STORAGE DEVELOPER CONFERENCE
SDC 21
BY Developers FOR Developers
Virtual Conference
September 28-29, 2021
A SNIA Event

A Quintuple Parity Error Correcting Code - a Game Changer in Data Protection

Presented by Marek Rychlik, Ph.D., CEO of Xoralgo, Inc.

Introductions

The speaker, Xoralgo Inc. and the patent

About Xoralgo, Inc.

- Xoralgo, Inc. is a University of Arizona start-up, 2018
- Technology is based on the US utility patent 10,997,024 awarded in May 2021
- Assignee: The Regents of The University of Arizona
- Priority date: January 24, 2017
- In addition to being the presenter, I am:
  - A professor at the U of A Mathematics Department
  - The co-inventor on the patent along with my student
  - The CEO of Xoralgo, Inc.

PentaRAID™and its Error Correcting Code

About Xoralgo’s RAID implementation

Undetected Disk Errors (UDE)

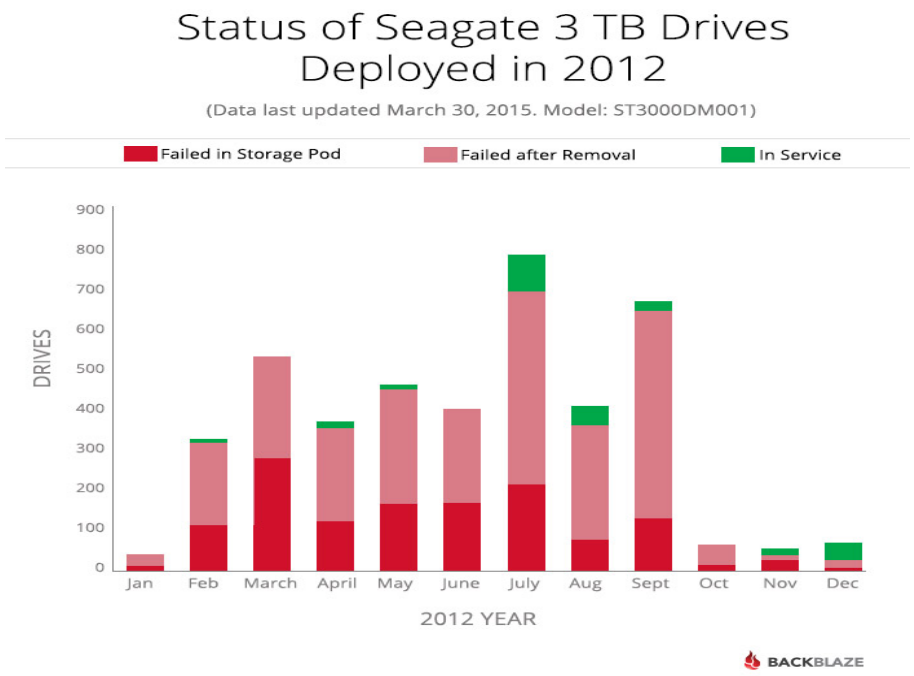
- Errors undetected by the disk controller of a hard drive
- Occur during normal operation due to the laws of physics
- Frequency is 1 error per 10^{14} or 10^{15} reads
- For data rate of 1GB/sec a UDE occurs in 10^5 seconds = 2 days
- RAID 6 with 1 failed disk operating in degraded mode takes weeks to recover with big drives of today, during which period every UDE becomes data loss
- In conclusion, data loss due to UDE is a common occurrence

Errors (= UDE) and Erasures

- We read previously stored data N blocks of data at a time
- An **error** occurs when one of the blocks is incorrect, but we do not know which one
- An **erasure** occurs when one of the blocks is incorrect, and we know which one
- An erasure is called that because we just may as well assume that the corresponding data is zeroed out, or erased
- UDE are synonymous with errors

How common are disk errors?

