Highly Reliable Two-Dimensional RAID Arrays for Archival Storage

Jehan-François Pâris
Computer Science Dept.
University of Houston
Houston, TX 77204-3010
jparis@uh.edu

Thomas J. E. Schwarz, S. J.
Depto. de Informática y
Ciencias de la Computación
U. Católica del Uruguay
11600 Montevideo, Uruguay
tschwarz@ucu.edu.uy

Ahmed Amer
Computer Engineering Dept.
Santa Clara University
Santa Clara, CA 95050
aamer@scu.edu

Darrell D. E. Long
Computer Science Dept.
University of California
Santa Cruz, CA 95064
darrell@cs.ucsc.edu

***Abstract*—We present a two-dimensional RAID architecture that is specifically tailored to the needs of archival storage systems. Our proposal starts with a fairly conventional two-dimensional RAID architecture where each disk belongs to exactly one horizontal and one vertical RAID level 4 stripe. Once the array has been populated, we add a *superparity* device that contains the exclusive OR of all the contents of all horizontal—or vertical—parity disks. The new organization tolerates all triple disk failures and nearly all quadruple and quintuple disk failures. As a result, it provides mean times to data loss (MTTDLs) more than a hundred times better than those of sets of RAID level 6 stripes with equal capacity and similar parity overhead.[1]**

Keywords-disk arrays, RAID arrays, fault-tolerance, storage system reliability.

I. INTRODUCTION

Archival storage systems constitute a new class of storage systems that were specifically designed to preserve fairly stable data over long periods of time. As a result, these systems differ from conventional storage systems in several important points.

First, the data they contain have to remain available over time periods that can span decades. Hence archival storage systems must be exceptionally reliable. Second, these data will remain largely unmodified once they are stored. For this reason, write rates are a much less important issue than in conventional storage systems. Finally, archival data are not likely to be accessed as frequently as active data. We can thus expect that many of the disks of an archival system will be powered down most of the time.

We propose here a low-redundancy storage solution that takes advantage of these access patterns to achieve much higher mean times to data loss than RAID level 6 architectures. While the archival store is populated, it operates as a conventional two-dimensional RAID array such as the one described on Fig. 1: each disk belongs to exactly one horizontal and one vertical RAID level 4 stripe. As we will show later, this organization protects data against all double disk failures and most triple and quadruple disk failures.

Once the array is filled with data, we add a *superparity* disk [19]. This disk will contain the exclusive OR (XOR) of all the contents of all horizontal—or vertical—parity disks. As we will see, the new organization tolerates all triple disk failures and most quadruple and quintuple disk failures. Should a significant fraction of the stored data need to be updated, we will disconnect the superparity disk and revert to the original disk organization until the update has completed.

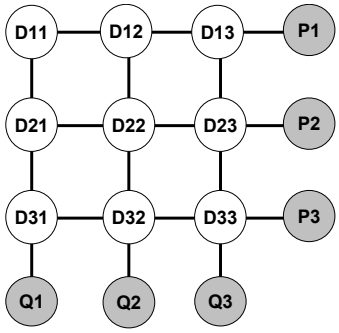

Fig. 1. A two-dimensional RAID array with 9 data and 6 parity disks.

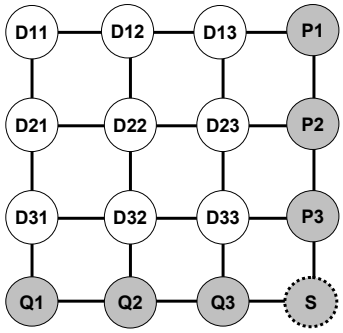

Fig. 2. The same array with a superparity disk (S).

The main advantage of our new organization is its excellent reliability: two-dimensional RAID arrays with superparity provide mean times to data loss (MTTDLs) more than a hundred times better than those of sets of RAID level 6 stripes with equal capacity and similar parity overhead. Its main limitation is its poor write throughput, which restricts its application to the storage of archival and other immutable data.

The remainder of this paper is organized as follows. Section II reviews previous work. Section III introduces our architecture and discusses its vulnerability to quadruple and quintuple failures. Section IV evaluates its reliability and compares it to that of RAID level 6 arrays

[1]Supported in part by Grant CCF-1219163, by the Department of Energy under Award Number DE-FC02-10ER26017/DE-SC0005417 and by the industrial members of the Storage Systems Research Center.

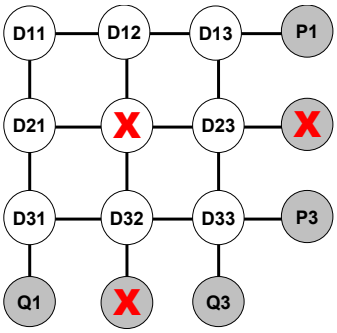

Fig. 3. A triple failure resulting in a data loss.

with the same space overhead. Section V discusses possible implementations and Section VI has our conclusions.

