ا

تعهدنامه

اینجانب **حسین سلامی** دانشجوی دوره دکتری/کارشناسی ارشد رشته **کامپیوتر- نرم افزار** دانشکده **مهندسی** دانشگاه فردوسی مشهد نویسنده پایان‌نامه **کاهش صدمات ناشی از خرابی کارها با استفاده از پیش بینی وضعیت آتی کار** تحت راهنمایی **دکتر حسین دلداری** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود و یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه فردوسی مشهد می باشد و مقالات مستخرج با نام "دانشگاه فردوسی مشهد" و یا "Ferdowsi University of Mashhad" به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند در مقالات مستخرج از رساله رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است، اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه فردوسی مشهد می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

II

پایان نامه ی کارشناسی ارشد

کاهش صدمات ناشی از خرابی کارها با استفاده از پیش بینی وضعیت آتی کار

دانشجو:

حسین سلامی

استاد راهنما:

جناب آقای دکتر حسین دلداری

استاد مشاور:

جناب آقای دکتر محمود نقیب زاده

مرداد ۱۳۹۰

بسمه تعالی

دانشگاه فردوسی مشهد

صورتجلسه دفاع از پایان نامه کارشناسی ارشد

جلسه دفاع از پایان نامه آقای/خانم **حسین سلامی** دانشجوی دوره کارشناسی ارشد رشته **کامپیوتر**گرایش **نرم افزار** در ساعت ..... روز ..... در محل ..... دانشکده ..... با حضور امضا کنندگان ذیل تشکیل گردید. پس از بررسی های لازم، هیأت داوران پایان نامه نامبرده را با نمره به عدد .....، به حروف ..... و با درجه ..... مورد تأیید قرار داد / نداد.

عنوان رساله

.....

هیئت داوران امضا

- **داور: دکتر .....**
  دانشیار گروه ..... دانشگاه .....
- **داور و نماینده تحصیلات تکمیلی: دکتر .....**
  .....گروه ..... دانشگاه .....
- **استاد راهنما: دکتر .....**
  .....گروه ..... دانشگاه .....
- **استاد مشاور: دکتر .....**
  .....گروه ..... دانشگاه .....
- **مدیر گروه: دکتر .....**
  .....گروه ..... دانشگاه .....

تقدیر و تشکر

مطالب این اوراق حاصل راهنمایی و پشتیبانی عزیزانی می باشد که با تلاش ها و تجربیات خود، بنده را در طی این مسیر یاری نموده اند. در این بین لازم می دانم از اساتید بزرگوار جناب آقای دکتر حسین دلداری و جناب آقای دکتر محمود نقیب زاده به خاطر تمام زحماتی که در دوره ی تحصیل برای اینجانب متحمل شده اند، تشکر نمایم. همچنین از پدر و مادر مهربانم که وجود گرانبهایشان، همیشه دلگرم کننده ی بنده بوده است، کمال تشکر را دارم. از تمام دوستان عزیزم، خصوصا دانشجویان آزمایشگاه پردازش موازی و توزیع شده که با راهنمایی های خود، زمینه ی هموار شدن مسیر انجام این پایان نامه را سبب شده اند، سپاسگذاری می نمایم. در اینجا تشکر ویژه ی خود را از جناب آقای مهندس حمید سعادتفر، دانشجوی دکترای مهندسی کامپیوتر دانشگاه فردوسی به خاطر تمام تلاش های بی دریغشان، ابراز می دارم.

چکیده:

