

ارایه تکنیک افزونگی نوین جهت افزایش حفاظت داده ای در سیستمهای با نیاز امنیتی بالا

محمد رضا میبدی^۱

دانشکده مهندسی کامپیوتروفناوری اطلاعات، دانشگاه صنعتی امیر کبیر
meybodi@ce.aut.ac.ir

جواد اکبری ترکستانی^۲

دانشکده مهندسی کامپیوتو، دانشگاه آزاد اسلامی اراک
akbari@jdmarkazi.ac.ir

چکیده

در این مقاله سعی داریم تکنیک ذخیره داده ای نوینی ارایه دهیم که قادر است تنها با بکارگیری کمی اطلاعات افزوده بیشتر، امنیت و حفاظت داده ای بالاتر، بازدهی و پنهانی باند بیشتر و دسترس پذیری بهتری را ارایه دهد. اغلب الگوریتم های مشابه قبلی علاوه بر هزینه افزوده فضایی نسبتاً بالا، در برخی از موارد قادر به تشخیص ویا تصحیح خطاهای بروز کرده در سطح سیستم نبوده و یا گاهای بخوبی قادر به احیاء داده از دست رفته نمی باشند. لازم بذکر است که این تکنیک قادر است تا در مواجه با بروز خطأ در سطح دیسکهای ذباله یا بلکهای داده ای سیستم، علاوه بر تشخیص دقیق تر خطأ، با قابلیت اطمینان بالاتری موقعیت آثرا شناسایی و بدون نیاز به بکارگیری بلکهای داده ای متناظر در سایر دیسکهای ذباله آنرا تصحیح نماید. تا در مواردی همچون بروز خطأ ویا بهنگام احیاء داده ای نیز سیستم از دسترس پذیری و توازنی بالاتری در سطح داده ای بر خوردار باشد.

واژه های کلیدی : افزونگی توزیع شده- حفاظت داده ای - تشخیص خطأ - RADD – Redundancy

۱- مقدمه :

سطوح مختلف RAID هر یک با بکارگیری الگوریتم ذخیره سازی و هزینه افزوده خاص خود ، سطوح مختلفی از قابلیت اطمینان ، تحمل پذیری نسبت به خطأ و بازدهی را برای کاربر فراهم می نماید. تشریح دقیق سطوح مختلف تکنیک RAID در [2] آمده است. با بررسی ویژگیها و ساختارهای خاص سطح پنج از این تکنیک می توان دریافت که این مدل یکی از بهترین و مناسب ترین مدلها برای پیاده سازی در سیستم های مرکز توزیع شده است(البته نسبت به سایر مدلها دیگر). در این مدل داده ها و اطلاعات افزوده را بمنظور کاهش گلوگاه عملیات دیسکی(عملیات خواندن و نوشتن داده ای) ، بر روی ذباله دیسکی بصورت افقی توزیع مینماییم [4]. این ویژگی در پیاده سازی بر روی سیستم های توزیع شده این امکان را میدهد که سیستم بهنگام دستیابی داده ای و یا بررسی بروز خطأ از توازن گرایی بالاتر و داده ها از دسترس پذیری بیشتری برخوردار باشند. این در حالیست که ما در سایر مدلها مجبوریم تا اطلاعات افزوده را بر روی یک سایت منفرد(دیسک منفرد) مرکز نماییم و بجهت آنکه هر گونه عملیات نوشتن ویا بروزسازی داده ای نیاز به دستیابی و بروز کردن اطلاعات داده ای سایت حاوی اطلاعات افزوده دارد، از این روی این سایت مرکز در واقع در حکم یک نقطه شکست برای سیستم و یک گلوگاه برای عملیات دیسکی(عملیات ورودی / خروجی) بشمار می رود. چرا که با از کار افتادن سایت مزبور امکان بررسی و تشخیص خطأ های داده ای تا هنگام باز سازی مجدد سایت در سراسر سیستم از بین می رود. اما با این نحوه توزیع داده های افزوده که ما نیز از آن برای پیاده سازی مدل پیشنهادیمان در سیستم های

^۱ عضو هیات علمی دانشکده مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد اراک

^۲ عضو هیات علمی دانشکده مهندسی کامپیوترو فناوری اطلاعات، دانشگاه صنعتی امیر کبیر