II. Previous Work

RAID arrays were the first disk array organizations to utilize erasure coding in order to protect data against disk failures [3, 12, 18]. While RAID levels 3, 4 and 5 only tolerate single disk failures, RAID level 6 organizations use $(n-2)$-out-of-n codes to protect data against double disk failures [2]. EvenOdd, Row-Diagonal Parity and the Liberation Codes are three implementations of RAID level 6 that use only XOR operations to construct their parity information [1, 3, 5, 14, 15]. Huang and Xu proposed a coding scheme correcting triple failures [7].

Two-dimensional RAID arrays, or 2D-Parity arrays, were investigated by Schwarz [17] in his dissertation. Hellerstein et al. [6] noted that these arrays tolerated all double disk failures but did not investigate how they reacted to triple or quadruple disk failures. More recently, Lee patented a two-dimensional disk array organization with prompt parity updates in one dimension and delayed parity updates in the second dimension [8]. Pâris et al. [10] discussed two-dimensional RAID arrays that reorganized themselves after a disk failure and noted that two-dimensional RAID arrays tolerated most triple failures.

Superparity devices were introduced by Wildani et al. [19] in order to increase the reliability of archival storage systems for very little cost. Their proposal partitions each disk into fixed-size "disklets," which are used to form conventional RAID stripes. These stripes are then grouped into larger units, called "supergroups," and one or more "superparity" devices are then added to each supergroup. Because superparity disklets experience a high update load, they are implemented using a faster, but more expensive technology such as flash drives. When failures occur, the system first tries to recover without using the superparity. If this is impossible, then the system shuts down normal access and recovers using superparity. The reason for the shut-down is the high read load that is necessary since supergroups are quite large and the system is not expected to support both workloads. Calculations show the effectiveness of introducing the supergroup, but also that the need to have recourse to it rarely arises.

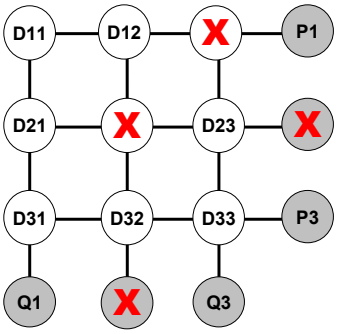

Fig. 4. A Type 1 quadruple failure resulting in a data loss.

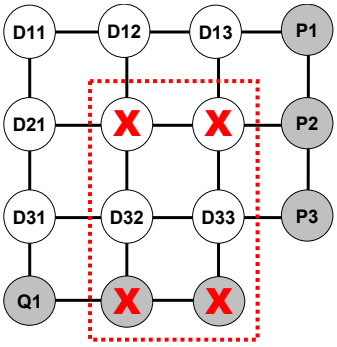

Fig. 5. A Type 2 quadruple failure resulting in a data loss.

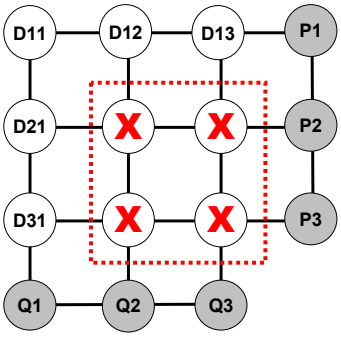

Fig. 6. A Type 3 quadruple failure resulting in a data loss.

III. Our Proposal

While two-dimensional RAID arrays protect data against all single and all double disk failures, some triple disk failures will result in a data loss. As seen in Fig. 3, these fatal triple failures consist of the failure of one data disk and its two parity disks. Consider now a two-dimensional array comprising n^2 data disks and $2n$ parity disks. Out of the $\binom{n^2+2n}{3}$ possible triple failures the array can experience, exactly n^2 will be fatal. As its size grows, the ratio between the number of fatal triple failures and the total number of triple failures

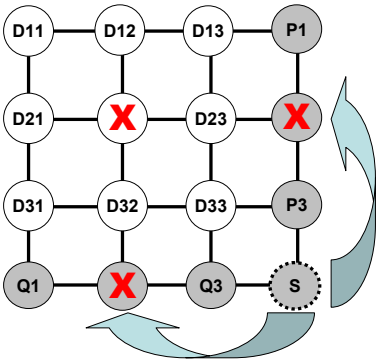

Fig. 7. Recovering from a previously fatal triple failure.

$$\frac{n^2}{\binom{n^2+2n}{3}}$$

will quickly decrease: it becomes less than 1 percent for $n \geq 4$ and less than 0.1 percent for $n \geq 8$.

Let us now consider which types of quadruple failures will be fatal. These types will include:

1. The failure of a data disk, its two parity disks and any other disk: Fig. 4 shows of these Type 1 failures.
2. The failure of two data disks in the same row or column and their parity disks in the other dimension: Fig. 5 shows one of these Type 2 failures;
3. The failure of four data disks forming a square: Fig. 6 shows one of these Type 3 failures.