در سالیان اخیر محیط های پردازشی توزیع شده مانند گریدهای محاسباتی، به یکی از مهمترین بسترها برای رفع نیازهای پردازشی کاربران مبدل شده اند. این محیط ها دارای توان بالقوه ای برای پاسخگویی به نیازهای کاربران می باشند اما مشکلات خاص خود را نیز بهمراه دارند. یکی از پدیده هایی که در این محیط ها قابل مشاهده است، بالا بودن درصد کارهایی است که علیرغم دریافت منابع درخواستی، به صورت ناموفقی به پایان رسیده و سبب تحمیل صدمات قابل توجهی می شوند. تلاش های متعددی برای چیره شدن بر مسئله ی اجرای ناموفق کارها انجام شده است که تا حدی نیز موثر بوده اند اما یکی از ضعف های عمده ی آنها، توجه به بعضی عوامل اجرای ناموفق کارها (عوامل سمت منابع) و در نظر نگرفتن عوامل دیگر (عوامل سمت کار) می باشد. بدین منظور، در این تحقیق، هدف ما ارائه ی راهکاری برای کاهش صدمات ناشی از اجراهای ناموفق کارها بوسیله ی در نظر گرفتن عواملی است که از سمت کارها ناشی می شوند تا از این طریق نشان دهیم که می توان با توجه به این عوامل نیز صدمات اجراهای ناموفق را کاهش داد. راهکار پیشنهادی، افزودن یک مولفه ی جدید است که وظیفه ی آن پذیرش نکردن کارهایی است که مستعد خرابی می باشند. رفتار این مولفه با استفاده از دو روش مبتنی بر آستانه و مدلی مبتنی بر تئوری بازی به نام قمار ریسکی پیاده سازی شده است. نتایج آزمایشات نشان داد که بکارگیری این مولفه ی جدید می تواند در کاهش صدمات ناشی از اجراهای ناموفق کارها نقش قابل توجهی را ایفا نماید.

فهرست مطالب

تقدیر و تشکر IV
چکیده V
فهرست مطالب VI
فهرست جدول ها VIII
فهرست شکل ها IX
فصل ۱- مقدمه ۱
۱-۱- گریدهای محاسباتی و خرابی کارها ۱
۱-۲- عوامل موثر در خرابی کارها ۴
۱-۲-۱- عوامل سمت منبع ۶
۱-۲-۲- عوامل سمت کار ۸
۱-۳- تدوین مساله ۱۱
۱-۴- ساختار پایان نامه ۱۱
فصل ۲- بررسی کارهای پیشین ۱۳
۲-۱- روش های پیش بینی وقایع ۱۳
۲-۲- دسته بندی روش های مقابله با خرابی ۱۴
۲-۲-۱- روش های مقابله با خرابی های سمت منابع ۱۵
۲-۲-۱-۱- روش های مبتنی بر اجتناب ۱۶
۲-۲-۱-۲- روش های زمانبندی آگاه از خرابی (اطمینان) ۱۸
۲-۲-۲- مقابله با خرابی های سمت کار ۱۹

فصل ۳- استفاده از پیش‌بینی وضعیت آتی کار برای کاهش تاثیرات نامطلوب ناشی از خرابی کارها ۲۵

۳-۱- مقدمات ۲۶

۳-۲- پیش‌بینی‌کننده‌ی وضعیت آتی کار ۲۸

۳-۳- تصمیم‌گیری در مورد پذیرش کارهای ارسالی ۳۰

۳-۳-۱- روش مبتنی بر آستانه ۳۱

۳-۳-۲- روش مبتنی بر قمار ریسکی ۳۳

۳-۳-۲-۱- مقدمه ای بر تئوری بازی و قمار ۳۳

۳-۳-۲-۲- تصمیم گیری با استفاده از قمار ۳۵

فصل ۴- معرفی محیط شبیه سازی و نتایج آزمایشات ۴۱

۴-۱- محیط شبیه سازی ۴۱

۴-۱-۱- داده های مورد استفاده ۴۲

۴-۱-۲- آماده سازی پیش بینی کننده ۴۴

۴-۱-۳- آماده سازی شبیه ساز ۴۵

۴-۲- معیارهای ارزیابی ۴۵

۴-۳- نتایج ۴۸

۴-۳-۱- کیفیت پیش بینی کننده ۵۰

۴-۳-۲- میزان بارکاری ارسالی ۵۶

۴-۳-۳- میزان اطمینان منابع ۶۰

فصل ۵- نتیجه گیری و کارهای آینده ۶۷

مراجع ۷۰

فهرست جدول ها

جدول ۱-۱: درصد خرابی در برخی از گریدها ۳

جدول ۲-۱: مقایسه ی روش های مقابله با خرابی کارها ۲۴

فهرست شکل ها

شکل ۱-۱: سهم عوامل تاثیرگذار در خرابی کارها در آورگرید و سیستم شماره ی ۲۰ آزمایشگاه ملی لس آلاموس ۱۰