توزيع شده استفاده میبریم هم امکان انجام عملیات ورودی / خروجی همزمان و موازی برای سایتهاي مختلف فراهم میايد که از ضروريات در يك سیستم توزيع شده بشمار میرود وهم آنکه سیستم كمتر دچار ازدحام گشته وعلاوه برآن نقطه شکست سیستم نيز مرتفع میگردد.از اين روی بازدهی وکارايی سیستم بمقدار قابل توجهی افزایش یافته ودر عین حال داده های سیستم دسترس پذير تر و قابلیت اطمینان سیستم هم افزایش میابد. با اين همه در اين تکنيک نيز ، بهنگام برسی خطاهای داده ای در يك بلاک از ديسک ، تمامی بلاکهای اطلاعات داده ای وافزوده همردیف (که در يك stripe واحد قرار دارند)در سایر ديسکها (از سایت محلی آن یا سایت های دیگر) نيز بايستی بطور همزمان در عملیات سراسری که تحت نظرلت نرم افزار سراسری (کنترلر نرم افزاری RAID) صورت می پذيرد، شرکت نمایند. ودر واقع در حین اجراء عملیات سراسری برسی داده ای هیچگونه عملیات بروز رسانی در هیچ یک از بلاکهای داده ای مزبور نمی تواند صورت گیرد(عملیات تحت پروتکل قفل دو مرحله ای صورت گرفته وپردازه های حاوی دستورات برخوردار پی در پی می شوند) وپردازه های مذکور بايستی تا پایان عملیات سراسری منتظر باقی بمانند^[7].[5] اين ویژگی بعنوان يك نقطه ضعف تلقی شده و مانع از انجام عملیات موازی دستیابی به داده های سیستم میگردد. علاوه بر آنکه عملیات فوق تمامی ديسکهاي دنباله وبا سایتها را درگير نموده وموجب اتلاف وقت سایر سایتها ، به تعویق افتادن برنامه ها وپردازه های با اولویت و افت کارايی و بازدهی سیستم میگردد.از اين روی ما سعی داريم تا در مدل پیشنهادي خود شرایط و امکاناتي را فراهم آوريم تا هر سایت خود بتواند در صورت لزوم، بصورت محلی وبطور مستقل از سایر ديسکها اطلاعات وداده های محلی خود را برسی و خطاهای بروز کرده را شناسایي نماید. بدین ترتیب ما قادریم حتی دیسکی را که در آن داده ها دچار خطأ شده اند را با يك برسی در داده های محلی آن دیسک شناسایي نماییم در حالی که در روشهای قبلی عملیات سراسری برسی شرایط خطاهای داده ای بدلیل آنکه از روی اطلاعات افزوذه افقی صورت میگرفت قادر به تعیین موقعیت دقیق بروز خطأ نبود. ما با تلفیق این دو امکان علاوه بر آنکه میتوانیم دقت تشخیص خطاهای داده ای را افزایش دهیم ، قادریم تا خطاهای داده ای بروز کرده را بصورت محلی وبدون دخالت سایر ديسکهاي دنباله وبا حفظ سطح توازی گرایی(اجراء موازی پردازه های مستقل) شناسایي وتصحیح نماییم^[6][1]. در ادامه این مقاله به بحثهای زیر پرداخته ایم . در بخش ۲ به تشریح مدل پیشنهادي خود پرداخته ایم . در بخش ۳ مدل ارایه شده را با سایر مدلهاي مشابه از جنبه های مختلفی همچون هزینه فضایي و زمانی مدل ، حداکثر پهنهای باند ، سطح توازی سیستم و هزینه احیاء داده ای مورد مقایسه وبررسی قرار داده ایم. در بخش ۴ نيز با ارایه نتایج بدست آمده از شبیه سازی مدل به ارزیابی و مقایسه آن با RAID 5 پرداخته ایم . در بخش ۵ نيز با ارایه يك نتیجه گیری به مقاله خاتمه داده ایم .

۲ - مدل پیشنهادي

با توجه به مطالب گفته شده سعی ما بر آنست تا مدلی را ارایه دهیم که حتی الامکان بتواند نیازهای اساسی مطرح در سیستم های توزيع شده را برآورده سازد. بهمین لحاظ مدل پیشنهادي ما با اعمال تغیيراتی درساختار سطح پنج از تکنيک RAID پایه ریزی شده است. اين اصلاح ساختاري برای پیاده سازی روی محیطهای محاسباتی توزيع شده جهت نیاز به قابلیت اطمینان³ بالا مناسب است^{[16][12][4]}[3]. علاوه بر اين فراهم آوردن دسترس پذيری بالا خصوصاً بهنگام بروز خطاهای داده ای، از ویژگیهای منحصر بفرد اين مدل می باشد . همانطور که در تصویر ۱ نيز هم مشاهده میگردد، مدل RAID پیشنهادي ما از ردیفي کردن داده ای در سطح بلاکي استفاده می نماید. همچنین بمنظور کاهش گلوگاه نوشتن بر روی ديسک توازن وعملیات بروز رسانی سایتها ، اطلاعات توازن افزوذه میان سایتهاي دنباله توزيع گردیده است. ما اين بلاکهای توازن توزيع شده را بلاکهای توازن افقی⁴ مینامييم. همانطور که در تصویر هم مشاهده می شود علاوه بر بلاکهای توازن افقی ، در هر ديسک بلاکهای توافق دیگري بطور اختصاصي برای هر ديسک موجود است. اين بلاکهای توازن که ما آنها را بلاکهای توافق عمودی⁵ مینامييم برای هر ديسک منحصر بفرد بوده ، بروی همان ديسک قرار دارد و تنها روي بلاکهای داده ای موجود درهمان ديسک تولید می گردد. اين بلاک توازن عمودی معمولاً در آخرین بلاک از هر ديسک ذخیره میگردد چرا که توزيع آن در سطح بلاکهای داده ای باعث نیاز به بروز کردن افزوذه در سطح بلاکهای توافق میگردد. البته ناگفته نماند که برای حل اين معزل می توان بلاکهای توافق افقی رانیز تنها از

³Reliability

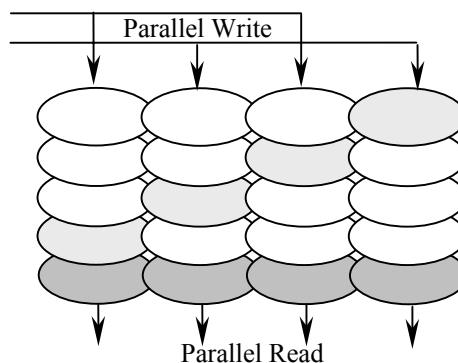
⁴Horizontal Parity Block

⁵Vertical Parity Block

سطح بلاکهای توازن افقی میگردد. البته ناگفته نماند که برای حل این معزل می توان بلاکهای توازن افقی رانیز تنها از روی بلاکهای داده ای متناظر در سایر دیسکها تولید کرد. بدین ترتیب هر عمل نوشتن نیاز به تنها ۲ عمل بروز سازی در سطح بلاکهای توازن عمودی و افقی دارد که بلاک نوشته شده محل تلاقي آندو می باشد. مزیت این روش نسبت به مدلهای پیشین آنست که بروز خطاها ای بلاکی بطور محلی تشخیص داده شده و با سرعت و دقیق بیشتری احیاء میگردد. نکته قابل توجه آنست که عمل احیاء بطور محلی صورت میگیرد و بهمین جهت بمیزان قابل توجهی نرخ پردازش همزمان عملیات /O I را بهنگام عملیات غیر معمول و بروز خطا افزایش میدهد. بهمین دلیل دسترس پذیری⁶ بیشتری را برای کاربر فراهم می نماید. از این رومدل مناسبی برای پیاده سازی بر روی محیطهای محاسباتی توزیع شده می باشد. چرا که پردازش موازی عملیات /O I ، دسترس پذیری و قابلیت اطمینان بیشتر ناشی از افزونگی عمودی از نیازمندیهای حیاتی یک محیط توزیع شده بشمار میروند[9][11].

