




**ORACLE®**

## **Eliminating Silent Data Corruption with Oracle Linux**

Martin K. Petersen, Linux Kernel Developer  
Sergio Leunissen, Sr. Director



The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.



## Oracle Linux - History



- Launched at Oracle Open World in 2006
- Compatible with Red Hat Enterprise Linux
- Freely available source and binaries
- Oracle offers Linux support for RHEL and Oracle Linux
- Oracle's base Linux development platform
- Oracle does not use or test on RHEL
- Customers can switch in minutes – no reinstall needed
- Applications run unchanged
- **No Red Hat compatibility bug has ever been reported to Oracle**

# More Than 5,500 Customers Use Oracle Linux



ORACLE



## Oracle: Enhancing Linux for Mission Critical Use

- Linux kernel maintainers at Oracle
- All work submitted to mainline kernel
- Focus on datacenter use of Linux

“Oracle's development work for the Linux kernel represents vital contributions to the open source community, which benefit anyone using Linux.”

– Andrew Morton, Linux Kernel Maintainer, Google



## **New: The Unbreakable Enterprise Kernel**

- Fast, modern, reliable and optimized for Oracle
- Used by Exadata and Exalogic for extreme performance
- Allows Oracle to innovate without sacrificing compatibility
  - Oracle Linux now includes **both** the Unbreakable Enterprise Kernel **and** our existing Red Hat Compatible Kernel
  - You choose at boot time: a system optimized for running Oracle software **or** strict Red Hat compatibility.

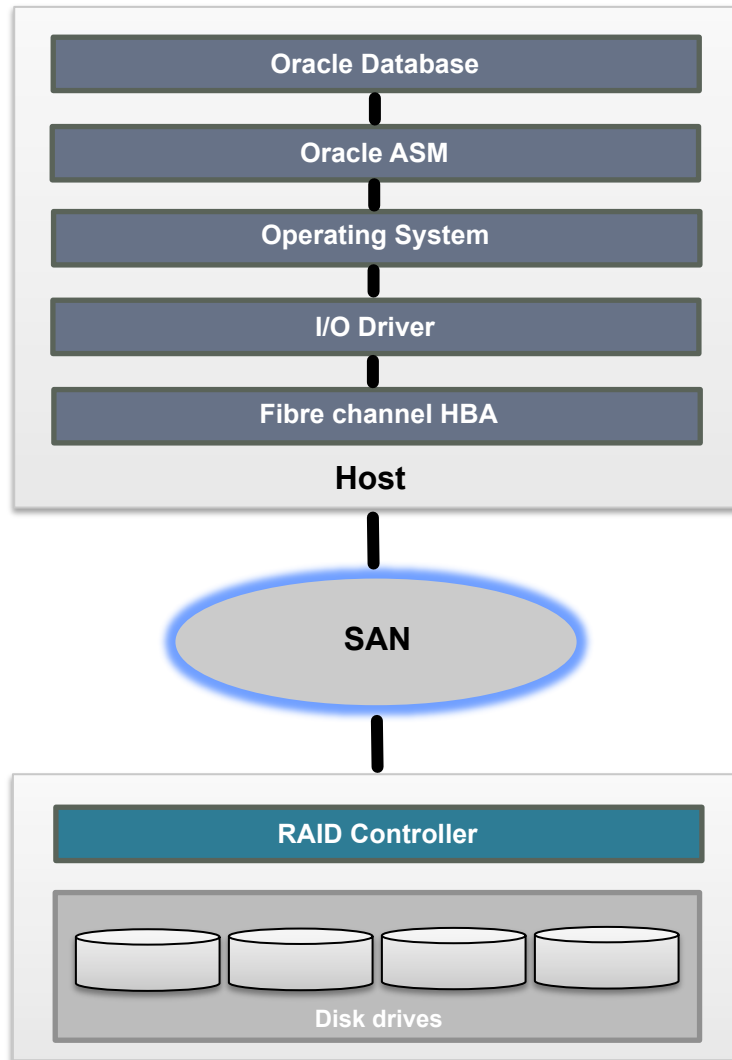
**Oracle now recommends only the Unbreakable Enterprise Kernel for all Oracle software on Linux**



## Silent Data Corruption

- Data corruption that goes unnoticed
  - No errors or warning
- Logical block checksum checking not enough to prevent silent data corruption
  - Often used at READ time, when it's already too late
- Requires end-to-end integrity checking to detect
- There are areas in the data path that can cause corruption
- End-to-end data protection **prevents** bad data from being written

# Potential Data Integrity Problems



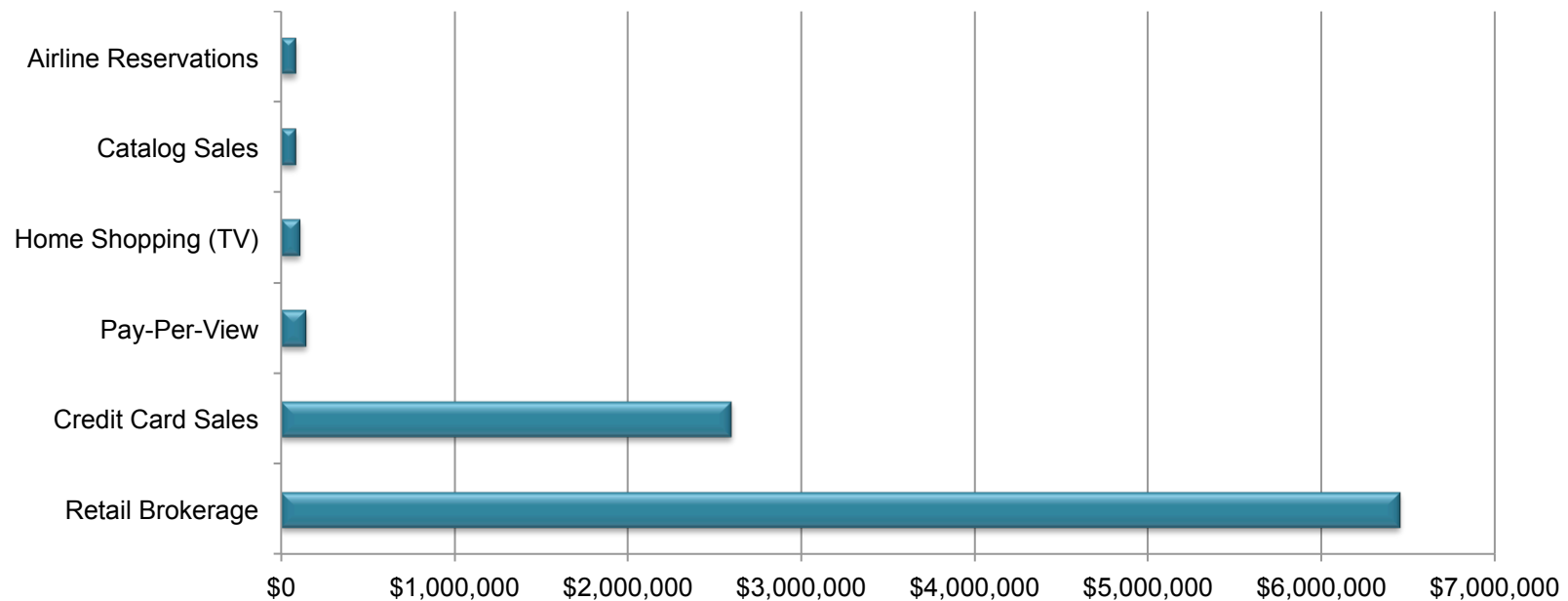
- Hardware
  - Memory, CPU, disk
- Firmware
  - RAID controller, disk
- Software
  - I/O Driver
  - OS: VM, block layer

Software and hardware components in the I/O stack

# The Impact of Silent Data Corruption

- System downtime
- Lost revenue
- Lack of regulatory compliance
  - Gramm-Leach-Bliley Act (GLBA)

Average financial impact per hour downtime by industry



Source: Gartner Group & Contingency Planning Research, Inc.



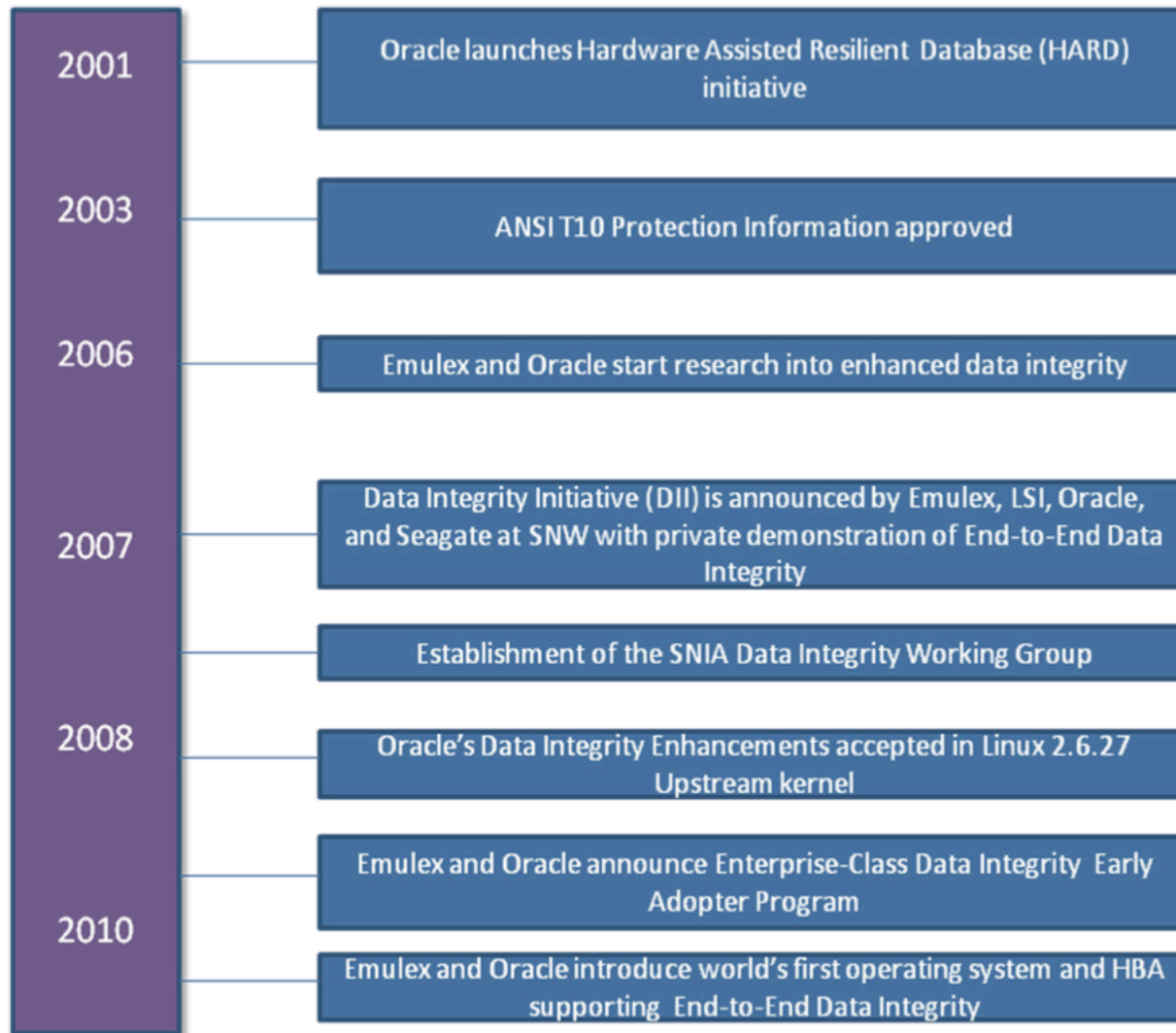
## Silent Data Corruption: How Common Is It?

