

End-to-end Data Protection and Tapes (07-373r2)

Kevin Butt

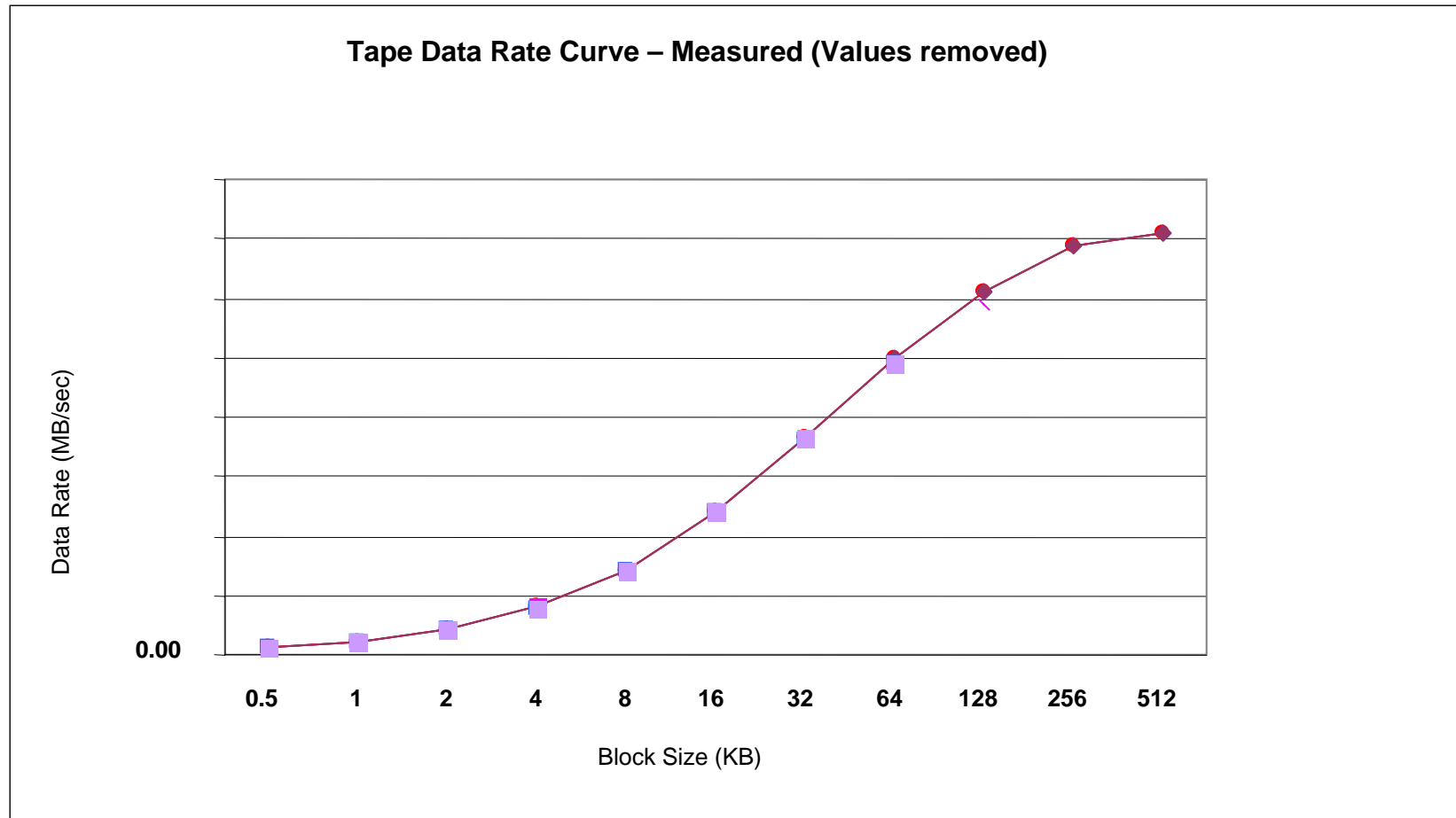
End-to-end data protection

- T10 standardized end-to-end data protection for blocks transferred to disk
 - This is getting adopted by an increasing number of vendors
 - An increasing number of customers are asking for this protection on tape
 - Tape solutions are losing to pure disk solutions because of a perceived data integrity advantage by disk systems
- T10 did not standardize end-to-end data protection for blocks transferred to tape
 - During the work on end-to-end for disk, the tape group was told a common solution for disk and tape was not going to happen and if tape wanted end-to-end data protection the tape group needs to do a unique solution for tapes.

Why Disk end-to-end solution does not work for Tapes

- Disk devices
 - measure performance by I/O operations per second
 - uses fixed block transfers
 - DIF information not included in block length
 - use 512 byte blocks
 - LBAs for blocks are known a priori
 - blocks are not expected to have their LBAs change
- Tape devices
 - measure performance by MB/s
 - use both variable length transfers (most common) as well as fixed block transfers
 - Data Protection information adds to the length of the data transferred and if not included in the transfer length for variable length transfers **will force changes in Transport Layer Protocols** before it can be supported. (e.g. FCP FCP_DL)
 - Performance varies by block size
 - LBA's are not known a priori
 - LBA is determined by where the medium is currently positioned
 - Backup data is expected to be duplicated on multiple copies of media
 - Backup data is often “refreshed” from one piece of medium to another

