Data Integrity Face to Face Meeting

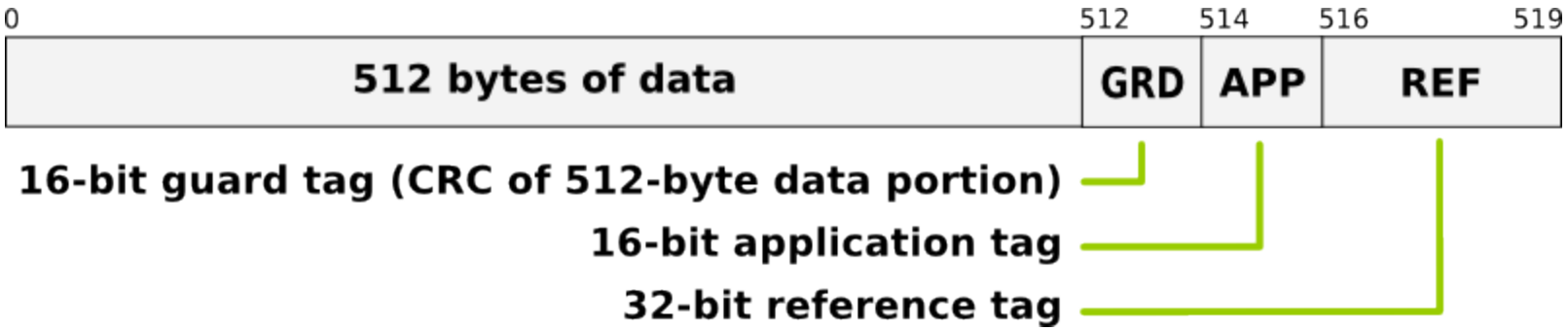
January 31st, 2008

SNIA

Status

- T10 Data Integrity Field
- Existing protection schemes
- I/O controller extensions
- Linux Data Integrity Framework

