مدل تطبیق‌پذیر مدیریت برنامه به‌منظور افزایش بهره‌وری بهینه منابع سرورها

آی‌ناز غفرانی[۱]، سمیه طاهری[۲]، مازیار گودرزی[۳]

[۱]گروه مهندسی کامپیوتر، دانشگاه علم و فرهنگ، تهران،
ghofrani.ainaz@gmail.com
[2]گروه مهندسی کامپیوتر،دانشگاه صنعتی شریف، تهران،
s.taheri32@gmail.com
[3]گروه مهندسی کامپیوتر،عضو هیأت علمی دانشگاه صنعتی شریف، تهران،
goudarzi@sharif.ir

چکیده

رایانش ابری یکی از جدیدترین راه‌حل‌های ارائه‌شده برای دوام سازمان‌ها در فضای رقابتی است. یک سازمان را در نظر بگیرید که قصد مهاجرت به زیرساخت رایانش ابری را دارد. حال مسئله این است که آیا سازمان می‌تواند به کارایی برنامه‌ها در ابر اطمینان کند؟ از چه راهکاری جهت نزدیک کردن منابع پردازشی موجود در ماشین فیزیکی با منابع پردازشی مورد نیاز برنامه‌ها می‌توان بهره برد تا بدین طریق مانع از بروز مشکلاتی چون عدم بهره‌وری بهینه از منابع پردازشی، عدم رضایت مشتریان، زمان پاسخ بالا و میزان گذردهی پایین برای اجرای برنامه‌ها شد.

در این مقاله با ارائه یک مدل تطبیق‌پذیر، گلوگاه کارایی را پیش‌بینی می‌نماییم تا بتوان قبل از وقوع گلوگاه اقدام مناسب انجام داده جلوی کاهش کارایی را گرفت. رویکرد پیشنهادی این است که مقدار منابع پردازشی مورد نیاز در ماشین‌های فیزیکی پیش‌بینی شود. سپس با محاسبه ارتباط میان میزان منابع پردازشی و هریک از سنجه‌های کارایی، مقدار زمان پاسخ و میزان گذردهی در ماشین‌های فیزیکی پیش‌بینی می‌گردد وسپس با شبیه‌سازی درخواست سرویس‌ها،داده‌های مورد نظر جهت محاسبه مدل کارایی تولیدمی‌شوند. سپس با مقایسه مقدار هر منبع با حد آستانه در نظر گرفته شده برای آن، گلوگاه کارایی پیش‌بینی می‌گردد. در انتها با استفاده از آزمایش‌های مختلف، اقدام به ارزیابی عملکرد مدل نمودیم. در انتهای مقاله، با بررسی نتایج به‌دست‌آمده و انجام تحلیل-های آماری لازم، نقاط قوت و ضعف مدل را بیان نموده و راهکارهایی برای ادامه پژوهش پیشنهادی ارائه می‌نماییم.

کلمات کلیدی

محاسبات ابری، مدل تطبیقی، کارایی برنامه، بهره‌وری منابع سرورها

۱- مقدمه

علی‌رغم تمایل شرکت‌ها و سازمان‌ها به استفاده از سرویس‌های بستر[۱] رایانش ابری، بسیاری از آن‌ها به دلیل نگرانی‌های مرتبط با کارایی، از انتقال کل برنامه‌های خود به ابر تردید دارند [۱]. ارائه‌دهندگان سرویس بستر جهت سرویس‌دهی برنامه‌های مشتریان خود، از سرویس‌ها و منابع زیرساخت استفاده می‌کنند. منابع با استفاده از فناوری مجازی‌سازی در میان برنامه‌های میزبانی‌شده به اشتراک گذاشته می‌شود؛ به این چارچوب، ساختار مجازی اشتراکی[۲] گفته می‌شود. در این شرایط به دلیل آن‌که الگوی مصرف منابع و بارکاری[۳] برنامه در طول زمان تغییر می‌کند [۲]. ممکن است مجموع منابع پردازشی موردنیاز کل ماشین‌های مجازی هم‌میزبان[۴]، بیشتر از منابع پردازشی ماشین فیزیکی شود بنابراین در عملبا همسایگان پرکار[۵] در ماشین‌های مجازی هم‌میزبان و رقابت آن‌ها بر سر منابع پردازشی مواجه هستیم که موجب کاهش کارایی کل ساختار می‌شود. همچنیندر مواقعی که با هجوم ناگهانی بارکاری روبرو شویم، زمان پاسخ[۶]افزایش یافتهو درنتیجه میزان گذردهی[۷]کاهش می یابد. بنابراین در این ساختار مجازی اشتراکی، تخصیص

Archive of SID
www.SID.ir

درست منابع از اهمیت بسیار بالایی در کارایی نهایی برنامه برخوردار بوده و تاثیر مستقیم بر میزان رضایت کاربران نهایی و مشتریان سازمان خواهد داشت.

در زمینه عیب‌یابی کارایی برنامه‌ها در ابر پژوهش‌های بسیاری صورت گرفته که یکی از مهم‌ترین آن‌ها محیط کاری Vperfguard[۸][۹] می‌باشد. این محیط از طریق یافتن رابطه و نسبت میانتغییرات مقادیر کارایی و منابع مصرفی،نقاط و گلوگاه‌های کاراییمانند منابع در حال رقابت، و میزبانی که رقابت روی آن شکل‌گرفته است را کشف و در اختیار ارائه‌دهنده سرویس قرار می‌دهند.و این نتایج سپس به منظور انجام عملیات اصلاحی، نظیر مهاجرت یا تغییر اندازه ماشین مجازی، مورد استفاده قرار می‌گیرد [۳]. بنابراین این محیط کاری می‌تواند به بهبود کارایی برنامه‌های در حال اجرا کمک کند. اما نقطه ضعف این مدل و محیط کاری Vperfguard آن استکه بر روی افزایش بهره‌وری[۱۰] منابع پردازشی و تخصیص بهینه[۱۱] منابعو نیز پیش بینی و درنتیجه پیشگیری از وقوع گلوگاهمتمرکز نشده است.

