30 May 2001

IBM TotalStorage^{tm} Peer-to-Peer Virtual Tape Server with Model B18 VTSs

By	*Aare Onton*	*Senior Engineer/Scientist*
	Jesse L Thrall	*Senior Programmer*

Contents

Introduction

This paper provides performance information on the IBM TotalStorage^{tm} Peer-to-Peer Virtual Tape Server (PtP VTS). The PtP VTS provides automated dual copy tape data management and storage through a single storage system image, one copy on each of two VTS model B18s (B18 VTSs). The PtP VTS physically comprises of two suitably configured B18 VTSs, each with four Extended Performance ESCON channels (EHPO) and the Performance Accelerator feature (PAF), interconnected with four Virtual Tape Controllers (AX0) which are unique to the PtP VTS. The performance related architecture of the 3494 model B18 and its performance is described in a separate performance white paper [1].

The two physical components of the PtP VTS (AX0 VTCs and B18 VTSs) can be physically adjacent or they can be separated by up to 26 km. The second copy of the virtual volume can be made immediately at volume close time (*immediate copy* mode), or its timing can be managed by the PtP VTS using customer set policies (*deferred copy* mode). The *deferred copy* mode

***Fig. 1**. Peer-to-Peer VTS.*

is useful at times when having a write bandwidth approaching that of two B18 VTSs for a period of time is an advantage.

Highlights

Product Description (PtP VTS)

Figure 1 shows the physical configuration of the PtP VTS. It shows two B18 VTS units connected via ESCON communication links. The PtP VTS comprises of (from left to right) an *IBM 3494 Tape Library*, extendable from the two unit model shown, the *B18 VTS*, and a frame which houses the *AX0 VTCs.* Having Virtual Tape Controllers at both VTSs is optional; they can all be housed with one VTS.

The ESCON connections are illustrated in the PtP VTS schematic in Fig. 2. This figure shows the interconnection scheme between S/390, the AX0, and the B18 VTSs. All the channels are ESCON; from the AX0 V there are eight channels to S/390 hosts, while there are four paths to each of the B18 VTSs. All data transmission between the two B18 VTSs occurs via the four paths through the AX0 VTCs.

PtP VTS with eight host ESCON channels

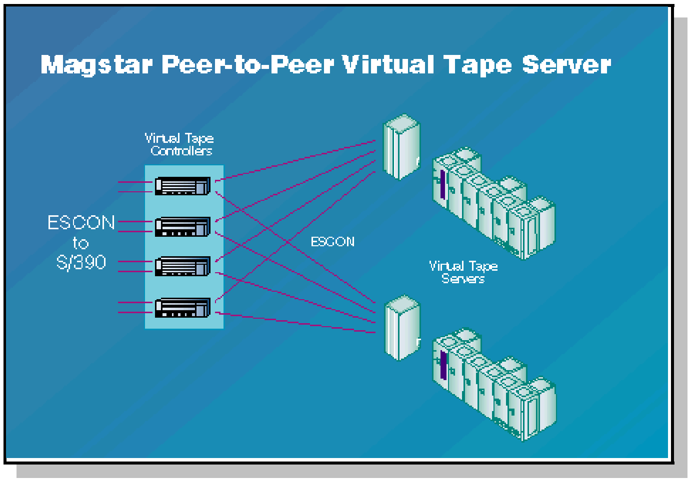

Fig. 2*. The Peer-to-Peer VTS showing ESCON interconnection with the Virtual Tape Controllers.*

Operational Modes

Immediate-Copy and Deferred-Copy Modes defined

The operational modes refer to how data written to the PtP VTS are handled. Data is initially directed to one of the two VTSs. Balancing algorithms keep the loads on the two VTSs approximately equal. A copy is made to the other VTS when the tape volume on the first VTS is complete in tape cache. The copy is made prior to completion of the logical rewind of the original (*immediate copy* mode) or it can be deferred to a later time (*deferred copy* mode). In *immediate copy* mode, when a volume completes *close* processing, it means that the PtP VTS has completed performing the copy. The deferred copy mode is provided to balance the PtP VTS workload when very high input rates must be sustained for periods of time while periods of lower input rate allow the later synchronization of the data on the two VTSs.

The immediate-copy and deferred-copy modes are the only user selectable operating modes available on the PtP VTS. In addition to this mode selection, the observed throughput performance can depend on the initial state of the tape volume cache (TVC), the write content and compressibility of the workload, and how long the operation has been sustained. For each of the operating modes we define additionally a *"peak"* throughput, observed at the beginning with either an empty TVC or one in which all new or updated volumes have been copied to the other VTS and physical tape. We define as a *"sustained"* throughput one that is observed after sustained operation with a given workload; after which it can be verified that the content of the TVC is in dynamic equilibrium. There can also be periods of other, intermediate throughput, in the transition from peak to sustained throughput.[1]

Regardless of the operating mode, the internal algorithms do as much of the background work (peer-to-peer copies and copies to physical tape) as possible with any excess bandwidth that is available. Thus, unless there is a strict requirement to keep the VTSs synchronized, the best performance can be observed with the PtP VTS in *deferred copy* mode, within the constraints detailed on page 6. When the maximum write input is occurring, potentially almost all asynchronous background copy work is suspended in order to handle read/write traffic with the host. The other extreme is the *immediate copy* mode in which all dual copies are being made. In either mode, the sustained throughput performance is approximately the same.

