka, the Traveling Salesman solution).

Estimating Locate Timings

Being able to estimate the time it would take to go from the end of one file on a tape to the beginning of another file on a tape is a crucial part of this problem. Due to the complexity of the internals of the tape drive, machine learning is leveraged to create a model that can be used to predict the time it would take to traverse from two arbitrary points on a tape.
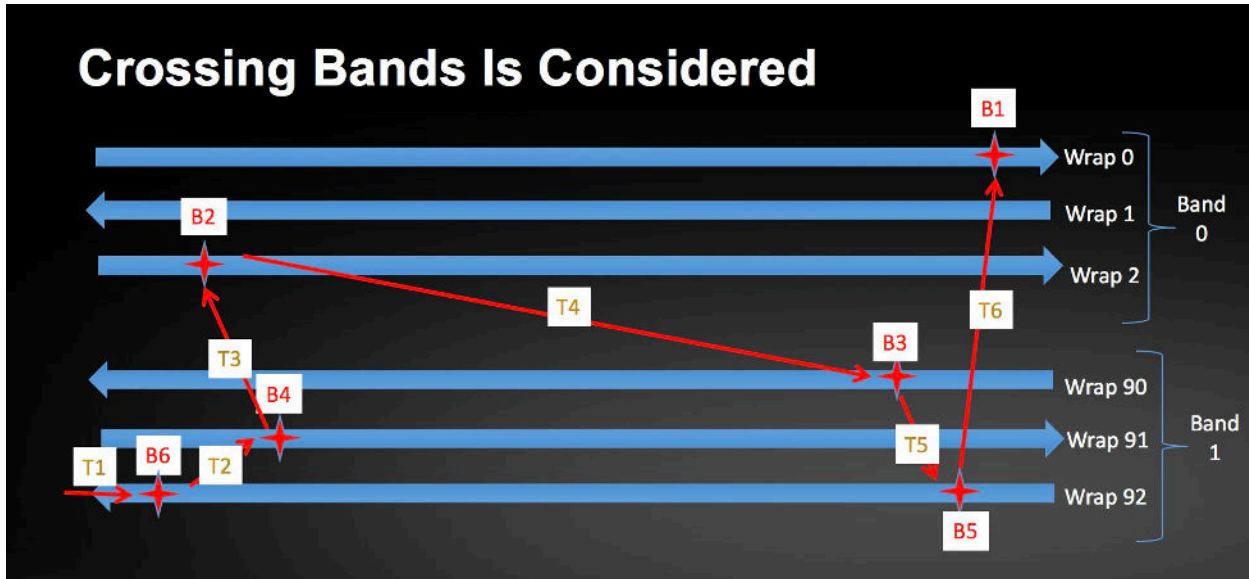
To create an accurate model, proper data collection was key. Over 100,000 data points were collected to feed into the machine learning model. Each data point consisted of 11 features including source band, destination band, linear distance traversed, how long the operation took, did the drive need to change directions, and more. Data for multiple drive types was collected in order to create models for the following tape drives: LTO-7 (full and half height), LTO-8 (full and half height) and IBM® TS1155.

The machine learning estimator chosen produced a model that resulted in an average error of 44.54 milliseconds, a standard deviation of error of 3.801 seconds, and an R^2 of .973586. With this model, estimating the time it takes to locate between two files can be done accurately. The following chart shows all of the data points that were collected in blue and fed into the machine learning model. The red line most closely represents the model to be used for LTO-8 tape drives.

Actual vs Predicted Seek Times with Machine Learning
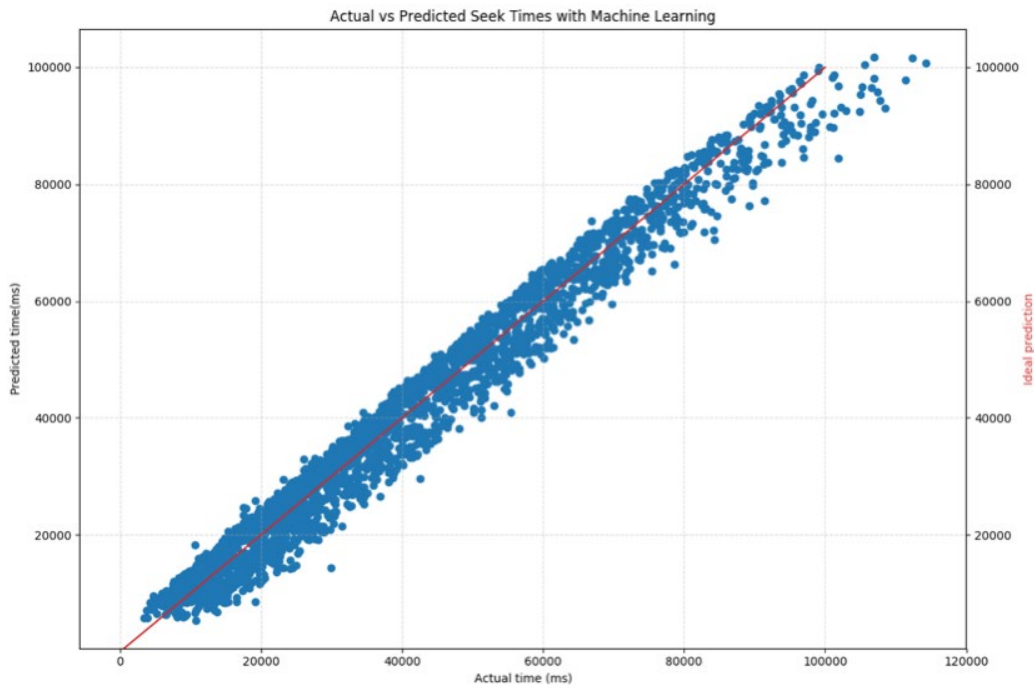
Ordering the Read Requests