بنابراین مسئله اصلی، عدم بهره‌وری مناسب از منابع پردازشی موجود در میزبان‌ها در شرایط رخداد گلوگاه می‌باشد. واضح است که پایین بودن میزان بهره‌وری منابع پردازشی موجب نارضایتی مشتری از کارایی برنامه خواهد شد. ازاین‌رو بهبود Vperfguard برای مدیریت بهتر و مناسب‌تر انواع مختلف برنامه اهمیت فراوانی دارد. این بهبود باید بتواند با پیش‌بینی گلوگاه کارایی، از آن پیشگیری کرده و موجب افزایش بهره‌وری منابع و مانع از تخطی از قرارداد سطح سرویس[۱۲] شود.برآورده کردن این موضوع، هدف و زمینه کاری این مقاله می باشد.

۲- تحقیقات مرتبط

در این بخش مروری بر کارهای انجام شده در حوزه کارایی برنامه‌ها در ابر خواهیم داشت.یکی از مهم‌ترین چالش‌های موجود در محاسبات ابری، بحث کارایی برنامه‌ها می‌باشد. اگر بخواهیم دسته‌بندی کاملی از کارهای انجام‌شده در این حوزه داشته باشیم، روش‌های پیشنهادی را می‌توان به سه دسته جعبه سفید، داده محور و vperfguard تقسیم کرد.حال به‌صورت مختصر مروری بر هر دسته خواهیم داشت.

۲-۱- شیوه جعبه سفید

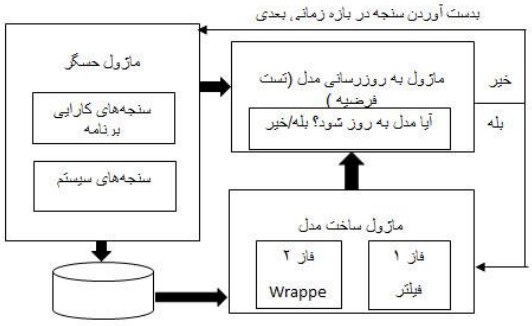
دراین شیوه ارائه‌دهندگان سرویس IT با تکیه‌بر تجربه تخصصی خود و با کمک ابزارهای نظارتو سیستم‌های ثبت وقایع و یا توسط روال‌هایی استاندارد در عیب‌یابی کارایی، گلوگاه‌های بالقوه کارایی را شناسایی و علت اصلی آن را می‌یابند[4] .به دلیل پویایی زیاد و مقیاس‌پذیری بالایمحیط رایانش ابری،ازاین‌روش نمی‌توان در عیب‌یابی کارایی در محیط ابر بهره برد.

۲-۲- شیوه داده محور

در این شیوه جهت مدیریت و عیب‌یابی کارایی به‌صورت خودکار از فن‌های آماری و سنجه‌های گردآوری‌شده از برنامه‌ها و سیستم‌ها استفاده می‌گردد. جهت شناسایی این سنجه‌هااز مدل‌های احتمالاتی استفاده می‌شود [۵]. در این شیوه به دلیل داده‌های غیر مرتبط بسیار زیاد و صرف زمان بسیار طولانی جهت بیان ارتباط میان آن‌ها، عیب‌یابی کارایی بدین روش نیازمند صرف زمان و هزینه زیادی است [۶].

۲-۳- Vperfguard

این محیط کاری، ارتباط بین سنجه‌های کارایی و سنجه‌های سیستم مرتبط با آن را توسط مدل‌های پیشگوییان می‌کند. مدل‌های پیشگو شامل linear regression، Regression tree، K nearest neighbor، approachBoosting می‌باشد که رگرسیون خطی به‌عنوان مدل کارایی پیش‌فرض انتخاب شده است.همان‌طور که در شکل ۱ نشان داده‌شده است، معماری VperfGuard متشکل از سه پیمانه به شرح زیر می‌باشد:



شکل۱: محیط کاری VperfGuard[3]

۱) پیمانه حسگر:این پیمانه به گردآوری سنجه‌های سیستم و سنجه‌های کارایی برنامه می‌پردازد. سنجه‌های کارایی برنامه مانند زمان پاسخ و گذردهی می‌باشند که به ترتیب با THR,MRT نمایش داده می‌شوند.

۲) پیمانه ساخت مدل:این پیمانه، توسط سنجه‌های سیستم گردآوری‌شده در پیمانه حسگر، مدل هر یک از سنجه‌های کارایی را ایجاد می‌کند. اما ازآنجاکه بسیاری از این سنجه‌ها وابسته به سنجه‌ی کارایی نمی‌باشند، موجب کاهش دقت پیشگویی مدل می‌شوند. یا بسیاری از آن‌ها تکراری اند که بر دقت پیشگویی مدل بی‌تأثیر اما موجب بزرگ شدن فضای جستجو و درنتیجه کند شدن الگوریتم یادگیری ماشین و گران شدن مدل ازلحاظ محاسباتی می‌شوند. لذا قبل از ایجاد مدل، این پیمانه از طریق الگوریتم‌های انتخاب خصیصه، در فاز Filter به انتخاب سنجه‌هایی وابسته به سنجه‌های کارایی و از میان آن‌ها در فاز wrapper به انتخاب سنجه‌های غیرتکراری می‌پردازد.

در الگوریتم انتخاب خصیصه، تمام سنجه‌ها، ارزیابی می‌شوند و تنها سنجه‌هایی انتخاب می‌گردند که طبق معیار ارزیابی باشند. در Filter معیار ارزیابی،همبستگی می‌باشد. ورودی این مرحله سنجه‌های سیستم گردآوری‌شده در پیمانه حسگر و خروجی آن سنجه‌هایی با ضریب همبستگی بالا با سنجه‌های کارایی می‌باشد. در فاز wrapperمعیار ارزیابی، مدل‌های پیشگوییمی‌باشد. ورودی این مرحله، خروجی Filter است و از میان آن‌ها زیرمجموعه‌ای از سنجه‌ها انتخاب می‌شوند که بالاترین دقت پیشگویی را در مدل کارایی موردنظر دارند. دقت پیشگویی توسط ضریب تعیین اندازه‌گیری می‌شود.