Performance Metrics

In the following sections we present the performance of the PtP VTS as viewed from the host. The metric we use is megabyte per second (MB/s) in each of the possible combinations of PtP VTS modes and B18 VTS operating manners. The data is presented as a function of data compressibility since data is compressed by the VTS. Data compressibility has a significant effect on the observed PtP VTS throughput.

All of the measurements and modeling of performance assume eight active host ESCON channels. The workload comprises sixty-four applications,

---

[1] The definition of *"peak"* and *"sustained"* operation here have been modified from those in PtP VTS performance white paper dated 28 July 2000. For the user observable cases, they describe the same operating regimes.

writing 800 MB tape volumes simultaneously. The block size used is 32 KB and the BUFNO parameter is set at 20. When there is a read component to the workload, it is assumed to be in volumes of 250 host MB. The workloads used in this paper are either "100% write" or "typical mix." The latter comprises 60% writes and 40% reads, with the reads having a 50% hit ratio in tape volume cache (TVC).

Performance in Local Operation

In this section we present data for the case when all PtP VTS components, AX0 VTCs and B18 VTSs, are local. Performance with any of the components remote (i.e., at greater than 1 km distance) is discussed beginning on page 10.

Figure 3 shows the expected performance of the PtP VTS under a 100% write workload. The curves represent data derived from a performance model calibrated to the measurements shown.

Write throughput as a function of:

- ***Data compression factor***
  - ***Mode of Operation***

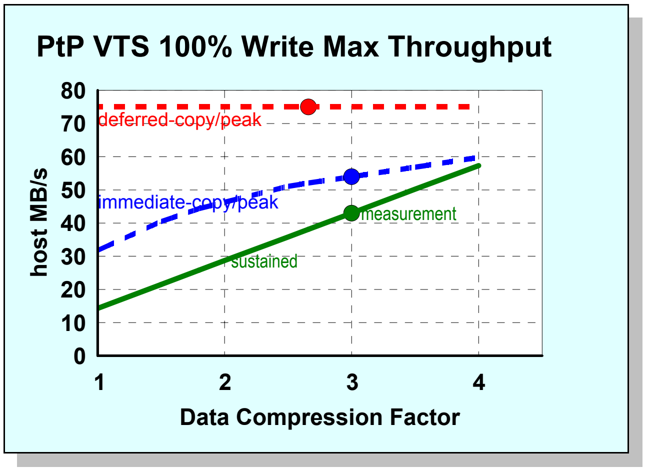

Fig. 3. *PtP VTS maximum throughput in host MB/s for a 100% write workload as a function of data compression factor. These data assume that all AX0 and B18 VTSs are local.*

The upper curve, marked deferred-copy/peak is the performance with copies from one VTS to the other being deferred in favor of maximum host write throughput.

The middle curve shows the performance in *immediate copy* mode. In this mode a copy is made on the second VTS before the *rewind/unload* completion is given to the host. The *immediate-copy/peak* curve describes performance in which copies to tape are not necessarily keeping up with host input. In the *sustained* write throughput, reached after an extended period of operation, the rate of copying of data from tape cache to physical tape is the same as the host input rate. The maximum host throughput is reduced by the work required to make the second copies concurrently with host write input. The *sustained* throughput is approximately the same for both *deferred copy* and *immediate copy* modes.

The *deferred* and *immediate copy* modes of operating performance represent the envelope of write throughput bandwidth. In the *deferred copy* mode, copies make up the balance of the PtP VTS workload if the host input is not at peak bandwidth. Thus, there is no substantial loss of PtP VTS resource utilization in the *deferred copy* mode if the host input should lapse. In effect, any time the host I/O rate drops below the maximum *deferred copy* mode level some of the background peer-to-peer copies and copies to the stacked tape are done. If the write input drops to about 40 MB/s (at a data compression factor, CF=3), then the background work in host MB/s approximately keeps up with the write input rate (i.e., the amount of deferred copy and copy-to-tape data in TVC stops increasing). This is different from the *immediate copy* mode in that response time to the end of tape *close* processing is quicker; but the copy process is executed at a lower priority. Thus, except for applications requiring the higher level of data availability and security against loss, the *deferred copy* mode makes efficient use of PtP VTS resources, offers the better response time, and the best response to host requirements for maximum throughput.