۳- مقایسه وارزیابی سطوح RAID

در این بخش ما به بررسی ، مقایسه و ارزیابی مدلها ای مختلف تکنیک RAID پرداخته و آنها را از جنبه های گوناگونی همچون هزینه فضایی افزوده مورد نیاز بمنظور تأمین امنیت و حفاظت داده ای مدل ، هزینه عملیات ورودی / خروجی مورد نیاز برای دستیابی به داده ها ، امکان کنترل همزمانی و هزینه احیاء خرابی دیسک و داده مورد ارزیابی قرار داده و در هر مورد مزایا و معایب مدل مورد بررسی را ارایه و در صورت امکان راه حل هایی را هم برای مرفوع ساختن ناکارامدیهای مدل در مواردی خاص ارایه میدهیم.



تصویر ۱ . مدل افزونگی افقی و عمودی(HVRAID)

۱-۳ - حداکثر پهنای باند

یکی از نتایج ارزشمندی که بواسطه 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. در این بخش ما حداکثر پهنای باند مؤثری را برای هر یک از مدلها محاسبه نموده و در نهایت آنها را با یکدیگر مقایسه مینماییم. برخی از نمادهای مورد نیاز در جدول ۱ آمده است .

جدول ۱. جدول نمادها

نماد	توضیحات
N	تعداد دیسک در یک سیستم RAID توزیع شده
B	حداکثر پهنای باند هر دیسک
S	اندازه هر بلاک از دیسک
R	متوسط زمان خواندن هر بلاک از دیسک
W	متوسط زمان نوشتن هر بلاک از دیسک
C _d	تعداد دیسکهای وارسی (Check block)
M	تعداد بلاکهای موجود در هر فایل

مفاهیم Large Read / Write و Small Read / Write 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 بر روی تمام یا قسمتی از یک

⁶Accessibility

بلاک داده ای تاثیر میگذارند و عملیاتهای موازی ورودی / خروجی از این نوع ، می توانند روی بلاکهای متفاوتی از یک Stripe منفرد یا بر بلاکهای متفاوت یا یکسان از Stripe های متفاوت اعمال شوند[3][2]. که هر یک از انواع فوق نیز در هر یک از تکنیکهای مورد بحث ما نتایج نتفاوتی را در پی دارد. با این همه دامنه عملیاتی آنها از یک بلاک فراتر نمی رود. واما منظور از خواندن ونوشتن های طولانی آن دسته از عملیات دیسکی را شامل می شود که در آن داده ها بر روی چندین بلاک متولی از یک یا چند Stripe متولی قرار داده شده وبا از آنها خوانده می شود[8].

لازم بذکر است که می توان به موارد خاصی اشاره کرد که پهنانی باند فراهم شده در سیستم مورد مطالعه از نتایج بدست آمده کمتر است چرا که ما حداکثر پهنانی باند را برای سیستم ها محاسبه نموده ایم . همانطور که در جدول ۲ هم مشاهده می شود، بغیر از سطوح RAID مورد بررسی برای عملیات خواندن کوتاه پهنانی باند B^* N را فراهم میاورند.لازم به توضیح است که در سطح ۲ تعداد C_d دیسک برای نگهداری کدهای همینگ ودر سطح ۴ نیز یک دیسک منمرکز برای نگهداری اطلاعات توازن افزوده از سایر دیسکها متمایز گردیده است و بهمین لحاظ پهنانی باند عملیات موازی کاهش می یابد.برای عملیات Large Read نیز در سطح صفر بدلیل استفاده از تکنیک ردیفی کردن داده ای پهنانی باند تمامی N دیسک می تواند مورد استفاده واقع شود. اما در سطح یک از این تکنیک، بهنگام اجراء $2/N$ پردازه موازی روی دیسکهای مختلف از پهنانی باند تمامی دیسکهای دنباله N دیسکی استفاده کامل میگردد[10].در سایر موارد برای انجام عملیات Large Read بدلیل آنکه در همگی مدلها از تکنیک Stripping بمنظور توزیع داده ای در سطح دیسکهای دنباله استفاده گردیده است می توان از فرمول کلی زیر استفاده کرد.

$$(N_{Total Disk} - N_{Check Disk}) * B_d \quad B_d : \text{پهنانی باند پایه دیسک} \quad N_{Check Disks} : \text{تعداد دیسکهای حاوی اطلاعات کنترلی} \quad N_{Total Disks} : \text{تعداد کل دیسکهای دنباله}$$

جدول ۲. حداکثر پهنانی باند عملیات دستیابی داده ای

Max.I/O Bandwidth	Small Read	Large Read	Small Write	Large Write
RAID 0	$N * B$	$N * B$	$N * B$	$N * B$
RAID 1	$N * B$	$N * B$	$(N/2) * B$	$(N/2) * B$
RAID 2	$(N - C_d) * B$			
RAID 4	$(N - I) * B$	$(N - I) * B$	B	$(N - I) * B$
RAID 5	$N * B$	$(N - I) * B$	$(N * B) / 2$	$(N - I) * B$
HVRAID	$N * B$	$(N - I) * B$	$(N * B) / 2$	$(N - I) * B$

برای انجام عملیات Small Write روی دنباله دیسکی در سطح یک از آنجا که بهنگام نوشتن هر داده یک کپی از آن نیز روی دیسک پشتیبان متناظر ذخیره می گردد ، از این روی پهنانی باند مؤثرسیستم به نصف تقلیل پیدا می کند . نکته جالب توجه در مورد سطح چهار آنست که بدلیل مرکز بودن اطلاعات کنترلی افزوده دریک دیسک منفرد ، این دیسک به گلوگاهی برای عملیات نوشتن موازی روی دیسکها اینگونه از عملیات را پی در پی مینماید و در هر لحظه تنها از پهنانی باند یک دیسک بطور مؤثر استفاده میگردد . با توزیع بلاکهای اطلاعات افزوده در سطح دنباله دیسکی این گلوگاه از بین رفته و هر عملیات Small Read تنها دو دیسک را درگیر مینماید. از این روی در مدلی همچون 5 RAID از پهنانی باند نیمی از دیسکهای دنباله بطور مؤثر استفاده میگردد[14][13].

