ارایهٔ الگوی انعکاس داده ای نوین بمنظور بهبود قابلیت اطمینان و تحمل پذیری نسبت به خطا

جواد اکبری ترکستانی[1]
دانشکدهٔ مهندسی کامپیوتر ، دانشگاه آزاد اسلامی اراک
akbari@jdmarkazi.ac.ir

محمد رضا میبدی[2]
دانشکدهٔ مهندسی کامپیوتر وفناوری اطلاعات،دانشگاه صنعتی امیر کبیر
meybodi@ce.aut.ac.ir

چکیده

این مقاله یک تکنیک ذخیره سازی داده ای نوین را بمنظور پیاده سازی بر روی سیستم های دنبالهٔ دیسکی توزیع شده با نیاز امنیتی بالا ارایه مینماید. هدف ازطراحی این الگوی داده ای ، افزایش حفاظت داده ای ، ارایهٔ پهنای باند افزوده ، سطح بالاتر اجراء عملیات موازی دستیابی داده ای و بهبود تحمل پذیری نسبت به خطا ست . این روش بگونه ای خاص از ترکیب دو تکنیک انعکاس داده ای و ردیفی کردن (در سطح بلاکی) از تکنیکهای RAID داده ها را بر روی دنباله توزیع مینماید. از این روی ، ما مدل پیشنهادیمان را درمقایسه با سایرمدلهای مشابه همچون RAID 5 , RAID 1+ 0 , RAID1 مورد ارزیابی قرار داده و دریافته ایم که پیاده سازی مدل مزبور موجب افزایش فابلیت اطمینان وسطح بازگرد پذیری سیستم خواهد شد . ویژگی بارز این مدل افزایش تحمل پذیری سیستم بهنگام بروز خطا وکاهش سطح فقدان داده ای وافزایش نرخ احیاء داده ایست .

واژه های کلیدی : افزونگی داده ای ، تحمل پذیری خطا ، انعکاس داده ای ، فابلیت اطمینان

۱- مقدمه

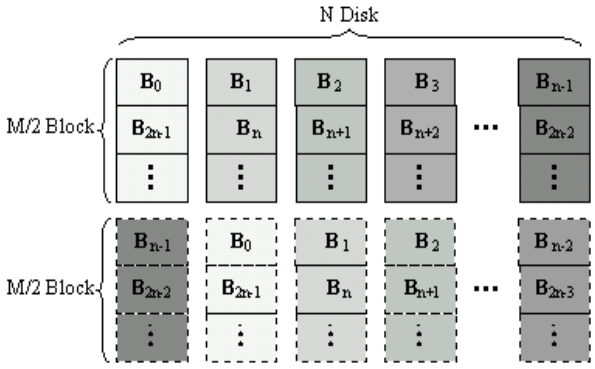
برخی از سیستم ها ، اطلاعات و داده های فوق العاده بحرانی را در بر داشته وبهمین لحاظ در اینگونه از سیستم ها ، حفاظت وامنیت داده ها از اهمیت بسزایی برخوردار است . هر یک از تکنیکهای ذخیره سازی داده ای با توجه به شرایط بحرانی داده درون سیستم و میزان امنیت داده ای مورد نیاز بروشهای مختلفی سطوح گوناگونی از امنیت را برای سیستم فراهم میا ورند. در تکنیک RAID نیز هر مدل با توجه به ویژگیهای ذاتی خود سطوح گوناگونی از امنیت داده ای را فراهم میاورد . از میان سطوح مختلف تکنیک ، RAID سطح یک دارای افزونگی داده ای نسبتا بالایی بوده اما در مقابل حفاظت ، صحت وامنیت داده ای را در سطح خوبی فراهم مینماید. ایراد اصلی این مدل در سیستم هایی که نیاز چندانی به امنیت داده ای ندارند ، در درجهٔ اول حجم بالای اطلاعات افزوده ای است که به سیستم تحمیل میگردد. اما در سیستم های با نیاز امنیتی بالاتر عدم توازی گرایی عملیاتهای داده ای سیستم ، عدم استفاده از حداکثر پهنای باند وحتی پایین بودن سطح امنیتی داده ها بهنگام بروز خرابی واحیاء داده ای از معایب مدل فوق میباشد. نا گفته نماند که حفاظت داده ای متناسب با میزان افزونگی هر مدل ،در سایر سطوح نیز فراهم است. برای مثال در سطح پنجم ازاین تکنیک با توجه به ویژگیهای خاص مدل امنیت داده ای در سطح خوبی برای سیستم فراهم میاید. در RAID 1+0 که از ترکیب دو سطح پایه ای (سطوح یک ودو) ایجاد میگردد تاحدودی سعی گردیده تا با ادغام دو تکنیک Stripping وMirroring ، سیستمی پیاده سازی گردد که علاوه بر سطح امنیتی بالا ، توازی گرایی وسایر فاکتورهای ذکر شده را نیز بهبود بخشد. این مدل نیز خود دارای کاستیهایی است که در بخشهای بعدی بطور مفصل به بررسی آنها میپردازیم[4][1] .

---

[1] عضو هیات علمی دانشکدهٔ مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد اراک

[2] عضو هیات علمی دانشکدهٔ مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیر کبیر

با توجه به ویژگیها، مزایا ، معایب و نقاط قوت وضعف مدلهای قبلی، ما سعی داریم مدلی ارایه دهیم که علاوه بر فراهم آوردن سطح بالایی از امنیت ، صحت و حفاظت داده ای بتواندحداکثر توازی گرایی ، حداکثرپهنای باند موثر بهنگام عملیاتـهای دسـتیابی داده[۳] (عملیاتهای ورودی /خروجی) و احیاء داده ای وحداکثرتحمل پذیری خطا را به گونه ای بـرای سیسـتم فراهـم آورد ، تـا هـم برای پیاده سازی روی سیستم های با نیاز امنیتی بالا مناسب باشد وهم بازدهی سیستم را به میزان خوبی افزایـش دهـد. در ادامـه مقاله ، در بخش ۲ به تشریح مدل پیشنهادی خود پرداخته و پس از آن پیاده سازی مدل را ارایه نموده ایم .در بخش۴ مـدل ارایـه شده را با سایر مدلهای مشابه ازجنبه های مختلفی همچون هزینهٔ فضایی وزمانی مدل ، حداکثر پهنای باند ، سطح توازی سیسـتم و هزینهٔ احیاء داده ای مورد مقایسه قرار داده و در بخشهای بعدی نیز با شبیه سازی مدل وارایه یک نتیجه گیری به مقالـه خاتمـه داده ایم [3][2].