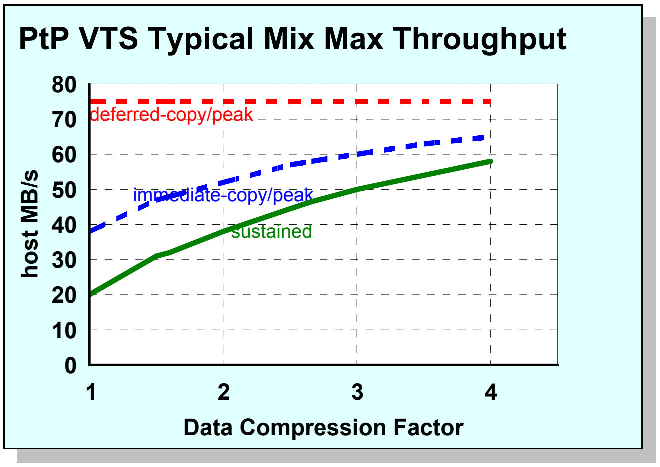

***Fig. 4.** Modeled PtP VTS maximum throughput in host MB/s for a "Typical Mix" workload as a function of data compression. These data assume that all AX0 and PtP VTS are local.*

A typical read/write/hit-ratio workload defined

The "Typical Mix" Workload throughput performance

The *typical mix* workload is defined here as a mix of 60% writes and 40% reads (in terms of host write/read MB). In addition it is assumed that 50% of the reads are tape volume cache hits and that the remaining reads require staging of data from physical tape. This mix represents approximately what is found in a grand average of operating statistics in actual application environments. Individual applications or time periods during operation can vary widely from this average.

The typical mix workload performance projected from modeling is shown in Fig. 4. The *deferred copy* mode maximum throughput is approximately the same as for 100% writes. In the *immediate copy* mode the maximum throughput is somewhat higher than with 100% writes. That results primarily from the fact that read I/Os do not require a copy operation between the peer VTSs. The sustained throughput for the mixed workload is somewhat lower than for 100% writes because read recalls are competing for the tape drives with the copy processes to physical tape. In Fig 4, the *sustained* curve, the relative number of tape drives allocated to copies and recalls in the model has been varied to obtain maximum throughput - such control is not available to the operator of the VTS.

Tape Volume Cache (TVC) Effects on Performance

The length of time for which a *deferred copy* write mode can be sustained is dependent on the host input rate and the tape volume cache capacity, because the mode requires the buffering of data in the TVC for later copy operations.

How long can I run in deferred-copy mode for a given tape cache capacity?

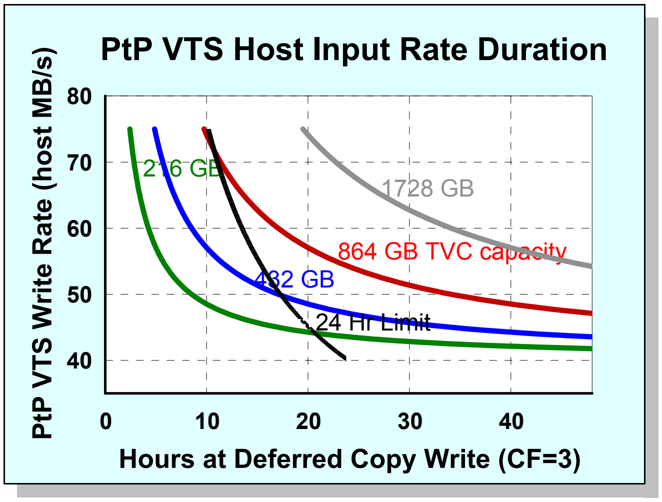

Fig. 5. *The PtP VTS host deferred copy write input rate maximum duration as a function of VTS TVC capacity and the "24 hour limit," if it is required that all data be copied within a twenty-four hour period. A data compression factor of three is assumed. The TVC sizes are given in physical GB per VTS.*

Figure 5 shows, for four TVC sizes (physical GB per VTS), the length of time that the PtP VTS can be operated in *deferred copy* write mode before it becomes necessary for the PtP VTS to utilize some of its bandwidth capability to make dual copies of logical volumes which have accumulated in the TVC. For example, if the PtP VTS is operated at the write rate of 70 MB/s host input (assume a data compressibility of three), it can operate in the *deferred copy / peak* mode for over eleven hours if the TVC capacity is 864 GB per VTS.

At lower input rates the time to fill the TVC to the maximum allowed un-copied data level is extended. For example, if the average deferred copy write rate is 60 MB/s, then the capacity criterion for the 864 GB TVC allows the accumulation of the input for about 17 hours. Note that these curves include an estimate of how much background peer-to-peer copy work is getting done at any given input rate.

In estimating the hours at a 60 MB/s deferred copy write input rate from Fig. 5, we start at the 60 MB/s level on the vertical axis and proceed to the right until we intercept the TVC size of the PtP VTS configuration being assessed. Note that, if the TVC size is 864 GB, we first cross the line labeled "24 Hr Limit" at about 13.5 hours. This limit means that if we want to manage the PtP VTS installation in such a way that there is no un-copied data remaining more than twenty-four hours, then we have to stop host input entirely at 13.5 hours for there to be enough time for all the un-copied data to be copied to the peer VTS in the remaining time of that twenty-four hour cycle.