Nearly all disks die in 3 years – in 2012



- Seagate advertises .35%/year failure rate w. 5 year warranty
- 1%/year failure rate is independently reported
- Institutions (e.g. university) retire disks in a few months

Time to the next Undetected Disk Error

Laptop/Desktop

30 hours

Server

1-3 hours

Self-driving car

1-10 minutes

UDE - source of **Silent Data Corruption!**

How many errors/erasures can PentaRAID™ correct?

- Z – number of erasures
- E – number of errors

$$Z + 2E \leq 4$$

- Up to 4 erasures (failed disks)
- Up to 2 errors (UDE)
- To be able to correct 2 errors, we must be able to correct 4 failed disks
- A similar law applies to all storage systems (4 is specific to us)

How many errors/erasures can RAID 6 correct?

- Z – number of erasures
- E – number of errors

$$Z + 2E \leq 2$$

- Up to 2 erasures (failed disks)
- Only 1 error (UDE)
- This is why RAID 6 **will lose data** operating in degraded mode

A description of PentaRAID™ for an Impatient Expert

- Based on a linear, systematic, forward error correcting code with $N \leq q - 2$ data words and fixed $K = 5$ parity words if Galois field $GF(q)$ is used
- Not a Maximum Distance Separable (MDS) code; $D = 5$
- For example, if GF(256) is used, 254 data disks are supported

Advantages of PentaRAID™

- Offers greatly superior data protection as compared to RAID 6, without increasing computational complexity
- Offers an extremely efficient syndrome decoding algorithm as compared to, e.g., Reed-Solomon coding
- The decoder does not use Chien Search
- Chien Search is a trial-and-error method of solving polynomial equations responsible for high computational complexity of most error correcting schemes

Mean Time To Data Loss (MTTDL)

- Mean Time To Data Loss (MTTDL) is a standard measure of the reliability of a storage system
- Realistic assumptions on the number of disks, UDE rate, etc. yields

one hundred quadrillion years

- Comparable to number of atoms in a gallon of milk…
- …or the number of stars in visible universe

PentaRAID™ Implementation

Xoralgo's first storage appliance

A reference software implementation

- RAID implementation in user space as a C library under Linux
- Exposed to the Linux OS using NBD protocol and NBD Kit (Red Hat)
  - Storage exposed to the OS as a block device
  - Can be used raw, partitioned (MBR/GPT), or used as a partition
  - Storage can be exposed as network storage using NBD
- PentaRAID™, along with RAID 0, 1 and 6 is available
- The user can format and partition the storage as if it were a single disk
- Original implementation – 2017, used NBD Kit version 1
- New implementation – 2021, uses NBD Kit version 2

A Xoralgo storage appliance and test bed

- Industry standard server
- RAID controller in JBOD mode
- Software PentaRAID™
- Performance of RAID 6
- Vastly superior data protection
- Commercially available in 2022

Testing PentaRAID™ Implementation

- Cannot wait 100 quadrillion years for an error… must speed things up!
- Simulation 1: Injection of multiple random errors into physical or virtual disks
- Simulation 2: Disk removal test
- Simulation 3: Running operating systems on PentaRAID™ storage