تصویر ۱.پیکره بندی کلی RAID-RMS

۲- مدل RAID-RMS

مدل RAID پیشنهادی ما که آنرا RAID-RMS نامیده ایم از ترکیـب دو تکنیـک ردیفـی کـردن [۴]داده ای در سـطح بلاکـی و انعکاس داده ای[۵]در سطح دیسکهای سیستم استفاده می نماید. اما تفاوت این مدل با سایر مدلها درآنسـت کـه ایـن دو تکنیـک را بگونه ای با یکدیگر در هم می آمیزد که موجب افزایش سطح اجراء عملیات موازی داده ای ، تحمل پذیری خطا و ایجاد پهنای باند بالاتر سیستم گردد. علاوه بر این،در مدل پیشنهادی، الگوی ذخیره سازی خاصی پیاده سازی می شود که قادر است ، قابلیت احیاء داده ای [۶]را افزایش ، داده بگونه ای که بهنگام بروزهر خرابی حداقل داده از دست رود ، و انجـام عملیـات احیـاء داده ای را تسـریع نموده وعلاوه برآن بهنگام بروز خرابی بازدهی عملیات دستیابی داده ای در سیستم کمترین افت را داشته باشد[4][3].

همانطور که در تصویر ۱ هم مشاهده میگردد، با استفاده از تکنیک انعکاس داده ای ، بلاکهای سیستم را به دو دسته بلاکهای داده ای وبلاکهای پشتیبان[۷] تقسیم بندی نموده ایم وبر خلاف مدلهای مشابه RAID 1 , RAID 1+0 کـه در آنـها بلاکـهای داده ای وپشتیبان در دیسکهای متمایزی قرار دارند،در این مدل داده های اصلی ونسخه های پشتیبان آنها بر روی دیسکهای یکسانی توزیع می شوند ودر واقع ما نیمهٔ بالایی هر دیسک را برای بلاکهای داده ای اصلی ونیمی دیگر از آنرا برای بلاکـهای داده ای پشـتیبان در نظر گرفته ایم .آنچه در این میان از اهمیت بالایی برخوردار است نحوهٔ قرار دادن وتوزیع داده ها بر روی دنبالهٔ دیسکی اسـت ، کـه ما در بخشهای بعدی برای این نحوهٔ توزیع داده ای را در قالب فرمولهای کلی ارایه مینماییم. لازم بذکراست که در این مدل بـرای توزیع داده ای از تکنیک انتقال براست[۸] بلاکهای داده ای وپشتیبان در هر بلاک وهر دیسک متناسب با سطح بلاک وشمارهٔ دیسک استفاده می شود[4][3][1]. در این مدل از توزیع داده ها ، بلاکهای داده ای مختلف ونسخه هـای پشـتیبان آنـها بـر روی دیسـکهای

---

[3] I/O operation

[4] Data Stripping

[5] Mirroring

[6] Data Recovery

[7] Mirrored Block

[8] Rotate – Right

مختلف دنباله بطریقی گسترده میشوند که بهنگام بروز خرابی در سطح دیسکی کمترین داده های سیستم مفقود گردد.وعلاوه برآن ، سرعت بازدهی سیستم بهنگام احیاء بلاکهای داده ای از دست رفته افزایش میابد. برهمین اساس برای پیاده سازی بر روی محیطهای ذخیره سازی داده ای که هم نیاز به امنیت وصحت داده ای نسبتا بالایی داشته وهم سرعت دستیابی داده ای وبازدهی سیستم در آنها بالاست مناسب است .

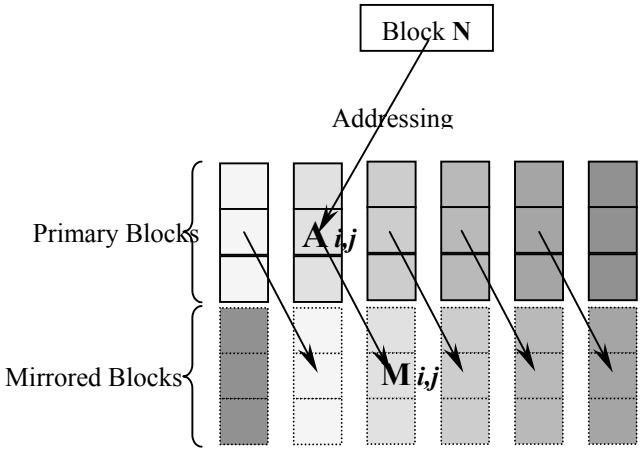
۳- پیاده سازی مدل RAID-RMS[9]

همانگونه که در بخشهای قبلی نیز گفته شد، مدل پیشنهادی از ادغام دو تکنیک انعکاس داده ای(Mirroring) وردیفی کردن داده ای (Stripping) بهمراه تکنیکهایی از انتقال بلاکهای داده ای استفاده مینماید. ما قصد داریم تا در ادامه نحوهٔ توزیع داده ها را دراین مدل برای حالات کلی بیان نموده و فرمولهایی را بکمک پارامتر های جدول ۱ برای ایجاد الگوهایی بمنظور برقراری انطباق میان داده های پیوستهٔ فضای حافظه با آدرسهای فضای فیزیکی مجازی روی دنبالهٔ دیسکی ارایه میدهیم . برای همین منظور به تجزیهٔ چگونگی آدرس دهی (نگاشت شمارهٔ بلاکهای داده ای بر روی آدرسهای دنبالهٔ دیسکی) در این مدل پرداخته وآنرا در قالب فرمولهایی بمنظور نگاشت آدرس فضای حافظهٔ تک بعدی (شمارهٔ بلاک داده ای) به آدرس فضای دو بعدی دنبالهٔ دیسکی ، محاسبهٔ آدرس بلاک داده ای وآدرس نسخهٔ انعکاس یافتهٔ آن (بلاک پشتیبان متناسب با آن) از روی بلاک داده ای بیان مینماییم .