- NetApp, Univ. of Wisconsin, and Univ. of Toronto study<sup>1</sup>
  - 41 month period
  - More than 1.5 million SATA and enterprise class fibre channel drives
- Silent data corruption detected:
  - 3,078 SATA drives
  - 760 fibre channel drives
- CERN study, 2007<sup>2</sup>
  - Write known data patterns to more than 3,000 nodes
  - 5 week period
  - 22 out of 33,700 files (8.7TB) corrupt
  - Nearly 1 in 1500 files

1) <http://www.pdsi-scidac.org/publications/papers/schroeder-fast08.pdf>

2) <http://indico.cern.ch/getFile.py/access?contribId=3&sessionId=0&resId=1&materialId=paper&confId=13797>

# Data Integrity Timeline





## Industry First: Application-to-SAN data integrity on Linux

- Emulex LightPulse® HBA with BlockGuard + Oracle Unbreakable Enterprise Kernel
- Implements T10 PI Model and Oracle's Data Integrity Extensions (DIX)
- Protects against data corruption in the software stack and HBA
- Support for Oracle Database (10gR2, 11gR1, 11gR2) using ASMLib
- Transparent to application





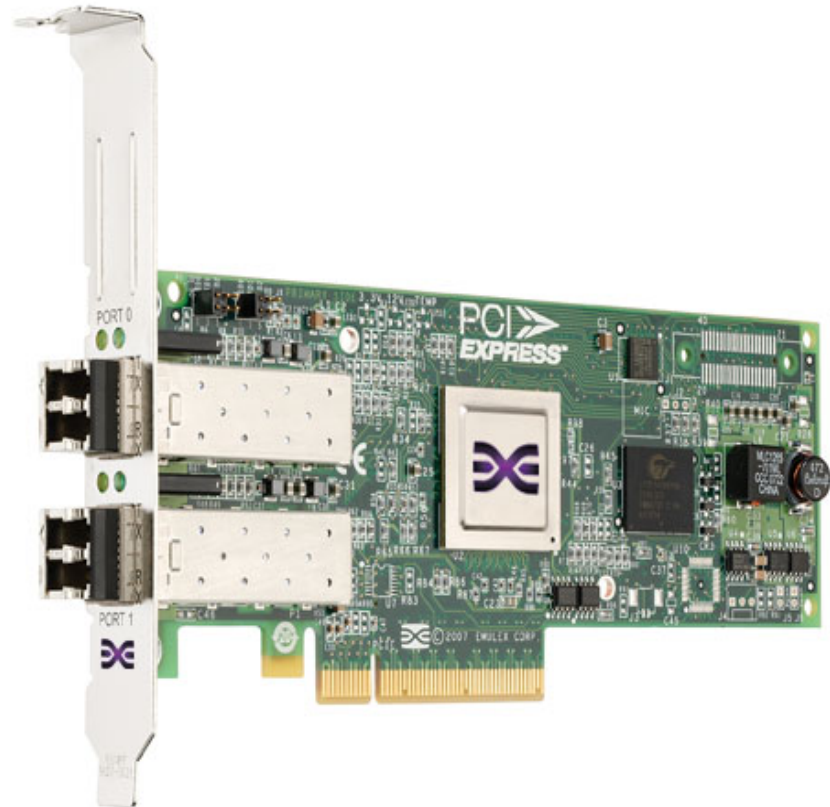
# Data Integrity

## Preventing Data Corruption with DIX and T10 DIF

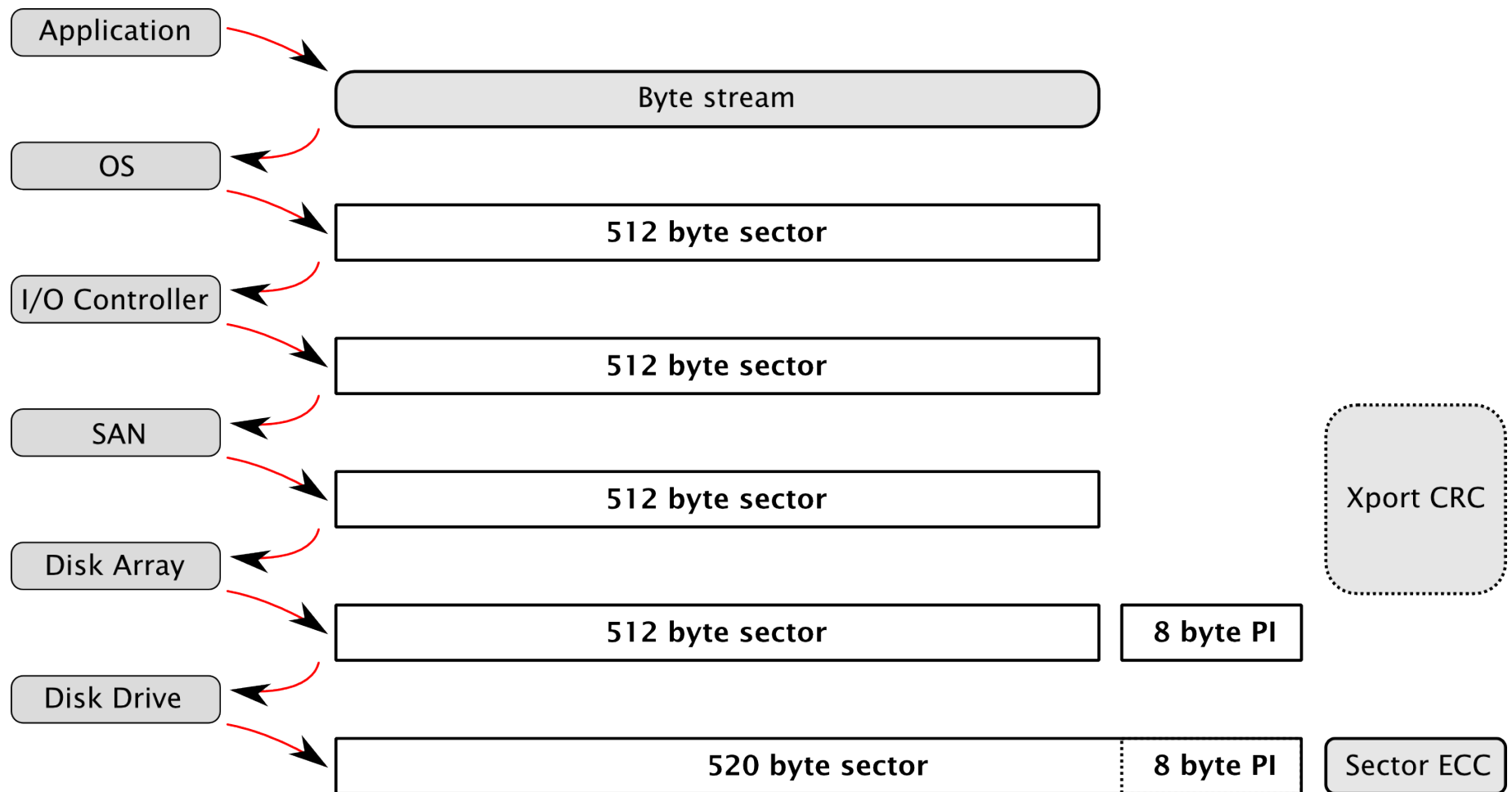
# Emulex LightPulse LPe1200x Fibre Channel HBA



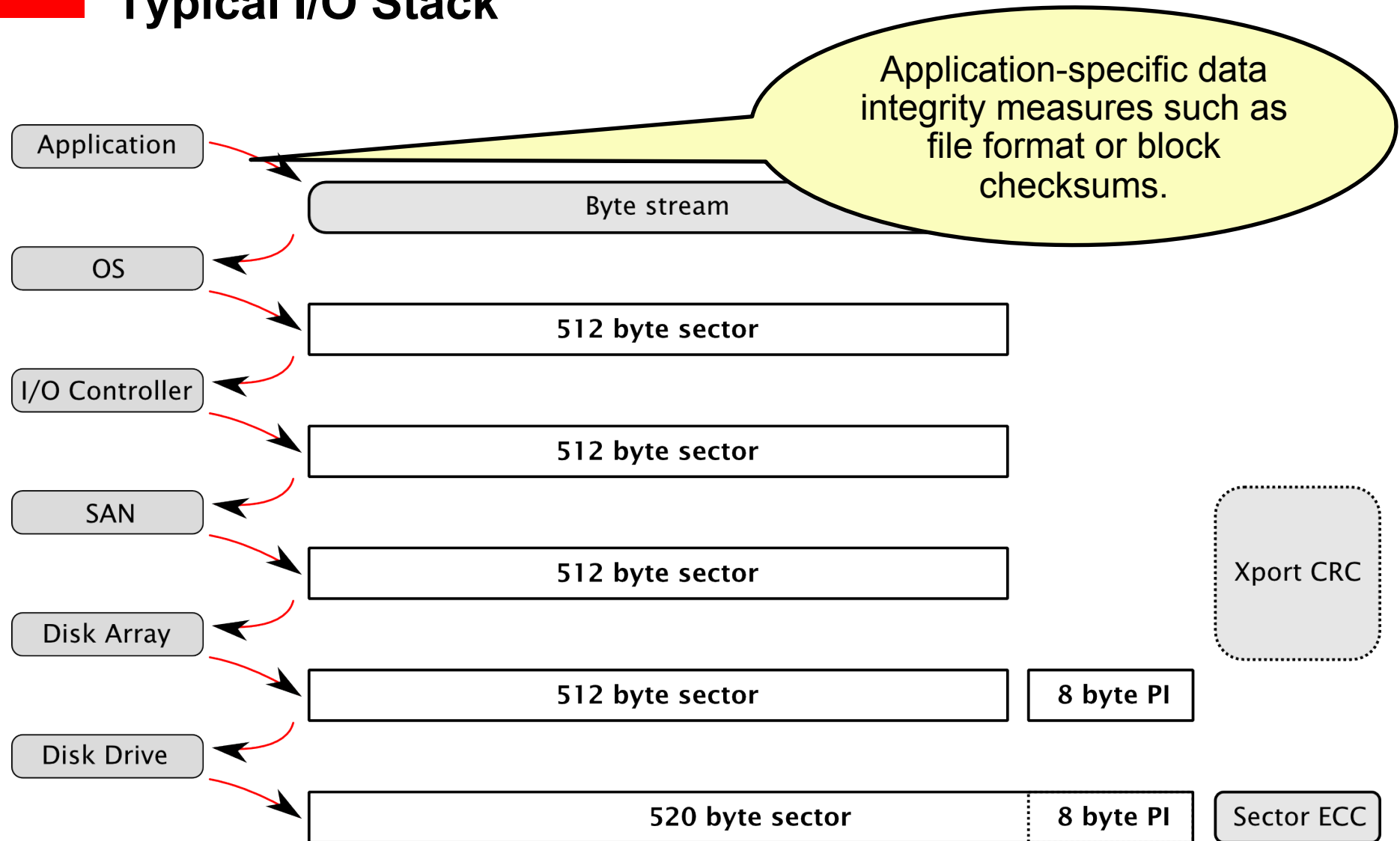
- 8 Gbps Fibre Channel host adapter
- PCI Express 2.0
- BlockGuard:
  - T10 Protection Information Model
  - Oracle Data Integrity Extensions