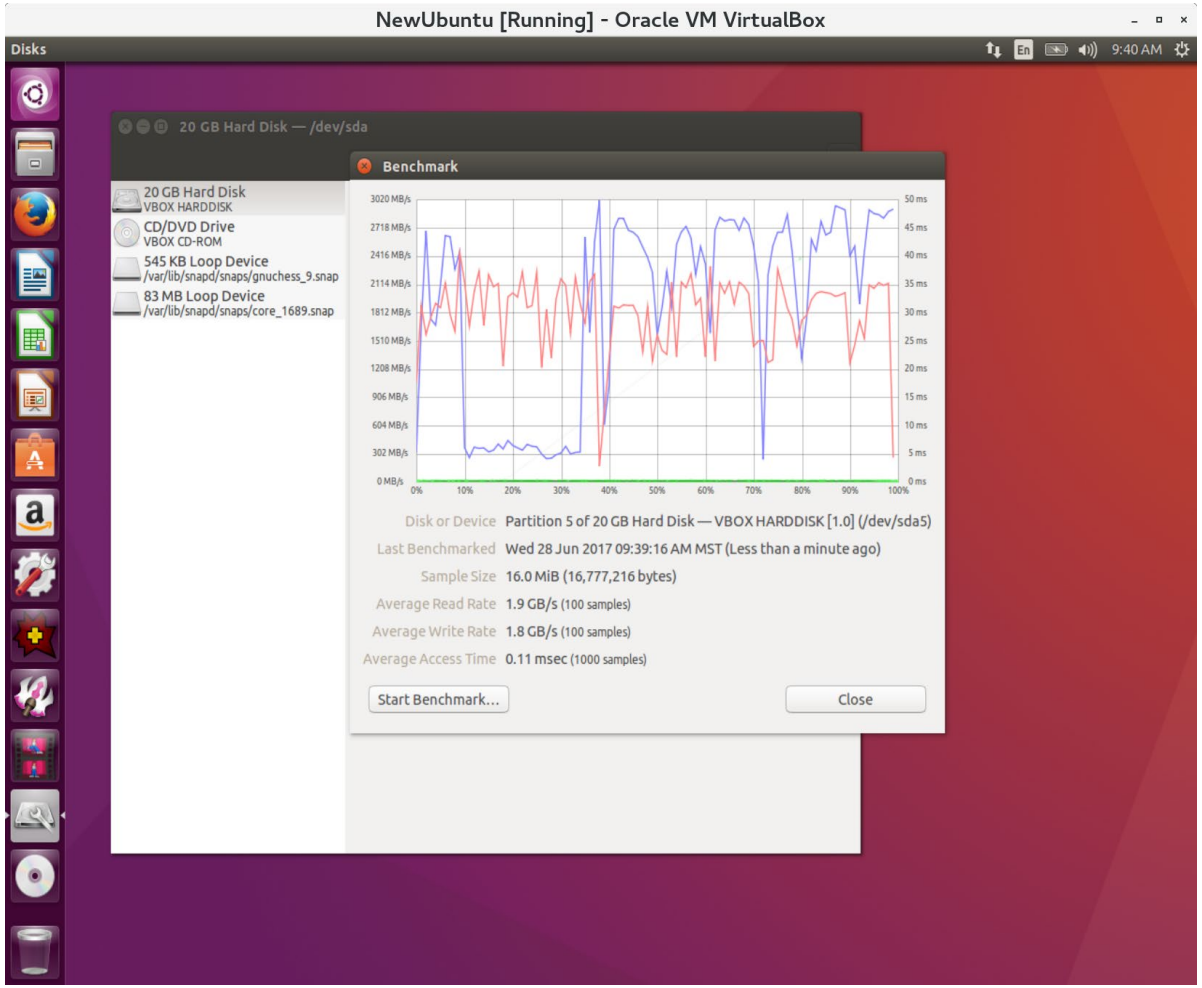
Disk Removal Testing (15 disks) – real results!

STATUS	REASON	REMOVED DEVICES	TOTAL
[OK]		2	1
[OK]		2 3	2
[OK]		2 7	2
[OK]		2 7 8	3
[OK]		2 6 7 8	4
[OK]		2 7 10	3
[OK]		2 7 11	3
[OK]		2 7 12	3
[OK]		2 7 13	3
[OK]		2 7 14	3
[OK]		2 10 11	3
[OK]		2 10 12	3
[OK]		2 10 13	3
[OK]		2 10 14	3
[OK]		2 11 12	3
[OK]		2 11 13	3
[OK]		2 11 14	3
[OK]		2 12 13	3
[OK]		2 12 14	3
[OK]		2 13 14	3
[OK]		10 11 12	3
[OK]		10 11 13	3
[OK]		10 11 14	3
[OK]		10 12 14	3
[OK]		11 12 13	3