جدول۱ .جدول نمادها

نماد	توضیحات
N	شمارهٔ بلاک داده ای
n	تعداد دیسکهای موجود در دنباله
m	تعداد بلاکهای موجود در هر دیسک
i	موقعیت افقی بلاک داده ای درون دنبالهٔ
j	موقعیت عمودی بلاک داده ای درون دنبالهٔ

همانطور که در تصویر ۲ نیز مشاهده میگردد بلاکهای داده ای را با استفاده از تکنیک ردیفی کردن داده ای در سطح بلاکی بر روی دیسکهای مختلف ومتوالی دنباله توزیع مینماییم که در هر سطح بلاکی از هر دیسک از تکنیک انتقال داده ای متناسب با همان سطح استفاده میگردد . از سوی دیگر نیز برای هر داده ای که بر روی یک بلاک داده ای از دنبالهٔ دیسکی کپی میگردد، نسخه ای بر روی یکی از بلاکهای پشتیبان که آدرس آن با توجه به فرمولهای زیرین بدست می آید قرار داده می شود و در واقع بدین نحو از تکنیک انعکاس داده ای نیز در این مدل استفاده میشود.پس از پیاده سازی الگوهای سخت افزاری ساختار مدل پیشنهادی ، بایستی یک نگاشت بمنظور برقراری ارتباط میان شمارهٔ بلاکهای داده ای وفضای آدرس دهی مجازی دنبالهٔ دیسکی تعریف کرد، تا بهنگام قراردادن بلاکهای داده ای بر روی آدرس دوبعدی دنبالهٔ دیسکی ، شمارهٔ بلاک متناظر با هر زوج (i , j) را بتوان بر اساس این نگاشت محاسبه کرد . تعداد بلاکهای موجود در هر دیسک را **m** وتعداد دیسکهای دنباله را **n** در نظر گرفته ایم. ما در این مرحله سه نوع نگاشت برای ذخیره سازی داده ها روی دنبالهٔ دیسکی محاسبه مینماییم که عبارتند از :



تصویر ۲. نحوهٔ توزیع بلاکهای داده ای بر روی دنبالهٔ دیسکی

[9]Rotating – Mirroring - Stripping

۱. محاسبهٔ نگاشت آدرس دنبالهٔ دیسکی به شمارهٔ بلاک داده ای $\mathbf{N}_{A(i,j)}$
۲. محاسبهٔ نگاشت شمارهٔ بلاک داده ای به آدرس دنبالهٔ دیسکی بلاک اصلی $\mathbf{A}_{(i,j)}$
۳. محاسبهٔ نگاشت شمارهٔ بلاک داده ای به آدرس دنبالهٔ دیسکی بلاک پشتیبان $\mathbf{M}_{(i,j)}$

ابتدا نگاشتی را محاسبه مینماییم که آدرس بلاک داده ای روی دنبالهٔ دیسکی را بکمک فرمولی به شمارهٔ بلاک ذخیره شونده در آن آدرس تبدیل نماید .

$$\mathrm{N}_{A(i,j)} = \begin{cases} (i \times n) + (j \bmod n) - i & ; i \le j \\ & 0 \le i \le \dfrac{m}{2} - 1, 0 \le j \le n-1 \\ (i+1) \times n + j - i & ; i \phi\ j \end{cases} \qquad \text{فرمول ۱}$$

در فرمول شمارهٔ ۱ ابتدا شمارهٔ نسخهٔ اصلی بلاک داده ای مورد نظربر اساس پارامترهای مختلف تعیین و سپس بکمک فرمولهای دیگری آدرس وموقعیت بلاک داده ای پشتیبان محاسبه میگردد بر همین اساس دامنهٔ تغییر پارامتر i از 0 تا m/2 می باشد چرا که نیمهٔ بالایی بلاکهای درون هر دیسک را برای نگهداری بلاکهای داده ای اصلی در نظر گرفته ایم. بجهت آنکه برای توزیع داده ها در سطح دنبالهٔ دیسکی پس از بکارگیری تکنیک انعکاس وردیفی کردن داده ای ، داده ها را در سطح بلاکی تحت یک چرخش براست قرار میدهیم .بهمین لحاظ بلاکهای داده ای که آدرس افقی آنها از آدرس عمودیشان بیشتر است به ابتدای آن سطح منتقل می شوند و در فرمول ، ما آنها را با شرط i>j از سایر بلاکها متمایز مینماییم .

بکمک فرمول ۱ می توان مشخص نمود که در هر موقعیت افقی وعمودی از دنبالهٔ دیسکی ، کدامین بلاک داده ای میبایستی ذخیره گردد. اما بهنگام انتقال داده ها بر روی دنبالهٔ دیسکی آنچه از اهمیت بالاتری برخوردار است محاسبهٔ یک نگاشت معکوس برای فرمول فوق میباشد که بر اساس آن بتوان موقعیت افقی وعمودی هر بلاک داده ای را بهنگام ذخیره سازی آن بر روی دنباله مشخص نموده ودادهٔ مورد نظر را درآن بلاک قرار داد. که بسادگی وبکمک فرمول۲ آدرس افقی (شمارهٔ بلاک درون دیسک) و آدرس عمودی (شمارهٔ دیسک درون دنباله) محاسبه میگردد.

$$A_{(i,j)} = \begin{cases} i = \left\lfloor \dfrac{N}{n} \right\rfloor \\ & 0 \le i \le \dfrac{m}{2} - 1, \quad 0 \le j \le n-1 \\ j = \left[\left\lfloor \dfrac{N}{n} \right\rfloor + (N \mod n) \right] \bmod n \end{cases} \qquad \text{فرمول ۲}$$

در این روش بمنظور افزایش سطح حفاظت وامنیت داده ای از تکنیک انعکاس داده ای (کپی های دوگانهٔ داده ای) استفاده میگردد. یعنی بهنگام نوشتن هر بلاک داده ای یک کپی از آن نیز بعنوان نسخهٔ پشتیبان بر روی دیسک نگهداری میگردد تا بهنگام بروز خطا در یکی از بلاکهای مورد نظر دادهٔ مزبور از روی نسخهٔ دیگر واکشی شود در این مدل پیشنهادی بلاکهای داده ای پشتیبان بر روی بلاکهای نیمهٔ پایینی هردیسک ***(m/2 <= i <= m - 1)*** و با استفاده از روابط حاکم در فرمول ۳ قرار داده می شوند.