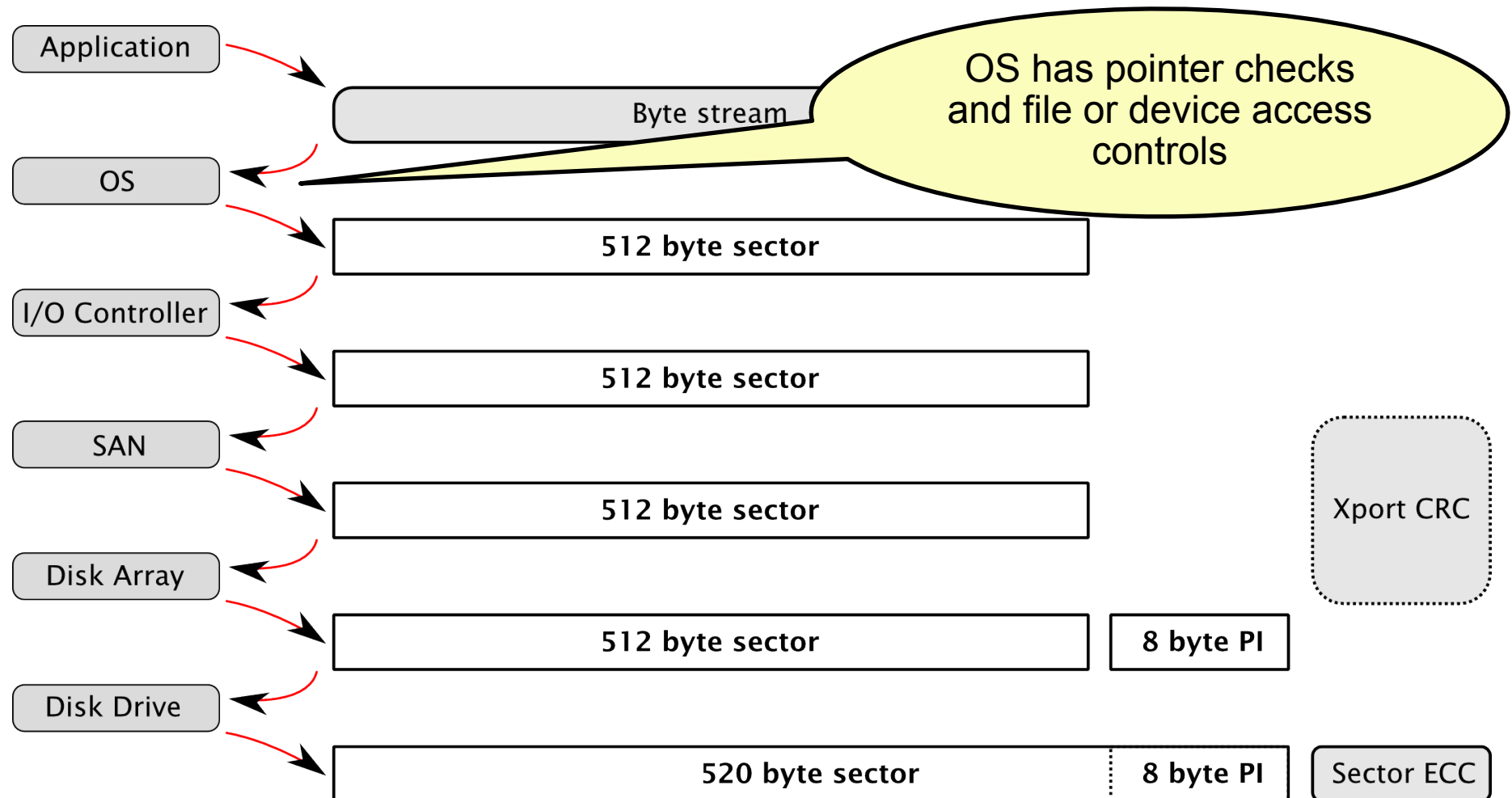
# Typical I/O Stack



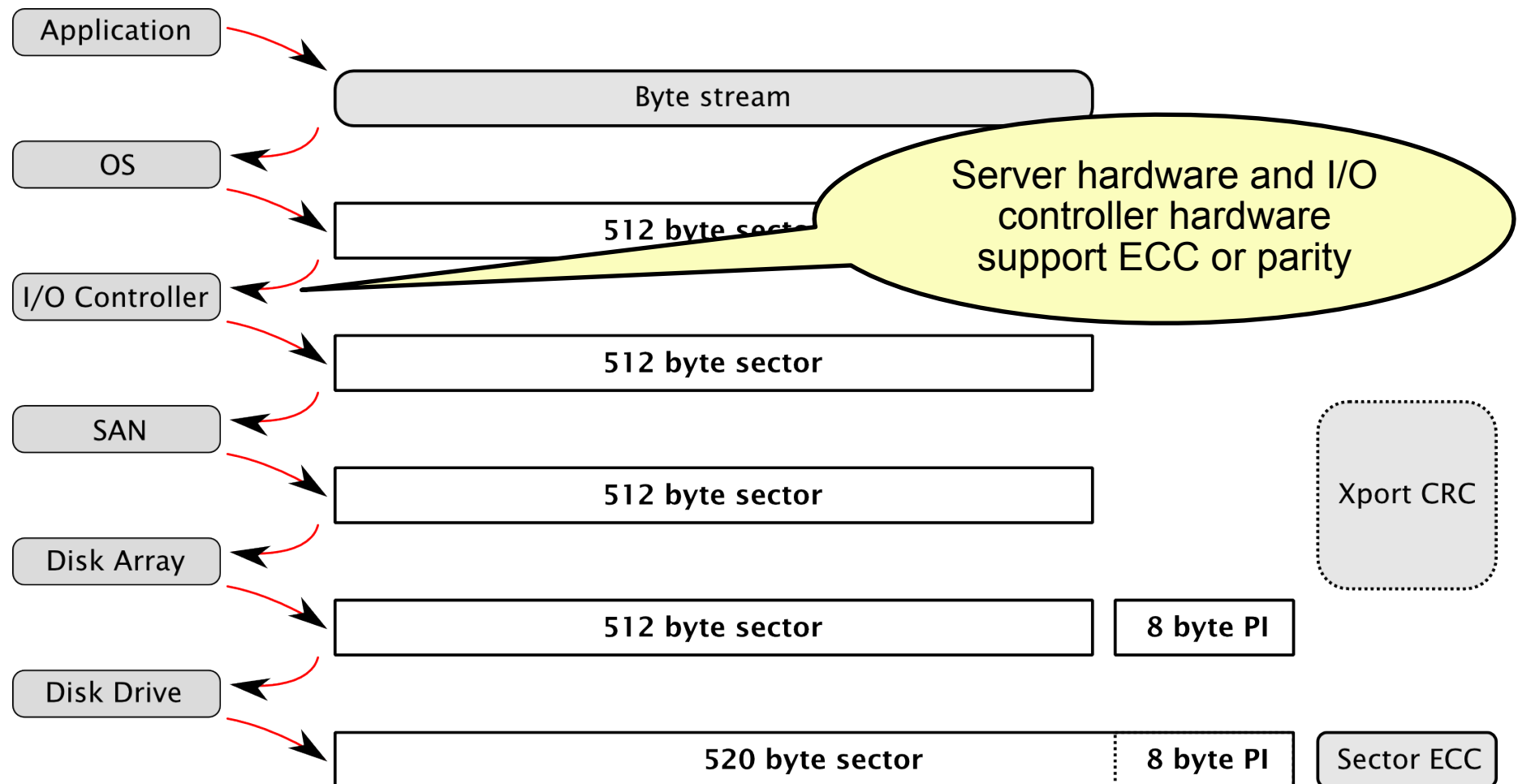
# Typical I/O Stack



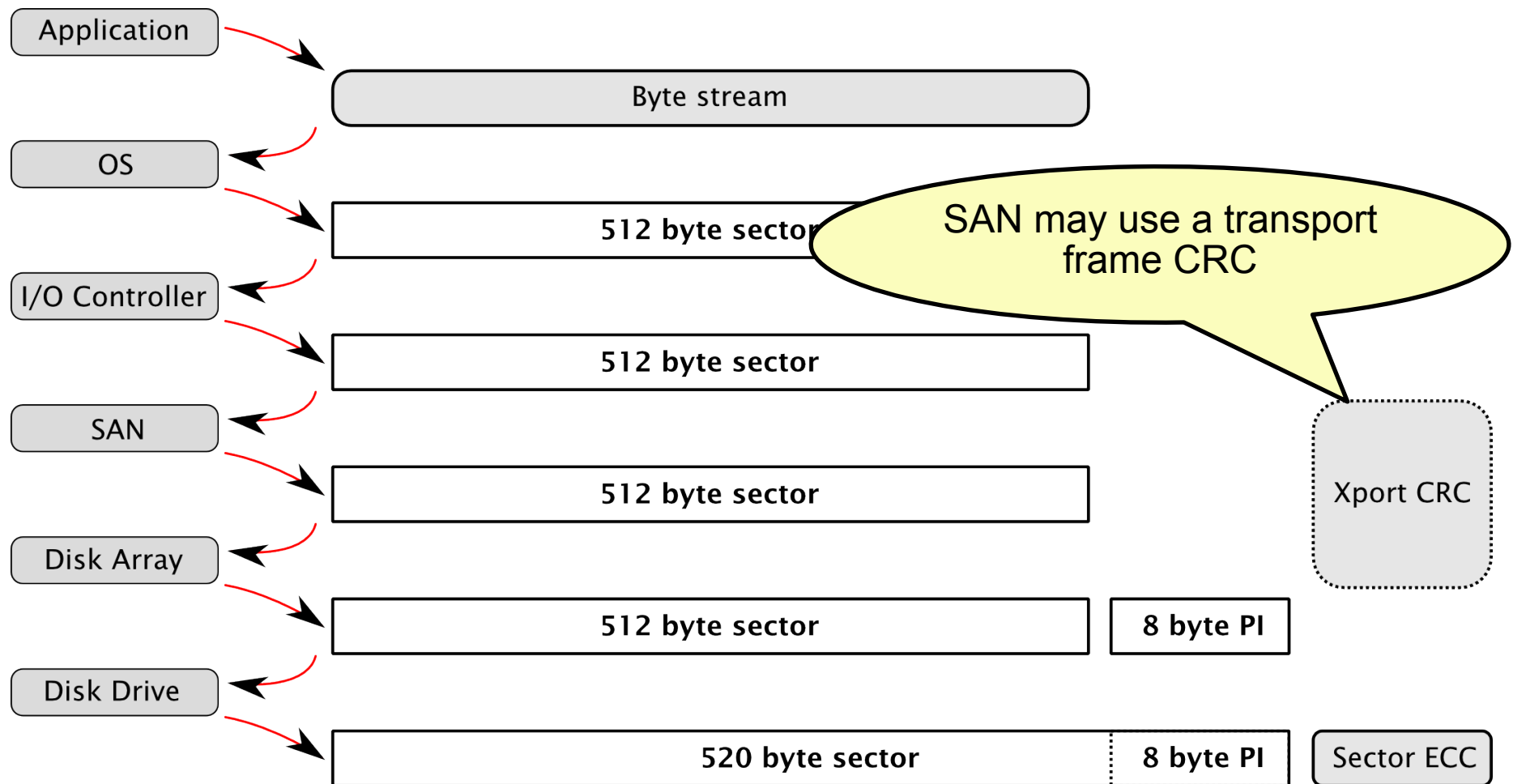
# Typical I/O Stack



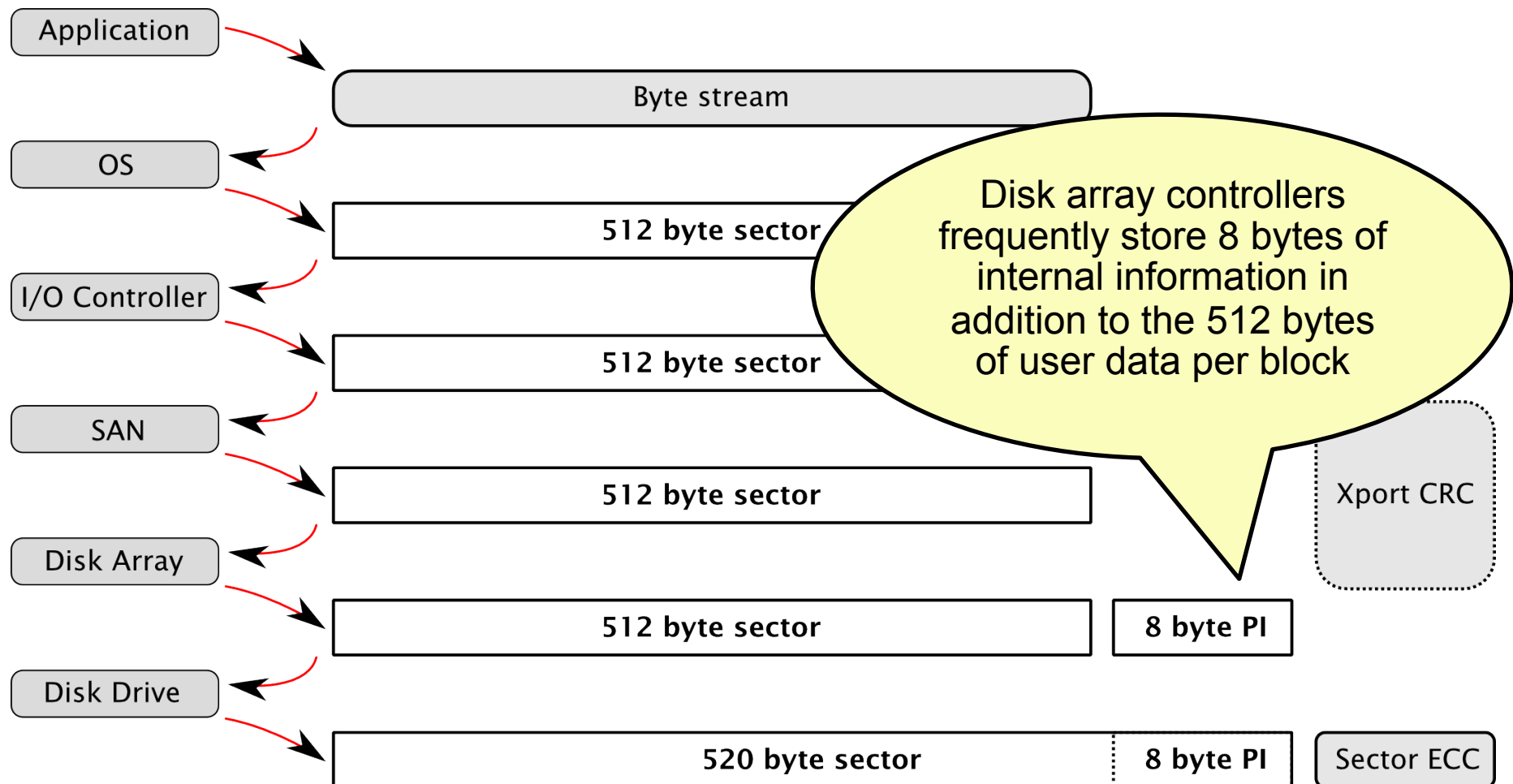
# Typical I/O Stack



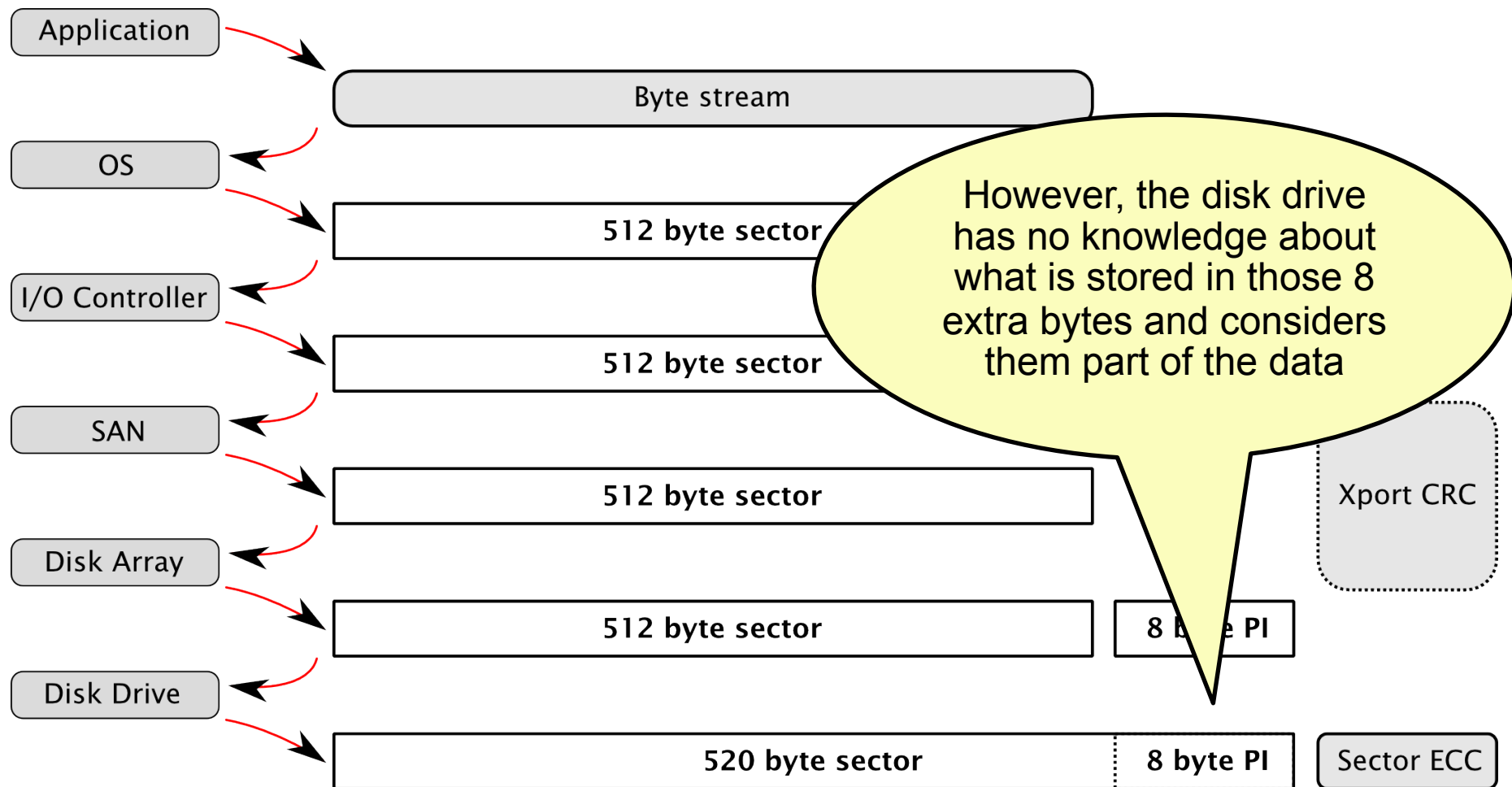