شکل ۲-۱: مدیریت ریسک در گریدهای سودمند ۱۷

شکل ۲-۲: نمونه ای از درخت تصمیم که با استفاده از الگوریتم **C4.5** ایجاد شده است ۲۰

شکل ۲-۳: درخت تصمیم ایجاد شده بر روی آرگومان های کار ۲۲

شکل ۳-۱: معماری سیستم در حضور مدیر ارسال کار و پیش بینی کننده ی وضعیت آتی کار ۲۵

شکل ۴-۱: سهم هریک از دلایل خرابی در سیستم های آزمایشگاه لس آلاموس ۴۳

شکل ۴-۲: نرخ عدم پذیرش کارها در کیفیت های مختلف پیش بینی کننده ۵۰

شکل ۴-۳: نرخ عدم پذیرش نادرست در کیفیت های مختلف پیش بینی کننده ۵۲

شکل ۴-۴: نرخ پذیرش نادرست کارها در کیفیت های مختلف پیش بینی کننده ۵۳

شکل ۴-۵: میانگین زمان انتظار کارها در کیفیت های مختلف پیش بینی کننده ۵۴

شکل ۴-۶: میزان بهره وری مفید منابع در کیفیت های مختلف پیش بینی کننده ۵۵

شکل ۴-۷: میانگین زمان انتظار کارها در بارهای کاری مختلف ۵۷

شکل ۴-۸: میزان بهره وری مفید در بارهای کاری مختلف ۵۸

شکل ۴-۹: نرخ عدم پذیرش کارها در روش قمار ریسکی در بارهای کاری مختلف ۵۸

شکل ۴-۱۰: نرخ عدم پذیرش نادرست کارها در روش قمار ریسکی در بارهای کاری مختلف ۵۹

شکل ۴-۱۱: نرخ پذیرش نادرست کارها در روش قمار ریسکی در بارهای کاری مختلف ۶۰

شکل ۴-۱۲: نرخ رد کردن کارها توسط زمانبند (روش **NSA**) در گریدهای متفاوت ۶۲

شکل ۴-۱۳: نرخ رد کردن کارها توسط مدیر ارسال کار در گریدهای متفاوت ۶۳

شکل ۴-۱۴: نرخ رد کردن نادرست توسط مدیر ارسال کار در گریدهای متفاوت ۶۳

شکل ۴-۱۵: نرخ پذیرش های نادرست مدیر ارسال کار در گریدهای متفاوت ۶۴

شکل ۴-۱۶: میانگین زمان انتظار در گریدهای متفاوت ۶۵

شکل ۴-۱۷: میزان بهره وری مفید در گریدهای متفاوت ۶۵

فصل۱– مقدمه

در ایـن فصـل، زمینـه‌ی تحقیـق ارائـه شـده در ایـن پایـان نامـه معرفـی شـده اسـت. در ابتـدا، گریـدهای محاسـباتی بـه عنـوان یکـی از بسـترهای نوظهـور بـرای انجـام محاسـبات بـا مقیـاس بـزرگ معرفی شده و سپس بـا معرفـی خرابـی کارهـا بـه عنـوان یکـی از مسـائل موجـود در ایـن محـیط هـا، بـه عوامل موثر در بـروز آنهـا خـواهیم پرداخـت. نهایتـا در انتهـای ایـن فصـل، سـاختار مطالـب بعـدی پایـان نامه بیان شده‌اند.