پس از بکارگیری فرمولهای ارایه شده بمنظور تعیین موقعیت بلاکهای داده ای ، اگر به نحوهٔ توزیع داده ای اصلی وپشتیبان نگاه نماییم بلاکهای داده ای پشتیبان متناظر با هر بلاک داده ای در نیمهٔ پایینی دیسک بعدی قرار میگیرد، که به این ترتیب در اثر از کار افتادن یک دیسک منفرد ، دنبالهٔ دیسکی باز هم می تواند بمنظور دستیابی داده ها با حداکثر پهنای باند فراهم شده توسط سایر دیسکهای باقیمانده در دنباله بکار خود ادامه دهد. ویا آنکه بهنگام بروز یک خطا در سطح دیسک ، داده های از دست رفته با سرعت بالاتری قابل احیاء بوده و حتی اگر دو دیسک متوالی از دنباله بطور همزمان دچار خرابی شوند، نسبت به سایر مدلهای دیگر کمترین دادهٔ موجود درون سیستم از دست خواهد رفت .

$$M_{(i',j')}=\begin{cases} i'=\left\lfloor \frac{N}{n} \right\rfloor+\frac{m}{2} \\ \\ \frac{m}{2}\le i'\le m-1 \quad , \quad 0\le j'\le n-1 \\ \\ j'=\left[\left\lfloor \frac{(N+1)}{n} \right\rfloor+((N+1) \mod n)\right] \bmod n \end{cases}$$

فرمول ۳

۴- ارزیابی مدل پیشنهادی

همانگونه که در بخشهای قبلی نیز به آن اشاره گردید مدل پیشنهادی ما از تلفیق تکنیک انعکاس داده ای با مدل ردیفی کردن داده ای در سطح بلاکی استفاده مینماید ودر واقع مدل مزبور برای پیاده سازی در سیستم هایی مناسب است که اطلاعات وداده های آنها نیاز به سطح امنیتی بالایی دارند . بنابراین بمنظور ارزیابی این مدل آنرا با سایر تکنیکهای ذخیره سازی داده ای مطرح در سیستم های امن مقایسه مینماییم . تکنیک سطح یک (RAID 1) بدلیل امنیت بالا و تشابه در استفاده از تکنیک Mirroring وسطح پنج از این تکنیک (RAID 5) بجهت بکارگیری تکنیک Stripping و تکنیک ترکیبی RAID 1+0 سه تکنیکی هستند که ما مدل پیشنهادیمان را با آنها مقایسه نموده ودر نهایت به بررسی نتایج بدست آمده وارزیابی نهایی مدل خوا هیم پرداخت.

در این بخش ما سعی داریم مدل پیشنهادی را با سایر مدلهای مشابه قبلی از جنبه های گوناگونی همچون هزینهٔ فضایی افزودهٔ مورد نیاز بمنظور تأمین امنیت وحفاظت داده ای[۱۰]مدل ، هزینهٔ عملیات ورودی/ خروجی مورد نیاز برای دستیابی به داده ها(خواندن ونوشتن داده) ، امکان کنترل همزمانی[۱۱]، تحمل پذیری مدل نسبت به خطا [۱۲] وهزینه احیاء خرابی دیسک وداده ارزیابی نموده و در هر مورد مزایا و معایب مدل مورد بررسی را ارایه کرده وآنها را با همدیگر مقایسه نموده ودر صورت امکان راه حل هایی را هم برای مرتفع ساختن ناکارامدیهای مدل در مواردی خاص ارایه میدهیم.

۴-۱- حداکثر پهنای باند[۱۳]

همانگونه که در بخشهای قبلی نیز بدان اشاره گردید یکی از نتایج ارزشمندی که بواسطهٔ پیاده سازی تکنیک RAID فراهم میگردد، پهنای باند بیشتر برای سیستم است . در این تکنیک سعی میگردد با بکارگیری روشهای مختلف توزیع داده ای، امکان انجام عملیات ورودی/ خروجی موازی را افزایش داده و بدین ترتیب به حداکثر پهنای باند برای انتقال داده ها در سیستم دست یافت.دراین بخش ما تکنیکهای مطرح شده را مورد بررسی قرارداده و حداکثر پهنای باند مؤثری را که هر یک از این مدلها برای سیستم فراهم میاورند محاسبه نموده ودر نهایت آنها را با یکدیگر مقایسه مینماییم.برای این منظور ابتدا برخی از نمادهای بکار رفته در مقایسات را در قالب جدول۲ ارایه میدهیم .

جدول ۲. جدول نمادها

نماد	توضیحات
N	تعداددیسک در یک سیستم RAID توزیع شده
B	حداکثر پهنای باند هر دیسک
S	اندازهٔ هر بلاک از دیسک
R	متوسط زمان خواندن هر بلاک از دیسک
W	متوسط زمان نوشتن هر بلاک از دیسک
M	تعداد بلاکهای موجود در هر فایل

پهنای باند فراهم آورده شده در هر یک از مدلهای مورد بررسی برای حالات مختلف وبازاء عملیاتهای مختلف دیسکی مقادیر متفاوتی را شامل میگردد ، از این روی ما در تحلیل پهنای باند مؤثر سیستم عملیاتهای خواندن ونوشتن داده ای را بطور جداگانه

---

[10] Data Protection

[11] Concurrency Control

[12] Fault tolerance

[13] Maximum Bandwidth

مورد بررسی قرار داده ایم وحتی در برخی موارد که عملیات دیسکی (ورودی/ خروجی) برای فایلها ومقادیر داده ای طولانی و کوتاه نتایج متفاوتی را در بر دارد. آنها را از یکدیگر متمایز نموده وتحلیل کرده ایم[4][3] .

مفاهیم Small Read / Write ویا Large Read / Write که در جدول ۳ مشاهده می شود بنابر نیاز تقسیم بندی عملیات دیسکی که در بالا گفته شد صورت گرفته است . منظور از خواندن یا نوشتن کوتاه آن دسته از عملیاتهای وردی/ خروجی داده ای هستند که تنها بر روی تمام یا قسمتی از یک بلاک داده ای تاثیر می گذارند ودر واقع عملیاتهای موازی ورودی / خروجی که از این نوع هستند می توانند روی بلاکهای متفاوت از یک Stripe منفرد اعمال شده ویا بر بلاکهای متفاوت یا یکسان از Stripe های متفاوت تاثیر گذارند . که هر یک از انواع فوق نیز در هر یک از تکنیکهای مورد بحث ما نتایج نتفاوتی را در پی دارند. با این همه دامنهٔ عملیاتی آنها از یک بلاک فراتر نمی رود. اما منظور از خواندن ونوشتن های طولانی آن دسته از عملیات دیسکی را شامل می شود که در آنها داده ها بر روی چندین بلاک متوالی از یک یا چند Stripe متوالی قرار داده شده ویا از آنها خوانده می شوند[12][11].