# Typical I/O Stack



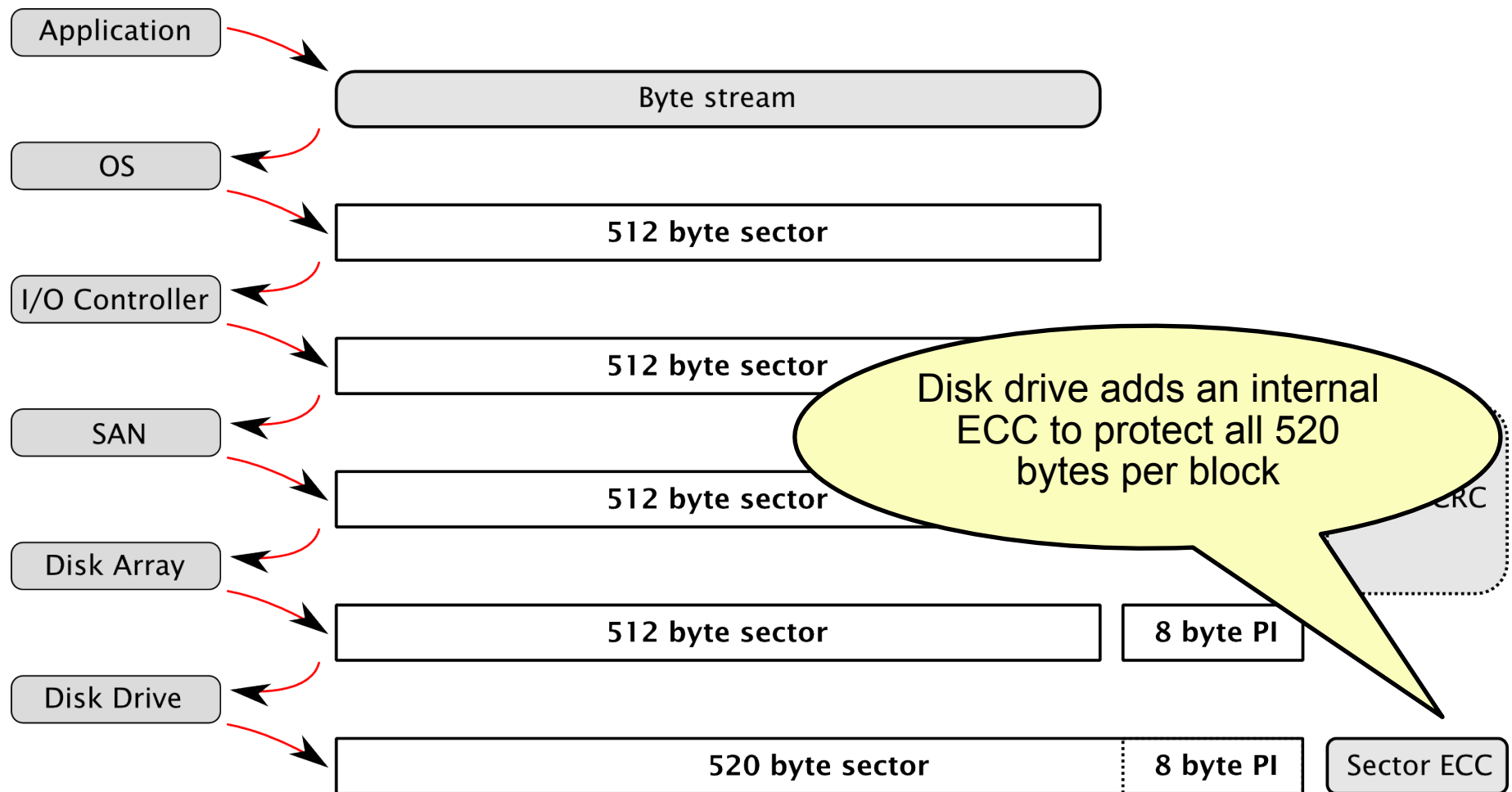
# Typical I/O Stack



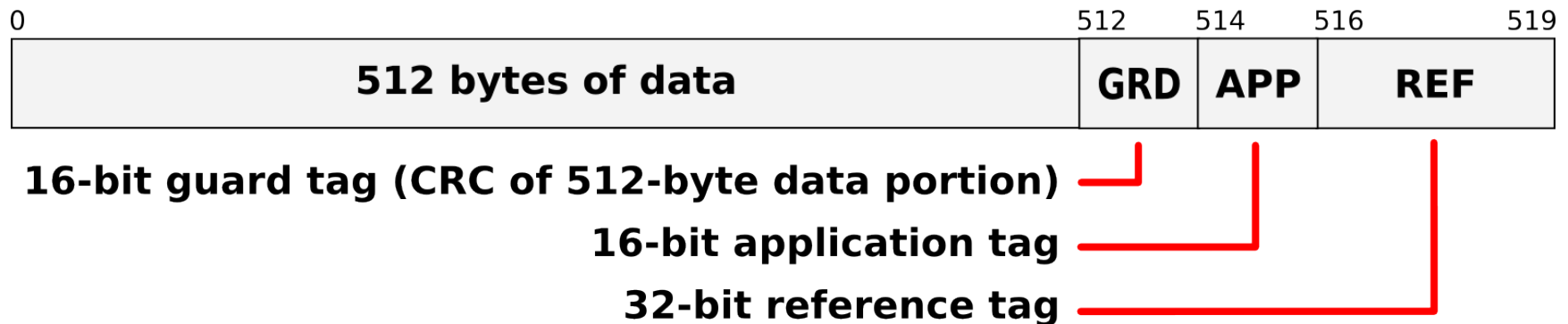
# Typical I/O Stack



# Typical I/O Stack

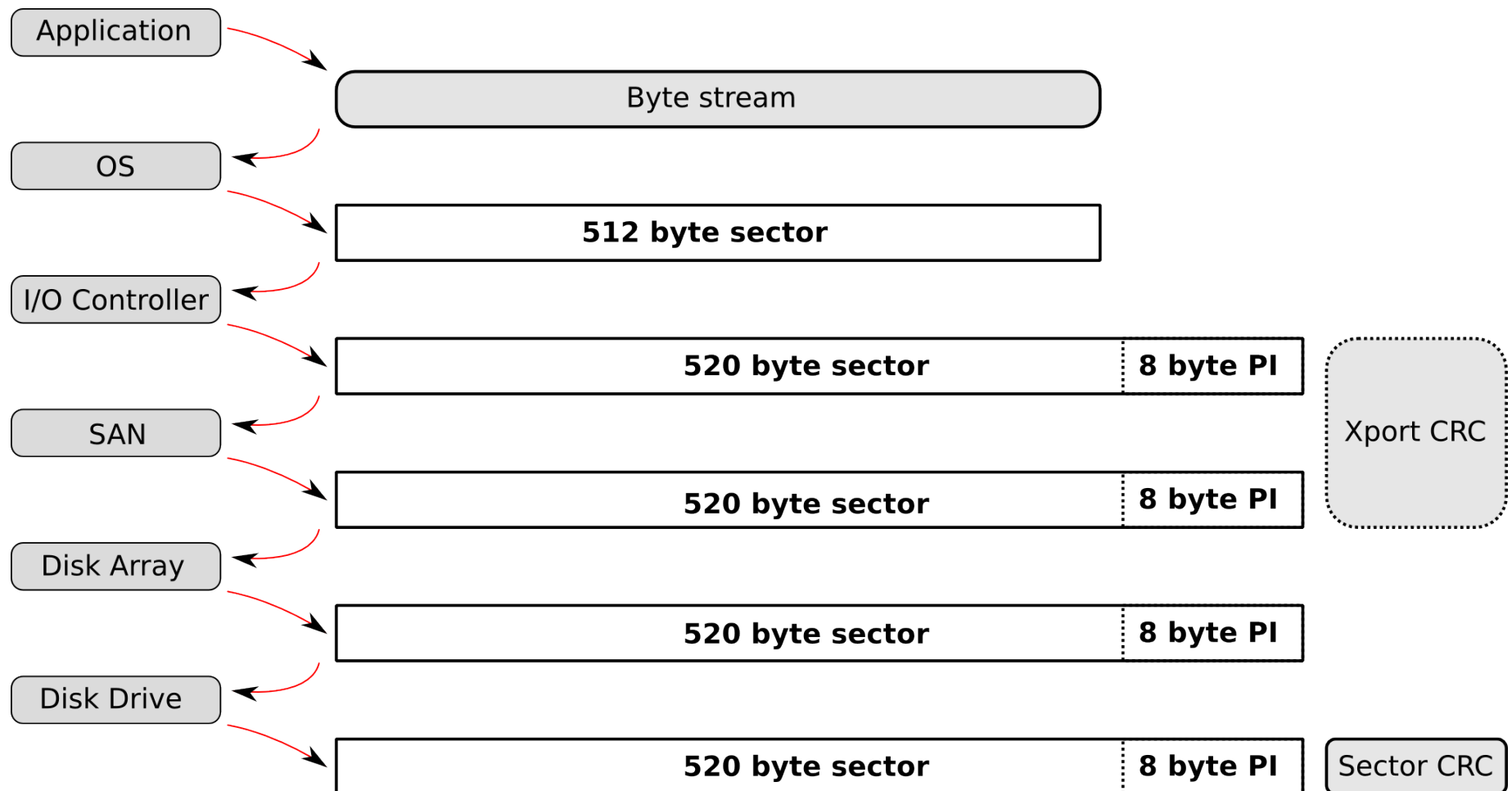


## T10 Protection Information Model

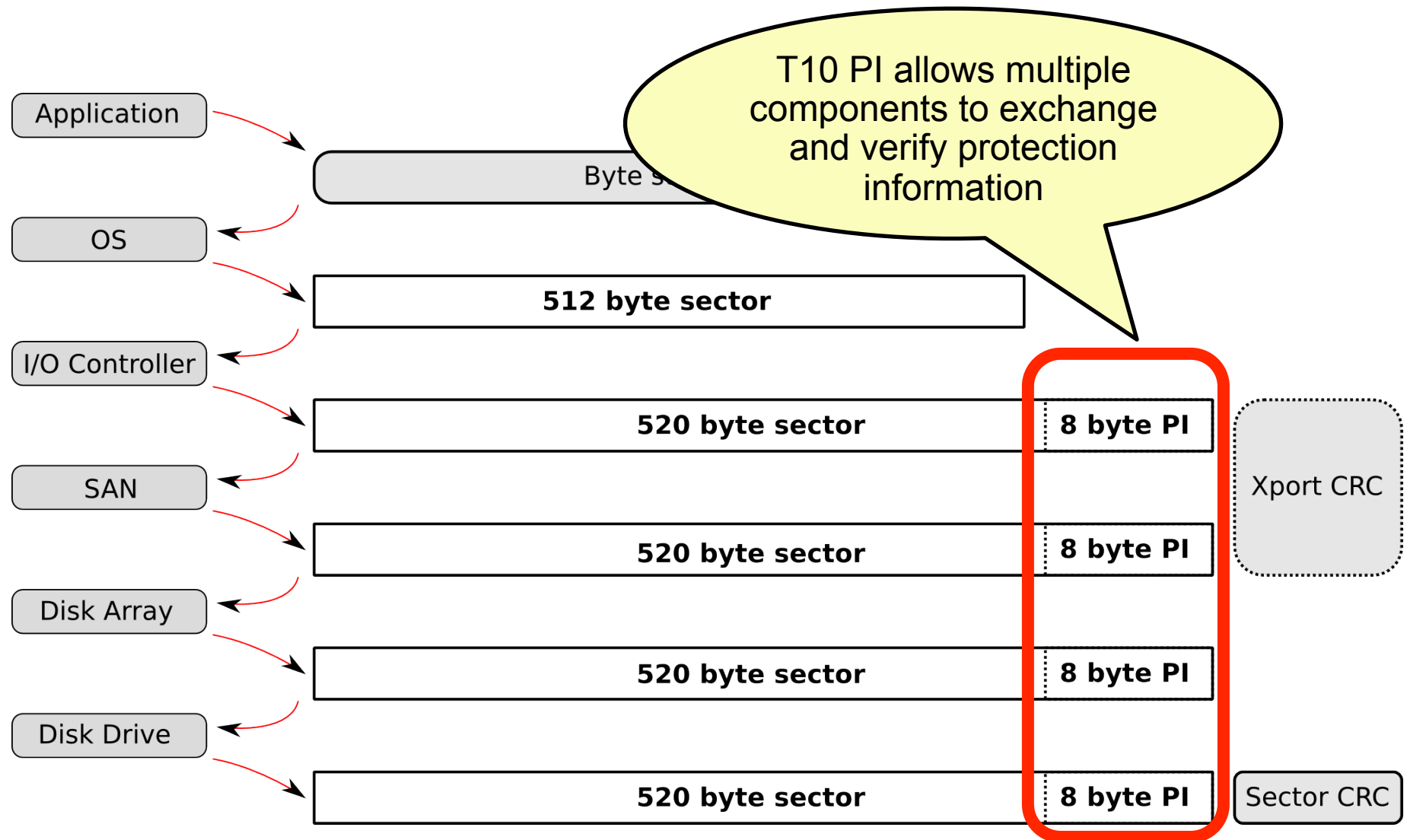


- Standardizes those extra 8 bytes
- Prevents content corruption and misplacement errors
- Protects path between HBA and storage device
- Protection information is interleaved with data on the wire, i.e. effectively 520-byte logical blocks