STATUS	REASON	REMOVED DEVICES	TOTAL
[OK]		11 12 14	3
[OK]		11 13 14	3
[OK]		12 13 14	3
[OK]		2 6 7 10	4
[OK]		2 6 7 11	4
[OK]		2 6 7 12	4
[OK]		2 6 7 13	4
[OK]		2 6 7 14	4
[OK]		2 7 10 11	4
[OK]		2 7 10 12	4
[OK]		2 7 10 13	4
[OK]		2 7 10 14	4
[OK]		2 7 11 12	4
[OK]		2 7 11 13	4
[OK]		2 7 11 14	4
[OK]		2 7 12 13	4
[OK]		2 7 12 14	4
[OK]		2 7 13 14	4
[OK]		2 10 11 12	4
[OK]		2 10 11 13	4
[OK]		2 10 11 14	4
[OK]		2 10 12 13	4
[OK]		2 10 12 14	4
[OK]		2 10 13 14	4
[OK]		2 11 12 13	4
[OK]		2 11 12 14	4
[OK]		2 11 13 14	4
[OK]		2 12 13 14	4
[OK]		11 12 13 14	4
[OK]		10 12 13 14	4
[OK]		10 11 13 14	4
[OK]		10 11 12 14	4
[OK]		10 11 12 13	4

Running Ubuntu Linux on PentaRAID™

- Testing swap partition performance
- Playing video games
- Testing app performance