ازآنجاکه فرایند استخراج کل زیرمجموعه‌های سنجه بسیار گران است از الگوریتم تپه‌نوردیاستفاده می‌گردد. این الگوریتم در هر تکرار، از میان سنجه‌هایورودی، سنجه‌ای با بیشترین افزایش دقت پیشگویی مدل کارایی موردنظر را در مجموعه‌ی خروجی قرار داده و از مجموعه ورودی حذف می‌کند و این الگوریتم زمانی پایان می‌یابد که این افزایش، کمتر

Archive of SID
www.SID.ir

از حد آستانه باشد. حال مدل کارایی موردنظر با سنجه‌های خروجی از فاز۲ ایجاد می‌گردد.

۳) پیمانه به‌روزرسانی مدل:مدل ایجادشده توسط پیمانه ساخت مدل، ممکن است در اثر هجوم ناگهانی بارکاری و رقابت بر سر منابع ارتباط بین کارایی و منابع سیستم را به‌درستی بیان نکند. که به آن نقطه تغییر می‌گویند. بنابراین نیاز است این نقطه، کشف و مدل به‌روز شود. کشف این نقطه، توسط آزمون فرضیه در این پیمانه انجام می‌گیرد سپس جهت به‌روزرسانی سنجه‌های پیشگو و مدل کارایی، پیمانه ساخت مدل، فراخوانی می‌شود. در آزمون فرضیه، ضریب همبستگی بین سنجه کارایی و سنجه سیستم بررسی می‌شود.

۲- ۲-٤. مقایسه شیوه‌های عیب‌یاب کارایی

هریک از روش‌های عیب‌یاب کارایی برنامه‌ها در ابر شرح داده شد و معایب و محدودیت‌های هریک بیان شد. اکنون در جدول ۱ از جنبه‌های مختلف به مقایسه روش‌ها با یکدیگر و مدل پیشنهادی پرداخته می‌شود. به دلیل پویایی زیاد و مقیاس‌پذیری بالایمحیط رایانش ابری، جهت عیب‌یابی کاراییدر محیط ابرازروش جعبه سفید نمی‌توان بهره برد. در شیوه داده محور، به دلیل وجود داده‌های غیر مرتبط بسیار زیاد و صرف زمان بسیار طولانی جهت بیان ارتباط میان آن‌ها، عیب‌یابی کارایی بدین روش نیازمند صرف زمان و هزینه زیادی است. محیط vperfgurdلحظه‌ی افت کارایی را کشف می‌کند. تفاوت اصلی مدل پیشنهادی با سایر مدل‌ها در پیش‌بینی و درنتیجه پیشگیری از وقوع گلوگاه کارایی است.

جدول ۱: مقایسه مدل پیشنهادی با روش‌های پیشین

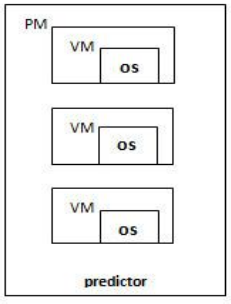
مرجع	روش‌های آنالیز کارایی	قابلیت استفاده در رایانش ابری		عیب یابی خود کار	مقرون به‌صرفه ازلحاظ هزینه و زمان	پیش-بینی گلوگاه کارایی
		مقیاس-پذیری	تطبیق-پذیری			
[7,8, 9,10]	جعبه سفید	X	X	X	X	X
[11]	داده محور	√	√	√	X	X
[3]	Vperfg uard	√	√	√	√	X
	مدل پیشنهادی	√	√	√	√	√

۳- رویکرد پیشنهادی

شیوایی و در این بخش رویکرد موردنظر برای حل مسئله را عنوان می‌کنیم.با توجه به تغییراتبارکاری برنامه‌ها در طول زمان وخاصیت ارتجاعی ماشین‌های مجازی، مقدار منابع پردازشی مورد نیازماشین‌های مجازیبه‌صورت پویا تغییر می‌کند.البته باید متذکر شد که حد آستانه‌ی حداکثری برای منابع پردازشی تخصیص یافته به ماشین‌های مجازی وجود دارد. ضمن آن‌که مشتری، تحلیلی از حداکثر میزان منابع پردازشی مورد نیاز برنامه‌هارا در اختیار ارائه‌دهنده سرویس قرار می‌دهد؛ لذا تخصیص برنامه‌ها به ماشین‌های مجازی براساس تأمین حداکثر میزان منابع پردازشی موردنیاز برنامه‌ها و حد آستانه حداکثر منابع پردازشی ماشین‌های مجازی صورت می‌گیرد. حال ازآن‌جاکه منابع پردازشی مورد نیاز برنامه‌ها در اکثر مواقع در نقطه‌ای کمتر از حداکثر منابع پردازشی مورد نیازشان می‌باشد. بنابراین در اکثر مواقع، مقدار منابع پردازشی اختصاص‌یافته به ماشین‌های مجازی کمتر از حد آستانه می‌باشد و مقداری از منابع پردازشی در ماشین فیزیکی بدون استفاده باقی می‌ماند؛ لذا ارائه‌دهندگان سرویس، جهت کسب سود بیشتر و استفاده حداکثر از منابع پردازشی، تخصیص ماشین‌های مجازی به ماشین فیزیکی را به‌جای آن‌که براساس حداکثر میزان بارکاری ماشین‌های مجازی انجام دهند، براساس میزان بارکاری‌ای که در اکثر مواقع رخ می‌دهد، انجام می‌دهند. بنابراین اگر میزان منابع پردازشی موردنیاز ماشین‌های مجازی هم‌میزبان در نقطه اوج خود باشند، میزان بارکاری تخصیص داده‌شده به ماشین فیزیکی با منابع پردازشی موجودش تطبیق نداشته و در اثر رقابت ماشین‌های مجازی هم میزبان بر سر منابع پردازشی با عدم بهره‌وری از منابع پردازشی و در نتیجه افزایش زمان پاسخ و کاهش میزان گذردهی مواجه خواهیم بود؛ که موجب تخطی از قرارداد سطح سرویس می‌شود و ارائه‌دهنده سرویس، مسئول پاسخ‌دهی کاهش کارایی به مشتریان می‌باشد.