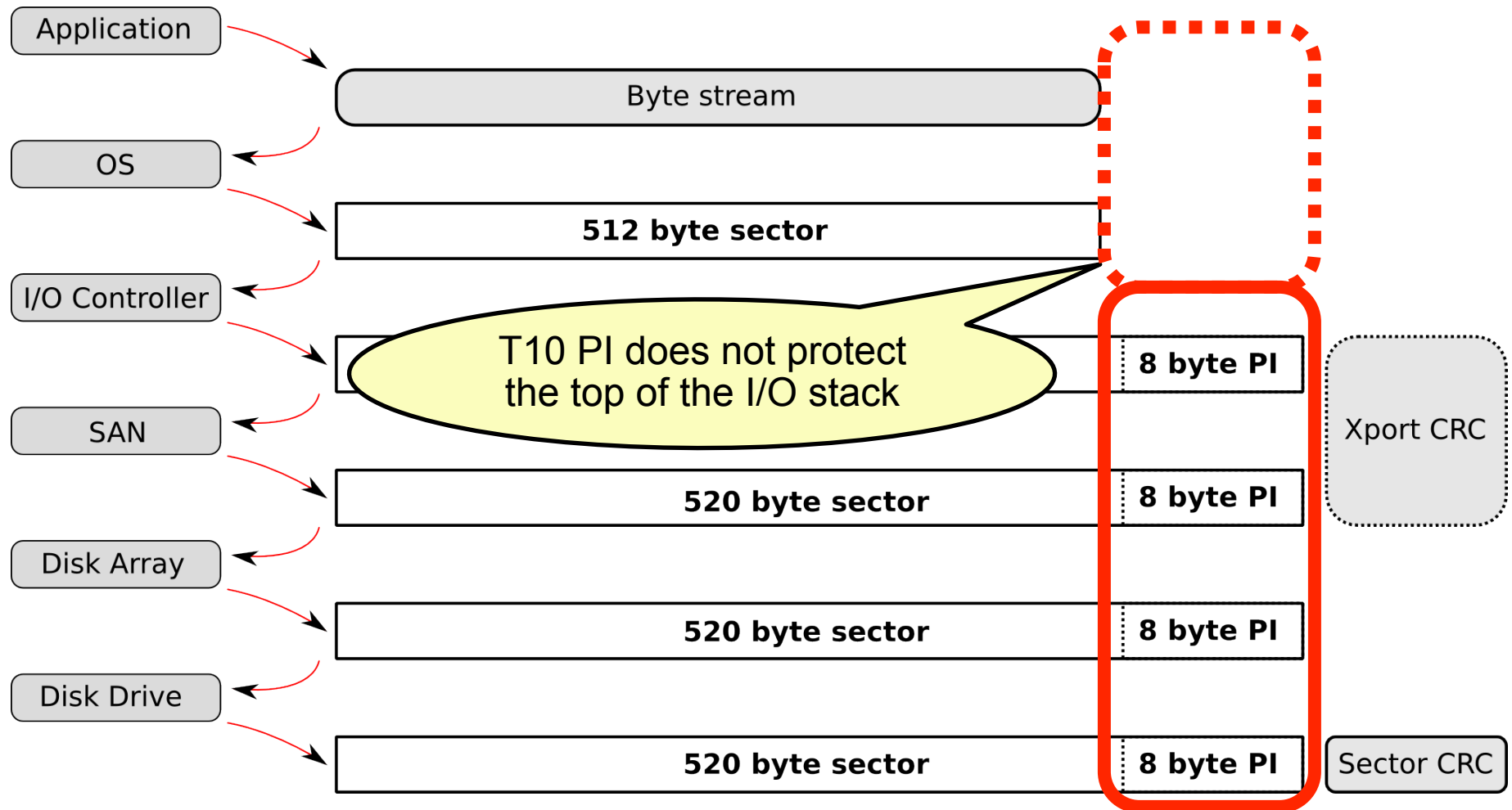
# T10 Protection Information Model



# T10 Protection Information Model



# T10 Protection Information Model

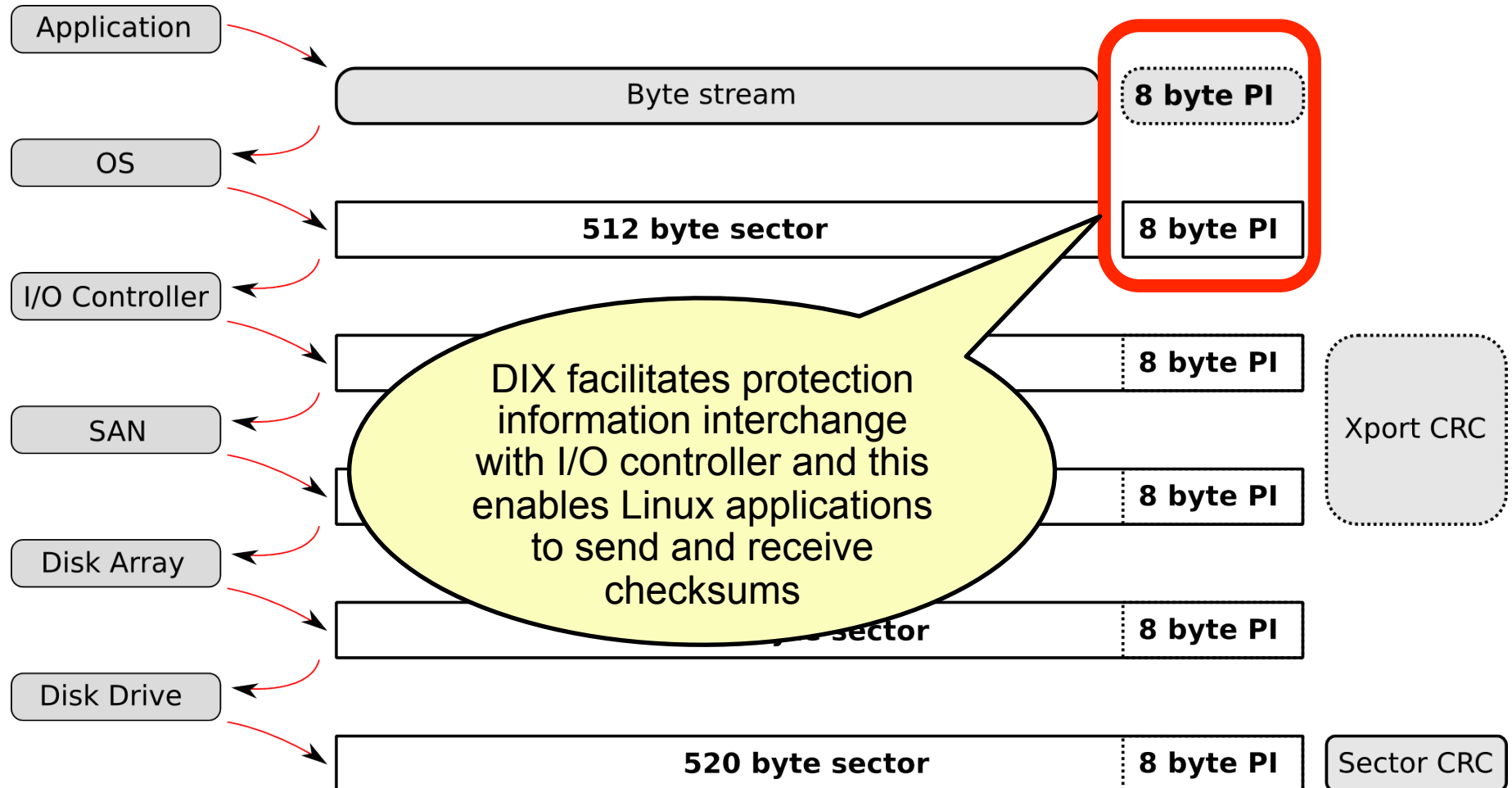




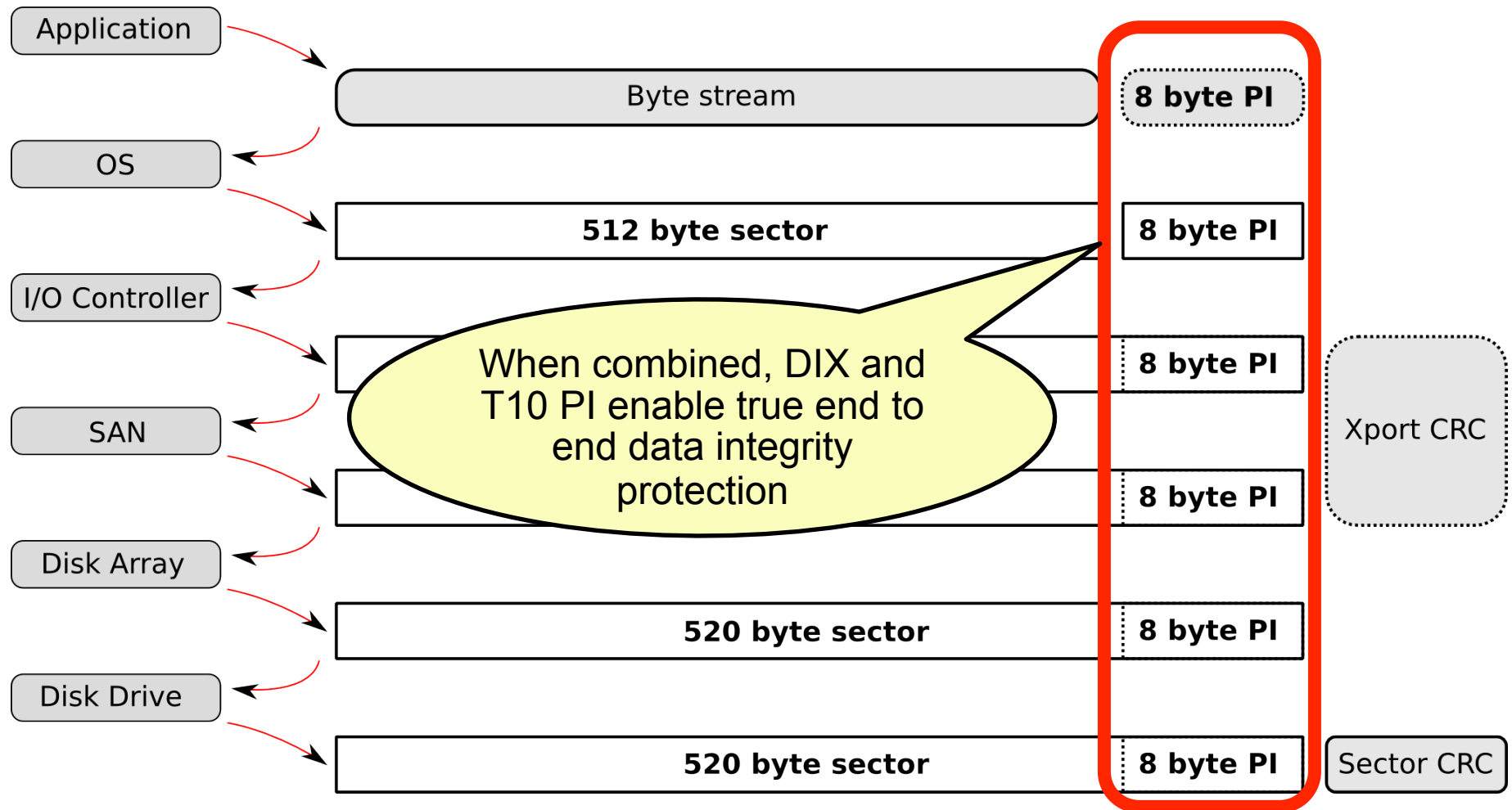
## Oracle Data Integrity Extensions

- Extends T10 PI all the way up to the application, enabling true end-to-end data integrity protection
- The Data Integrity Extensions (DIX)
  - Enable DMA transfer of protection information to and from host memory
  - Separate data and protection information buffers to avoid inefficient 512+8+512+8+512+8 scatter-gather lists
  - Provide a set of commands that tell HBA how to handle the I/O: *Generate, Strip, Forward, Verify, etc.*

# Data Integrity Extensions + T10 PI



# Data Integrity Extensions + T10 PI



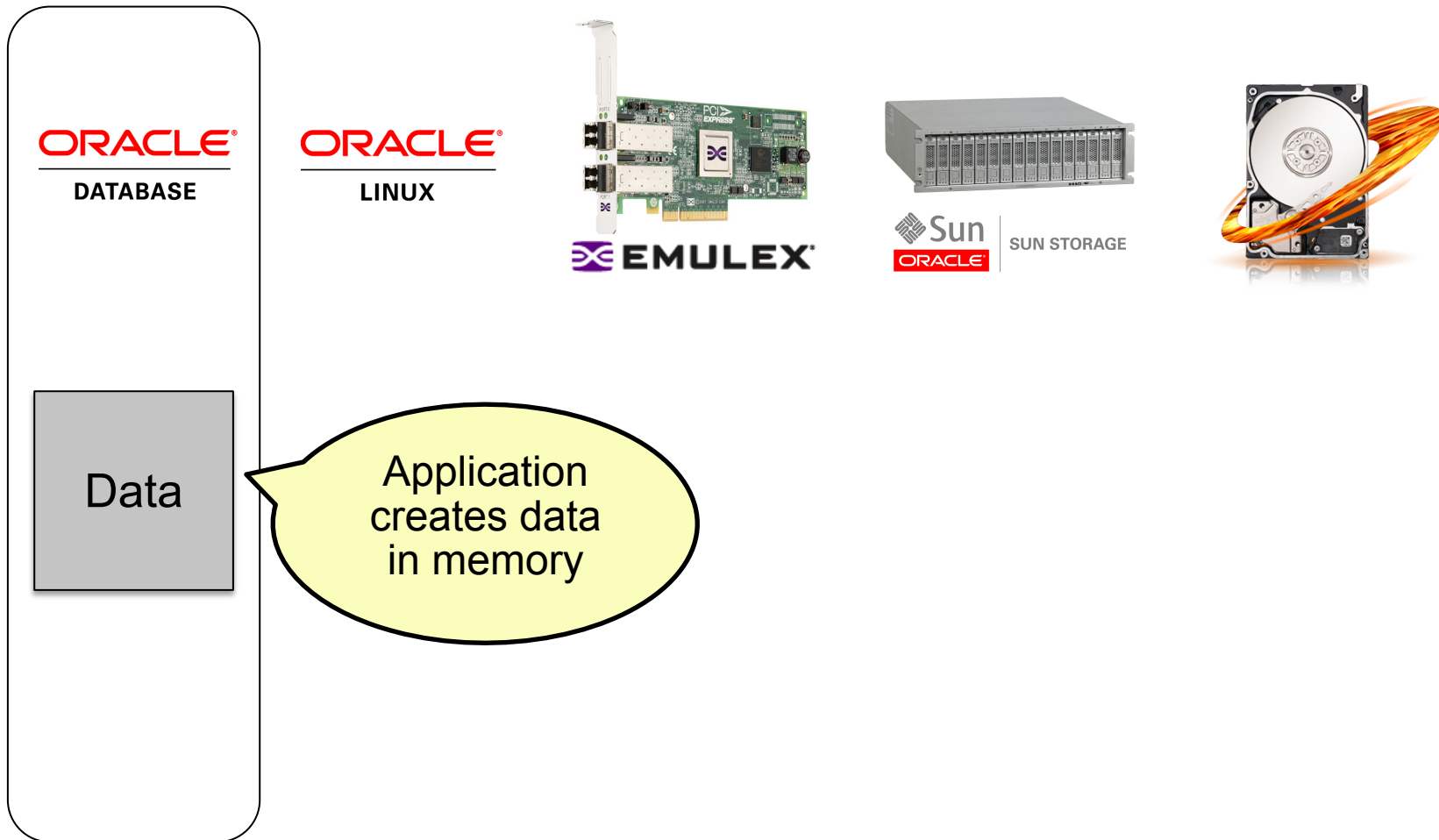
# Preventing Data Corruption with DIX and T10 DIF

