Tape write efficiency improvements in CASTOR

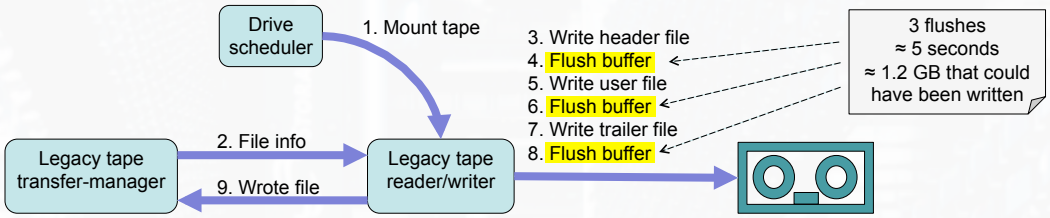
DSS Data Storage Services

CERN IT Department

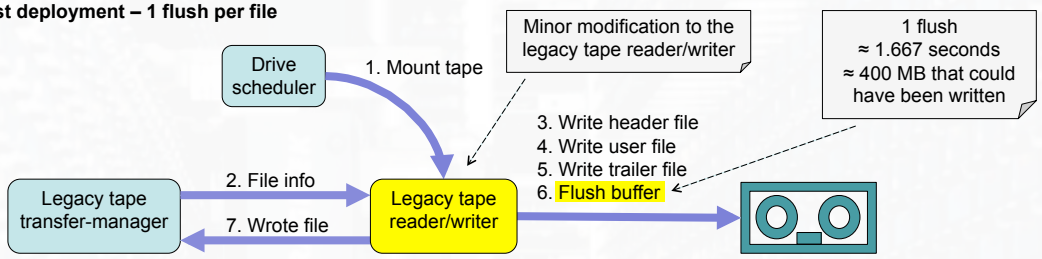
Primary author: MURRAY Steven (CERN) Co-authors: BAHYL Vlado (CERN), CANCIO German (CERN), CANO Eric (CERN), KOTLYAR Victor (Institute for High Energy Physics (RU)), LO PRESTI Giuseppe (CERN), LO RE Giuseppe (CERN) and PONCE Sebastien (CERN)

The CERN Advanced STORage manager (CASTOR) is used to archive to tape the physics data of past and present physics experiments. The current size of the tape archive is approximately 61PB. For reasons of physical storage space, all of the tape resident data in CASTOR are repacked onto higher density tapes approximately every two years. The performance of writing files smaller than 2GB to tape is critical in order to repack all of the tape resident data within a period of no more than 1 year.

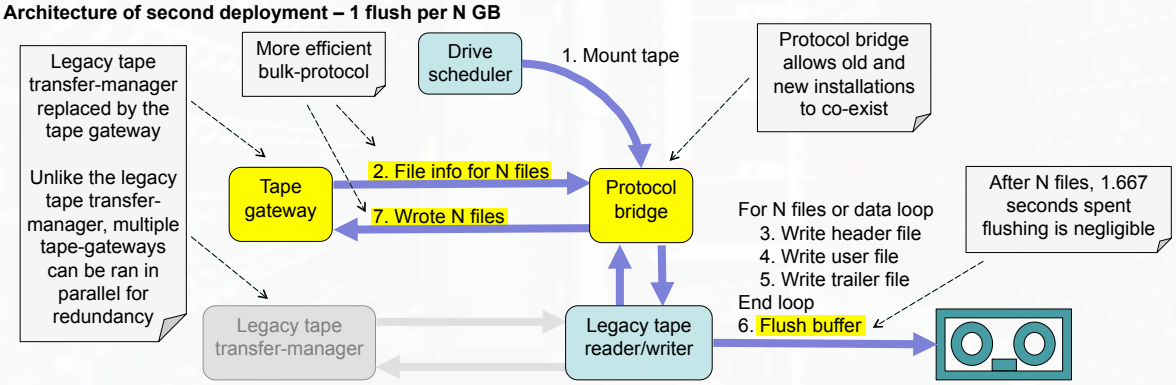
Original architecture before improved write efficiency – 3 flushes per file



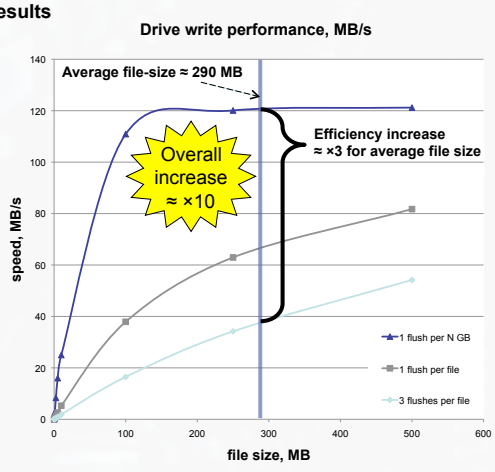
Architecture after first deployment – 1 flush per file



Architecture of second deployment – 1 flush per N GB



Results



Implementing the delayed flushing of the tape-drive data-buffer

- Implemented using immediate tape-marks from version 2 of the SCSI standard
- CERN worked with the developer of the SCSI tape-driver for Linux to implement support for immediate tape-marks
- Support for immediate tape-marks is now available as an official SLC5 Kernel-patch and is in the vanilla Linux-Kernel as of version 2.6.37

Methodology used when modifying legacy code

- The legacy tape reader/writer is critical for the safe storage and retrieval of data
- Modifications to the legacy tape reader/writer were kept to a bare minimum
- It is difficult to test new code within the legacy tape reader/writer, therefore unit-tests were used to test the new code separately
- Used the unit-testing framework for C++ named CPPUnit
- Developed 189 unit-tests so far
- Tests range from simple object instantiation through to testing TCP/IP application-protocols

CERN IT Department
CH-1211 Genève 23
Switzerland
http://cern.ch/it-dss