The corresponding "24 Hr Limit" at the maximum deferred copy write input rate of 70 MB/s is at about 11 hours, somewhat before the maximum un-copied data limit of the 864 GB cache has been reached.

Another parameter affecting deferred-copy write performance is the *deferred copy priority threshold* that specifies the maximum age of un-copied data in the TVC (in integral 0 to 24 hours). Once this *hours* age has been reached by a tape volume its priority for being copied is increased.

This parameter exists to minimize the time that the age of un-copied data will exceed the *hours* value. The way it can affect performance is that when the *hours* threshold is reached the copy load is added to the existing host I/O load. For example, if the PtP VTS is operating in deferred-copy mode and un-copied data in TVC reaches the specified *hours* threshold, the observed write rate will drop because of the additional copy load. Thus, for best performance, the *deferred copy priority threshold (hours)* parameter should be set to exceed the anticipated critical peak write requirement period duration. It also requires that there was no old un-copied data in the TVC at the beginning of the critical peak write period.

If there is no host I/O input, the PtP VTS can copy approximately 65 physical GB/hr between the PtP VTSs (32.5 GB/hr in each direction). For a PtP VTS with two 864 GB caches, only half of which is allowed to fill with un-copied data, the time to copy all the data then is about 13.3 hours.

All of the TVC size curves in Fig. 5 are asymptotic to the 40 MB/s level. At a deferred copy write rate of about 40 MB/s there is no net accumulation of un-copied data in the TVC.

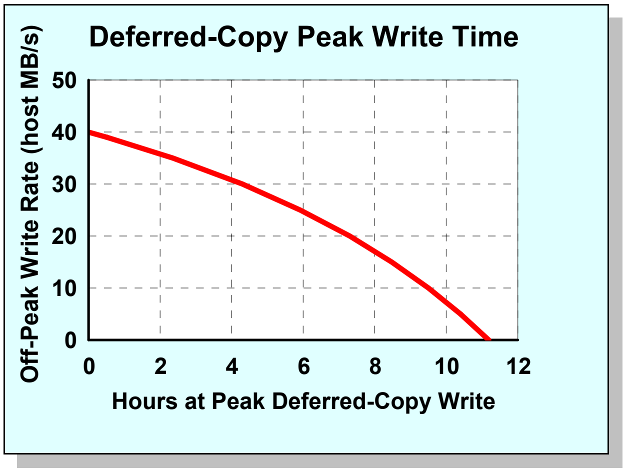

***Fig. 6.** An approximate empirical guideline for estimating the number of hours one can operate the PtP VTS at a 70 MB/s deferred copy write rate in a twenty-four hour period given an off-peak write input shown on the vertical axis. The TVC size must be large enough to hold the un-copied data for the peak period (see Fig. 5) and the deferred copy priority threshold* (hours) *has to allow it. This figure applies for data with CF=3.*

If I run at the maximum deferred-copy write rate for some time, what write throughput can I expect during the recovery period and still have the TVC fully copied to tape at the start of the next 24-hour cycle?

If there is some write input during the off-peak time, the peak deferred copy write window under a twenty-four hour recovery constraint is reduced as shown in Fig. 6. For example, if the average write rate is expected to be about 20 MB/s during the off-peak time, then the maximum allowable peak period rate should be no more than about 7 hours. The formula for the curve in Fig. 6 is:

$$HrAtPeak = \frac{24}{\left(\frac{m}{1-y}\right)+1},$$

where m = *Pk*/(CF*18.06) and y = 1- *Wr*/(CF*13.33), CF is the data compression factor, and *Pk* and *Wr* are the "net" deferred-copy peak and off-peak write rates, respectively. "Net" refers to the fact that only about

90% of the maximum deferred copy write input is accumulating in cache; about 10% of the write input is getting copied to the peer VTS at the maximum write input rate.

Based on the foregoing analysis shown in Figs. 5 and 6, we can conclude that in order to have access to the full range of deferred copy write duration in a twenty-four hour cycle, a 864 GB TVC per VTS is necessary, assuming CF=3. This allows for storage of the peak deferred copy write data as well as extra capacity, thus increasing the read hit ratio. Smaller TVC capacity is possible in environments with a more limited peak deferred-write requirement. Larger TVC capacity can extend peak deferred copy write time over twenty hours and/or increase performance through an increased read hit probability.

All of the data and recommendations in this section should be taken as planning guideline approximations. It is suggested that, in using the deferred copy periods given in Figs. 5 and 6, a safety factor be included as some variation from these numbers can be expected with usage patterns and the specific workload.

Read throughput and response time performance is significantly better if the I/O can be served from TVC (read hit), versus requiring a recall from physical tape. As a result performance planning should take into account an allowance of TVC capacity for read hits. In order to improve the TVC capacity available for read hits, the PtP VTS attempts to keep only one copy of a particular logical volume in cache at any given time. This makes the effective cache size for the purpose of read hits approximately equal to the combined size of the TVCs of the two VTSs.