With an estimated time to go between each file on tape, a Nearest Neighbor algorithm is applied to the list of files in order to create an ordered list. Further optimization is applied to this ordered list to reduce the overall time needed to recall all the files requested. Note that with multiple bands available, that must also be taken into account for an optimum implementation.

Using TAOS

The TAOS algorithm runs on a library processor within the Spectra tape drive sled.

TAOS processors located inside the drive sled housing and connected to fibre channel input ports
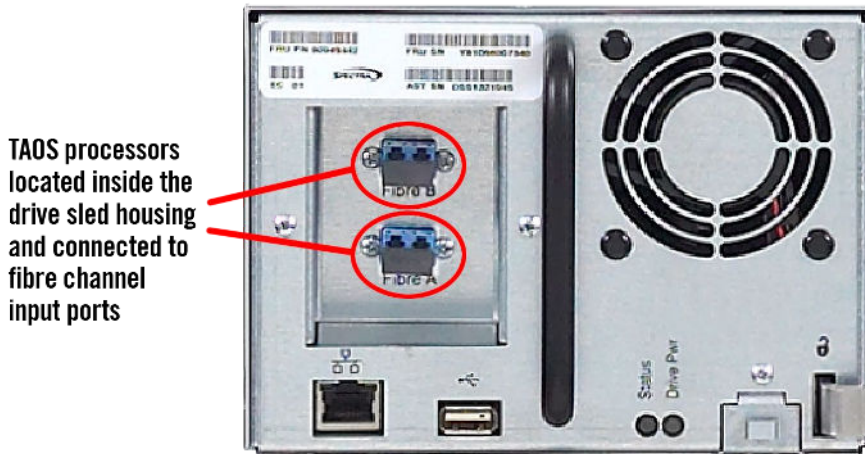
This implementation allows direct access to tape information upon load and provides a simple interface mechanism to the higher level HSM software. Each LTO-7 and newer tape drive will expose an additional SCSI device on the same physical connection on LUN 1. This second SCSI device will be listed as a 'Spectra TAOS' device with a serial number that matches the tape drive. Matching the serial numbers allows for backup software to easily know which Spectra TAOS device should be communicated with for file recalls.

The Spectra TAOS SCSI device supports the Generate Recommended Access Order (GRAO) SCSI command and the Receive Recommended Access Order (RRAO) SCSI command. Up to 3,000 files can be sent for reordering at a time. In comparison, the IBM® TS115x drive series support 2,730 files at a time.

The returned list of files from the RRAO command can then be sent on by the tape application software such as HPSS and DMF to the SCSI Tape Drive device on LUN0.

TAOS is only supported on LTO-7 and newer tape drives, with any compatible media, and in the Spectra T950 and TFinity ExaScale libraries (additional libraries will be added later based on demand). TAOS will also be supported with the upcoming Spectra Swarm™ Ethernet / RoCE drive interface solution.

Spectra expects that TAOS will be supported by HPSS in their fall 2018 release, and DMF in an upcoming release in late 2018 or early 2019. It is anticipated that other HSM software will follow and support the TAOS feature.