همانطور که در جدول۳ هم مشاهده می شود، تمامی سطوح RAID مورد بررسی برای عملیات خواندن کوتاه پهنای باند **N * B** را فراهم میاورند چرا که تعداد **N** عملیات داده ای Small Read قادرند بطور همزمان روی بلاکهای **N** دیسک مختلف انجام گیرند.برای عملیات Large Read نیز در RAID 1 بدلیل استفاده از تکنیک Mirroring داده ها وکپی آنها روی بلاکهای متوالی دیسک قرار داده میشوند واین شرایطی را فراهم میاورد تا هر پردازه که عملیات Large Read را انجام میدهد بتواند بلاکهای داده ای متوالی را بطور یک در میان از روی دیسک اصلی وپشتیبان بطور موازی وهمزمان بخواند وبدین ترتیب بهنگام اجراء **N / 2** پردازهٔ موازی روی دیسکهای مختلف از پهنای باند نیمی از دیسکهای دنباله **N** دیسکی استفادهٔ کامل میگردد . اما در سایر موارد بدلیل استفاده از تکنیک Stripping به مدل امکان انجام عملیات داده ای موازی به مقدار قابل توجهی افزایش میابد .

در RAID 1 انجام هر عملیات Large Write بدلیل آنکه داده ها ی منطقا متوالی روی بلاکهای فیزیکی متوالی هر دیسک گسترده شده اند تنها دو دیسک را درگیر کرده وپهنای باند **B * 2** را مورد استفاده قرار میدهد. اما برای اجراء عملیات Large Write پردازه های موازی که هر یک بر روی یک دیسک از دنباله اثر میکنند پهنای باند سیستم تا **(N / 2)*B** نیز افزایش میابد. بغیر از مدل RAID 5 که بدلیل Stripping در سطح بلاکی واستفاده از یک بلاک توازن در مقابل **N - 1** دیسک برای نگهداری اطلاعات داده ای ، حداکثر پهنای باند **(N – 1) * B** میباشد سایر مدلها پهنای باندی همچون RAID 1 خواهند داشت [5][2][1] .

۴-۲- هزینهٔ زمانی عملیات ورودی/ خروجی[14]

دراین مرحله به ارزیابی هزینه های زمانی اجراء عملیات ورودی / خروجی روی دنبالهٔ دیسکی پرداخته وهمچون مراحل قبلی هزینهٔ مدلهای گوناگون را بازاء حالات مختلف عملیات دیسکی مورد بررسی قرار میدهیم . تمامی سطوح مورد بررسی برای انجام عملیات Small Read بایستی هزینهٔ زمانی **R** را بپردازند.اما در صورتیکه بخواهیم یک فایل یا یک دنبالهٔ داده ای بطول **M** را بر روی دنبالهٔ دیسکی توزیع شده درج نماییم در مواردی که بلاکهای داده ا ی بر روی دنبالهٔ دیسکی Stripe شده باشند بسادگی از فرمول زیر برای تعیین هزینهٔ زمانی مدل برای عملیات Large Read می توان استفاده کرد .

$$(L_{Data} * T_R) / (N_{Total\ Disk} - N_{Check\ Disk})$$

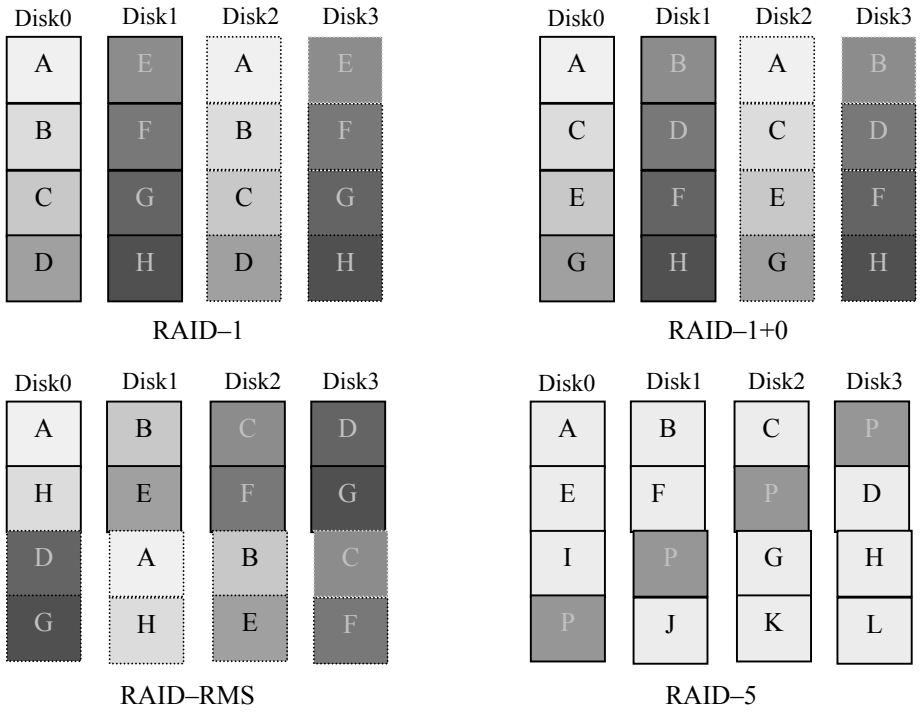
$N_{Total\ Disks}$: تعداد کل دیسکهای دنباله $N_{Check\ Disks}$: تعداد دیسکهای حاوی اطلاعات اقزوده

L_{Data} : طول فایل یا دنبالهٔ داده ای برحسب بلاک T_R: هزینهٔ زمانی متوسط عملیات خواندن یک بلاک

در این میان تنها مورد استثناء RAID سطح یک است [8]که با استفاده از تکنیک Mirroring بلاکهای داده ای منطقا متوالی را بر روی بلاکهای متوالی فیزیکی بصورت دوگانه قرار می دهد. در این روش امکان ذخیره سازی موازی بلاکهای متوالی بلحاظ ساختار ذخیره سازی موجود فراهم نمیباشد و تمامی یک دادهٔ **M** بلاکی بطور پی در پی در یک دیسک ذخیره ونگهداری می شود وبهمین لحاظ هزینهٔ زمانی عملیات متوالی خواندن یک دادهٔ **M** بلاکی بصورت **M*R** محاسبه میگردد.