Out of the $\binom{n^2+2n}{4}$ possible quadruple failures the array can experience, we can enumerate:

1. $n^2(n^2+2n-3)$ Type 1 failures;
2. $2n\binom{n}{2}$ Type 2 failures;
3. $\binom{n}{2}^2$ Type 3 failures.

As we observed for triple failure, the percentage of fatal quadruple failures also decreases with the size of the array: it becomes less than 4 percent for $n \geq 4$ and less than 0.4 percent for $n \geq 8$.

One obvious way to increase the reliability of two-dimensional arrays would be to eliminate all fatal triple failures and reduce as much as possible the number of fatal quadruple failures. Let us show how it can be done at a fairly low cost.

Assume that we add to the array an additional superparity device containing the parity of all data disks. Once the archive is stable, this parity could be easily computed by computing the parities of either:

1. All horizontal parity disks, that is, disks P_1 to P_3 in Fig. 1 to 6, or

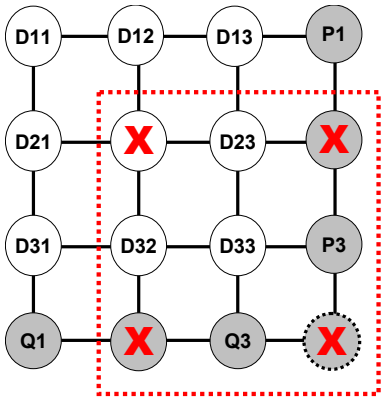

Fig. 8. A Type 1 quadruple failure resulting in a data loss.

2. All vertical parity disks, that is, disks Q_1 to Q_3 in Fig. 1 to 6.

In other words we would have

$$S = P_1 \oplus P_2 \oplus P_3 = Q_1 \oplus Q_2 \oplus Q_3 .$$

Note that the superparity device will only be used to recover the contents of failed parity disks. While Wildani et al. mentioned that their organization would have to read all data disk in a super-group when using its super-parity devices [19], our two-dimensional array will perform all its recovery operations on a row or a column basis.

Let us consider how the new array would handle the failure of a data disk and its two parity disks. As Fig. 7 shows, the remaining parity disks can now reconstitute the contents of the two failed parity disks with the help of the superparity disk S by doing

$$\begin{aligned} P_2 &= P_1 \oplus S \oplus P_3 \\ Q_2 &= Q_1 \oplus S \oplus Q_3 \end{aligned}$$

While all fatal triple failures would thus be eliminated, some quadruple failures would remain fatal. These failures include:

1. The failure of a data disk, its two parity disks and the superparity disk: Fig. 8 shows one of these Type 1 failures;
2. The failure of two data disks in the same row or column and their parity disks in the other dimension: these failures are identical to the Type 2 failures that discussed before ;
3. The failure of four data disks forming a square: these failures are identical to the Type 3 failures that we discussed before.

Out of the $\binom{(n+1)^2}{4}$ possible quadruple failures the array can experience, we can thus enumerate:

1. n^2 Type 1 fatal failures,
2. $2n\binom{n}{2}$ Type 2 fatal failures,

3. $\binom{n}{2}^2$ Type 3 fatal failures,

for a total of $\binom{n+1}{2}^2$ fatal quadruple failures.

As we observed before, the fraction of fatal quadruple failures also decreases with the size of the array: it becomes less than 1 percent for $n \geq 4$ and less than 0.1 percent for $n \geq 8$. This is four times less than the percentages of fatal failures we observed on the original array without a superparity device. This reduction is largely due to the lower number of Type 1 fatal quadruple failures, which were $n^2(n^2+2n-1)$ for the original array but only n^2 for the array with a superparity device.

Let us now turn our attention to quintuple failures and enumerate which ones will result in a data loss. These fatal failures will consist of all quadruple fatal failures plus any other disk. Out of the $\binom{(n+1)^2}{5}$ possible quadruple failures the array can experience, we can thus enumerate $\binom{n+1}{2}^2((n+1)^2-4)$ fatal failures.

It should come as no surprise that the percentage of fatal quintuple failures also decreases with the size of the array: it becomes less than 4 percent for $n \geq 4$ and less than 0.4 percent for $n \geq 8$.

IV. Reliability Analysis

Estimating the reliability of a storage system means estimating the probability $R(t)$ that the system will operate correctly over the time interval $[0, t]$ given that it operated correctly at time $t = 0$. Computing that function requires solving a system of linear differential equations, a task that becomes quickly intractable as the complexity of the system grows. A simpler option is to use instead the mean time to data loss (MTTDL) of the storage system, which is the approach we will take here. As our aim is to measure the impact of the superparity device on two-dimensional array reliability, we will start by evaluating the reliability of two-dimensional arrays without superparity devices.