۲-۳ - هزینه زمانی عملیات ورودی / خروجی

در این مرحله به ارزیابی هزینه زمانی انجام عملیات های ورودی/ خروجی روی دنباله دیسکی میپردازیم . تمامی سطوح مورد بررسی برای انجام عملیات Small Read بایستی هزینه زمانی R را بپردازند. اما در صورتیکه بخواهیم یک فایل یا یک دنباله داده ای بطول M را بر روی دنباله دیسکی توزیع شده درج نماییم در مواردی که بلاکهای داده ای بر روی دنباله دیسکی شده باشند بسادگی از فرمول زیر برای تعیین هزینه زمانی مدل برای عملیات Large Read می توان استفاده کرد .

$$(L_{Data} * T_R) / (N_{Total Disk} - N_{Check Disk}) \quad T_R : \text{هزینه زمانی متوسط عملیات خواندن یک بلاک} \quad L_{Data} : \text{طول فایل یا دنباله داده ای بر حسب بلاک}$$

در این میان تنها مورد استثناء RAID سطح یک است که با استفاده از تکنیک Mirroring بلاکهای داده ای منطقاً متواالی را بر روی بلاکهای متواالی فیزیکی بصورت دوگانه قرار می‌دهد. در این روش امکان ذخیره سازی موادی بلاکهای متواالی بلحاظ ساختار ذخیره سازی موجود فراهم نمی‌باشد و تمامی یک داده M بلاکی بطور پی در پی در یک دیسک ذخیره و نگهداری می‌شود و بهمین لحظه هزینه زمانی عملیات متواالی خواندن یک داده M بلاکی بصورت $M * R$ محاسبه می‌گردد. نتایج تحلیلی در جدول ۳ آمده است.

در هنگام انجام عملیات Large Write نیز با بهره گیری از امکان اجراء موادی عملیات ورودی / خروجی هزینه زمانی عملیات را می‌توان به مقدار قابل توجهی کاهش داد. هزینه سطوح صفر، یک ودو با توجه به هزینه عملیات Large Write قابل توجیه است تنها با این تفاوت که هزینه پایه ای نوشتن در نظر گرفته می‌شود. برای سطوح دیگر نیز از فرمول زیر می‌توان هزینه را محاسبه کرد [5][10].

$$(L_{Data} * T_W) + (L_{Data} * T_{LW}) / (N_{Total\ Disk} - N_{Check\ Disk})$$

: هزینه زمانی متوسط عملیات نوشتن یک بلاک T_W

جدول ۳. هزینه زمانی عملیات ورودی / خروجی

<i>Read / write Time (parallel)</i>	Small Read	Large Read	Small Write	Large Write
RAID 0	R	$M * R / N$	W	$M * W / N$
RAID 1	R	$M * R$	W	$M * W$
RAID 2	R	$M / (N - C_d) * R$	$R + W$	$M / (N - C_d) * W$
RAID 4	R	$M / (N - I) * R$	$R + W$	$M * (R + W) / (N - I)$
RAID 5	R	$M * R / (N - I)$	$R + W$	$M * (R + W) / (N - I)$
HVRaid	R	$M * R / (N - I)$	$R + W$	$M * (R + W) / (N - I)$

۳-۳ - هزینه عملیات ورودی / خروجی^۷

در این بخش ما عملیات ذخیره وبازیابی داده ای سیستم را از نظر تعداد عملیات ورودی / خروجی مورد ارزیابی قرار میدهیم. در یک سیستم متمرکز تمامی عملیاتهای داده ای بصورت محلی صورت می‌پذیرد. اما برای پیاده سازی مدل‌های مورد بررسی در یک سیستم توزیع شده نیاز است تا عملیاتهای داده ای راه دور را نیز تعریف نماییم. در RAID سطح پنج بهنگام نوشتن داده بر روی یک بلاک داده ای بایستی بلاک توازن افقی منتظر با آن بلاک داده ای نیز بروز گردد. در چنین شرایطی بازه هر عمل نوشتن باید اصلاح بلاک توازن جدید صورت گیرد [16]. اما در مدل پیشنهادی ما بجهت آنکه برای افزایش صحت و حفاظت داده ای از اطلاعات افزوده تری (بلاک توازن عمودی) نسبت به مدل دیگر استفاده شده، بمنظور نوشتن یک بلاک داده ای بایستی بلاکهای توازن افقی و عمودی منتظر با آن بلاک داده ای بروز گردد. در چنین شرایطی بازه هر عمل نوشتن بایستی سه عمل خواندن، بلاک داده ای قدیمی و بلاکهای توازن قدیمی، یک عمل نوشتن بلاک داده ای جدید و دو عمل اصلاح بلاک توازن جدید صورت گیرد [3]. ما هزینه عملیات ورودی / خروجی بازه هر خواندن از نوشتن روی یک بلاک داده ای را برای مدل‌های مختلف RAID محاسبه کرده و نتایج بدست آمده برای هر یک را بطور جداگانه در جدول ۴ به اختصار آورده ایم.