Process order:

1. HSM software orders requests based on tape affinity and commands tape load
2. HSM software generates GRAO command to LUN 1 for all objects on that tape
3. HSM software requests proper order using RRAO command to LUN1
4. HSM software generates the read requests on LUN0 for that tape in the order provided by RRAO
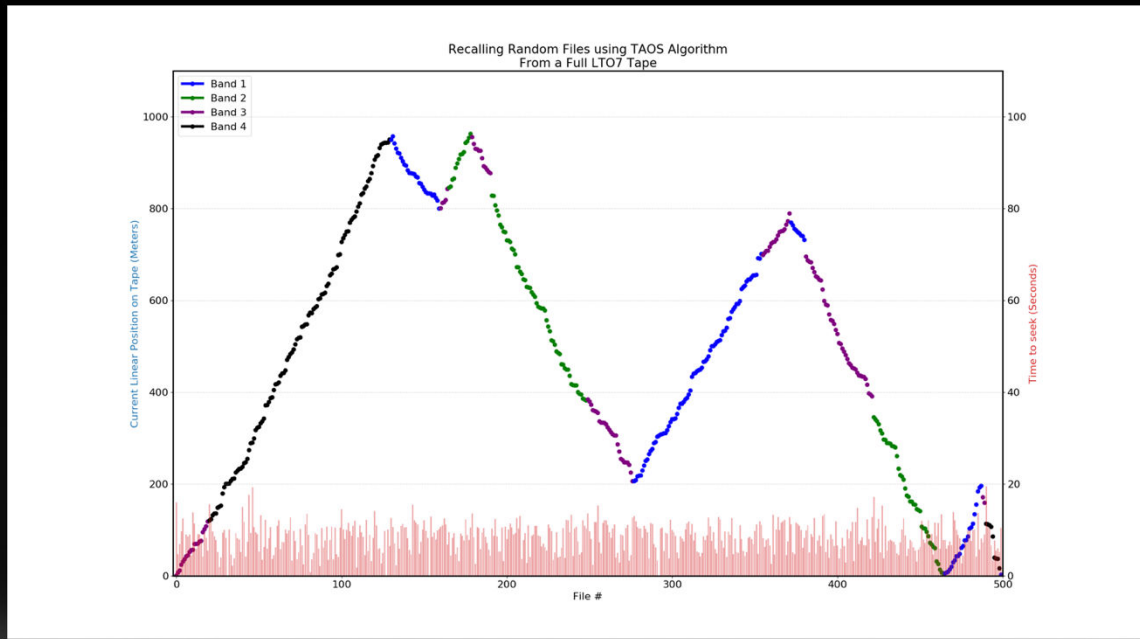
Results

As expected, without using TAOS, unordered requests generate excessive tape movement.
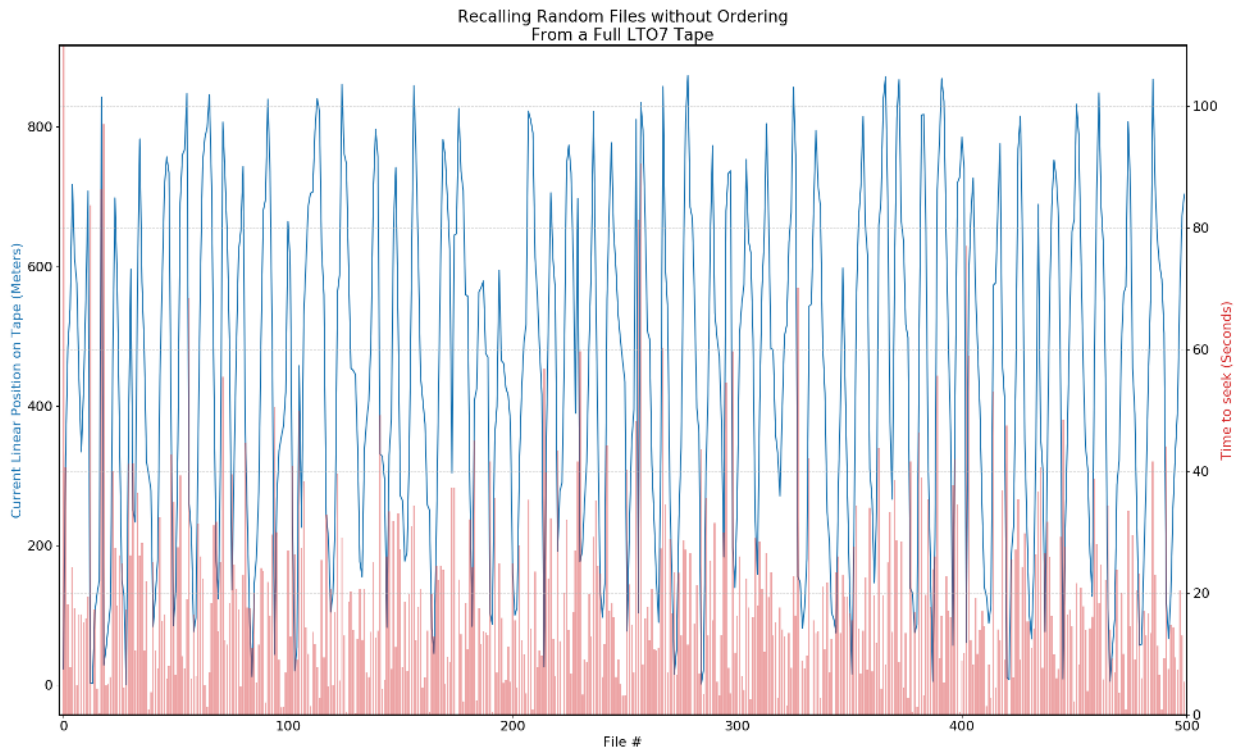
The blue graph shows unordered access while the red shows the seek time between reads using standard unordered request.