A. Our Model

Our system model consists of an array of disks with independent failure modes. Whenever a disk fails, a repair process is immediately initiated for that disk. Should several disks fail, the repair process will be performed in parallel on those disks. We assume that disk failures are independent events and are exponentially distributed with mean λ. In addition, we require repairs to be exponentially distributed with mean μ. Both hypotheses are necessary to represent our system by a Markov process with a finite number of states.

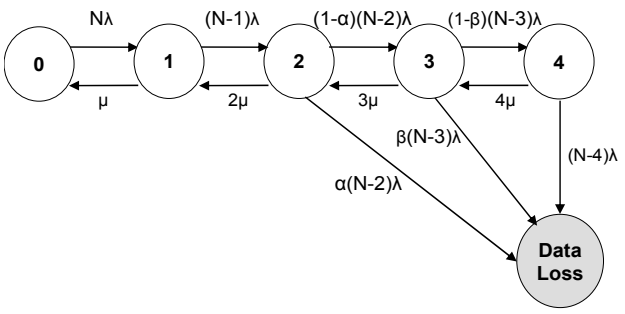

Fig. 9. State transition probability diagram for a two-dimensional RAID array with N= n^2 +2n disks.

B. Arrays without superparity devices

Fig. 9 displays the state transition probability diagram for a two-dimensional RAID array with n^2 data disks and $2n$ parity disks for a total of $N = n^2 + 2n$ disks. State <0> is the original state where all N disks are operational. Should one of the disks fail, the system would move to state <1> with an aggregate failure rate $N\lambda$. A second failure would bring the system to state <2>. A third failure would could either bring the system to state <3> or result in a data loss. Given that n^2 of the $\binom{n^2+2n}{3}$ possible triple failures will result in a data loss, the two failure transitions from state <2> are:

1. A transition to the failure state with rate $\alpha(N-2)\lambda$ where

$$\alpha = \frac{n^2}{\binom{n^2+2n}{3}}$$

2. A transition to state <2> with rate $(1-\alpha)(N-2)\lambda$.

Similarly, the two failure transitions from state <3> will be:

1. A transition to the failure state with rate $\beta(N-3)\lambda$ where

$$\beta = \frac{n^2(n^2+2n-3)+2n\binom{n}{2}+\binom{n}{2}^2}{\binom{n^2+2n}{4}}$$

2. A transition to state <4> with rate $(1-\beta)(N-3)\lambda$.

As we did not take into account the possibility that the array could survive a quintuple failure, there is a single failure transition leaving state <4>.

Recovery transitions are more straightforward: they bring the array from state <4> to state <3>, then from state <3> to state <2> and so on until the system returns to its original state <0>.

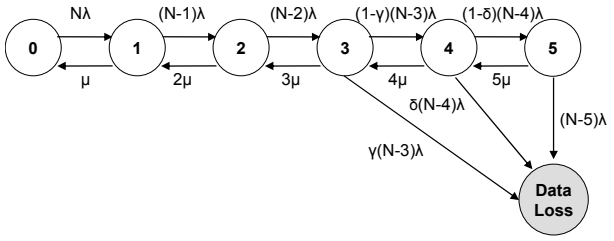

Fig. 10. State transition probability diagram for a two-dimensional RAID array with N= $(n+1)^2$ disks.

The Kolmogorov system of differential equations describing the behavior of the array is

$$\begin{aligned}
\frac{dp_0(t)}{dt} &= -N\lambda p_0(t) + \mu p_1(t) \\
\frac{dp_1(t)}{dt} &= -((N-1)\lambda+\mu)p_1(t) + N\lambda p_0(t) + 2\mu p_2(t) \\
\frac{dp_2(t)}{dt} &= -((N-2)\lambda+2\mu)p_2(t) + (N-1)\lambda p_1(t) + \\
&\quad 3\mu p_3(t) \\
\frac{dp_3(t)}{dt} &= -((N-3)\lambda+3\mu)p_3(t) + (1-\alpha)(N-2)\lambda p_2(t) + \\
&\quad 4\mu p_4(t) \\
\frac{dp_4(t)}{dt} &= -((N-4)\lambda+4\mu)p_4(t) + (1-\beta)(N-3)\lambda p_3(t)
\end{aligned}$$

where $p_i(t)$ is the probability that the system is in state <i> with the initial conditions $p_0(0) = 1$ and $p_i(0) = 0$ for $i \neq 0$.