---

[14] Input / Output Cost



تصویر۳ . ساختار توزیع داده ای در مدل پیشنهادی وسایرمدلهای مشابه

۴-۳- هزینهٔ عملیات احیا، داده ای[۱۵]

همانگونه که در بخشهای ابتدایی نیز بدان اشاره گردید، مزیتهای مدل پیشنهادی را نسبت به سایر مدلهای مشابه می توان در مواردی همچون ، تحمل پذیری بالاتر نسبت به خطا ، صحت وحفاظت داده ای بیشتر ، توازی گرایی بالاتر بهمراه حداکثر پهنای باند موثر بهنگام انجام عملیاتهای دستیابی ویا احیاء داده ای خلاصه کرد. دراین بخش ما قصد داریم مدلهای مورد نظر را در موارد فوق بررسی وبا یکدیگر مقایسه نماییم[16][14][13].

به لحاظ صحت و حفاظت داده ای ، مدل پیشنهادی ما از هر سه مدل دیگر محیط امن تری را برای داده های درون سیستم فراهم میاورد . چرا که در دو مدل RAID 1 و RAID 1+0 بهنگام بروز خرابی در بیش از یک دیسک ، بدلیل آنکه تمامی بلاکهای داده ای پشتیبان دیسک مزبور دقیقا بر روی یک دیسک دیگر متمرکز شده اند بنابر این ، در یک سیستم دنبالهٔ دیسکی با Nدیسک ، با احتمال $1/(N-1)$ داده های از دست رفته قابل بازگرداندن نمیباشد. این در حالی است که در این دسته از سیستم ها در برخی از حالات خاص نیز حتی تا $N/2$خطای دیسکی نیز قابل پوشش بوده وداده های از دست رفته قابل احیاء میباشد. در مدل RAID 5 نیز بهنگام بروز خرابی در یک دیسک، تشخیص خطا ویا تعیین موقعیت خطاهای بروز کرده تا حدود زیادی دشوار وگاها در مواردی همچون تعداد خطاهای زوج غیر ممکن است ودر واقع این مدل نسبت به دو مدل قبلی نیز امنیت کمتری دارد . اما در مدل پیشنهادی بجهت استفاده از نوعی روش انتقال بلاکی خاص که با دو تکنیک انعکاس داده ای وریفی کردن داده ای همراه شده ، داده ها بگونه ای بر روی دنبالهٔ دیسکی توزیع می شوند که بهنگام بروز هر نوع خطا ی داده ای ودیسکی در سطح دنباله کمترین داده از دست خواهد رفت . نکتهٔ جالب در این روش آنست که بدلیل توزیع مورب نیمه های دیسکی اگر خطا های بروز کرده فقط بر روی دیسکهای با شمارهٔ زوج ویا فقط بر روی دیسکهای با شمارهٔ فرد گسترده شده باشند، هرنوع وهر تعداد از خرابیها و خطا های دیسکی(حتی تا $N/2$ خطا نیز) قابل بازگرداندن بوده وموجب از دست رفتن هیچ داده ای نمی شود. در این روش نرخ احتمال خطای غیر قابل برگشت کاهش میابد و با احتمال $1/((N-1)*(N-2))$ داده های از دست رفتهٔ یک دیسک قابل احیا، نمیباشد (به [4] مراجعه گردد) .

[15]Recovery Cost

در یکی از بخشهای قبلی به ارزیابی پهنای باند سیستم پرداخته وحداکثر پهنای باند موثر را درحالتی که سیستم عملکردی عادی دارد(حالتی که خطا در سیستم بروز نکرده) بازاء عملیاتهای داده ای کوتاه وطولانی برای مدلهای مختلف محاسبه نموده ایم . اما آنچه را که ما دراین قسمت مورد توجه قرار میدهیم، حداکثر پهنای باندی است که سیستم در هنگام بروز خطا بطور مؤثر برای پردازه ها فراهم میاورد. ودر خلال این بررسی به دو سؤال زیرپاسخ میدهیم.

۱. بهنگام بروز خرابی در یک یا عده ای از دیسکهای دنباله، مدل بکار رفته قادر است تا چه مقدار از پهنای باندسایردیسکهای دنباله را برای انجام عملیات دیسکی مورد استفاده قراردهد؟

۲. بهنگام بروز خرابی چه مقدار از پهنای باند مؤثر سیستم برای انجام عملیات احیاء دسترس پذیر است؟

در مدل RAID 1 ، در شرایط عادی، هر پردازه تنها قادر است برای انجام عملیات Large Read تنها از پهنای باند دودیسک استفاده نماید که این در مقابل مدل پیشنهادی ما بسیار ضعیف است اما در چنین شرایطی RAID 1+ 0 پهنای باندی مشابه مدل پیشنهادی ما دارد . بهنگام بروز خطای منفرد پهنای باند مؤثری که برای انجام عملیات Small Read مورد استفاده قرار میگیرد در تمامی مدلها مشابه یکدیگر وبرابر است با :

(N – 1) * B

برای عملیات خواندن فایلها بزرگ ویا داده ها ی طولانی پهنای باند مدل پیشنهادی ما نتیجهٔ جالب توجهی را ارایه کرده وپهنای باندی حتی بیش از مدل RAID 5 را فراهم میاورد. نتایج بررسی ومقایسهٔ مدلهای مختلف در جدول ۳ آورده شده است.

پس از بروز هر خرابی در یک سیستم دنبالهٔ دیسکی بلافاصله بایستی داده هایی که برروی دیسک مزبور قرار داشته اند احیاء وتا هنگام تصحیح ویا جایگزینی دیسک خراب بر روی یک دیسک یدکی نگهداری شود. در خلال عملیات احیاء داده ای ، انجام عملیات عادی سیستم متوقف میگردد. بنابر این هر مدلی که مدت زمان کوتاه تری را صرف انجام عملیات احیاء نماید مناسب تر بوده وبازدهی بالاتری را برای سیستم فراهم میاورد. در تکنیک RAID 1+ 0 , RAID 1 بهنگام عملیات احیاء دیسکی تنها از پهنای باند یک دیسک استفاده میگردد وعملیات احیاء بصورت سریال انجام میگیرد[9][7][6]. اما در RAID – RMS بدلیل نحوهٔ خاص توزیع داده ها بر روی دنبالهٔ دیسکی برای عملیات احیاء **N – 1** دیسک باقیمانده قادرند بطور موازی وهمزمان داده های دیسک خراب را احیاء نمایند[18][17][10].