**ORACLE®**  
DATABASE

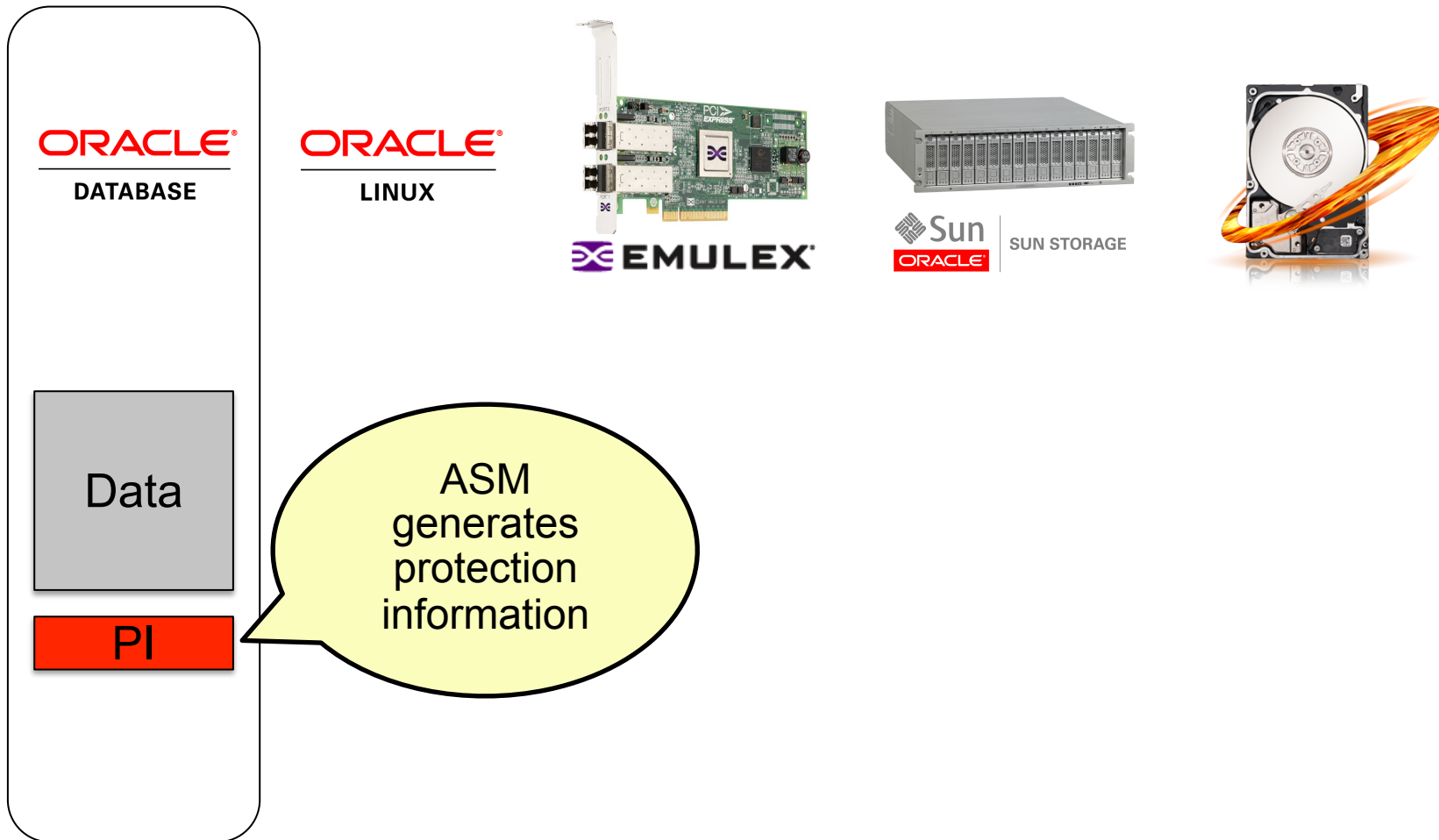
**ORACLE®**  
LINUX



# Preventing Data Corruption with DIX and T10 DIF



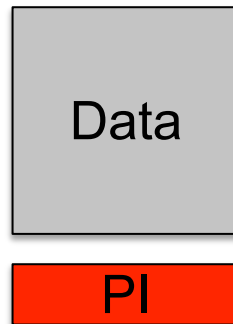
# Preventing Data Corruption with DIX and T10 DIF



# Preventing Data Corruption with DIX and T10 DIF

**ORACLE®**  
DATABASE

**ORACLE®**  
LINUX

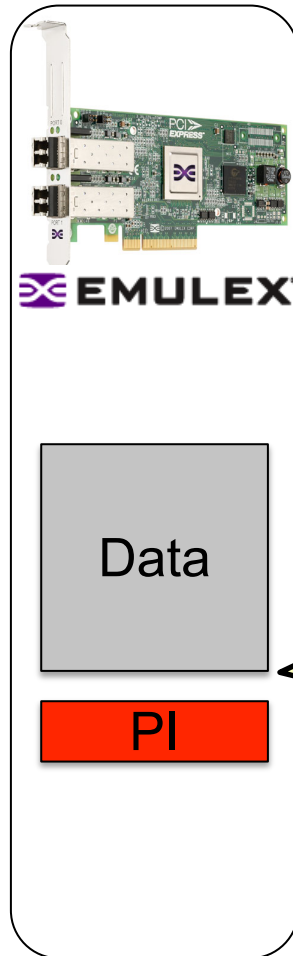


Linux kernel forwards  
data and protection  
information to Emulex  
HBA using DIX

# Preventing Data Corruption with DIX and T10 DIF

**ORACLE®**  
DATABASE

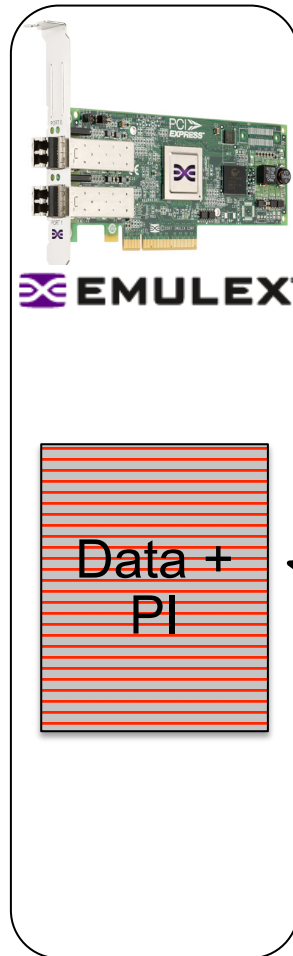
**ORACLE®**  
LINUX



# Preventing Data Corruption with DIX and T10 DIF

**ORACLE®**  
DATABASE

**ORACLE®**  
LINUX



Emulex HBA interleaves data and protection information and transmits 520-byte sectors to storage

# Preventing Data Corruption with DIX and T10 DIF

**ORACLE®**  
DATABASE

**ORACLE®**  
LINUX



Storage array controller  
verifies that data,  
protection information,  
and target location  
match