The Laplace transforms of these equations are

$$\begin{aligned}
sp_0^*(s) - 1 &= -N\lambda p_0^*(s) + \mu p_1^*(s) \\
sp_1^*(s) &= -((N-1)\lambda+\mu)p_1^*(s) + N\lambda p_0^*(s) + 2\mu p_2^*(s) \\
sp_2^*(s) &= -((N-2)\lambda+2\mu)p_2^*(s) + (N-1)\lambda p_1^*(s) + \\
&\quad 3\mu p_3^*(s) \\
sp_3^*(s) &= -((N-3)\lambda+3\mu)p_3^*(s) + \\
&\quad (1-\alpha)(N-2)\lambda p_2^*(s) + 4\mu p_4^*(s) \\
sp_4^*(s) &= -((N-4)\lambda+4\mu)p_4^*(s) + \\
&\quad (1-\beta)(N-3)\lambda p_3^*(s)
\end{aligned}$$

Observing that the mean time to data loss (MTTDL) of the array is given by

$$MTTDL = \sum_{i=0}^{4} p_i^*(0),$$

we solve the system of Laplace transforms for $s = 0$ and a fixed value of N then use this result to compute the MTTDL of our system.

C. Adding a superparity device

Fig. 10 displays the state transition probability diagram for a two-dimensional RAID array with n^2 data disks, $2n$ parity disks and a superparity disk for a total of $N = (n+1)^2$ disks. As we can see, there is an additional state <5> and no failure transitions leaving state <2>.

The two failure transitions from state <3> are:

1. A transition to the failure state with rate $\gamma(N-3)\lambda$ where

$$\gamma = \frac{\binom{n+1}{2}^2}{\binom{(n+1)^2}{4}}$$

2. A transition to state <4> with rate $(1-\gamma)(N-3)\lambda$.

Similarly, the two failure transitions from state <4> will be:

1. A transition to the failure state with rate $\delta(N-4)\lambda$ where

$$\delta = \frac{\binom{n+1}{2}^2((n+1)^2-4)}{\binom{(n+1)^2}{5}}$$

2. A transition to state <5> with rate $(1-\delta)(N-4)\lambda$.

The Kolmogorov system of differential equations describing the behavior of the array is

$$\begin{aligned}
\frac{dp_0(t)}{dt} &= -N\lambda p_0(t) + \mu p_1(t) \\
\frac{dp_1(t)}{dt} &= -((N-1)\lambda+\mu)p_1(t) + N\lambda p_0(t) + 2\mu p_2(t) \\
\frac{dp_2(t)}{dt} &= -((N-2)\lambda+2\mu)p_2(t) + (N-1)\lambda p_1(t)+) \\
&\quad 3\mu p_3(t) \\
\frac{dp_3(t)}{dt} &= ((N-3)\lambda+3\mu)p_3(t) + (N-2)\lambda p_2(t) + \\
&\quad 4\mu p_4(t) \\
\frac{dp_4(t)}{dt} &= -((N-4)\lambda+4\mu)p_4(t) + (1-\gamma)(N-3)\lambda p_3(t) + \\
&\quad 5\mu p_5(t) \\
\frac{dp_5(t)}{dt} &= -((N-5)\lambda+5\mu)p_5(t) + (1-\delta)(N-4)\lambda p_4(t)
\end{aligned}$$

Using the same techniques as before, we obtain an algebraic expression of the system for a fixed value of N.

D. Comparing the two organizations

Fig. 11 displays on a logarithmic scale the MTTDLs achieved by two-dimensional RAID arrays with 64 data disks for average repair times varying between half a day and seven days. The lower curve refers to an array with 16 parity blocks while the upper curve refers to the same array with an additional superparity disk. We selected this array size because it corresponded to a space overhead of 20 percent for the array with 64 data and 16 parity disks, an overhead that we assumed to be reasonable.