بنابراین رویکرد پیشنهادی برای حل مسئله به این صورت است که اگر ما بتوانیم منابع پردازشی مورد نیاز ماشین‌های مجازی واقع در یک ماشین فیزیکی را پیش‌بینی نماییم، توانسته‌ایم با مجموع منابع پردازشی آن‌ها، منابع پردازشی موردنیاز ماشین فیزیکی را پیش‌بینی نماییم و با استفاده از آن و ارتباطی که میان منابع پردازشی و سنجه‌های کارایی در محیط کاری Vperfguard بیان‌شده، سنجه‌های کارایی هریک از ماشین‌های فیزیکی را در بازه‌های زمانی مشخص در آینده را بسنجیم و طی آن رخداد گلوگاه کارایی را پیش‌بینی نماییم. لذا ارائه‌دهندگان سرویس می‌توانند قبل از رسیدن به زمان رخداد گلوگاه، تصمیمی در خصوص تخصیص بارکاری، متناسب با منابع موجود ماشین فیزیکی در بازه مورد نظر بگیرند.

جهت انجام عملی این رویکرد از واحدی به نام پیشگو استفاده می‌شودکه بر روی هر ماشین فیزیکی در حال اجراست.همانطور که می‌دانیم هر ماشین فیزیکی میزبانی چند ماشین مجازی را بر عهده دارد. هر ماشین مجازی، سیستم عامل مربوط به خود را دارد که داده‌های مربوط به میزان منابع پردازشی مورد نیاز ماشین‌های مجازی را در اختیار دارد. سیستم‌عامل‌های ماشین‌های مجازی هم‌میزبان این داده‌ها را از طریق ماشین فیزیکی به واحد پیشگو انتقال می‌دهند. این واحد از طریق داده‌های مربوط به گذشته، داده‌های مربوط به آینده را پیش‌بینی می‌کند که مدل مفهومی آن در شکل۲ قابل مشاهده است. از آن‌جا که پیشگویی در هر ماشین‌فیزیکی به صورت مجزا صورت می‌گیرد بنابراین سربار بسیار کم است.



شکل ۲: مدل مفهومی واحد

Archive of SID
www.SID.ir

٤- پیاده‌سازی و اجرا

در این بخش سعی در پیاده‌سازی راه‌حل پیشنهادی خواهیم داشت. این کار با استفاده از آزمایش‌های مختلف صورت می‌گیرد. هر آزمایش را به‌طور دقیق بیان کرده و تمام جزئیات موردنیاز را بررسی خواهیم کرد.از آن‌جا که میزان منابع پردازشی تخصیص‌یافته به ماشین‌های مجازی با توجه به تعداد درخواست‌های برنامه‌های در حال میزبانی تغییر می‌کند. لذا محک مناسب جهت پیاده‌سازی مدل پیشنهادی، محکی است که در آن مشخصات ماشین-های مجازی در طول زمان تغییر کند. از آن‌جا که محکی با مشخصات مورد نظر در دسترس نبود، از محک دست‌ساز استفاده کردیم. برای تولید محکهای دست ساز، فرض کردیمواحد پیشگوپروفایل بارکاری ماشین‌های مجازی تا ۲۰۰ ثانیه بعد را در اختیار دارد. این پروفایل توسط رفتار ماشین‌های مجازی یعنی میزان منابع پردازشی مورد نیاز و زمان آغاز و پایان اجرا توصیف شده است. منابع پردازشی شامل پردازنده، حافظه و پهنای باند می‌باشد. ماشین‌های مجازی، میزبانی تعدادی وظایف را به عهده دارند و هریک از وظایف در بازه زمانی مشخص اجرا می‌شود. تعداد ماشین‌های مجازی ۱۰۰ در نظر گرفته-شده‌است. برای منابع پردازشی ماشین‌های مجازی حدبیشینه‌ایدر نظر می-گیریم که در جدول ۲ شرح داده‌شده‌است.

جدول ۲: حد حداکثر مشخصات ماشین‌های مجازی

نوع منبع پردازشی	پردازنده GHZ	حافظه MB	پهنای باند MB/s
میزان منابع	1.75	8192	20.027

جهت تخصیص ماشین‌های مجازی به ماشین‌های فیزیکی، به تعداد کافی ماشین فیزیکی همگن موجود می‌باشد. مشخصات ماشین فیزیکی در جدول ۳ شرح داده‌شده‌است.

جدول۳: مشخصات ماشین‌های فیزیکی[12]

نوع منبع پردازشی	پردازنده GHZ	حافظه MB	پهنای باند MB/s
میزان منابع	۳.۵	16384	40.054

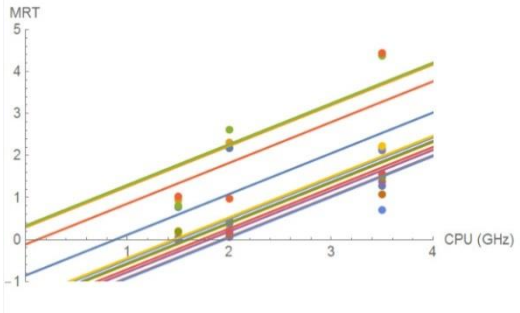
جهت گرداوری داده‌های اولیه، نیاز است با تخصیص ماشین‌های مجازی به ماشین‌های فیزیکی و اجرای ماشین‌های مجازی برای ۳ دوره،منابع پردازشی مورد نیاز ماشین‌های فیزیکی را در بازه‌های زمانی مختلف به دست آوریم. این تخصیص توسط مسئله کلاسیک بسته‌بندی صورت داده‌شده‌است کهسیاست-های مختلفی از قبیل قبیل اولین جایگذاری[13] و بهترین جایگذاری[14] و چرخشی[15] و ... جهت حل آن وجود دارد [13]. در این مقاله از روش اولین جایگذاری استفاده شده‌است.خروجی ۲۶ ماشین فیزیکی بارکاری تخصیص-یافته در بازه‌های زمانی در ماشین‌های فیزیکی می‌باشد.حال باید مدل‌های کارایی که همان رابطه میان میزان گذردهی و میزان منابع پردازشی موردنیاز و همچنین رابطه میان زمان پاسخ و میزان منابع پردازشی موردنیاز است را به دست آوریم. همان‌طور که بیان شد این مدل‌های کارایی، همان معادله رگرسیون می‌باشند که جهت محاسبه‌ی آن‌ها نیازمند در اختیار داشتن پایگاه داده‌ای با فیلدهای میزان گذردهی، زمان پاسخ و میزان هریک از منابع پردازشی موردنیاز هستیم.لذا با شبیه‌سازی بار ترافیکی که شامل ۳۶ سناریو برای وضعیت‌های مختلف مجموع منابع پردازشی ماشین‌های مجازی هم-میزبان می‌باشد و محاسبه متوسط زمان پاسخ و میزان گذردهی در هر سناریو، پیش نیاز لازم جهت محاسبه رگرسیون را در اختیار داریم.نتایج حاصل از این آزمایش در جدول 4 قابل‌مشاهده است.

