sina-pub.ir

تخصیص تکنیک های منابع بهینه و غیر بهینه در مراکز داده محاسبات ابر

چکیده

ارائه دهندگان خدمات ابر تحت فشار مداوم برای بهبود عملکرد هستند، گزینه های مختلف استقرار منابع بیشتری ارائه و قابلیت حمل نرم افزار را افزایش می دهند. برای دستیابی به اهداف عملکرد و هزینه، ارائه دهندگان نیاز به یک سیستم تخصیص منابع جامع دارند که هر دو منابع محاسباتی و شبکه را مدیریت می کند. روش جدیدی برای حل مسئله تخصیص منابع مرکز داده کافی به درخواست رزرو ماشین مجازی مشتری (VM) و درخواست ارتباط برنامه ریزی معرفی می شود. این امر در هنگام دستیابی به اهداف ارائه دهندگان و به حداقل رساندن نیاز برای جابه جایی VM باید انجام شود. در این کار، مشکل تخصیص منابع در مراکز داده محاسبات ابری به عنوان یک مشکل بهینه سازی و حل شده فرمول بندی شده است. علاوه بر این، مجموعه ای از راه حل های اکتشافی معرفی شده به عنوان سیاست های رزرو VM و تنظیم ارتباط مورد استفاده قرار می گیرد. یک راه حل ساده غیر بهینه مبتنی بر تجزیه مسئله اصلی نیز ارائه شده است. نتایج آزمایش برای مجموعه متنوع بار شبکه نشان می دهد که راه حل ساده به سطوح امیدوار کننده برای میانگین تاخیر درخواست ارتباط رسیده است. راه حل پیشنهادی قادر به رسیدن به سطوح عملکرد بهتر از راه حل های اکتشافی بدون بار هستند که ساعتهای طولانی در حال اجرا است. این باعث یک نامزد احتمالی برای حل مسائل با تعداد بسیاری از درخواست ها و محدوده های اطلاعاتی گسترده تر در مقایسه با راه حل بهینه می باشد.

کلمات کلیدی: ابرها، تخصیص منابع، مدل های تحلیلی، شبیه سازی سیستم ها، سیستم ارتباطی ترافیک، عملیات سیستم ارتباطات و مدیریت، خدمات وب و اینترنت، ماشین های مجازی، طراحی سیستم های راه حل