۱–۱– گریدهای محاسباتی و خرابی کارها

افـزایش روزافـزون کـاربرد کامپیوترهـا بـرای رفـع نیازهـای محاسـباتی در زمینـه هـای مختلـف، حقیقـت غیرقابـل انکـار در دنیـای امـروزی اسـت. در همـین حـال، افـزایش حجـم داده‌هـا و ظهـور پدیـده‌ی انفجـار محاسـباتی، موجـب شـده اسـت تـا روش‌هـای سـنتی بـرای انجـام محاسـبات، ماننـد اسـتفاده از کامپیوترهـای غـول‌پیکـر و گـران قیمتـی همچـون اَبَرکامپیوترهـا[1]، بـه سـبب عـدم توانـایی در فـراهم آوردن کیفیـت سـرویس مـورد انتظـار کـاربران و همچنـین تحمیـل هزینـه‌هـای قابـل توجـه، کـاربرد چنـدانی نداشـته باشـند. بـا ایـن حـال، پیشـرفت‌هـا در زمینـه‌هـای ارتبـاطی، ظهـور فنـاوری‌هـای جدیـد شـبکه‌ای و توسـعه‌ی سیسـتم‌هـای توزیـع شـده، نویـد دهنـده‌ی بسـترهای جدیـد برای انجـام محاسـبات مبتنـی بـر محـیط هـای شـبکه‌ای اسـت. ایـن بسـترهای محاسـباتی بـا در اختیـار گـرفتن منـابع گسـترده شـده در محـیط شـبکه و توزیـع پـردازش‌هـای کـاربران در بـین آنهـا، قـادر بـه رفـع بسـیاری از نیازهـای محاسـباتی کـاربران مـی باشـند. گریـدهای محاسـباتی یکـی از ایـن بسـترهای

[1] SuperComputer

محاسباتی است. گریدهای محاسباتی، زیرساخت‌های سخت‌افزاری و نرم‌افزاری‌ای هستند که امکان اشتراک‌گذاری توان محاسباتی منابع موجود در محیط‌های شبکه‌ای را فراهم می‌آورند. حاصل این اشتراک، قدرت محاسباتی قابل ملاحظه‌ای است که با بکارگیری آن می‌توان بسیاری از نیازهای محاسباتی که در گذشته انجام آنها امکان پذیر نبوده است را رفع نمود. گریدهای محاسباتی برای کاربردهای متعددی بکار می‌روند که از آن جمله می‌توان به اَبَرمحاسبات توزیع‌شده[1]، محاسبات با گذردهی بالا[2]، محاسبات برحسب نیاز[3]، محاسبات بر روی حجم زیاد داده‌ها[4] و محاسبات مشترک[5] اشاره نمود.

گریدها دربرگیرنده‌ی تعداد زیادی از منابعی می باشند که در سطح شبکه گسترده شده اند. این تعداد از منابع، پتانسیل پردازشی قابل توجهی را برای محققان و دانشمندان علوم مختلف، که نیازمندی های پردازشی با مقیاس بزرگ دارند، نوید می دهد. با این حال گریدها هنوز نتوانسته اند این رویا را به طور کامل محقق نمایند. یکی از مشکلات اساسی در این راه، مدیریت پیچیده‌ای است که گرید به آن نیازمند است. عوامل متعددی در ایجاد این پیچیدگی سهیم می‌باشند که از جمله با اهمیت‌ترین آنها می‌توان به مقیاس بسیار بزرگ، پویایی و ناهمگونی منابع، و نیز ناهمگونی سیاست ها در عرضه ی منابع اشاره نمود. علاوه بر این، وجود بعضی مسائل باعث استفاده‌ی غیر بهینه از منابع در این محیط می‌گردد. یکی از این مسائل، مسئله‌ی خرابی کارها می‌باشد. خرابی کارها در بسیاری از بسترهای محاسباتی وجود دارد، اما این پدیده در گریدهای محاسباتی به صورت قابل توجهی دارای نرخ بالاتری است. در محیط

[1] Distributed SuperComputing
[2] High-Throughput Computing
[3] On-Demand Computing
[4] Data-Intensive Computing
[5] Collaborative Computing