جدول ٤: نتایج حاصل از اجرای آزمایش دوم با کمک سناریوهای مختلف

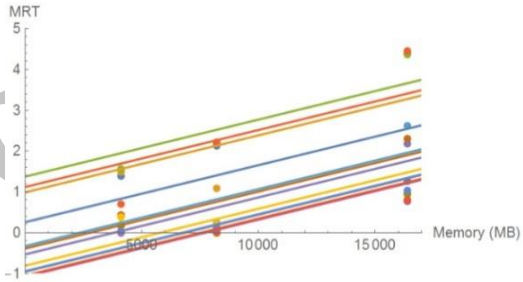
سناریو	مجموع منابع پردازشی ماشین-های مجازی هم‌میزبان			متوسط زمان پاسخ	درصد افزایش زمان پاسخ نسبت به انتظار	مقدار گذردهی	درصد تعداد درخواست‌های بی‌پاسخ
	قدرت پردازشی موردنیاز GHZ	حافظه موردنیاز MB	پهنای باند موردنیاز MB/s				
1	3.5	16384	10	253	1.271	7	0.632
2	3.5	8192	20	232	1.082	10	0.474
3	3.5	4096	40	189	0.696	14	0.263
4	2	4096	20	161	0.445	15	0.211
5	2	16384	40	220	0.974	8	0.579
6	2	8192	30	138	0.239	17	0.105
7	2	4096	10	130	0.167	18	0.053
8	1.5	4096	20	133	0.194	17	0.105
9	1.5	8192	40	125	0.122	18	0.053
10	1.5	16384	30	201	0.804	12	0.368
11	3.5	16384	20	602	4.403	8	0.579
12	3.5	16384	30	598	4.367	5	0.737
13	3.5	16384	40	607	4.448	2	0.895
14	3.5	8192	10	348	2.123	16	0.158
15	3.5	8192	30	356	2.195	16	0.158
16	3.5	8192	40	359	2.222	13	0.316
17	3.5	4096	10	266	1.387	18	0.053
18	3.5	4096	20	277	1.486	17	0.105
19	3.5	4096	30	286	1.567	17	0.105
20	2	4096	30	130	0.167	19	0.000
21	2	4096	40	155	0.391	18	0.053
22	2	16384	10	354	2.177	17	0.105
23	2	16384	20	369	2.312	16	0.158
24	2	16384	30	403	2.617	13	0.316
25	2	8192	10	119	0.068	19	0.000
26	2	8192	20	121	0.086	19	0.000
27	2	8192	40	135	0.212	19	0.000
28	1.5	4096	10	110	0.000	19	0.000
29	1.5	4096	30	115	0.032	19	0.000
30	1.5	4096	40	114	0.023	18	0.053
31	1.5	8192	10	101	0.000	19	0.000
32	1.5	8192	20	106	0.000	19	0.000
33	1.5	8192	30	113	0.014	19	0.000
34	1.5	16384	10	198	0.777	17	0.105
35	1.5	16384	20	217	0.948	15	0.211
36	1.5	16384	40	226	1.028	15	0.211

Archive of SID
www.SID.ir

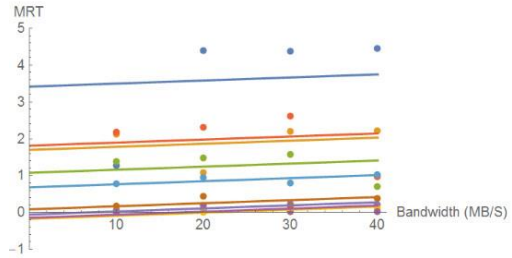
حال نمودار پراکندگی را برای نقاط زمان پاسخ و منابع پردازشی سناریوهاو همچنین برای نقاط میزان گذردهی و منابع پردازشی سناریوها ترسیم می‌کنیم و با برازش خطی آن‌ها مدل‌های کارایی را به دست می‌آوریم. نمودار پراکندگی و برازش خطیمربوط به نقاط زمان پاسخ و منابع پردازشیدر نمودار-های 1و2و3 نشان داده شده است.



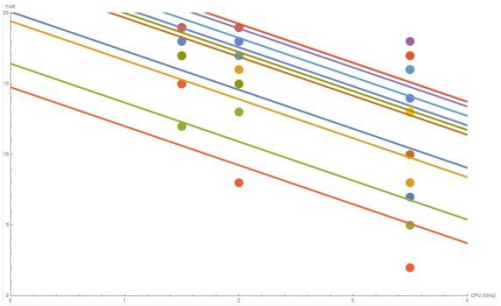
نمودار ۱: نمودار پراکندگی و برازش خطی برای مجموع نقاط زمان پاسخ و پردازنده



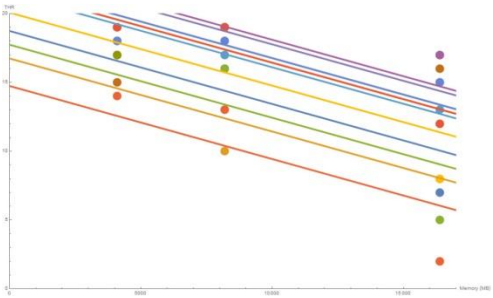
نمودار ۲: نمودار پراکندگی و برازش خطی برای مجموع نقاط زمان پاسخ و حافظه



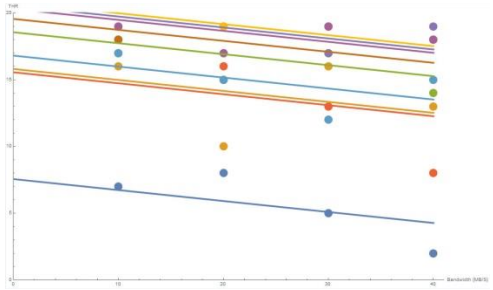
نمودار ۳: نمودار پراکندگی و برازش خطی برای مجموع نقاط زمان پاسخ و پهنای باند



نمودار ٤: نمودار پراکندگی و برازش خطی برای مجموع نقاط گذردهی و پردازنده



نمودار ٥: نمودار پراکندگی و برازش خطی برای مجموع نقاط گذردهی و حافظه



نمودار ٦: نمودار پراکندگی و برازش خطی برای مجموع نقاط گذردهی و پهنای باند

مدل‌های کارایی شامل درصد افزایش زمان پاسخ در فرمول۱ و درصد تعداد درخواست‌های پاسخ داده نشده در فرمول ۲ نشان داده شده است.