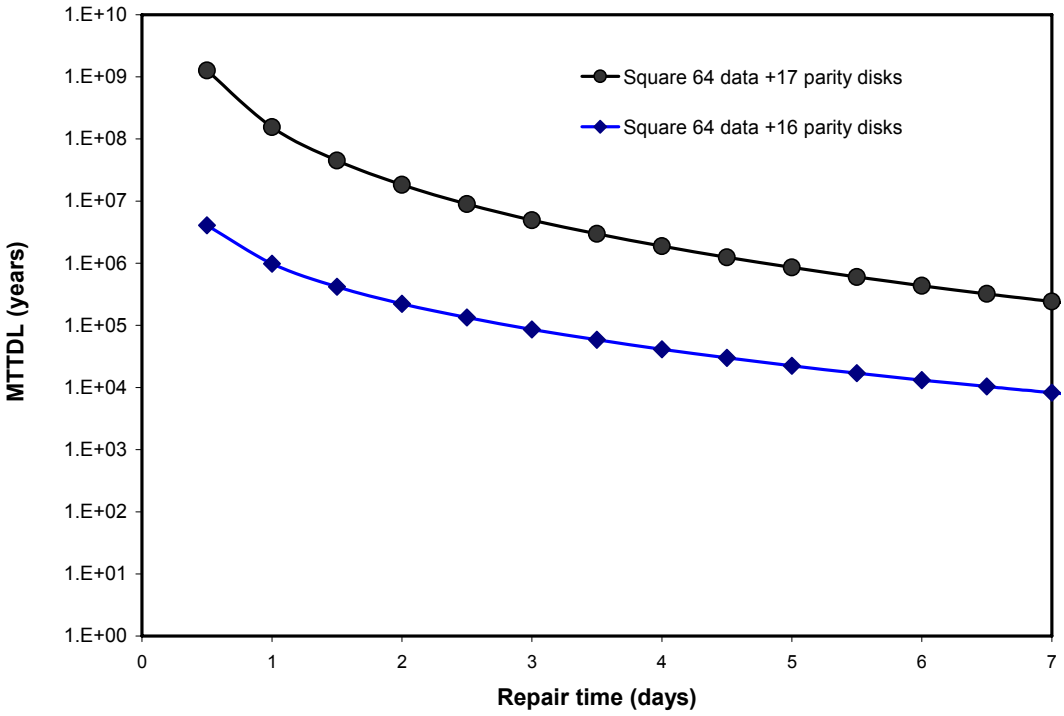

Fig. 11. MTTDL of two-dimensional RAID arrays with and without a superparity disk (17 vs. 16 parity disks respectively).

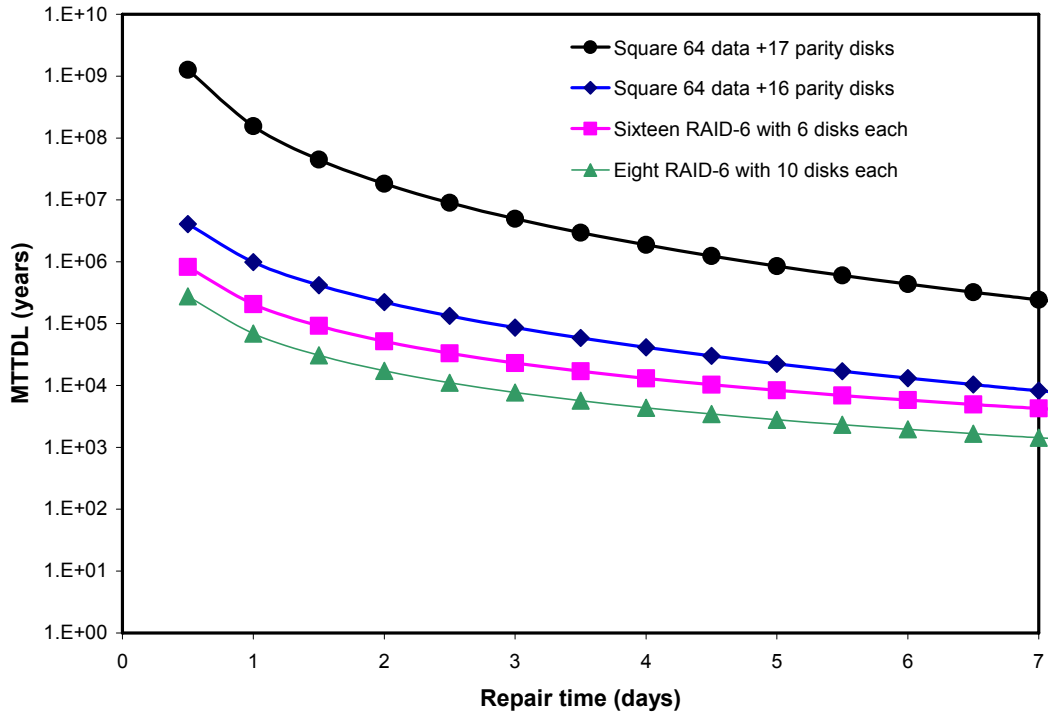

Fig. 12. Comparing the MTTDLs achieved by two-dimensional RAID arrays and those achieved by sets of independent RAID level 6 strings.

TABLE I. RELATIVE MTTFS OF THE TWO "SQUARE" ORGANIZATIONS WITH REGARDS TO THOSE OF A SET OF 8 RAID LEVEL 6 ARRAYS WITH 10 DISKS EACH.

Repair Times	Square 64D +17 P	Square 64D +16P	8×10 RAID 6
Half a day	4589.381	14.760	1
One day	2252.041	14.289	1
Two days	1056.169	12.862	1
Half a week	521.670	10.295	1
One week	169.018	5.746	1

We assumed a disk failure rate λ of one failure every one hundred thousand hours, which is slightly less than one failure every eleven years. This rate is at the high end of the failure rates observed by Pinheiro et al. [13] as well as Schroeder and Gibson [16]. MTTDLs are expressed in years and repair times in days.

As we can see, adding a superparity disk to the original array results in dramatic increases in its MTTDL. This effect is most pronounced for short repair times. Adding a superparity disk to the array multiplies its MTTDL by a factor of at least 50 when the repair times remain below three days and half but only by a factor of 28 to 30 when the repairs take one week to complete.