Typical Data Rate vs Block Size (Tape)



End-to-end data protection is needed on tape devices

- Potential data corruption between the application and the HBA's (same on tape as disk)
- Potential data corruption on the interfaces to the data delivery subsystem on both ends of the wire (same on tape as disk)
- Potential data corruption internal to the device (same on tape as disk)
- Potential data corruption as block is transferred between intermediate devices (e.g., protocol bridges)

Is Protection Embedded by Application Sufficient?

- Many backup applications embed a CRC or other form of protection into their data stream
 - Allows the application to detect bad data stream when the data is read
 - Does not guarantee that the data written to medium is not corrupted
 - Generating this information adversely effects performance because it is done in software

What does end-to-end data protection need to cover?

- Application to HBA
 - Needs supported same as in disk solution
 - Likely will not be supported by applications for a long time
- HBA through service delivery subsystem to tape device
 - Needs supported to cover memory buffers in source and destination
- Tape drive interface through internal workings of tape drive and onto medium
 - To get true end-to-end logical block protection the protection information needs to be saved on medium with the logical block

Likelihood of quick adoption

- Tape drives today save protection information with data blocks.
 - ECC
 - CRC
- Making the protection information fit into existing schemes in use today would greatly accelerate adoption and increase the likelihood of adoption
- To increase likelihood of adoption, the Protection Information should not cause a change in Transport Layer Protocols