MRT-inc٪=(3.3959*(درصد پردازنده موردنیاز)+ 2.2868*(درصد پهنای باند موردنیاز) + 0.3326*(حافظه موردنیاز) -2.6654

فرمول۱

Not-THR=(0.5081*(درصد پردازنده موردنیاز)+0.4574*(درصد پهنای باند موردنیاز)+0.1734*(حافظه موردنیاز) -0.5149

فرمول۲

حال با در اختیار داشتن مدل‌های کاراییمی‌توان در هریک از بازه‌های زمانی در ماشین فیزیکی، مقدار هریک از سنجه‌های کارایی را به دست آورد. برای هریک از سنجه‌های کارایی حد آستانه‌ی ۰.۷۵ در نظر گرفته شده‌است که با مقایسه‌ی هریک از مقادیر به‌دست‌آمده با آن می‌توان گلوگاه کارایی را پیش-بینی کرد.

۵- ارزیابی

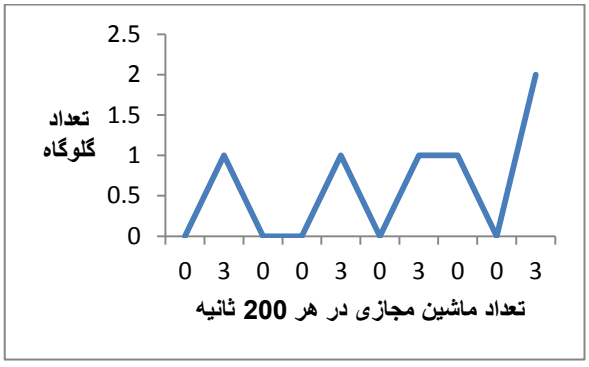
در این بخش آزمایش‌هایی به‌منظور ارزیابی و سنجش صحت عملکرد مدل پیشنهادی صورت می‌گیرد.بدین منظورابتدا بررسی می‌نماییم که آیاتعدادگلوگاه‌های پیش‌بینی‌شده در بازه‌های زمانی در یک ماشین فیزیکی با تعداد ماشین‌های مجازی میزبانی‌شده در آن بازه متناسب است یا خیر. این

Archive of SID
www.SID.ir

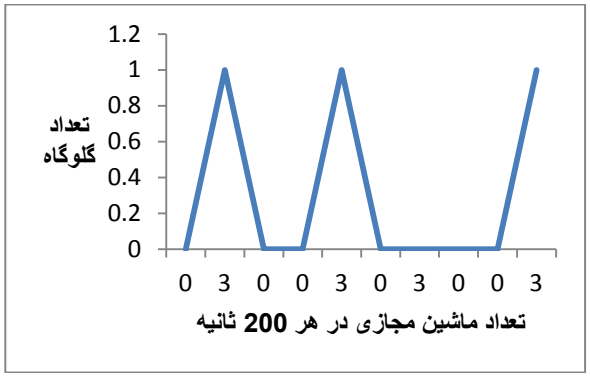
آزمایش بر روی یکی از ماشین‌های فیزیکی انجام شده‌استکه مشخصات ان در جدول ۵ قابل مشاهده است. همان‌طور که پیش از این بیان شد، فرض آن است که ۲۰۰ ثانیه‌ی آینده از پروفایل بارکاری را در اختیار داریم. لذا تعدادماشین‌های مجازی میزبانی‌شدهو همچنین تعداد گلوگاه‌هاپیش‌بینی‌شده در هریک از سنجه‌های کارایی در ماشین فیزیکی مورد نظر در هر ۲۰۰ ثانیه تعیین می‌شود.ضمن آن‌که حد آستانه ۰.۷۵ در نظر گرفته شده است. با تعیین محور عمودی به عنوان تعداد گلوگاه کارایی و محور افقی به عنوان تعداد ماشین مجازی در هر ۲۰۰ ثانیه و با رسم نمودار نقطه‌ای دو شکل۳و۴ بدست امده استهمان‌طور که مشاهده می‌شود تعداد ماشین‌های مجازی تقریبا ثابت است اما تعداد گلوگاه‌های کارایی با گذر زمان افزایش یافته است. بنابراین می‌توان نتیجه گرفت که تنها دلیل افزایش گلوگاه کارایی، ازدیاد ماشین‌های مجازی نیست.

جدول ۵: مشخصات یک ماشین فیزیکی

t0	t1	cpu used% GHZ	Memory requirement% Mb	bandwidth% Mb/s	Mrt	0.75	Thr	0.75
229	260	0.155413	0.222473	0.214287	0.000	0	1	0
260	267	0.405971	0.574707	0.666672	905.159	1	0.930154	0
267	500	0.529576	0.802185	1.000008	1696.762	1	0.705502	1
500	1274	0.834263	0.839966	0.80953	4325.898	1	0.566438	1
1274	1400	0.630301	0.512024	0.380955	4753.206	1	0.894386	0
1400	1502	0.771205	0.674927	0.261907	5099.964	1	0.768924	0
1502	1917	0.380022	0.54248	0.428575	6508.324	1	0.999365	0
1917	1932	0.311105	0.414673	0.238097	0.000	0	1	0
1932	1938	0.51423	0.551208	0.595243	6579.948	1	0.898282	0
1938	1945	0.469587	0.774231	0.500004	6603.691	1	0.835469	0
1945	1992	0.785993	0.919617	0.928579	6764.071	1	0.533889	1
1992	1995	0.478516	0.267151	0.80953	0.000	0	1	0
1995	2000	0.32952	0.198853	0.452385	0.000	0	1	0