جدول۳.جدول مقایسهٔ هزینه های مدل پیشنهادی وسایر مدلهای مشابه

	RAID Level	**Small Read**	**Large Read**	**Small Write**	**Large Write**
Max.I/O Bandwidth	**RAID 1**	*N * B*	*(N / 2)* B or 2*B*	*(N / 2) * B*	*(N / 2)* B or 2*B*
	RAID 1+0	*N * B*	*N * B*	*(N / 2) * B*	*(N / 2) * B*
	RAID 5	*N * B*	*(N – 1) * B*	*(N / 2) * B*	*(N – 1) * B*
Read / write Time (parallel)	**RAID 1**	*R*	*M * R*	*W*	*M * W*
	RAID 1+ 0	*R*	*M / N * R*	*W*	*M * W / (N / 2)*
	RAID 5	*R*	*M * R / (N – 1)*	*R + W*	*M * (R + W) / (N – 1)*
	RAID-RMS	*R*	*M * R / N*	*W*	*M * W / (N / 2)*
Recovery Max.I/O Bandwidth	**RAID 1**	*(N – 1) * B*	*B*		
	RAID 1+0	*(N – 1) * B*	*(N / 2) * B*		
	RAID 5	*(N – 1) * B*	*(N – 2) * B*		
	RAID-RMS	*(N – 1) * B*	*(N – 1) * B*		

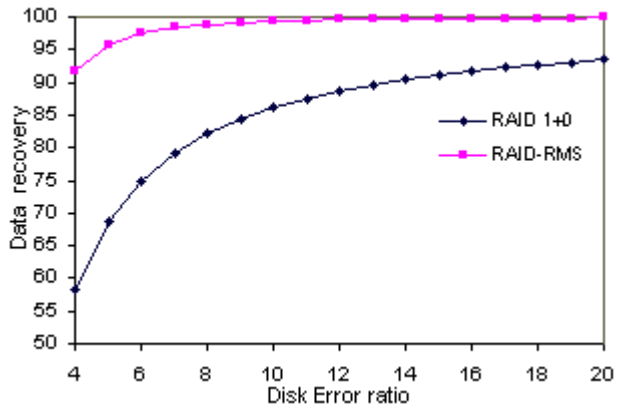
۵- شبیه سازی مدل پیشنهادی

در این فرایند شبیه سازی بمنظور ارزیابی دقیقتر نتایج و داده های حاصل از بررسیهای تحلیلی انجام شده بر روی مدل پیشنهادی و سایر مدلهای مشابه ، و مطالعهٔ نحوهٔ عملکرد و رفتار واقعی مدلهای تحت بررسی در مقابل تغییرات پارا مترهای مختلف ، مدل پیشنهادی خود را بهمراه تکنیک RAID 1+ 0 , RAID 1 در یک محیط برنامه سازی سطح بالا شبیه سازی نموده ایم . بدلیل تشابه RAID 1+ 0 , RAID 1 و آنکه در پاره ای از موارد RAID 1+0 به لحاظ احیا، داده ای نتایج بهتری را ارایه میدهد، در شبیه سازی از این مدل استفاده گردیده است. مدلهای تحت بررسی را در معرض گونه های مختلف خطا های داده ای در سطح بلاکی ودیسکی قرار داده ، رفتارها ، نحوهٔ عملکرد ، میزان حفاظت و امنیت داده ای و تحمل پدیری مدلها نسبت به گونه ها ی مختلف خطا بررسی ونتایج حاصل ، در قالب نمودارهایی ارایه گردیده است .

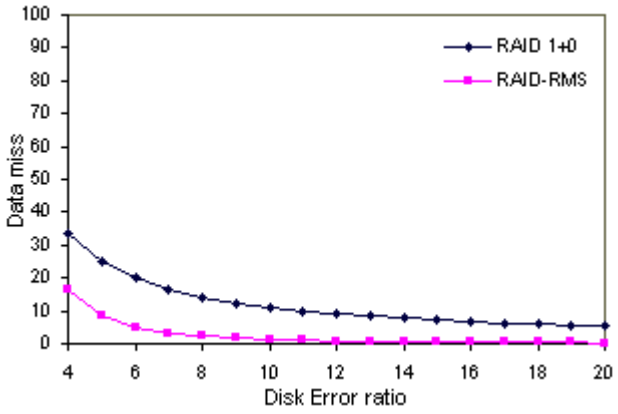
در مرحلهٔ نخست هر دو مدل را در معرض گونه های مختلفی از خطا ها قرار داده و پاسخ سیستم را بـرای دامنـه ای از خطـا هـای بروز کرده در سیستم مشخص نموده و بمنظور افزایش سطح دقت در تعیین نتایج، انجام آزمون را بـرای هـر مـدل تـا ۱۰۰۰۰ بـار تکرار نموده ایم . در حقیقیت در این مرحله ازآزمون قصد داریم مشخص نماییم که هر مدل قادر است تا چند درصد از خطا هـای بروز کرده را بطور کامل احیا، نماید.

همانگونه که در نمودار ۱ هم مشاهده می شود ، مدل پیشنهادی در شرایط یکسان (بازا، تعداد خطا و دیسک) نرخ احیا، داده ای بسیار بالاتری را داراست . هر دو مدل در مقابل تغیرات افزایشی تعداد دیسکهای درون دنبالـه تغـیرات افزایشـی مشـابهی را از خود نشان میدهند . با توجه به نمودار۱ میتوان دریافت که مدل RAID1+0 بازا، تعداد دیسکهای بیشتر درون دنباله نرخ احیا، بسیار خوبی را فراهم میاورد. اما مدل پیشنهادی بازا، تعداد دیسکهای کمتر نیز مؤفق عمل مینماید .

در مرحلهٔ بعد رفتار ونحوهٔ عملکرد هر دو مدل را در مقابل نرخ تغیرات تعداد مشابه خطـا و تعـداد دیسـکهای درون دنبالـه آزموده و نتایج حاصله را در نمودار ۲ ار ایه کرده ایم . با کمی دقت میتوان گفـت کـه در مـدل پیشـنهادی در بدتریـن حـالت وبـا بیشترین نسبت تعداد خطا به دیسک درون دنباله کمتر از ۲۰٪ داده از دست خواهند رفت . اما در هر دو مدل با افزایش نسبتعداد دیسک در مقابل تعداد خطا های یکسان ، تغیرات کاهشی مشابهی را از خود نشان میدهند ونرخ داده های از دسـت رفتـه بمقـدار قابل توجهی کاهش خواهد داشت . با بررسی نتایج حاصل از شبیه سازی مدل پیشنهادی و انطباق آنها با نتایج تحلی که در مراحل قبلی در قالب فرمولهایی ارایه شده می توان گفت مدل پیشنهادی در شرایط یکسان ودر مقایسه با سایر مدلهای مشابه نـرخ احیـا، داده ای بسیار بالاتر ی را فراهم میاورد.