Remote Dual Copy Performance

The performance cited in the prior section is with all of the hardware "local," i.e., within a computer facility complex (< 1 km). With dual copy tape such as is provided by the PtP VTS, having one of the copies at a remote location provides additional protection against data loss and availability. The throughput performance of the PtP VTS, however, is affected by the distance and the nature of the connection to the remote VTS. Fig. 7 shows the base "local" configuration (1), a common remote configuration (2), and a configuration with most of the hardware “local” except for one of the VTSs at a disaster recovery site (3). (In config. (3) VTS-B may be configured with four standby AX0 VTCs at the remote location.)

The underlying reason affecting performance at a distance is the data propagation time from the origin of the data to the remote location. The propagation rate is that of the speed of light reduced somewhat by the dielectric properties of the transmission fiber. This delay is augmented by the fact that periodically the transmitter has to wait for an acknowledgment from the receiver that the last data block has been received without error. Thus, for every block transmitted, there is a round trip delay before the next block can be sent. The resulting ESCON data rate is determined by a combination involving the distance, transmission and noise characteristics of the line (determines the number of re-transmissions required due to data error), buffer sizes at the source and destination, as well as the logical and transmission block sizes. None of these parameters is user selectable except the choice of distance and channel extender, when required.

Another factor that affects performance at a distance is the mode in which the AX0 VTCs are set. They can operate in “preferred” VTS or “no preference" mode regarding to which B18 VTS they send the host I/O they are handling. In preferred mode the host I/O is sent to the B18 VTS specified, usually the local one (there is no guarantee, however, that the I/O will actually occur at the B18 VTS specified). In no preference mode, the AX0 VTC sends the I/O to the B18 VTS having the smaller number of I/Os outstanding; in effect, it tries to balance the load on the two B18 VTSs.
The choice of "no preference" or "preferred" VTS mode is made at the AX0 VTC. It is a static choice requiring a power-down of the AX0 VTC.

Remote operation configurations

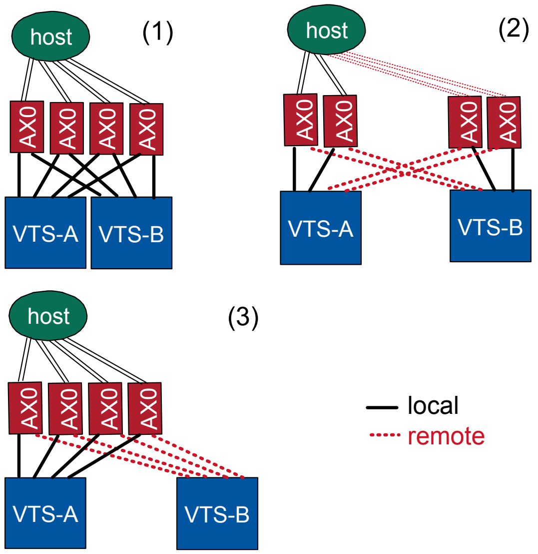

Fig. 7. *Common PtP VTS configurations. All remote distances are assumed to be mediated via IBM 2029 Fiber Savers.*

The following remote VTS performance projections are based on transmission characteristics obtained from a measurement at a distance of

25 km, where the remote connection between the AX0 VTCs and the B18 VTSs is via IBM 2029 Fiber Savers. The configuration used was (3) in Fig. 7

The principal result is that for configuration (3) in which the workload is balanced over four AX0 VTCs. In configuration (2) the workload can be skewed between the local and remote AX0 VTCs if the host to AX0 VTC distance is large enough (more than about 10 km for deferred copy write). Functionally the AX0 VTC/B18 VTS performance of configurations (2) and (3) are equivalent. It is only in the fact that configuration (2) can have an input skew among the AX0 VTCs that the configurations differ.

The remote configurations in Fig. 7 can also be achieved using ESCON Directors (IBM 9032) instead of the IBM 2029 Fiber Savers. Such configurations have been tested to a distance of 26 km. Their performance is expected to be similar to that described here using the Fiber Savers.

The elapsed time (as viewed from an AX0 VTC) is generally shorter for I/O operation on the local VTS than on the remote VTS. This *response time* advantage will be apparent for read and deferred copy mode write I/Os if they are directed preferentially to the local VTS. However, as described below, operation in the “preferred VTS” mode will generally result in overall reduced *throughput* performance.

The at-distance PtP VTS throughput performance is shown in Fig. 8. The characteristics shown apply to both configurations (2) and (3), except for the deferred copy write mode where throughput curves for both configurations are shown. All of the data have been obtained with 32 KB blocking and a a BUFNO=20 (a smaller BUFNO will yield a somewhat smaller throughput rate).

The throughput of the "no preference" mode is always higher than that of the "preferred" mode. This is because the “no preference” mode has the ability to shift new work to balance the work at the two VTSs. This tends to make uniform use of the PtP VTS resources. Specifying a "preferred" VTS at an AX0 VTC can leave one path to a VTS underutilized while the other is operating at maximum throughput, for example.