گرید، نیازهای محاسباتی در قالب کارها[1] ارسال می‌شوند و گرید با در اختیار قرار دادن منابع لازم، زمینه‌ی اجرای آنها را فراهم آورده و نتیجه را به کاربر(ان) ارسال کننده‌ی کار، باز می‌گرداند. با این حال برخی از کارهای ارسالی نمی‌توانند اجرای خود را به پایان برسانند، در واقع، اجرای آنها ناتمام باقی می ماند، به این دسته از کارها، کارهای خراب شده[2] گویند. متاسفانه بخش قابل توجهی از کارهایی که به گریدها ارسال می شوند، دچار خرابی می شوند. برای نمونه در جدول ۱-۱، نرخ کارهای خراب‌شده در چند محیط مختلف آورده شده است. خرابی کارها در محیط گرید به دلایل متعددی رخ می دهند، به عبارت دیگر عوامل متعددی سبب می شوند تا یک کار دچار خرابی گردد. مطمئنا آشنایی با این عوامل می تواند در ارائه روش مناسب برای مقابله با آنها، تاثیر بسزائی داشته باشد. از اینرو، ابتدا در بخش بعد به عواملی می‌پردازیم که در محیط گرید می‌توانند در بروز خرابی یک کار اثرگذار باشند.

نام محیط	درصد خرابی کارها
DAS2	بیش از ٪۱۰
Tera Grid	٪۴۵-۱۰
Grid3	٪۲۷
LCG	٪۳۳-۲۵

جدول ۱-۱: درصد خرابی در برخی از گریدها

[1] Job
[2] Failed Job

۱-۲- عوامل موثر در خرابی کارها

ارائه‌ی راهکارهای مناسب برای جلوگیری، و یا کاهش صدمات ناشی از بروز خرابی کارها، نیازمند آگاهی از عوامل پدیدآورنده ی آنها می باشد. شناسایی عوامل ایجادکننده‌ی یک پدیده، به داده‌های کافی در مورد آن، و نیز تحلیل دقیق داده‌های بدست آمده نیازمند می‌باشد. تا چند سال قبل، بسیاری از محیط‌های گرید، امکانات و ابزار مناسب برای جمع‌آوری داده‌ها را در اختیار نداشته‌اند و یا در صورت وجود چنین امکاناتی، به دلیل مسائل امنیتی، از در اختیار قرار دادن عمومی چنین اطلاعاتی ممانعت می‌نموده‌اند. اما در چند سال اخیر، ارائه‌ی عمومی پرونده‌های ثبت وقایع محیط‌های گرید، سبب شده است تا محققان زیادی به این پرونده‌ها برای مشخص کردن رفتارهای گریدها مراجعه نموده و از آنها استفاده نمایند. برای نمونه پیش بینی میزان بارکاری گریدهای آورگرید[1] و Grid5000 [Wu2010]، پیش بینی نرخ ورود کارها در EGEE [Chr2008]، نمونه هایی از چنین تلاش هایی می باشد.

دلایل متعددی برای بوقوع پیوستن خرابی کارها توسط محققان مختلف گزارش شده است. این امکان وجود دارد که یک کار با ماشین مشخصی سازگاری نداشته باشد، بیشتر کارها نیازمند CPU و یا سیستم عامل خاصی می باشند. برخی دیگر از خرابی ها ممکن است به سبب فعالیت های مدیریتی رخ دهد، برای نمونه، دسترسی یک کاربر به منبعی که قبلا توانایی دسترسی به آن را داشته است، سلب می شود [Cie2006]. خرابی یک کار ممکن است به سبب وجود یک اشکال[2] در خود برنامه (کار) ارسالی کاربر رخ دهد، و یا ممکن است که به سبب

[1] AuverGrid
[2] Bug

درخواست گواهینامه ای[1] که طول اعتبار آن منقضی شده است، خرابی کار بروز نماید. همچنین ممکن است که خرابی کار به سبب خرابی منبعی (منابعی) رخ دهد، که بر روی آنها در حال اجرا بوده است [Dua2006]. دسته بندی های متعددی برای خرابی ها در محیط گرید موجود است که نمونه هایی از آنها در [Li2006]، [Neo2007] و [Zei2008] ارائه شده اند. در [Li2006] لی و همکارانش، خرابی کارها را در پرونده‌های ثبت وقایع گرید LCG بررسی و تحلیل نمودند. آنها عوامل تاثیرگذار در خرابی کارها را در ۳ دسته زیر قرار دادند:

- مشکلات سیستمی مانند خرابی نودهای پردازشی، دیسک و شبکه.
- خطاهایی که در سطوح مختلف پشته‌ی نرم‌افزاری از جمله میان‌افزار محیط، بسته‌های نرم‌افزاری منابع (کلاسترها) و یا خود پردازش کاربر رخ می‌دهند.
- مسائل انسانی و سیاستی

در این نوشتار، دسته بندی دیگری را پیشنهاد می نمائیم. عوامل فوق را در این دسته بندی جدید می توان به دو دسته تقسیم نمود. عوامل دسته ی اول را عوامل سمت منبع[2] و عوامل دسته ی دوم را، عوامل سمت کار[3] می نامیم. در دو زیر بخش بعدی، به طور مفصل تری به هر یک از این دو دسته از عوامل می پردازیم.

[1] Certificate
[2] Resource-Side
[3] Job-Side

۱-۲-۱- عوامل سمت منبع

برخی از خرابی کارها، به سبب ازکارافتادگی منابع اجرا کننده ی آنها رخ می دهند. خرابی های سخت افزاری، مانند خرابی پردازشگر اصلی و یا دیسک سخت، خرابی های نرم افزاری، مانند بروز خطاهای مهلک[1] در نرم افزارهای سیستمی و مدیریتی همچون سیستم عامل و یا بسته های مدیریتی منبع، خرابی هایی که به سبب عوامل محیطی رخ می دهند، مانند قطع برق، خارج شدن منبع از محیط به صلاحدید مالک آن، و یا قطعی ارتباط با منبع، به سبب قطعی ارتباطات شبکه ای از جمله دلایلی است که سبب می شوند تا کارهایی که بر روی منبع (منابع) در حال اجرا بوده اند، دچار خرابی گردند. براساس تحلیل شرودر و گیبسون [Sch2006] در محیط های محاسباتی با مقیاس بزرگ، سخت‌افزار، بزرگترین عامل در بوجود آمدن مشکلات سمت منابع می‌باشد و سهم آن در حدود ۳۰ تا ۶۰ درصد است. نرم‌افزار در جایگاه بعدی قرار دارد و حدود ۵ تا ۲۴ درصد را به خود اختصاص داده است. علاوه بر سخت‌افزار، نرم‌افزار و شبکه، برخی دیگر از عوامل در مشکلات سمت منابع دخیل می‌باشند که دلیل ریشه‌ای آنها نامشخص است، این عوامل در حدود ۲۰ تا ۳۰ درصد از مشکلات سمت منابع را سبب می‌شوند.

مقابله با خرابی هایی از کارها که به سبب عوامل سمت منبع (منابع) رخ می دهند به عنوان مسئله ی ایجاد اطمینان[2] شناخته می شود. تلاش هایی که به منظور فراهم آوردن اطمینان در محیط گرید صورت گرفته اند، اکثرا بدنبال روش هایی برای ایجاد قابلیت تحمل

[1] Fatal Error
[2] Reliability

خرابی[1] بوده است. قابلیت تحمل خرابی، به معنای فراهم آوردن شرایطی است که در آن، اجرای کارها علیرغم بروز خرابی ها ادامه می یابد [Dab2009]. اکثر روش های تحمل خرابی بکار گرفته شده در گریدهای فعلی، مبتنی بر روش هایی است که در سایر فناوری های سیستم های توزیع شده بکار گرفته می شوند. اما با توجه به رشد سیستم های گرید، محققان در حال توسعه ی روش های تحمل خرابی ای هستند که تکنیک های فعلی را گسترش داده تا بتوانند در شرایط مقیاس بزرگتر، ناهمگونی و پویایی به صورت مناسبی عمل نمایند. برای نمونه، تکنیک check-pointing یکی از مهمترین تکنیک هایی است که برای ایجاد تحمل خرابی در اکثر محیط ها مورد استفاده قرار می گیرد. در این تکنیک، وضعیت کار، به صورت دوره ای در ذخیره گاه های مطمئن ذخیره سازی می شود. در صورتی که منبع یا منابعی که اجرای کار را برعهده داشته اند، خراب شوند، اجرای کار را می توان از آخرین وضعیت ذخیره شده ی آن (در عوض اجرای آن از ابتدا)، ادامه داد. اما بکارگیری این تکنیک با توجه به مقیاس بسیار بزرگ محیط گرید، سبب ایجاد پیچیدگی هایی به خصوص در مورد همگام سازی فرآیندهای همزمانی که دارای ارتباطاتی با یکدیگر هستند، می گردد [Dab2009]. علاوه بر این، سربارهای تحمیلی این تکنیک ها کاملا قابل ملاحظه می باشد. برای مثال برای یک کار با ۱۰۰ ساعت نیاز پردازشی، تخمین سربار عمل check-pointing با استفاده از روش های فعلی، برابر با ۱۵۱ ساعت می باشد [Fu2010]. علاوه بر روش های تحمل خرابی، که رفتاری واکنشی[2] دارند، بعضی تلاش ها نیز صورت گرفته اند که رفتار پیشگیرانه[3] را پیشنهاد می نمایند. برخلاف رهیافت واکنشی که انجام اعمال مناسب به بعد از وقوع خرابی موکول می شود، در رهیافت