Data +  
PI

# Preventing Data Corruption with DIX and T10 DIF

**ORACLE®**  
DATABASE

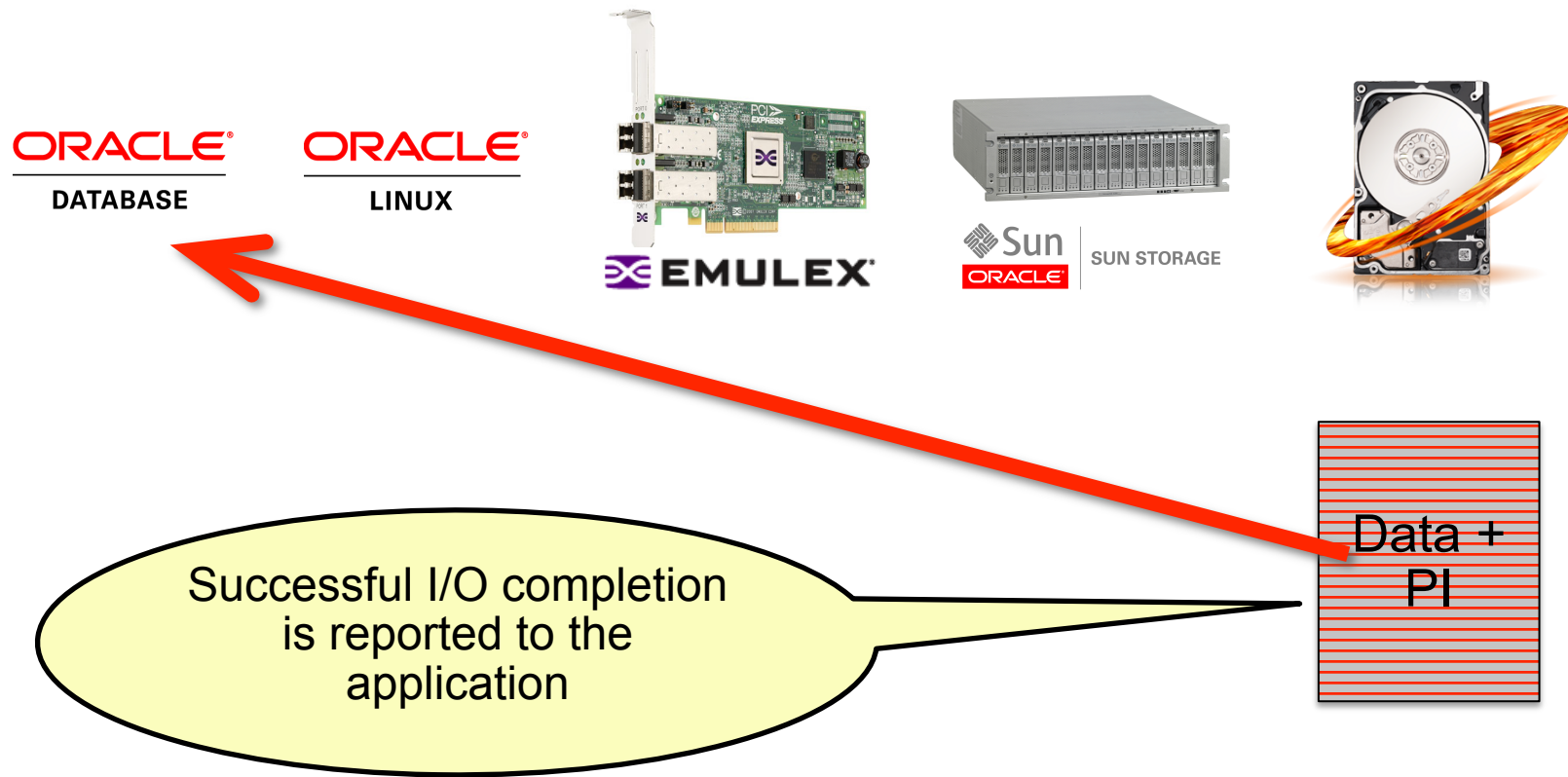
**ORACLE®**  
LINUX



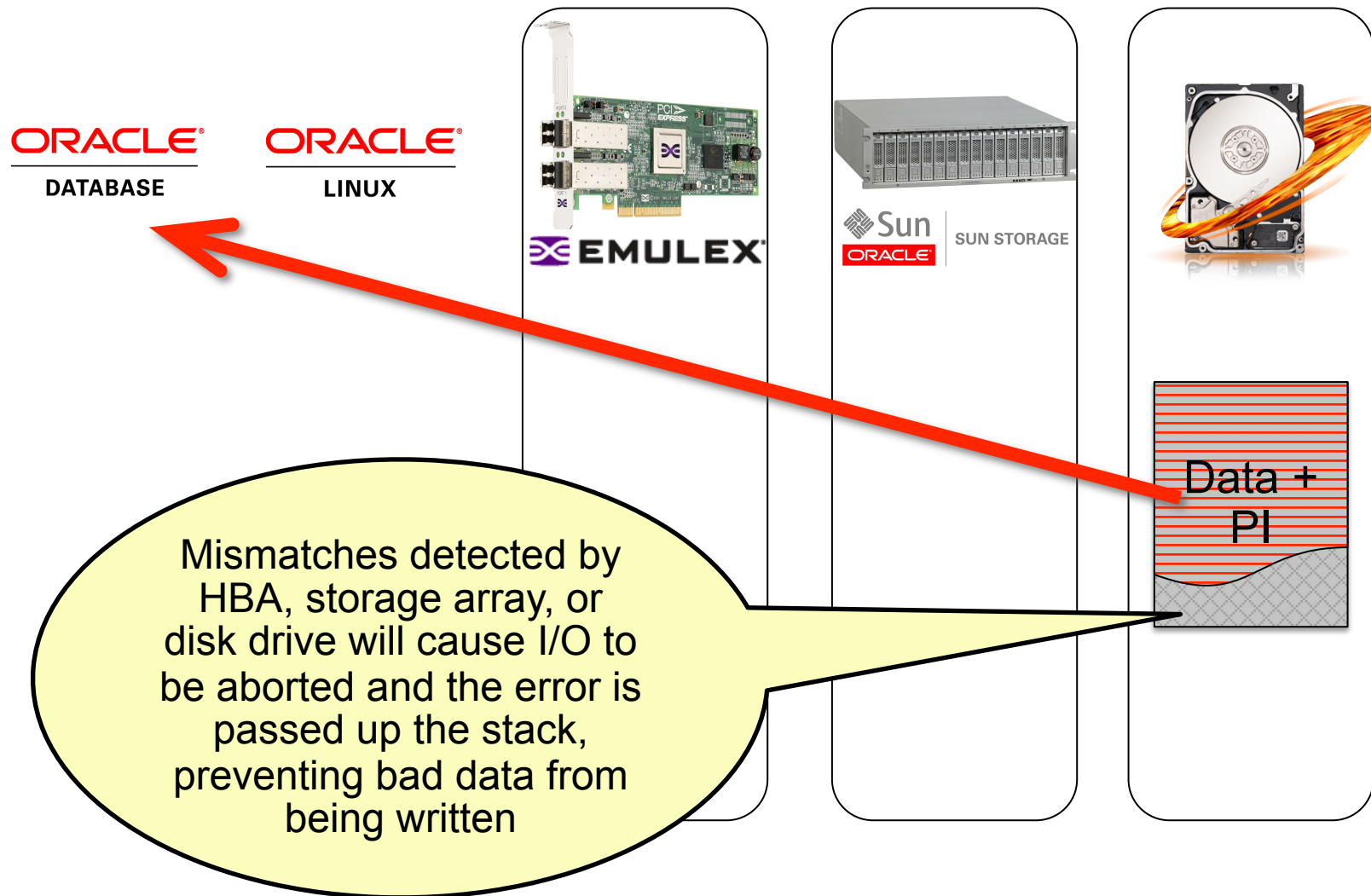
Disk drive firmware  
verifies that data,  
protection information,  
and target location  
match

Data +  
PI

# Preventing Data Corruption with DIX and T10 DIF



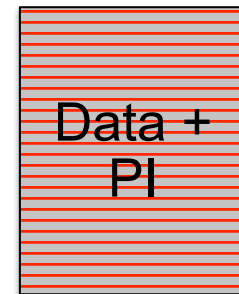
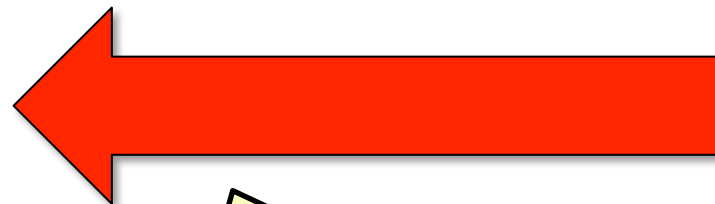
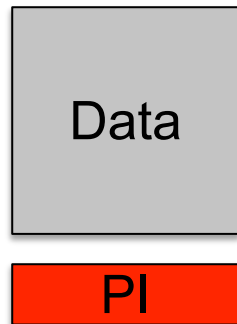
# Preventing Data Corruption with DIX and T10 DIF



# Preventing Data Corruption with DIX and T10 DIF

ORACLE<sup>®</sup>  
DATABASE

ORACLE<sup>®</sup>  
LINUX



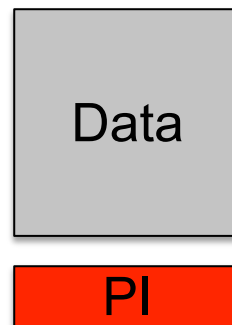
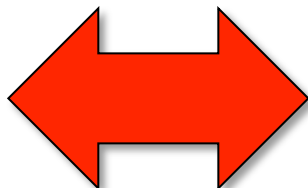
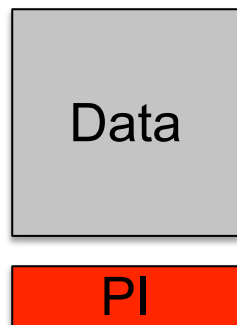
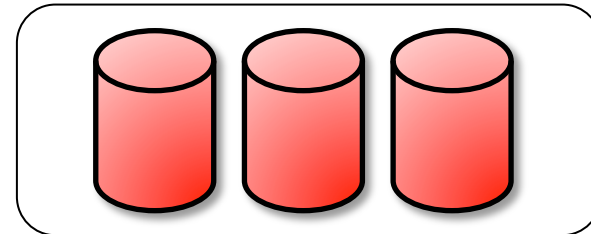
Protection information is also transmitted for read requests and ASM will verify I/O before signaling completion to the application

ORACLE<sup>®</sup>

# Preventing Data Corruption with DIX and T10 DIF

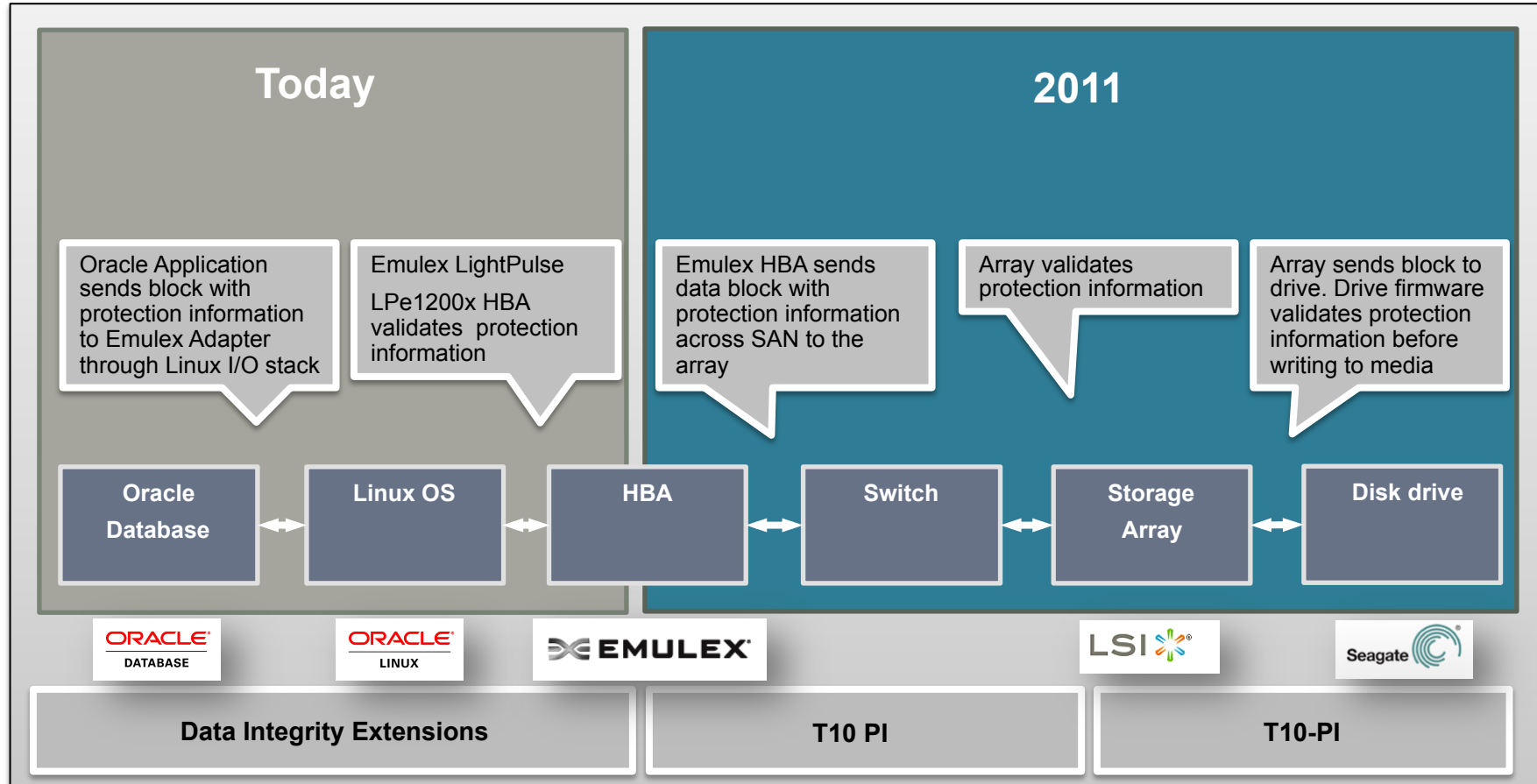
ORACLE<sup>®</sup>  
DATABASE

ORACLE<sup>®</sup>  
LINUX



Protection information exchange is dynamically negotiated. Protection is automatically enabled between application and HBA when using legacy storage

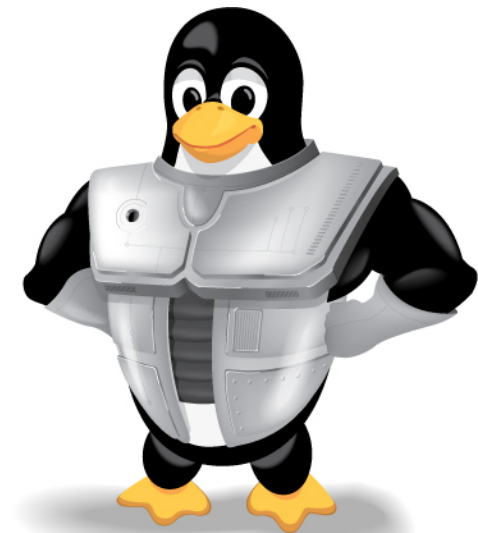
# What's Available Today



Application-to-Disk Data Integrity: current and future support

## Resources

- **Oracle Linux Information**  
[oracle.com/linux](https://oracle.com/linux)
- **Follow us on Twitter**  
[@ORCL\\_Linux](https://twitter.com/ORCL_Linux)
- **Free Download: Oracle Linux**  
[edelivery.oracle.com/linux](https://edelivery.oracle.com/linux)
- **Free Download:**
  - **Unbreakable Enterprise Kernel**  
[public-yum.oracle.com](https://public-yum.oracle.com)
- **Data Integrity Project Information**  
[oss.oracle.com/~mkp/](https://oss.oracle.com/~mkp/)



Unbreakable  
**Linux**  
ORACLE