The “no preference” mode throughput is reduced and becomes asymptotic to the “preferred” mode at large distances; for at very large distances the best work balance for performance is to have most of it done on the local VTS.

The "preferred" mode lines are straight because the throughput, at the distances shown, is determined principally by the local ESCON paths, which remain constant in length. They have about half the throughput of the “no preference” mode because each AX0 VTC essentially has only a single ESCON channel to the “preferred” VTS.

The reason the “no preference” deferred copy mode write performance is lower for configuration (2) than for configuration (3) is that the host to AX0 VTC ESCON distance causes the input to the remote AX0s to fall below that which the AX0s can handle.

As a result of the performance characteristics exhibited in Fig. 8, **the general recommendation is to operate the PtP VTS in "no preference" mode.** Only in cases where the VTSs are split at two separated sites should one consider “preferred” mode operation. In that case one will clearly want the local input to be preferentially targeted first to the local VTS. Even then, there is no *throughput* advantage to “preferred” mode; there is a throughput penalty. The principal advantage is in *response time* performance; namely, by having all tape I/O served locally, the deferred copy mode writes and reads will be more likely to have a shorter open time.

The effect of distance on remote operation Write Throughput

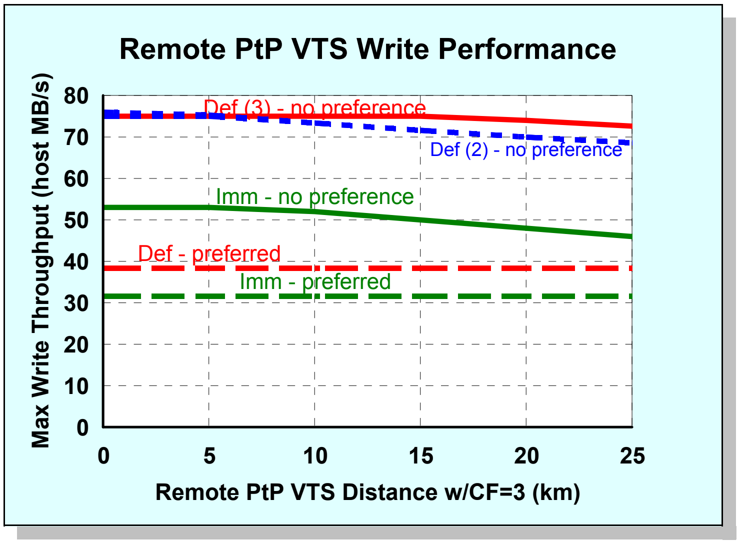

Fig. 8. *The remote* Immediate Copy *and* Deferred Copy *Mode write throughput performance of the PtP VTS using the IBM 2029 Fiber Saver / IBM 9032 ESCON Director. This figure applies to data with CF=3.*

Note that in the "Imm/no preference" mode a fraction of the I/Os incur a double "distance hit." Namely, if a remote VTS is chosen as the primary target for a write by the AX0 VTC, then the peer-to-peer copy has to travel the distance back again. From a throughput point of view this is still better than making all primary writes local. The preferred mode leaves the extended distance ESCON channels underutilized at distances within 25 km.

For practical workloads that involve a mix of read and write I/Os we have modeled the throughput performance of a **Typical Mix workload** (60% writes, 50% read hits). These results again show performance reduction for configuration (2) as shown in Fig. 9. The similarity of these results to those in Fig. 8 arise from an approximate balancing of the effects of less write copy traffic and the fact that the ESCON bandwidth for reads is somewhat smaller than for writes. The results are also sensitive to the size of the recall volume, here assumed to be 250 MB. The "no preference" mode throughput is always greater than the corresponding "preferred" mode case.

Typical Mix Workload remote operation throughput

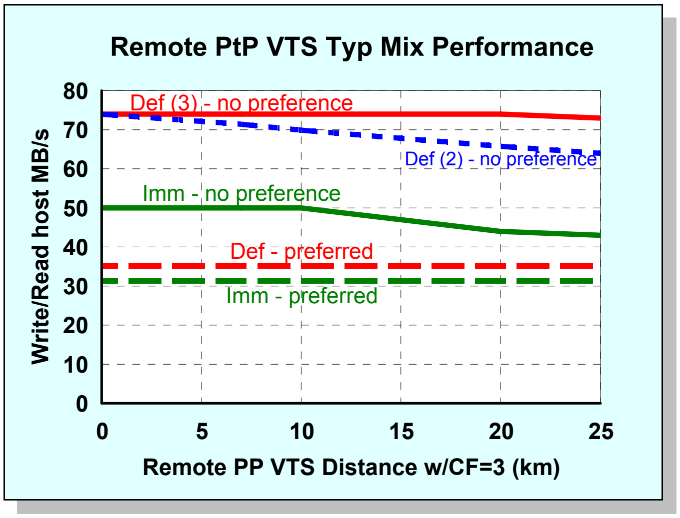

Fig. 9. *The remote Typical Mix (60% Write, 50% Read Hit) throughput performance of the PtP VTS in Immediate Copy and Deferred Copy modes using the IBM 2029 Fiber Saver / IBM 9032 ESCON Director.*