These improvements are truly remarkable as they were achieved by adding a single parity disk to an array already counting 64 data and 16 parity disks.

E. *Comparing the two organizations with conventional RAID arrays*

Finally, we compared our two array organizations with sets of RAID level 6 strings with similar storage capacities and comparable space overheads. These two organizations are:

1. Eight RAID level 6 arrays with ten disks each;
2. Sixteen RAID level 6 arrays with six disks each.

Their space overheads are 20 percent for the first organization and 33 percent for the second.

Observing that the MTTDL of a RAID level 6 with n disks is

$$MTTDL(n) == \frac{(3n^2 - 6n + 2)\lambda^2 + (3n-2)\lambda\mu + 2\mu^2}{n(n-1)(n-2)\lambda^3},$$

we obtain the MTTDL of an organization comprising m such arrays using the formula

$$MTTDL(m;n) == \frac{MTTDL(n)}{m}$$

Fig. 12 summarizes our results. As we can see, both two-dimensional RAID arrays achieve much better MTTDLs than the two sets of RAID level 6 strings even though one of them has a much higher space overhead than our two-dimensional array.

As before, we observe that our organization performs best at short repair times, that is, at low λ/μ ratios. The ratio between the MTTDL afforded by our organization and that of sixteen RAID level 6 arrays with six disks each varies between 4,530 when the repair process takes just half a day and 57 when the process takes a whole week. This is a very impressive achievement when we realize that the two dimensional array with a superparity device has a space overhead of 21 percent while the space overhead of the RAID organization is 33.3 percent.

The gap is even wider when we compare the performance of our organization with that of the RAID organization comprising eight RAID level 6 stripes with ten disks each. As Table I shows, the ratio between the MTTDLs afforded by the two organizations varies between 4,589 when the repair process takes just half a day and 169 when it takes a whole week.

F. *Impact on data survival*

We might now ask how the addition of a superparity disk to a two-dimensional RAID array would impact the data survival rate of the array. Assuming that disk arrays decay at an exponential rate, the probability that no data loss would occur over a time interval of duration t is

$$R(t) = \exp(-\frac{1}{MTTDL})$$

Two factors limit however the accuracy of this result. First, the MTTDL characterizes fairly well the behavior of a hypothetical disk array that would remain in service until they experience a data loss. In reality, disk arrays are typically replaced five to seven years after their initial deployment. In a previous study of disk array reliability**,** we observed that the disk array reliability estimates obtained using their MTTDLs significantly underestimated the actual reliability of these arrays every time the disk repair rate μ felt below one thousand times the disk failure rate λ [11].

Second, we have neglected so far to take into account the impact of irrecoverable read errors. Given the complexity of the issue, a thorough analysis of the impact of these read errors on disk arrays would be outside the scope of this paper. We can only state here that any disk array tolerating a fraction f of all simultaneous failures of n disks will always tolerate the same fraction f of all simultaneous failures of $n - m$ disks and irrecoverable read errors on m of the remaining disks.

V. IMPLEMENTATION ISSUES

As we mentioned earlier, the main limitation of our new scheme is its poor write throughput: since the superparity device contains the parity of all data disks, it must be updated each time any data disk is updated. Hence the write throughput of the whole array is equal to the throughput of its superparity device.

We already proposed one way to mitigate this problem: it consists of disabling superparity device updates while the archive is populated as well as every time a significant fraction of the stored data must be updated. The superparity device would then play the role of a lock securing the archived data. We would unlock the archive whenever it is being updated just as we open a conventional file cabinet

when we want to access its contents. Once the update is completed, we would update the parity device and return the archive to its locked state. Even then, the data would remain protected against all possible double-disk failures and most triple failures, which is still better than any RAID level 6 solution. As a result, the procedure would have a minimal impact on the MTTDL of the array as long as it remains protected by its parity device most of the time.

Other options are possible. First, we could use storage class memory for the parity device. These new devices are expected to have access times of the order of 100 ns and access rates between 200 and 1,000 megabytes per second [9]. Their superior performance would greatly reduce the need for disabling superparity device updates. Unfortunately, the higher reliability of these devices will not have a direct impact on the MTTDL of the array because most fatal quadruple failures are failures of Type 2 and 3 that do not involve the superparity device. We would still observe a small increase in the MTTDL of the array because storage class memories are immune to irrecoverable read errors.

A second solution applies to all archives requiring more than one two-dimensional RAID array. We could form groups of k arrays such that the superparity device of each array would contain $1/k$ of the superparity data of each array. This would multiply by k the maximum write throughput of each array but not the maximum write throughput of the whole archive.