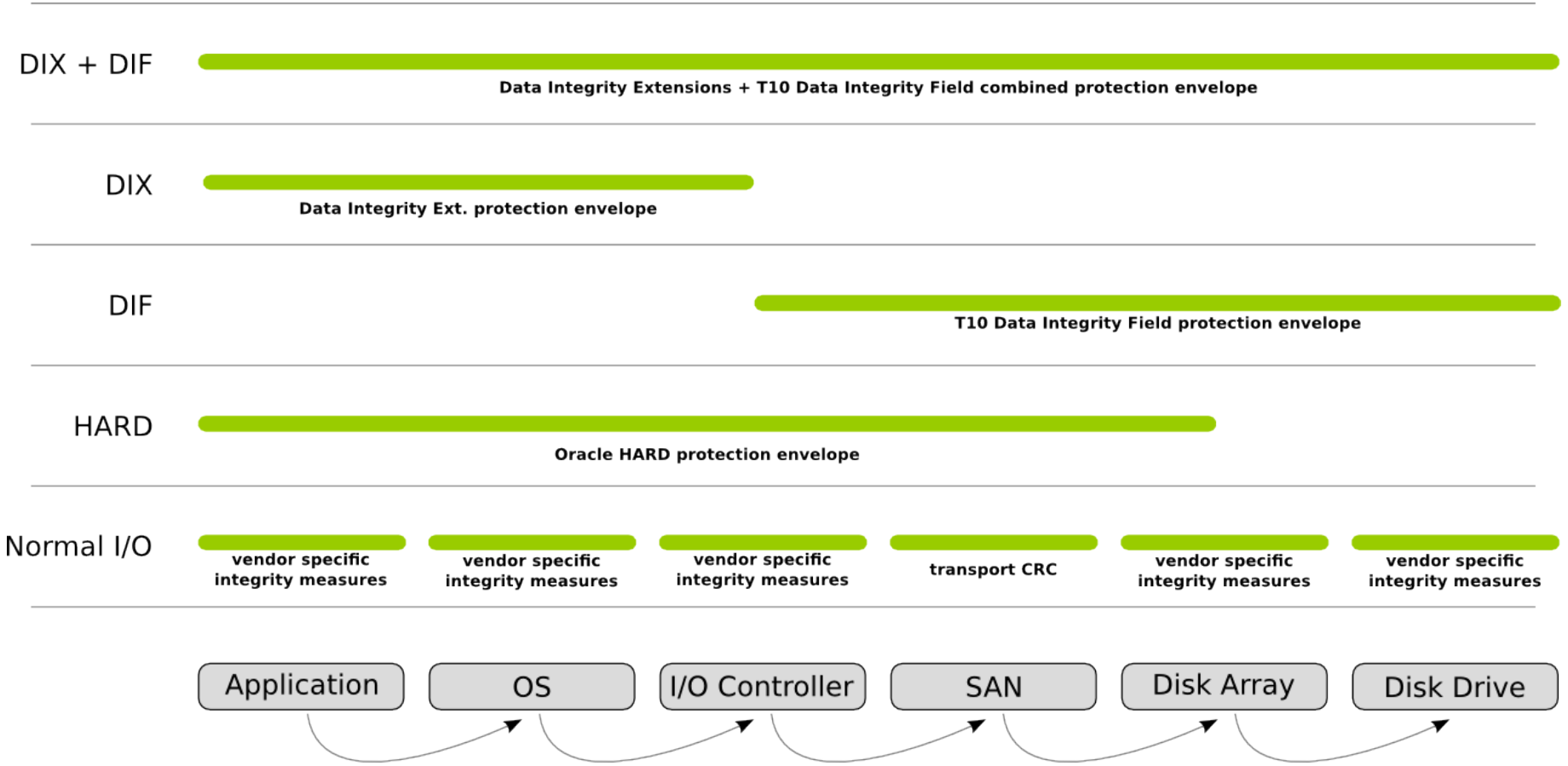
T10 Data Integrity Field



- Interleaved with data sectors on the wire
- Three protection schemes: Type 1, 2 and 3
  - All have guard tag defined
  - Type 1 reference tag is lower 32 bits of target sector
  - Type 2 reference tag is seeded in 32-byte CDB