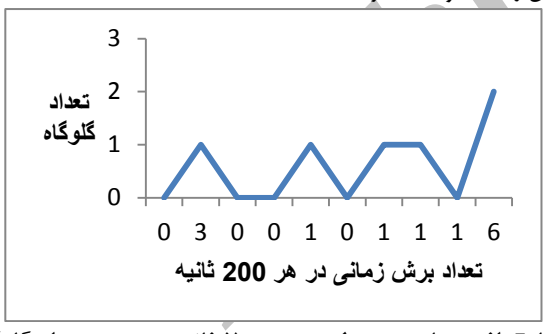


شکل3: اثر تغییرات حجم درخواست‌ها بر روی تعداد گلوگاه زمان پاسخ

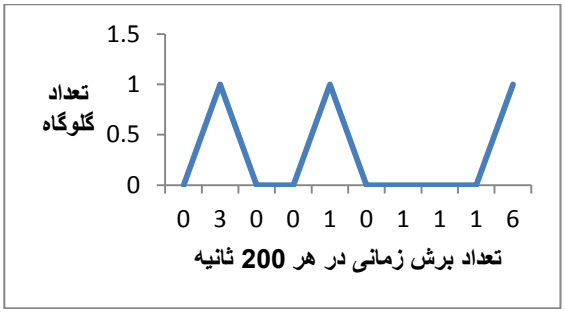


شکل4: اثر تغییرات حجم درخواست‌ها بر روی تعداد گلوگاه میزان گذردهی

حال به بررسی ارتباط میان تعداد گلوگاه کارایی و پویایی وضعیت ماشین فیزیکی می‌پردازیم. ماشین‌های مجازی هم‌میزبان در طول اجرا ممکن است دارای بازه زمانی همپوشان باشند لذا میزان منابع پردازشی مورد نیاز در یک ماشین فیزیکی باید در بازه‌های زمانی همپوشان و غیرهمپوشان محاسبه گردد. جهت انجام این بررسی، ماشین فیزیکی آزمایش قبل را در نظر می-گیریم و در هر ۲۰۰ ثانیه تعداد برش‌های زمانی و همچنین تعداد گلوگاه‌های کارایی محاسبه شده‌است. ضمن آن‌که حد آستانه ۰.۷۵ در نظر گرفته شده است. با تعیین محور عمودی به عنوان تعداد گلوگاه کارایی و محور افقی به عنوان تعداد برش زمانی در هر ۲۰۰ ثانیه و با رسم نمودار نقطه‌ای دو شکل ۵و۶ بدست امده است همان‌طور که مشاهده می‌شود هرچه تعداد برش‌های زمانی در بازه‌های ۲۰۰ ثانیه‌ای بیشتر شود، تعداد گلوگاه نیز افزایش یافته است. بازه زمانی (۲۰۰۰و۱۸۰۰) بیشترین تعداد برش‌های زمانی و همچنین بیشترین تعداد گلوگاه کارایی را دارد. بنابراین هرچه پویایی محیط بیشتر باشد احتمال رخداد گلوگاه بیشتر است.



شکل5: اثر پویایی محیط در هر ۲۰۰ ثانیه بر روی تعداد گلوگاه زمان پاسخ



شکل6 : اثر پویایی محیط در هر ۲۰۰ ثانیه بر روی تعداد گلوگاه میزان گذردهی

Archive of SID
www.SID.ir

جهت ارزیابی پیش‌بینی منابع پردازشی مورد نیاز یک ماشین فیزیکی به روش history-base از محک Ganglia[14] و سرویس Swifteqiad و سرور ms-be1013.eqiad.wmnet استفاده شده است. بارکاری سرور در ۴ فریم ۲۰۰ ثانیه‌ای از ۰ تا ۸۰۰ در اختیارمان قرار گرفته است و در جدول ۶ نشان داده شده است. جهت انجام این آزمایش، به پیش‌بینی منابع پردازشی مورد نیاز فریم چهارم یعنی (۸۰۰–۶۰۰) در سه مرتبه می‌پردازیم. بدین صورت که در مرتبه اول از یک فریم گذشته یعنی (۶۰۰–۴۰۰) و در مرتبه دوم از دو فریم گذشته یعنی (۶۰۰–۴۰۰) (۴۰۰–۲۰۰) و در مرتبه سوم از سه فریم گذشته یعنی (۶۰۰–۴۰۰) (۴۰۰–۲۰۰) (۲۰۰–۰) استفاده می‌کنیم. پیش‌بینی بارکاری براساس تاریخچه،از طریق میانگین بارکاری فریم‌های گذشته بدست می‌آید. نتایج حاصل در جدول ۷ نشان داده شده است. با به دست آوردن خطاهای پیش‌بینی در هر مرحله و مقایسه آن‌ها با یکدیگر مشاهده می‌شود که میزان خطا در پیش‌بینی با یک فریم کمتر از دو فریم و آن نیز کمتر از سه فریم می‌باشد ضمن آن‌که خطا در هر ۳ مرحله کم می‌باشد. بنابراین ورودی معادلات کارایی جهت پیش‌بینی مقادیر سنجه‌های کارایی قابل اطمینان می‌باشند لذا گلوگاه‌های کارایی به درستی پیش‌بینی می‌شوند.

جدول٦: بار کاری سرور ms-be1013.eqiad.wmnet در ٤ فریم ۲۰۰ ثانیه‌ای

	۰-۲۰۰	۲۰۰-۴۰۰	400-600	600-800
Memory	12795101275	12870338187	12883685412	12898468787
cpu	27.76311111	27.93733333	28.02	28.04844444
Bandwidth	42183036.34	42379232.36	43635420.92	43988579.11

جدول۷: مقایسه پیش‌بینیhistory-base منابع پردازشی مورد نیاز یک ماشین فیزیکی در۱ فریم، ۲ فریم و ۳ فریم گذشته

	۱ فریم	خطا	۲ فریم	خطا	۳ فریم	خطا
Memory	12883685412	0.1	12877011800	0.2	12849708291	0.4
Cpu	28.02	0.1	27.97866667	0.2	27.90681481	0.5
Bandwidth	43635420.92	0.8	43007326.64	2.2	42732563.2	2.9