[1] Fault-Tolerance
[2] Reactive
[3] Proactive

پیشگیرانه، هدف، پیشگیری از وقوع خرابی است و در نتیجه، عکس العمل روش هایی که از این رهیافت تبعیت می نمایند، قبل از رخداد خرابی صورت می گیرد. روش های زمانبندی آگاه از خرابی (اطمینان)[1] از جمله ی این روش ها می باشند. در فصل بعد، بررسی کارهای پیشین، به برخی از چنین تلاش هایی خواهیم پرداخت.

۱-۲-۲- عوامل سمت کار

در مورد عواملی از خرابی کارها که از کارافتادگی منابع در بروز آنها دخالتی ندارند، می توان نمونه هایی به شرح زیر را مشاهده نمود. وجود اشکالات نرم افزاری در برنامه (کار) ارسالی توسط کاربر [Dua2006]، تخطی کار از میزان سهمیه‌ی اعطایی منبع، که توسط سیاست‌های آن منبع مشخص می‌گردد، انتظار بسیار زیاد در صف مدیر بارِ کاری، اتمام اعتبارنامه ی[2] کاربر، لغو[3] کردن اجرای کار توسط کاربر ارسال کننده در حین اجرا [Li2006].

با دقت در مثال های فوق می توان نقطه ی مشترکی را یافت. این نقطه ی مشترک، "کار" است. در واقع در نمونه های ذکر شده، کار به طور مستقیم و یا غیر مستقیم دخیل می باشد. به همین دلیل، در ادامه ی این نوشتار، چنین عواملی از خرابی کار را، عوامل سمت کار می نامیم. دو کلاس مختلف از دلایل ریشه ای برای خرابی های سمت کار را می توان شناسایی نمود.

- کلاس اول، شامل دلایلی است که به ویژگی های منبع اجرا کننده ی کار مرتبط است. عدم سازگاری کار با منبع و یا ناکافی بودن فضای آزاد دیسک سخت یا حافظه ی

[1] Failure (Reliability)- Aware Scheduling
[2] Credential
[3] Cancel

اصلی منبع، نمونه هایی از این کلاس می باشند. رفع این نوع از خرابی ها، با انتخاب منبع مناسب برای اجرای کار امکانپذیر است.