Using Spectra's TAOS algorithm also takes into account needed band changes, which for LTO drives takes around 2.75 seconds to complete. The following graph shows a typical tape access including band changes with each band access color coded to highlight the intelligent method in which the tape is traversed.

Using TAOS to create an optimized list of file reads results in up to a 4 times decrease in overall time needed to retrieve all of the files and a 13 times reduction in the length of tape run across the drive head. This reduction in the length of tape has a direct relationship with reducing the amount of tape drive cleans and tape drive wear.
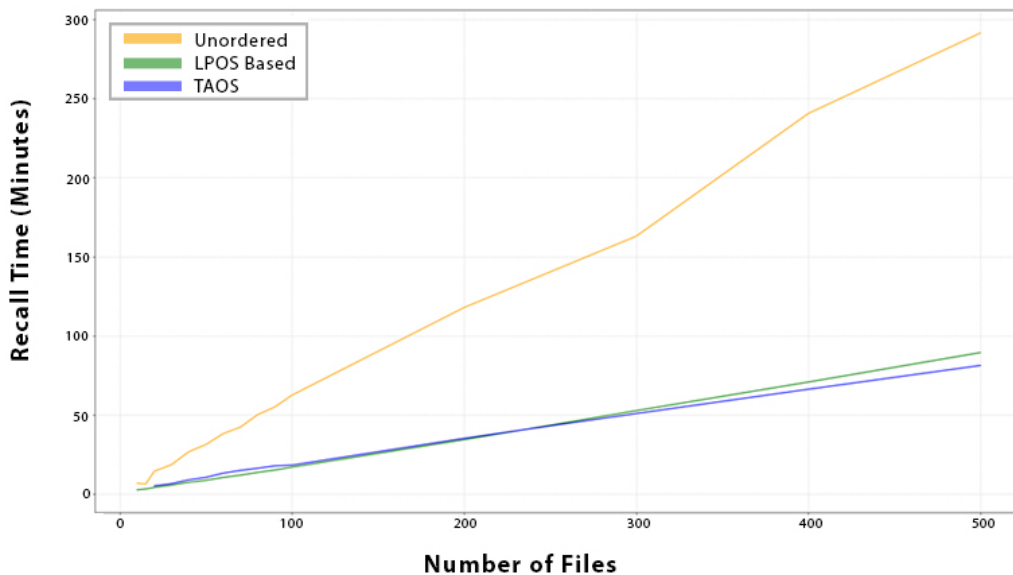
Conclusion

TAOS is a unique Spectra advancement that provides up to a 4 times improvement in overall access speed on reads and up to a 13 times reduction in tape movement across the drive head. The latter also results in reduced tape and drive wear providing the secondary benefit of improving overall system reliability and reducing cost.

With multiple decades of experience in robotic tape innovation, Spectra continues to advance tape library technology.   In addition to introducing TAOS, the new release from Spectra will include Zoning and TeraPack® Affinity. These new features, combined with its High Performance Transporter (HPT), deliver twice the robotic mount performance of a single-partition ExaScale TFinity running without these features. ASM Platinum is also now generally available to all enterprise tape customers, allowing them to perform all major service operations independently, speeding the time to repair and lowering overall support costs. Moreover, the most recent revision of Spectra Certified Media maximizes the capacity of every tape cartridge sold by Spectra Logic, including both LTO and IBM's TS tape technology, providing lifetime guarantees. These and many other tape innovations in development demonstrate the long-term commitment to tape archive technology that Spectra Logic provides to its enterprise customers worldwide.

About Spectra Logic

Spectra Logic develops data storage and data management solutions that solve the problem of long-term digital preservation for organizations dealing with exponential data growth. Dedicated solely to storage innovation for 40 years, Spectra Logic's uncompromising product and customer focus is proven by the adoption of its solutions by leaders in multiple industries globally. Spectra enables affordable, multi-decade data storage and access by creating new methods of managing information in all forms of storage—including archive, backup, cold storage, private cloud and public cloud. To learn more, visit www.SpectraLogic.com.