Existing Protection Schemes

SNIA
Advancing storage & information technology

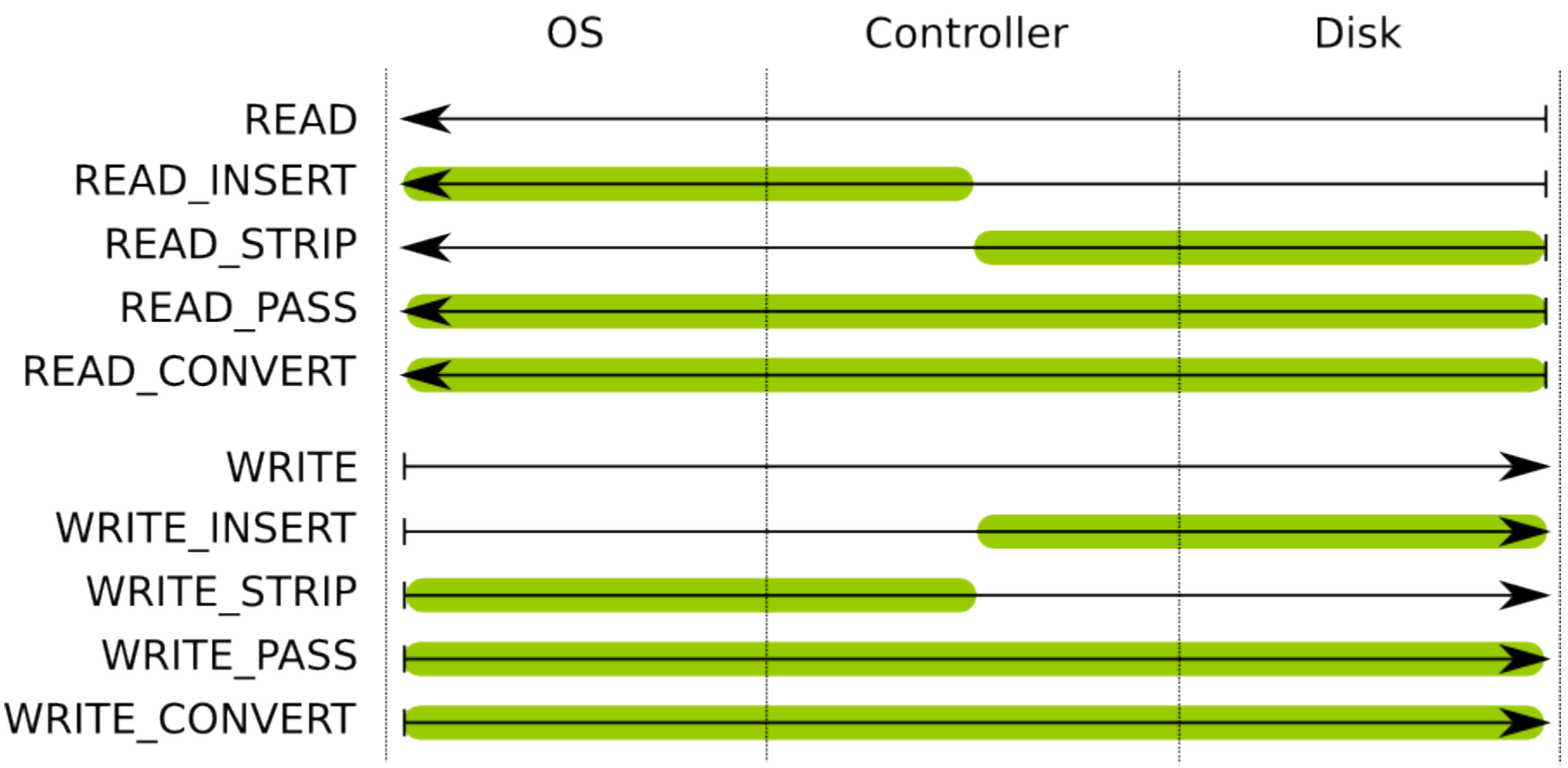


I/O Controller Extensions

- Separate protection scatter-gather list
  - 520-byte sectors are hard for the operating system to deal with
  - Separation of integrity metadata allows data DMA straight to memory
- DIF tuple endianness
  - Application tag must be portable across little and big-endian systems
- Checksum conversion
  - CRC16 is slow to calculate
  - IP checksum is cheap
  - Strength is in data and integrity metadata buffer separation
  - CRC32C in intel’s Nehalem microarchitecture
  - 4KB sectors

I/O Controller Extensions

SNIA
Advancing storage & information technology



Linux Data Integrity Framework

- Attempting to submit changes upstream in current merge window
- Block layer:
  - Full support for attaching opaque integrity metadata to block layer I/O requests
  - Supports merging, splitting and layered devices like software RAID and LVM
  - Filesystems can leverage application tags. Tagging on logical block level.
  - Oracle ASM
- Page cache interface is work in progress

Linux Data Integrity Framework

- SCSI layer:
  - Disk driver has been enhanced so it can prepare DIF I/Os
  - Midlayer allows a separate scatter-gather list for protection information to be mapped in scsi_cmnd
  - Low level drivers can use the protection scatter-gather list and a selection of helper functions to prepare the I/O

http://oss.oracle.com/projects/data-integrity/

New Developments

- T13 External Path Protection
- T10 SSC-3 Tape data integrity proposal from IBM
  - http://www.t10.org/ftp/t10/document.07/07-373r1.pdf
  - http://www.t10.org/ftp/t10/document.07/07-374r1.pdf
- Work on generic user application interface has begun
  - Building upon the new generic async syscall interface in Linux
  - Interface must be orthogonal to synchronicity
  - Extended error reporting is also orthogonal
  - Many levels of dealing with integrity metadata
    - Explicit: fsck, mkfs
    - Opaque: Databases
    - Transparent: Everything else

Userland vs. Integrity Metadata