Single Host Job Performance

Throughput is generally defined in terms of the aggregate number of MB that can be written and read by a large number of jobs acting concurrently during a period of time; and reduced to MB/s or GB/hr. This aggregate throughput capability is not directly useful for estimating the data transfer rate of a single job because a single job can only use one host channel, passes through one AX0 VTC, and can only use one VTS to handle the primary input. In addition, while with parallel processes a given resource

Single host job throughput can be significantly lower than the aggregate rate for multiple jobs

can be kept near 100% busy, with a single job some resources are utilized to process a packet at a time with a latency period between packets. For this reason, it is sometimes useful to know single job data rates to estimate job completion time.

On writes, in Fig. 10, we see that single job throughput is unaffected by distance up to 25 km as long as the channel extension (Fiber Savers or ESCON Director) is between the host and AX0 VTCs. Putting the channel extension between the AX0s and B18 causes the PtP VTS single job throughput to decrease systematically with distance.

For reads, having the channel extension between the host and AX0 VTC also has minimal throughput impact at 5 km. However, as the distance increases the read throughput is affected significantly more than the write throughput. By the time the host to AX0 VTC channel extension reaches 25 km, it has (slightly) more impact on read throughput than having the extension between the AX0 VTCs and the B18 VTSs.

Single host job read/write throughput in local and remote operation

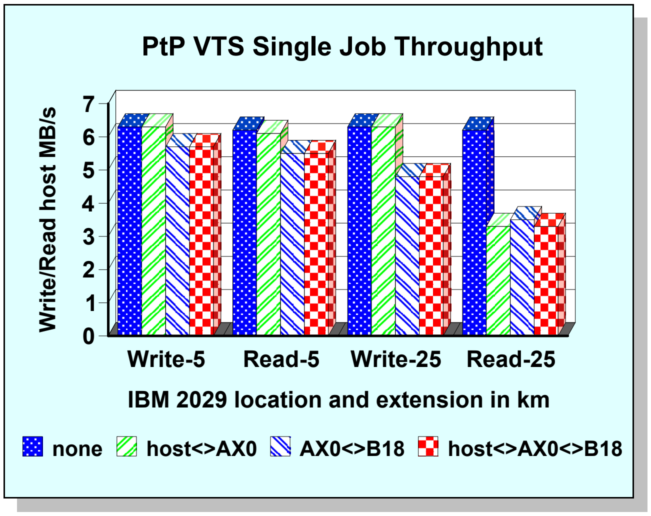

Fig. 10. *The measured single job write or read throughput on a PtP VTS in a remote configuration at 0, 5, and 25 km IBM 2029 extension. The lozenge in each configuration description in the legend represents the location of the IBM 2029. The zero distance ("none" = no IBM 2029) throughput is shown at the left of each configuration/distance group. The data are for 32 KB blocking and BUFNO=6. The data used had an average compressibility of 2.66.*

Support of Remote Operation

Current product support for extended distance (remote) operation with Fiber Savers is for up to 25 km (26 km using ESCON Directors) within the PtP VTS (i.e., between AX0s and B18s) and up to 75 km between the host and AX0 VTCs using IBM 2029 Fiber Savers (43 km with ESCON directors). For most PtP VTS applications this means support is limited to 25 (or 26) km. Configurations over longer distances are considered custom.

These longer extensions can in some cases be achieved with IBM 2029s or other vendor channel extension products. Generically there are two types of such extension devices: (1) signal repeaters and (2) virtual control units.

Some details on Remote Operation

Signal repeaters take an incoming signal, amplify it, and send it along in the direction it was propagating. There is minimal propagation delay through the repeater, but the full round trip propagation delay is experienced by the sender before an acknowledgment is received that a block of data arrived without error. At extended distances this delay per block can significantly affect the data rate of the channel. This kind of channel extension technology has been found to be acceptable in some applications (at reduced throughput) for distances up to about 100 km. An example of such a signal repeater for ESCON channel extension is the IBM 2029 Fiber Saver. (Note that for PtP VTS operation at distance more is required than simply a minimum acceptable data rate. The synchronization timing between the peer VTSs also puts requirements on maximum signal delay.)

A virtual control unit operating as a channel extender works in pairs of identical hardware devices. One device is at the sending location and the other is at the remote location. Each device buffers incoming data and immediately returns an acknowledgment to the local ESCON connected to it. Thus communication over such an extension appears to be local. Of course, the extended distance link between the virtual control unit devices still introduces response time delays. However, as long as the distance link bandwidth is adequate, it can appear as if the ESCON channel has been extended without affecting its bandwidth. This benefit is expected to extend at a reduced level even to single job throughput. Such extension can, in principle, occur over continental distances (say 3000 miles).