٦- نتیجه

در این بخش با جمع‌بندی نتایج به‌دست‌آمده از پیاده‌سازی مدل و تحلیل عملکرد مدل، نتیجه کلی این پژوهش را اعلام خواهیم کرد. و در انتها به کارهای آتی موردنیاز در این پژوهش و اندک مسائل مبهم یا حل‌نشده اشاره خواهیم کرد.همان‌طور که در آزمایش‌ها اشاره شد، در این مقاله به منظور طراحی مدل تطبیق‌پذیر خودکار، با تولید داده‌های اولیه مورد نیاز و در اختیار داشتن بارکاری ماشین‌های مجازی در بازه‌ای در آینده نزدیک و با محاسبه مدل‌های کارایی، اقدام به محاسبه سنجه‌های کارایی در ماشین فیزیکی در بازه‌های زمانی در آینده و پیش‌بینی گلوگاه کارایی کردیم. بنابراین قبل از رخداد گلوگاه کارایی می‌توان اقداماتی جهت عدم رخداد آن صورت داد. لذا بارکاری تخصیص داده شده به ماشین فیزیکی متناسب با منابع پردازشی آزاد آن می‌باشد.براساسآزمایش‌های صورت گرفته جهت ارزیابی مدل پیشنهادی ملاحظه کردیم که با بررسی اثر تغییرات درخواست‌ها بر روی تعداد گلوگاه-های کارایی در یک ماشین فیزیکی به این نتیجه رسیدیم که تنها ازدیاد ماشین‌های مجازی باعث افزایش گلوگاه کارایی نخواهد شد بلکه به دلیل پویایی در وضعیت ماشین فیزیکی که نشأت گرفته از همپوشانی ماشین‌های مجازی و درخواست مکرر آن‌ها در زمان‌های کوتاه برای مصرف منابع ماشین فیزیکی به صورت اشتراکی است؛ تعداد گلوگاه کارایی به صورت چشمگیری افزایش یافته است. همچنین با ارزیابی پیش‌بینی منابع پردازشی مورد نیاز در یک ماشین فیزیکی به این نتیجه رسیدیم که ورودی معادلات کارایی جهت پیش‌بینی مقادیر سنجه‌های کارایی قابل اطمینان می‌باشند لذا گلوگاه‌های کارایی به درستی پیش‌بینی می‌شوند.

تلاش زیادی برای حل مسئله حاضر صورت پذیرفت،اما آنچه مسلم است به دلیل گستردگی مسائل مطرح دراین حوزه نمی‌تواند درقالب یک پژوهش و مدل پیشنهادی تمام چالش‌ها وکاستی‌ها راحل نمود. درمدل پیشنهادی،فقط به پیش‌بینی گلوگاه پرداخته شده است .درادامه می‌توان درکارهای آتی به چیدمان ماشین‌ها–ی مجازی براساس نتایج حاصل ازاین بخش و باهدف پیشگیری از رخداد گلوگاه پرداخت و ازاین طریق بهره‌وری منابع را افزایش داده و رضایت مشتریان را ارتقاء دهیم. همچنین اشاره کردیم پیش‌بینی درآینده نزدیک صورت گرفته است چراکه فقط درآینده نزدیک است که ازمنبع بحرانی مطلع هستیم. درواقع تنهادراین شرایط است که مطمئن هستیم که ارتباط بین سنجه‌های کارایی وسنجه‌های سیستم حفظ شده است. ازطرفی می‌توان باپیش–بینی نقطه تغییر، ازتغییر منبع حیاتی مطلع شد و گلوگاه را در آینده دور نیز پیش‌بینی کرد. علاوه بر این در پژوهش حاضر تنها معیارهای زمان پاسخ و گذردهی برنامه‌ها درنظر گرفته شد، درحالی‌که می‌توان سنجه‌های کارایی بیشتری را در نظرگرفت.

مراجع

[1] Zhu, X. (2014). Application performance management using learning, optimization, and control.in Proceedings of the 5th ACM/SPEC international conference on Performance engineering, 271-271.

[2] Padala, P., Hou, K.-Y., Shin, K. G., Zhu, X., Uysal, M., Wang, Z. (2009). Automated control of multiple virtualized resources.in Proceedings of the 4th ACM European conference on Computer systems,13-26.

[3] Xiong, P., Pu, C., Zhu, X., Griffith, R. (2013). vPerfGuard: an automated model-driven framework for application performance diagnosis in consolidated cloud environments. in Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering, 271-282.

[4] Jain, R. (1991). The Art of Computer Systems Performance Analysis. Wiley-Interscience, New York.

[5] Cohen, I., Chase, J. (2004). Correlating instrumentation data to system states: A building block for automated diagnosis and control.OSDI.4, 16-16.

[6] http://www.correlsense.com/blog/six-application-performance-challenges-facing-it/

[7] Padala, P., Hou, K., Zhu, X., Uysal, M., Wang, Z., Singhal, S., Merchant, A., Shin, K. (2007). Adaptive control of virtualized resources in utility computing environments.In Proc of EuroSys.

Archive of SID
www.SID.ir

[8] Doyle, R.P., Chase, J.S., Asad, O.M., Jin, W., Vahdat, A. (2003). M. Model-based resource provisioning in a web service utility. In Proc.Of USITS, 5.
[9] Bodik, P., Goldszmidt, M., Fox, A., Woodard, D.B., Andersen, H. (2010). Fingerprinting the datacenter: automated classi_cation of performance crises.In Proc of EuroSys, 111-124.
[10] Shen, K., Zhong, M., Li, C. (2005). I/O system performance debugging using model-driven anomaly characterization.In Proc of FAST.
[11] Cohen, I., Goldszmidt, M., Kelly, T., Symons, J., Chase, J.S. (2004). Correlating instrumentation data to system states: A building block for automated diagnosis and control.In Proc of OSDI.
[12] http://azarinweb.com/dedicated-servers/
[13] Sotomayor, B., Montero, R., Llorente, I., Foster, I. (2009). An Open Source Solution forVirtual Infrastructure Management inPrivate and Hybrid Clouds.Internet Computing, IEEE. (13)5, 14 - 22.
[14] https://ganglia.wikimedia.or

زیرنویس‌ها

[1] Platform
[2] Shared virtualized infrastructure
[3] Workload
[4] Co-hosted VM
[5] Noisy Neighbor
[6] Response time
[7] Throughput
[8] Framework
[9] VPerfGuard stands for: virtual Performance Guard
[10] Utilization
[11] Optimization
[12] Service Level Agrement
[13] First fit
[14] Best fit
[15] Round robin

Archive of SID

www.SID.ir