VI. Conclusion

We have presented a two-dimensional RAID architecture that is specifically tailored to the needs of archival storage systems. Our proposal starts with a fairly conventional two-dimensional RAID architecture where each disk belongs to exactly one horizontal and one vertical RAID level 4 stripe. Once the array has been populated, we add a superparity device that contains the exclusive OR of all the contents of all horizontal—or vertical—parity disks. The new organization tolerates all triple disk failures and nearly all quadruple and quintuple disk failures. As a result, it provides MTTDLs:

- At least thirty times better than those on a two-dimensional array without superparity device;
- At least 169 times better than those of a set of RAID level 6 stripes with equal capacity and comparable parity overhead.

Our next task will be to confirm our results through discrete simulation. This would allow us to consider time-dependent failure rates and arbitrary repair time distributions and would provide us with confidence intervals for the probability than a given two-dimensional RAID array will operate during its useful lifetime without experiencing a data loss.

References

[1] M. Blaum, J. Brady, J. Bruck, and J. Menon, EvenOdd: An efficient scheme for tolerating double disk failures in RAID architectures, *IEEE Trans. Computers* 44(2):192–202, 1995.

[2] W. A. Burkhard and J. Menon. Disk array storage system reliability. *Proc. 23rd International Symposium on Fault-Tolerant Computing*, pp. 432–441, June 1993.

[3] P. Corbett, B. English, A. Goel, T. Grcanac, S. Kleiman, J. Leong, and S. Sankar, Row-diagonal parity for double disk failure correction, *Proc. 3rd USENIX Conference on File and Storage Technologies*, pp. 1–14, 2004.

[3] P. M. Chen, E. K. Lee, G. A. Gibson, R. Katz and D. A. Patterson. RAID, High-performance, reliable secondary storage, *ACM Computing Surveys* 26(2):145–185, 1994.

[5] W. Gang, L. Xiaoguang, L. Sheng, X. Guangjun, and L. Jing, Generalizing RDP codes using the combinatorial method, *Proc. 7th IEEE International Symposium on Network Computing and Applications*, pp. 93–100, July 2008.

[6] L. Hellerstein, G. Gibson, R. M. Karp, R. H. Katz, and D.A. Patterson. Coding techniques for handling failures in large disk arrays. *Algorithmica*, 12(3-4):182-208, June 1994

[7] C. Huang and L. Xu, STAR: an efficient coding scheme for correcting triple storage node failures, *Proc. 4th USENIX Conference on File and Storage Technologies*, pp. 197–210, Dec. 2005.

[8] W. S. Lee, *Two-dimensional storage array with prompt parity in one dimension and delayed parity in a second dimension*, US Patent #6675318 B1, 2004.

[9] S. Narayan, Storage class memory a disruptive technology, *Disruptive Technologies Panel: Memory Systems of SC '07*, Reno, NV, Nov. 2007.

[10] J.-F. Pâris, T. J. Schwarz and D. D. E. Long. Self-adaptive archival storage systems. *Proc. 26th International Performance of Computers and Communication Conference*, pp. 246–253, Apr. 2007.

[11] J.-F. Pâris, T. J. Schwarz, D. D. E. Long and A. Amer, When MTTDLs Are Not Good Enough: Providing Better Estimates of Disk Array Reliability. *Proc. 7th International Information and Telecommunication Technologies Symposium*, pp. 140–145, Dec. 2008.

[12] D.A. Patterson, G. Gibson, and R. H. Katz, A case for redundant arrays of inexpensive disks (RAID). *Proc. 1988 SIGMOD International Conference on Data Management*, pp. 109–116, June 1988.

[13] E. Pinheiro, W.-D. Weber and L. A. Barroso, Failure trends in a large disk drive population, *Proc. 5th USENIX Conference on File and Storage Technologies*, pp. 17–28, Feb. 2007.

[14] J. S. Plank, A new minimum density RAID-6 code with a word size of eight, *Proc. 7th IEEE International Symposium on Network Computing and Applications*, pp. 85–92, July 2009.

[15] J. S. Plank, The RAID-6 liberation codes, *Proc. 6th USENIX Conference on File and Storage Technologies*, pp. 1–14, Feb. 2008.

[16] B. Schroeder and G. A. Gibson, Disk failures in the real world: what does an MTTF of 1,000,000 hours mean to you? *Proc. 5th USENIX Conference on File and Storage Technologies*, pp. 1–16, Feb. 2007.

[17] T. J. E. Schwarz, *Reliability and Performance of Disk Arrays*, PhD Dissertation, Department of Computer Science and Engineering, University of California, San Diego, 1994.

[18] T. J. E. Schwarz and W. A. Burkhard. RAID organization and performance. *Proc. 12th International Conference on Distributed Computing Systems*, pp. 318–325 June 1992.

[19] A. Wildani, T. J. E. Schwarz, E. L. Miller and D. D. E. Long, Protecting against rare event failures in archival systems, *Proc. 17th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, pp. 246–256, Sep. 2009.