Existing vendor-specific methods

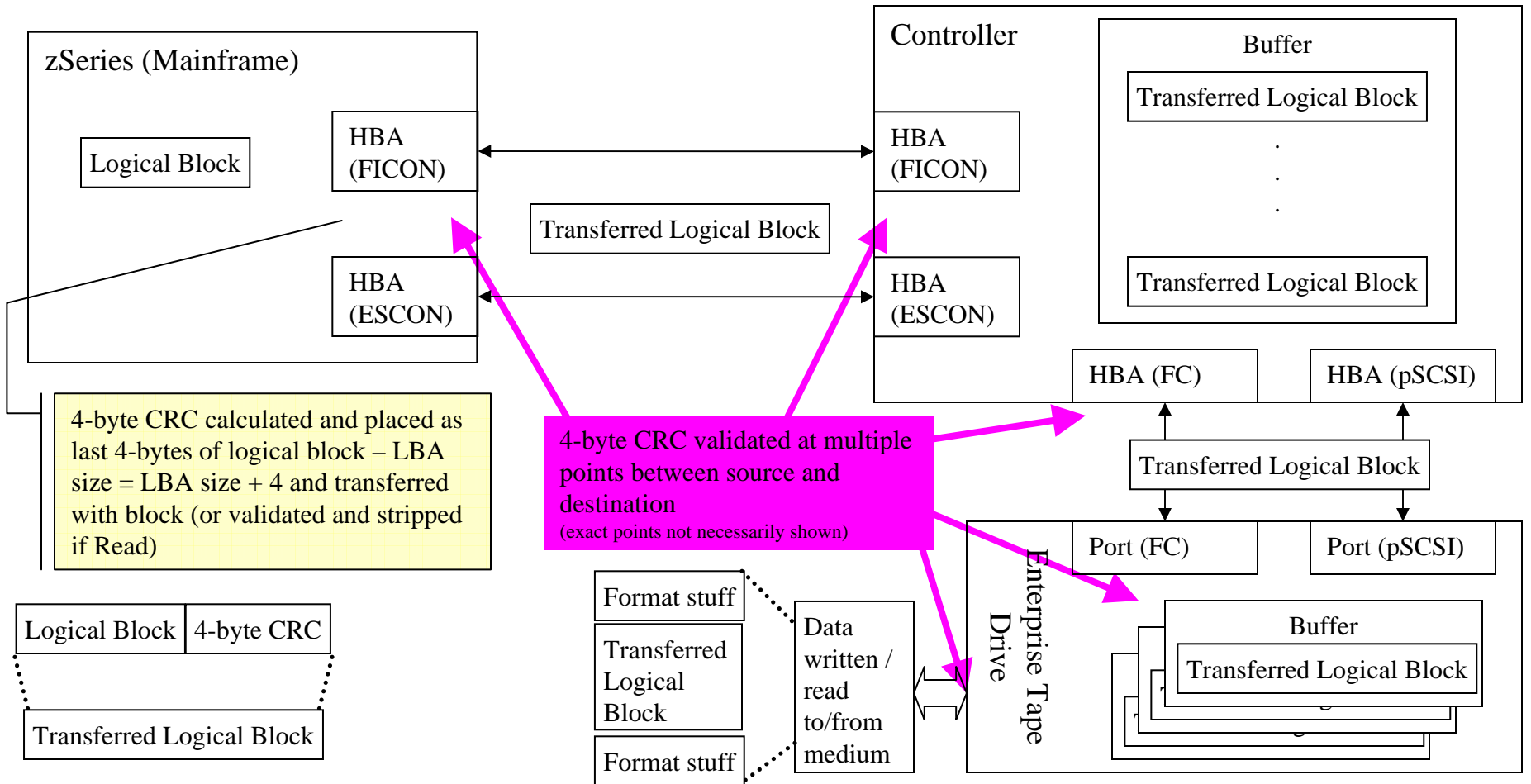
- IBM has been providing end-to-end logical block protection for over twelve years in its enterprise tape drives.
 - Began with 3480 Tape drives
 - Continued in 3590 and 3592
 - 4-byte CRC placed on logical block and transferred with the block
 - Writes: generated at host; validated at drive and written to medium with logical block
 - Reads: read from medium and validated at the drive; validated at the host and stripped from block for application use
 - Validated at multiple points along the path

Benefits to IBM and the customer with this solution

- Prior to solution
 - Data Integrity Issues
 - Difficult to find where the problem occurred
- After solution
 - Quickly find where problems occur
 - Data Integrity issues disappeared (Integrity was assured)
 - When there is bad memory or such, it is discovered before host completes its transaction and believes the data is on the medium

Diagram of IBM Solution

(logical representation)



Proposal – Implement end-to-end logical block protection on tape

- There is a proven method that has been in use for more than twelve years
- Leverage this proven solution
- Make sure that options are available to meet all vendors needs
- For proposal against SSC-3 see (<http://www.t10.org/ftp/t10/document.07/07-374r2.pdf>)