جدول ۴. هزینه تعداد عملیات ورودی / خروجی

Redundancy	Read Cost	Write Cost
RAID0	$1 * N_r$	$1 * N_w$
RAID1	$1 * N_r$ or $1 * N_{RR}$	$1 * N_w + 1 * N_{RW}$
RAID2	$1 * N_r + ?N_{RR}$	$1 * N_w + ?N_{RW}$ ^۸
RAID4	$1 * N_r + 1 * N_{RR}$	$1 * N_w + 1 * N_{RW}$
RAID5	$1 * N_r + 1 * N_{RR}$	$1 * N_w + 1 * N_{RW}$
HVRaid	$2 * N_r + 1 * N_{RR}$	$2 * N_w + 1 * N_{RW}$

در RAID2 بدلیل استفاده از کدهای همینگ ممکن است نیاز باشد تا افزونگی داده ای بر روی چندین دیسک قرار گیرد و در واقع بهنگام خواندن یا نوشتن چندین عملیات خواندن راه دور با نوشتن راه دور بطور همزمان صورت گیرد.

⁷I/O Cost

۴-۳ - هزینه فضایی^۸ عملیات

در این مدل بازاء هر گروه N دیسکی از دنباله RAID یک دیسک منفرد در نظر گرفته می شود، بگونه ای که بلاکهای توازن افقی را در خود نگهداری نماید. که بمنظور افزایش انعطاف پذیری و کاهش گلوگاه دیسک توازن، اطلاعات توازن افقی در میان $N+1$ دیسک توزیع میگردد. که تا این مرحله در هر دیسک $\frac{1}{N+1}$ درصد از افزونگی افقی در هر دیسک قرار داده می شود. علاوه بر این، در هر دیسک بطور محلی اطلاعات توازن عمودی هم افزوده میگردد، که این بخش سرباره فضایی افزوده ای را نسبت به RAID5 ایجاد مینماید. ما در ادامه نرخ افزونگی و سرباره فضایی سایر مدلها RAID را ارزیابی نموده و با مدل پیشنهادی مقایسه مینماییم. در RAID0 هیچگونه افزونگی داده ای وجود ندارد و درواقع این مقدار برای مقادیر مختلف داده ای همواره صفر درصد میباشد. در RAID1 همانطور که در یخشاهی اشاره گردید بدلیل استفاده از تکنیک انعکاس داده ای هزینه فضایی افزوده، ۱۰۰ درصد میباشد. برخی از نمادهای مورد نیاز در جدول ۵ آمده است.

جدول ۵. جدول نمادها

نماد	توضیحات
B	تعداد بلاک در هر دیسک
N	تعداد کل دیسکهای دنباله
r	نرخ سرباره در RAID سطح ۲
m	طول داده در RAID سطح ۲

در RAID2 با استفاده از کدهای تصحیح خطابهای کنترل و تشخیص خطابهای اضافی بترتیب زیر محاسبه میگردد. برای یک داده m بیتی هزینه فضایی افزوده بمنظور آنکه سیستم کد حاصل، تمامی خطابهای منفرد را تشخیص دهد بایستی 2^r بیت سرباره در نظر گرفته شود [15][13].

با توجه به فرمول فوق سرباره فضایی برابر است با $\frac{r}{m+r} \times 100\%$ و نسبت مقدار داده ای افزوده به کل داده سیستم $\frac{r}{m} \times 100\%$ است. در RAID سطوح سه، چهار و پنج برای هر دنباله N دیسکی یک دیسک سرباره در نظر گرفته می شود. که در هر سطح متناسب با تکنیک بکار رفته اطلاعات افزوده بگونه ای ذخیره میگردد. در تمامی سطوح مذکور اگر تعداد بلاکهای موجود در هر دیسک را با B و تعداد دیسکهای داده ای را N در نظر بگیریم با فرض یکسان بودن اندازه هم دیسکها می توان گفت که تعداد بلاکهای داده ای $N*B$ و تعداد بلاکهای توازن B میباشد. نسبت بلاکهای افزونگی به بلاکهای داده ای برابر است با: $\frac{B}{N * B} \times 100\%$ و درصد افزونگی برابر است با:

$$\frac{B}{(N + 1) * B} \times 100\%$$

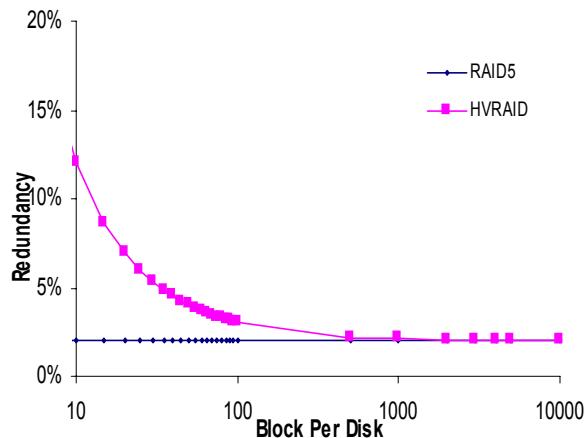
در مدل پیشنهادی مادرصد بلاکهای داده ای درون سیستم برابر است با:

$$\frac{N * B - N}{(N + 1) * B} \times 100\%$$

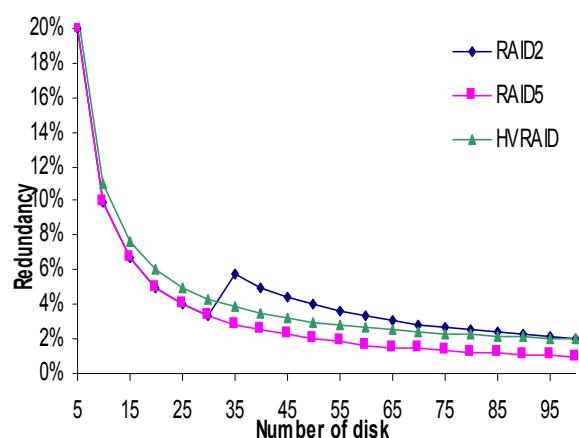
ونرخ سرباره افقی و عمودی مجموعاً $\frac{B + N}{N * B} \times 100\%$ میباشد [12][9][8].