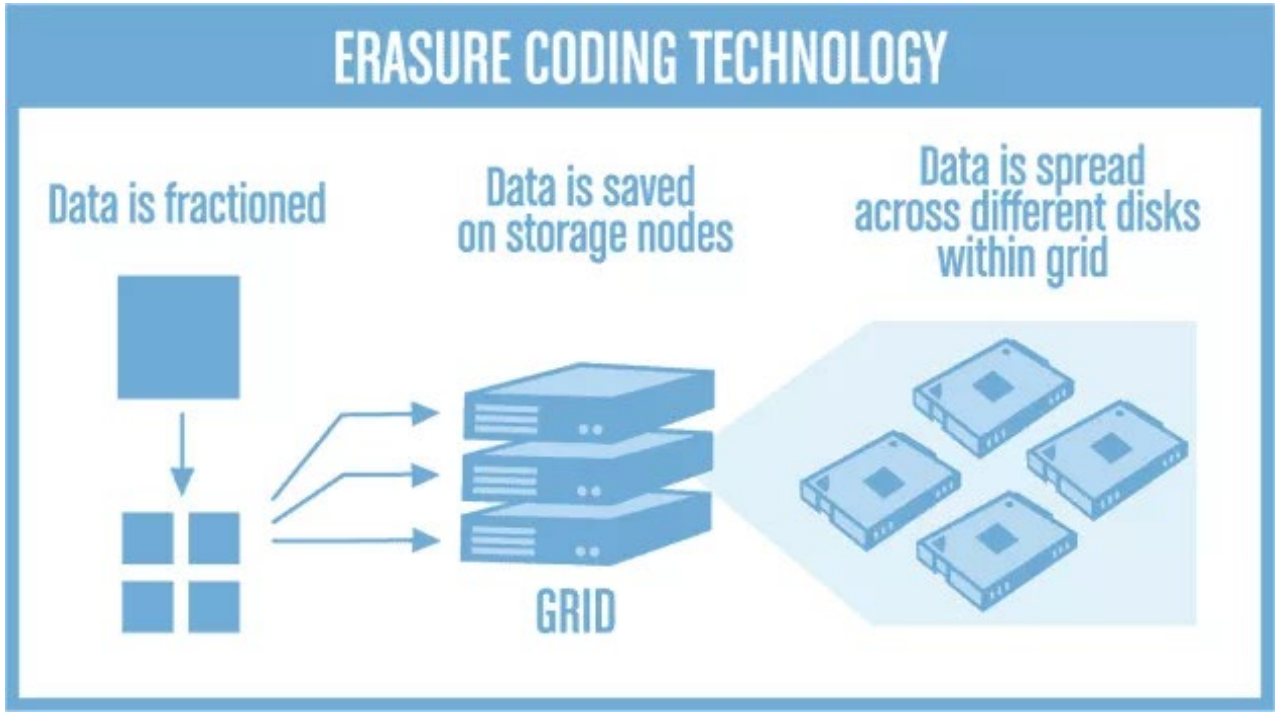
Future high-performance implementation

- Kernel module implementation for Linux
- Windows 10/11 driver
- FPGA based PentaRAID™ controller
- ASIC
- Licensing to hardware/software vendors

Partnerships welcome!

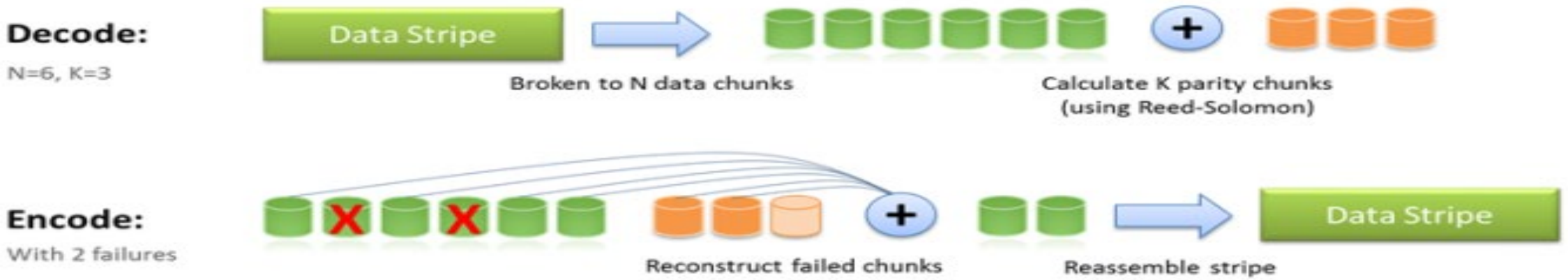
PentaRAID™ fits in “Erasure Coding” workflow

- Storage nodes use ECC to reduce replication
- PentaRAID™ can be that ECC and excel at this task



Graphic source: stonefly.com/blog

PentaRAID™ as a drop-in replacement



Graphic source: stonefly.com/blog

PentaRAID™ as a drop-in replacement

- K=5 and N=10 results in the same redundancy overhead
- Data protection is vastly increased

PentaRAID™ as a “Game Changer”

What can it do for the storage industry?

Case for PentaRAID™ at Large Data Centers

- There is no longer a case for storing data in triplicate vs. using ECC
- High computational complexity argument against ECC defeated by low complexity decoding algorithm of PentaRAID™
- Reducing number of spinning disks by 30% and increased data protection is an easy target
- Retiring old disk later, e.g. doubling time-in-service, will result in further savings

Case for PentaRAID™ at Small Businesses

- Data protection of a large data center in a small storage appliance
- Any IT manager who is able to manage RAID 6 will be able to manage PentaRAID™ (as simple as: “if disk enclosure blinking, replace disk”)
- Eliminates latency accessing one’s business data associated with cloud (example: a small, independent video producer)
- Significant cost-of-storage reduction

Contact information

- E-mail: Rychlik@Arizona.edu
- Xoralgo's Website: xoralgo.com

References

- PentaRAID™ White Paper: https://arxiv.org/abs/1806.08266
- Peter Anvin's RAID 6 algorithm exposition – followed by new RAID 6 implementations
- Sarah Mann's dissertation (directed by me) – a 2013 exposition of Reed-Solomon coding with computational complexity analysis

Please take a moment to rate this session.

Your feedback is important to us.