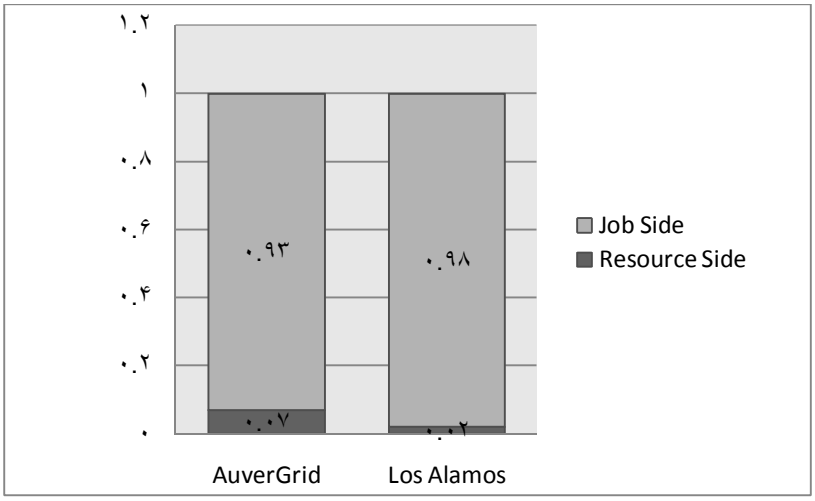
- کلاس دوم، شامل دلایلی است که به ویژگی های منبع اجرا کننده ی کار ارتباطی ندارد و در واقع مستقل از آنها می باشند. برای مثال، در [Cie2008]، خرابی یک الگوریتم داده کاوی[1] گزارش شده است. پس از بررسی های متعدد، مشخص شد که دلیل خرابی، تولید پرونده ی[2] خروجی حجیم توسط الگوریتم بوده است. در واقع به سبب حجیم بودن مجموعه داده ی[3] ورودی، پرونده ی خروجی تولیدی، بزرگتر از حداکثر اندازه ی یک پرونده در سیستم پرونده[4] گشته، و لذا اجرای الگوریتم دچار خرابی می شده است. از نمونه های دیگر در این کلاس می توان به وجود اشکالات نرم افزاری در کد برنامه ی ارسالی و یا لغو نمودن اجرای برنامه توسط کاربر را اشاره نمود. در این دسته از خرابی ها، ویژگی های منبع اجرا کننده نمی توانند مقصر باشند، در واقع اجرای این کارها در هر ماشینی، سبب اجرای ناموفق آنها می گردد [Cie2008].

یکی از صدمات اصلی خرابی هایی که به سبب عوامل سمت کار ایجاد می شوند، هدر دادن زمان منابعی است که برای اجرا به آنها اختصاص داده شده اند [Shr2008]. پیامد اصلی هدر رفت منابع، ایجاد تاخیر در آغاز اجرای سایر کارهایی است که در سیستم حضور دارند (افزایش زمان انتظار کارها). اگرچه تحقیقات گسترده ای در مورد روش های مقابله با خرابی

[1] Data Mining
[2] File
[3] Data Set
[4] System File

هایی از کارهـا کـه بـه سـبب عوامـل سـمت منبـع رخ داده انـد، انجـام شـده اسـت، امـا در مـورد عوامـل سمت کار، تلاش های به مراتب کمتری صورت گرفته است.

بیـان ایـن نکتـه خـالی از لطـف نیسـت کـه هـر یـک از عوامـل سـمت منبـع و سـمت کـار، بخشی از خرابی کارها را در محـیط سـبب مـی شـوند، امـا درصـد سـهم آنهـا در کـل خرابـی هـا متفـاوت اسـت. در شـکل ۱-۱، سـهم هـر یـک از عوامـل فـوق در خرابـی کارهـا در آورگریـد و یکـی از سیسـتم های آزمایشـگاه ملـی لـس آلامـوس نشـان داده شـده اسـت. در ایـن دو مثـال، سـهم عوامـل سـمت کـار در خرابی کارها، بیش از عوامل سـمت منبـع مـی باشـد، امـا ایـن امـر بـدین معنـی نیسـت کـه در همـه ی محیط هـا، عوامـل سـمت کـار، غالـب باشـند. در محیطـی بـا منـابع مطمـئن و نـرخ پویـایی پـایین، و یا محیطی با روش های تحمـل خرابـی کـارآ، انتظـار ایـن اسـت کـه عوامـل سـمت منبـع تـاثیر کمتـری را در بـروز خرابـی کارهـا داشـته باشـند، در حـالی کـه در چنـین محـیط هـایی، عوامـل سـمت کـار، نقش اصلی را در این امر ایفا خواهند نمود.



شکل ۱-۱: سهم عوامل تاثیرگذار در خرابی کارها در آورگرید و سیستم شماره ی ۲۰ آزمایشگاه ملی لس آلاموس