در ادامه این بخش ما بر اساس نتایج نحلیلی بدست آمده در مراحل قبلی به ارزیابی رفتار مدلها مختلف خواهیم پرداخت. در نمودار ۱ هزینه سرباره فضایی هر مدل را بازاء تغییرات افزایشی تعداد دیسکها و یا ایستگاههای کاری درون سیستم محاسبه مینماییم. با کمی دقت میتوان مشاهده کرد که RAID5 همواره دارای کمترین سطح افزونگی می باشد. RAID2 نیز بازاء تعداد دیسکهای کمتر از ۳۰ رقابت بسیار نزدیکی را با RAID5 دارد ولی یکباره تغییر غیر قابل پیش بینی را از خود نشان میدهد و بیشترین مقدار افزونگی را در مراحل بعدی داراست. مدل HVRAID نیز بازاء تعداد دیسکهای کمتر از ۳۰ تا ۳۵ بیشترین سطح افزونگی را داشته و پس از آن گوی رقابت را از RAID2 میبرد. در کل میتوان گفت رفتار تمامی سطوح مشابه بوده و با افزایش تعداد دیسکها درون سیستم نرخ افزونگی داده ای کاهش میابد.

⁸Space cost



نمودار ۲ . مقایسه مدل پیشنهادی با دنباله دیسکی سطح پنج



نمودار ۱ . مقایسه سطوح مختلف دنباله دیسکی

در نمودار ۲ ما بطور دقیقتر مدل پیشنهادی را بازاء تغییرات BPD^۹ مورد بررسی قرار میدهیم . افزونگی داده ای در RAID5 از تغییرات BPD مستقل است . در عوض میزان افزونگی در مدل پیشنهادی تغییرات محسوسی را از خود نشان میدهد . با مشاهده نمودار ۲ مشخص می شود که با افزایش BPD نرخ افزونگی داده ای بمقدار قابل توجهی کاهش پیدا میکند . این کاهش در مقدار حدی خود به سمت مقدار ثابت افزونگی در RAID5 میل مینماید . این رفتار حدی را بدین ترتیب می توان توجیه کرد که میزان افزونگی افقی همواره مقدار ثابتی است اما نسبت بلاکهای افزونگی عمودی به تعداد بلاکهای درون سیستم بطور معکوس با افزایش BPD کاهش میابد . نکته قابل توجه آنست که در مدل پیشنهادی میزان افزونگی بازاء تعداد دیسک ها و BPD کم هزینه قابل توجهی است . اما با افزایش این مقادیر ، به مقدار قابل قبولی کاهش میابد [15].

۵-۳ - هزینه عملیات احیاء^{۱۰}

در مدل پیشنهادی بهنگام بروز یک خطأ در سطح بلاکی ، کنترل RAID قادر است تا بکمک اطلاعات افزوده عمودی و افزونگی افقی ، خطأ را تشخیص دهد . بمنظور احیاء تنها با استفاده از اطلاعات توازن عمودی ، برای هر دیسک بصورت محلی و بدون دخالت بلاکهای داده ای متناظر در سایر دیسکها به بازیافت اطلاعات بپردازد . مزیت عمدئ مدل پیشنهادی نسبت به سایر مدلهای قبلی آنست که در این مدل امکان تشخیص و تصحیح خطأ های زوج نیز فراهم است و با اطمینان بیشتری میتوان موقعیت دقیق خطاهای بروزکرده را به لحاظ آنکه در کدام سایت یا بر روی کدام دیسک دنباله قرار دارند مشخص کرد .

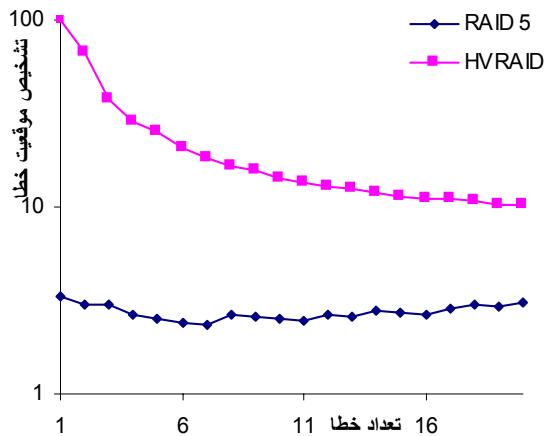
در سطح یک بدلیل استفاده از تکنیک Mirroring و افزونگی داده ای نسبتا بالای این مدل ، حفاظت داده ای در سطح خوبی فراهم می آید . در این مدل بدلیل وجود یک نسخه پشتیبان از هر داده ، حتی اگر نیمی از دیسکهای هم نوع (داده ای یا پشتیبان) هم دچار خرابی شوند سیستم قادر است تا داده ها و اطلاعات از دست رفته را احیاء نماید . اما در صورتی که یک دیسک داده ای و دیسک پشتیبان آن هر دو دچار خرابی شوند اطلاعات مورد نظر قابل احیاء نمی باشد . در مدل پیشنهادی برخلاف سطوح قبلی امکان احیاء خطأ های داده ای در سطح بلاکی تا $N-1$ خطأ در هر Stripe افزایش پیدا میکند ، علاوه برآنکه خطأ های زوج نیز در این مدل بادقت بالاتری قابل تشخیص است [2] .