نمودار ۱. مقایسهٔ نرخ احیاء داده ای



نمودار۲. مقایسهٔ میزان فقدان داده

۶- نتیجه گیری

با توجه به مطالعات وبررسیهای انجام شده و بر اساس نتایج تحلیلی بدست آمده در بخشهای قبل ، میتوان گفـت کـه سـایر مدلهای مشابه همچون RAID 1, RAID 1+ 0 با آنکه هزینه هایی مشابه با مدل پیشنهادی را به سیستم تحمیل مینمایند ، امـا قـادر به قراهم آوردن پهنای باند مناسبی بمنظور افزایش سطح دسترس پذیری خصوصا بهنگام عملیات احیا، داده ای نبوده ویا بـهنگام بروز خطا در سطح دنبالهٔ دیسکی، بخوبی نمی توانند صحت وامنیت داده ای را درون سیسـتم حفـظ نمـایند. مزیـت عمـدهٔ مـدل پیشنهادی ما آنست که بدون هیچگونه هزینهٔ افزوده وتنها بکمک الگوریتمـهای خـاص توزیـع داده ای در سـطح دنبالـه ، موجـب افزایش سطح اجراء عملیات داده ای موازی ، بهبود تحمل پذیری نسبت به خطا و ایجاد پهنای باند بالاتر در سیستم میگردد. علاوه بر این ، مدل RAID-RMS ، الگوی توزیع داده ای خاصی را پیاده سازی می کند که قـادر اسـت ، نـرخ احیـاء داده ای [16]را بگونـه ای افزایش دهد که ، بهنگام بروز هر خرابی در سطح دنباله دیسکی ، احتمال از دست رفتن یک داده بمقدار قابل توجهی کاهش پیـدا

[16]Data Recovery

کند . حفظ پهنای باند مؤثر در سطح سیستم بهنگام بروز خطا ، یکی دیگر از ویژگیهای بارز این مدل بشمار مـی رود کـه موجـب افزایش سطح دسترس پذیری داده ای و در نتیجه تسریع انجام عملیات احیاء داده ای میگردد.

مراجع

[1] P. M. Chen, E. K. Lee, G. A. Gibson, R. H. Katz and D. A. Patterson; "RAID: High-Performance,Reliable Secondary Storage", ACM Computing Surveys, Vol.26, No.2, June 1994, pp.145-185.

[2] G. Gibson, D. Nagle, K. Amiri, F. Chang, H. Gobioff, E. Riedel, D. Rochberg and J. Zelenka, "A Cost-effective, High-bandwidth Storage Architecture", Proc. of the 8th Conf. on Architectural Support for Programming Languages and Operating Systems, 1998,pp.97-106.

[3] J.Akbari and A.T.Haghighat . "A New Redundancy Algorithm For Distributed Environment", Operating system & Security Conference-OSSC 2003, Sharif university of technology, 24 -25 Dec,pp.70-81.

[4] T. Cortes, "Software RAID and Parallel Filesystems", in High Performance Cluster Computing–Architectures and Systems, Rajkumar Buyya (ed.), Prentice Hall PTR, 1999, pp.463-496.

[5] Michael Stonebraker ,Gerhard A. Schloss. "Distributed Raid – A New Multiple Copy Algorithm" , University of California,Berkeley, CA 94720,2000

[6] T. Anderson, M. Dahlin,D. Patterson,and R. Wang. "Serverless Network FileSystems", ACM Trans.on Computer Systems, Jan. 1996, pp.41-79.

[7] S. Asami, N. Talagala, and D. A. Patterson, "Designing a self-maintaining storage system", Proceedings of 16th IEEE Symposium on Mass Storage Systems, March 1999, pp. 222-233.

[8] L. F. Cabrera, and D. E. Long, "Using Distributed Disk Striping to Provide High I/O Data Rates", Proceedings of USENIX Computing Systems, Fall 1991, pp.405-433.

[9] P. Cao, S. B. Lim, S. Venkataraman, and J. Wilkes, "The TickerTAIP Parallel RAID Architecture", ACM Trans. on Computer System, Vol.12, No.3, August 1994, pp.236-269.

[10] P. F. Corbett, D. G. Feitelson, J.-P. Prost, and S. J. Baylor. "Parallel Access to Files in the Vesta File System". Proceedings of Supercomputing'93, 1993.

[11] T. H. Cormen and D. Kotz, "Integrating Theory and Practice in Parallel File Systems",Proceedings of DAGS '93 Symposium, June 1993, pp. 64-74.

[12] I. Foster, D. Kohr, Jr., R. Krishnaiyer, and J. Mogill, "Remote I/O: Fast Access to Distant Storage".Proc. of the Fifth Workshop on I/O in Parallel and Distributed Systems, November 1997, pp.14-25.

[13] M. Harry, J. M. del Rosario, and A. Choudhary, "VIP-FS: a VIrtual, Parallel File System for High Performance Parallel and Distributed Computing", Proceedings of the 9th International Parallel Processing Symposium (IPPS'95), April 1995, pp. 159-164.

[14] R. S. Ho, K. Hwang, and H. Jin, "Design and Analysis of Clusters with Single I/O Space",Proceedings of 20th International Conference on Distributed Computing Systems (ICDCS 2000), April 2000, Taiwan, pp.120-127.

[15] J. H. Howard, M. L. Kazar, S. G. Menees, D. A. Nichols, M. Satyanarayanan, R. N. Sidebotham, and M. J. West, "Scale and Performance in a Distributed File System". ACM Trans. on Computer System, Vol.6, No.1, pp.51-81, February 1988.

[16] H. I. Hsiao and D. DeWitt, "Chained Declustering: A New Availability Strategy for Multiprocessor Database Machines", Proc. of 6 th International Data Engineering Conf., 1990, pp.456-465.

[17] Y. Hu, Q. Yang and T. Nightingale. "RAPID-Cache --- A Reliable and Inexpensive Write Cache for Disk I/O Systems", Proceedings of the 5th International Symposium on High Performance Computer Architecture (HPCA-5), Orlando, Florida, Jan. 1999, pp. 204 – 213.

[18] J. Huber, C. L. Elford, D. A. Reed, A. A. Chien, and D. S. Blumenthal, "PPFS: A High Performance Portable Parallel File System", Proceedings of the 9th ACM International Conference on Supercomputing, Barcelona, July 1995, pp.385-394.