At this time, using channel extension technology within the PtP VTS other than the IBM 2029 to 25 km (or IBM 9032 to 26 km) is considered to be a custom installation not covered by the standard installation agreement. There are also no implied performance guarantees for such custom installations.

Small Volume Effects on Throughput Performance

The performance information presented up to this point has been based on measurements performed with full 3480E logical volumes (800 host MB), and 250 host MB logical volumes in the modeling for the Typical Mix workload. There are special performance considerations that need to be made if the average volume size in the workload is smaller.

A number of effects can reduce the PtP VTS throughput if the tape volumes are small

There are three VTS internal per-volume overheads that become significant for small volumes. All host volume sizes are quoted in MB before a data compression by a factor of three is applied (CF=3):

- There is about a 1.5 second Library Manager overhead in cataloging volumes in the tape library. This translates to a rate of about 2400 volume mounts/hr or a minimum host volume size of about 114 MB required to achieve the maximum PtP VTS write throughput.
- There is a fixed overhead associated with freeing space of the volumes in the TVC which have been copied to physical tape. For small volumes, on the order of 100 host MB or less, the rate at which this can be done will limit the throughput of the VTS.
- There is a latency associated with the beginning of data transfer in copying volumes from the TVC to physical tape. The effect is that tape volumes need to be at least 300 host MB in size (or 100 host MB in size at CF=1) to achieve the maximum sustained 100% write throughput described earlier in this paper. For example, if the host volume size is reduced from 300 MB to 150 MB the sustained write throughput is reduced to about 60% of maximum.

Performance Tools

Help in capacity planning and resolving performance issues

Tape Magic is a high-level tape subsystem configurator available to IBM customer representatives that is intended to give an initial prediction of a tape configuration that would satisfy a customer's tape processing needs. Tape Magic predicts both native and volume-stacking configurations. Input to Tape Magic is answers to a half-dozen or so simple questions about basic customer tape workload characteristics, typically entered via a Thinkpad on a visit to the customer's location. Because Tape Magic does not directly process any host-processor statistical data, such as MVS SMF records, it is also useful for host platforms that do not provide data that can be input to IBM's more detailed configuration tools.

A more accurate assessment of a VTS configuration than possible with Tape Magic can be made by a detailed analysis of the customer's workload as represented in SMF records, RMF data, and tape management system data. The current tool available to IBM representatives is called Consul Batch Magic and provides a detailed analysis of existing customer tape workload characteristics and projects the required VTS configurations for a subset of that workload. CBM uses as input, selected raw SMF records (14,15,21,30) to provide basic tape workload characteristics such as mount and drive allocation activity as well as input and output tape data transfer activity by hour. To project a VTS configuration, the user first uses the extensive filtering capabilities of CBM to identify certain tape activity, such as output files destined for trucking to a remote vault and tape activity that already efficiently utilizes native tape, that will not be volume-stacked. CBM then projects required VTS and native drive configurations based on the current workload. CBM also provides numerous statistics on expected VTS cache performance.

VTS generates data that is transmitted each hour to the host processor, where the data is embodied in an SMF type 94 record. This SMF record also contains information on library performance associated with native tape drives. Information provided in the SMF type 94 record includes logical and physical drive usage, number of fast-ready (virtual scratch), read-hit, and recall mounts, channel and tape input and output data transfer activity, and cache usage statistics. IBM provides routines that give hourly and daily reports on these VTS statistics. This allows the customers to understand the level of activity of their VTS subsystems, and allows customers to also, with assistance provided by IBM field personnel, to determine when the limits of the VTS subsystems are being reached.

Conclusions

The PtP VTS, featuring a new level of data protection, has an ancestry of continuous performance improvement

The virtual tape server, beginning with the VTS model B16, has demonstrated a clear customer requirement for consolidated tape data management and automation, while taking advantage of technological advances that reduce hardware and floor-space requirements. The model B18 VTS, built on the B16 base offers significantly improved throughput performance using data compression (EHPO). The introduction of the *Performance Accelerator* feature provided significantly more power and efficiency in data handling. The Peer-to-Peer VTS builds on this continued improvement by implementing an automated dual copy capability together with a hardware configuration that has the ability to maintain access to data after component failures. The PtP VTS can be split among multiple locations to help ensure continuous data availability even in the event that a disaster at one site makes that hardware completely unavailable. This extension of the VTS tape storage solution technology reflects the IBM storage modular *Seascape* architecture in which technological improvements in components can be quickly incorporated and proven building blocks can be combined to offer new functionality.

References

1. IBM Corporation, "IBM Magstar 3494 Virtual Tape Server, Performance White Paper, Version 4.0," 28 July 2000.

Disclaimers

The performance information contained in this document was derived under specific operating and environmental conditions. While the information has been reviewed by IBM for accuracy under the given conditions, the results obtained in specific operating environments may vary significantly. Accordingly, IBM does not provide any representations, assurances, guarantees or warranties regarding performance. Please contact your IBM marketing representative for assistance in assessing the performance implications of the product in your specific environment.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

IBM Storage System Group
9000 S. Rita Road
Tucson, AZ 85744