۴ - شبیه سازی مدل پیشنهادی

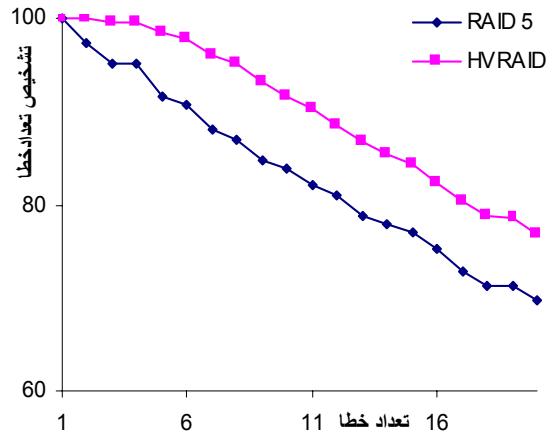
در این فرایند شبیه سازی سعی نموده ایم دو مدل RAID 5 ، HVRAID را از نظر حفاظت و امنیت داده ای مدل در مقابل بروز هرگونه خطأ در سطح دنباله دیسکی پیاه سازی نماییم . ما قصد داریم تا بکمک این شبیه سازی میزان دقت مدلها را بمنظور نمودار

⁹ Block Per Disk

¹⁰ Recovery Cost

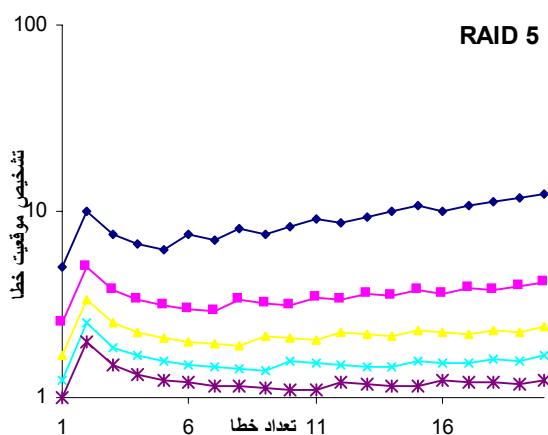


نمودار ۳ . مقایسه دقت در تشخیص موقعیت خطا

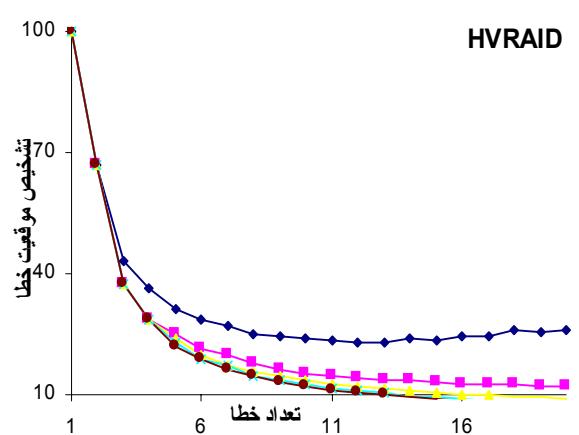


نمودار ۴ . مقایسه سطح تشخیص تعداد خطا

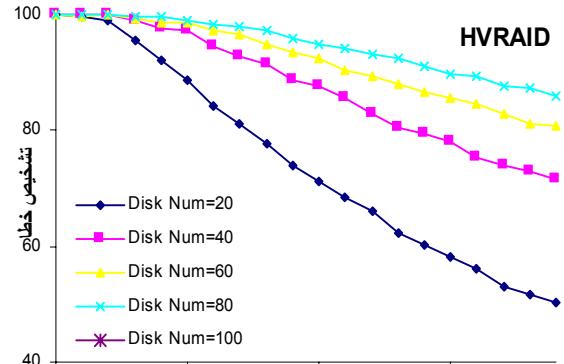
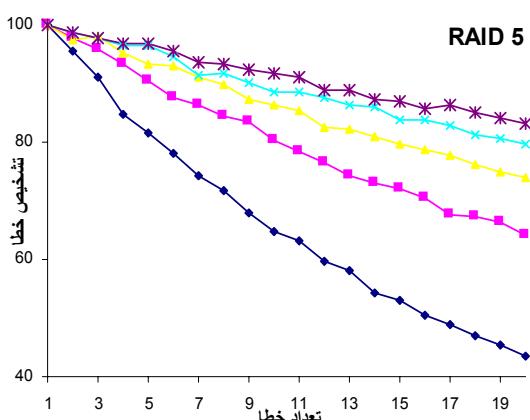
تشخیص خطاهای بروز کرده در سطح سیستم بررسی نموده و علاوه بر آن قابلیت ودقت مدلها را در تعیین موقعیت خطا (در سطح دیسکی و بلاکی) آزموده و مقدار احتمالی صحت تشخیص خطا و تعیین موقعیت فضایی خطا مشخص نماییم. بهمین منظور عملیات شبیه سازی را در دو مرحله مختلف انجام داده ایم. در مرحله نخست هر دو مدل را در معرض خطاهای مختلفی قرار داده و تعداد خطاهای بروز کرده در سیستم را از یک خطاب تا بیست خطاب افزایش داده و پاسخ سیستم را بازاء هر حالت مشخص نموده ایم. بمنظور افزایش سطح دقت در تعیین نتایج، انجام آزمون را برای هر مدل تا ۱۰۰۰۰ با تکرار نموده ایم. در حقیقت در این مرحله از آزمون قصد داریم مشخص نماییم که هر مدل قادر است تا چند درصد از خطاهای بروز کرده را تشخیص دهد. همانگونه که در نمودار ۳ نیز مشاهده میگردد هر دو مدل را بازاء تعداد ۲۰ خطاب مورد بررسی قرار داده ایم و دیده می شود که هر دو



RAID 5



RAID 5



HVRAID

نمودار ۶ . مقایسه دقت و سطح تشخیص خطاهای بروز کرده در RAID5,HVRAID

مدل در مقابل افزایش تعداد خطای رفتار مشابهی از خود نشان داده و سطح دقت آنها در تشخیص خطای های با افزایش تعداد خطای کاهش میابد. با یک نگاه در نمودار ۳ می توان مشاهده کرد که مدل پیشنهادی ما بازاء تمامی حالات برتری خود را در تشخیص دقیقتر خطای های حفظ مینماید. بازاء خطای های منفرد در سیستم، هر دو مدل با دقت ۱۰۰٪ خطای بروز کرده را تشخیص می دهند. اما درصد تشخیص تعداد خطای های HVRAID بازاء افزایش تعداد خطای های در سیستم با سرعت کمتری نسبت به مدل دیگر کاهش میابد و درصد اختلاف، میان سطح دقت در تشخیص خطای های در دو مدل مورد بررسی، بازاء تغییرات تعداد خطای بین ۸٪ تا ۱۰٪ متغیر است و بهمین لحاظ منحنی هر دو مدل، در نمودار ۶ از شبکه نزولی نسبت مشابهی برخوردارند.

از جمله ایرادات مدل های قبلی همچون مدل RAID 5 در آنست که قابلیت تشخیص موقعیت خطای در سطح دنباله دیسکی در این مدلها بسیار ضعیف بوده و با احتمال بسیار کمی می توان موقعیت واقعی خطای بروز کرده را بدسترسی تشخیص داد. همانگونه که در نمودار ۶ هم می توان مشاهده کرد، در این مدل امکان تعیین موقعیت خطای تا حدود زیادی از تعداد خطای های در سطح دنباله مستقل است و شبکه نمودار بازاء تغییرات تعداد خطای های تغییر محسوسی نداشته و همواره با احتمالی کمتر از ۳ درصد قادر به تشخیص موقعیت خطای های در مقابل مدل HVRAID قادر است تا ۱۰۰٪ درصد موقعیت فضایی خطای های منفرد را درون دنباله دیسکی بدسترسی تعیین نماید. اما در این مدل نیز با افزایش تعداد خطای های بدلیل بالارفتن میزان رخداد خطای زوج، سطح دقت در تشخیص خطای های تعیین موقعیت خطای های در تمامی موارد و بازاء افزایش تعداد خطای های به نسبت یکسانی کاهش پیدا نمی کند. یا بعبارتی دیگر شبکه نمودار در تعیین موقعیت خطای های در تمامی قسمتها یکسان نبوده و بازاء تعداد خطای های کاهش کاهش پیدا میکند اما با افزایش تعداد خطای های سرعت در روند کاهشی آن کاسته می شود بطوریکه با افزایش بیشتر تعداد خطای های شبکه نمودار به موازات با سطح افق، میل مینماید. بطور کلی می توان گفت که، مدل پیشنهادی در تعیین موقعیت خطای های درون دنباله دیسکی، نسبت به مدل دیگر خصوصا بازاء تعداد خطای های پایین، بسیار موفقتر عمل نموده و نتایج بهتری را ارایه میدهد.

۵- نتیجه گیری

پس از بررسی دقیق و مدل سازی تحلیلی ساختار پیشنهادی و با توجه به نتایج بدست آمده می توان گفت، مدل پیشنهادی دارای ویژگی های خاصی است که آنرا از سایر مدلها دیگر خصوصا برای پیاده سازی در سیستم های نامت مرکز متمایز مینماید. ویژگی بارز این مدل استفاده از بلاکهای توازن (افزونگی) افقی و عمودی توزیع شده است، که این نوع افزونگی علاوه بر افزایش دقت در تشخیص خطای های قابلیت اطمینان مدل، سرباره ناچیزی را نیز از نظر فضایی و تعداد عملیات ورودی/خروجی به سیستم اعمال می نماید. که این سرباره در محیطهای شبکه ای و توزیع شده که با تعداد زیادی از ایستگاههای کاری و مجموعه ای از دیسکها روبرو هستیم بمقدار قابل توجهی کاهش یافته و در حد سایر مدلها قرار خواهد گرفت. همانگونه که از فرمولها و مقایسات تحلیلی و نتایج حاصل از شبیه سازی مدل بر می آید، مدل پیشنهادی ما قادر است تا بکمک اطلاعات افزوده ناچیزی نسبت به بهترین مدلها قبلی، صحت و امنیت داده ای بالاتر، بازدهی و پهنای باند بیشتر و دسترسی پذیری بهتری را برای داده های بحرانی درون سیستم فراهم آورده و در مجموع سطح حفاظت داده ای و قابلیت اطمینان سیستم را بهبود بخشد.

در مجموع می توان گفت، مدل پیشنهادی ما در مقایسه با سایر مدلها و همچنین در مقایسه با سایر مزایای آن، این افزونگی نیز قابل چشم پوشی است. بکمک تغییرات اعمال شده در مدل RAID و معماری محیط محاسباتی توزیع شده عملیات احیاء دیسک نیز با هزینه و اثرات جانبی کمتری نسبت به سایر مدلها مشابه قابل انجام است. که برخی از ویژگیهای ذاتی محیطهای توزیع شده نیز این امر را تسهیل مینمایند.

مراجع

- [1] 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.

- [2] J.Akbari and A.T.Haghigat . "A New Redundancy Algorithm For Distributed Environment", Operating system & Security Conference-OSSC 2003, Sharif university of technology, 24 -25 Dec,pp.70-81.
- [3] Michael Stonebraker ,Gerhard A. Schloss. "Distributed Raid – A New Multiple Copy Algorithm" , University of California,Berkeley, CA 94720,2000,pp.83-92
- [4] T. Anderson, M. Dahlin,D. Patterson, and R. Wang. "Serverless Network FileSystems", ACM Trans. on Computer Systems, Jan. 1996, pp.41-79.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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,pp.101-111
- [9] T. H. Cormen and D. Kotz, "Integrating Theory and Practice in Parallel File Systems",Proceedings of DAGS '93 Symposium, June 1993, pp. 64-74.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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, February 1988, pp.51-81
- [14] 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.
- [15] K. Hwang, H. Jin, E. Chow, C. L. Wang, and Z. Xu , "Designing SSI Clusters with Hierarchical Checkpointing and Single I/O Space". IEEE Concurrency Magazine, March 1999, pp.60-69.
- [16] N. Muppalaneni and K. Gopinath, "A Multi-tier RAID Storage Syste with RAID1 and RAID5 " ,Proceedings of 14th International Parallel and Distributed Processing Symposium , Mexico, May 2000,pp.163-174.