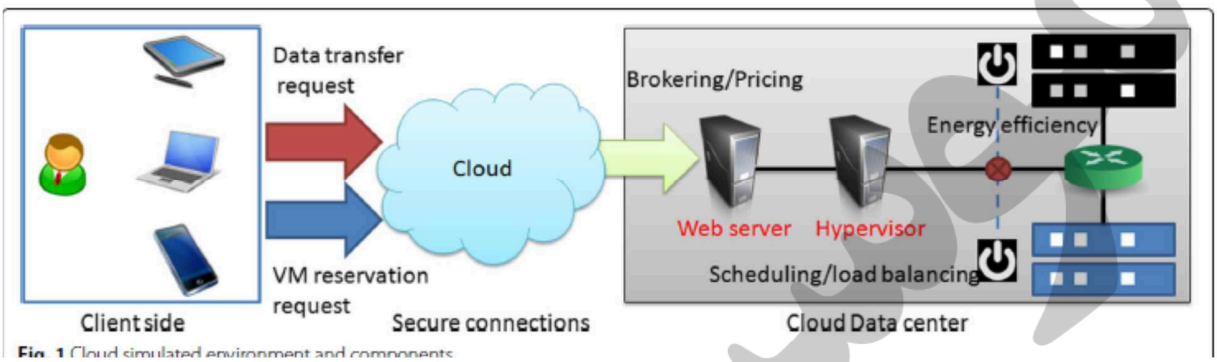
مقدمه

sina-pub.ir

تقاضای محاسبه ابر برای مشتریان از وعده تبدیل زیرساخت های محاسباتی به یک کالا یا خدمات نشئت می گیرد که سازمان ها دقیقا به همان اندازه ای که آنها استفاده می کنند، برای آن هزینه پرداخت می کنند. این ایده، رویا اجرایی یک شرکت فناوری اطلاعات است. همانطور که تحلیلگر گارتنر یعنی داریل پلومر بیان می کند: "رهبران محصولات تجاری در همه جا، دور از مراکز فناوری اطلاعات به دنبال دریافت برنامه ها ابر هستند .. و برای آنها هزینه پرداخت می کنند مانند اشتراک مجله. هنگامی که این سرویس دیگر مورد نیاز نیست، آنها می توانند این اشتراک را لغو کنند بدون اینکه تجهیزات استفاده شده باشد" [1]این ایده که محاسبات متمرکز بر شبکه، در آینده است، در اوایل سال 1997 برای رهبران صنعت روشن بود. هیچکسی به غیر از استیو جابز نگفت: " به هارد دیسک در کامپیوترم نیاز ندارم اگر بتوانم سریعتر به سرور دسترسی پیدا کنم . با حمل و نقل این کامپیوترهای همراه در مقایسه مربوط به روم شرقی می باشد[1] . این نیز برای خرید سازمان ها و برنامه ریزی مراکز داده بزرگ مورد استفاده قرار می گیرد.

به هر حال، عملکرد همچنان عامل مهمی است. اگر -در هر نقطه ای - شک و تردید بیش از توانایی ارائه دهنده به ارائه خدمات با توجه به موافقت نامه های امضا شده سطح خدمات(SLA) رایج می شود، مشتریان ارائه دهندگان دیگری را بررسی و انتخاب می کنند. آنها حتی می توانند بازگشت به مدل خرید و نگهداری را بررسی کنند. ارائه دهندگان تحت فشار ثابت برای بهبود عملکرد هستند، گزینه های استقرار منابع متنوع تر، بهبود قابلیت استفاده از خدمات و افزایش قابلیت حمل برنامه را ارائه می دهند. یک سلاح اصلی در اینجا یک سیستم تخصیص منابع کارآمد است. همانطور که در شکل 1، در سناریوی ابر، مشتریان قادر به اجاره ماشین های مجازی (VMs) از ارائه دهندگان ابر هستند. ارائه دهندگان چندین مدل استقراری ارائه می دهند که در آن پیکربندی VM در محاسبات قدرت، حافظه، ظرفیت ذخیره سازی و پلت فرم فقط چند عامل را نامگذاری می کند.

sina-pub.ir



شکل 1: اجزاء و محیط شبیه سازی ابری

در طول دوره اجاره، مشتریان نیاز به قابلیت های شبکه دارند. مشتریان اطلاعات را اغلب در میان ستاد مرکزی مشتری (یا ابر خصوصی) و VM ها یا بین دو مشتری VMs تبادل می کنند. هدف در اینجا برنامه ریزی برای برنامه های رزرو VM و درخواست اتصال در سریعترین راه ممکن است در حالی که از منابع مرکز داده بطور بهینه استفاده می کنند. این امر با ظهور مفاهیم داده بزرگتر سخت تر می شود. IBM چالش داده های بیشتری را به 4 ابعاد مختلف به Vs: 4 یعنی حجم، سرعت، تنوع و صداقت خلاصه می کند.[2] با اکثر شرکت هایی که حداقل 100 TB داده ذخیره شده دارند و با 18.6 میلیارد شبکه ارتباطی که فعلا در حال حاضر موجود برآورد می شود[2]، بازده تخصیص منابع هرگز اهمیت زیادی نداشته است.

هنگامی که روش تخصیص منابع با وظیفه طراحی مواجه می شود، چالش های داخلی و خارجی بسیاری را باید در نظر گرفت. تلاش برای خلاصه کردن این چالش ها در [3] یافت می شود. چالش های خارجی عبارتند از چالش های مقررات، جغرافیا و همچنین خواسته های مشتری مرتبط با ذخیره داده ها و مدیریت آنها می باشد. این محدودیت ها منجر به محدودیت در محل VM های محدود شده و محدودیت ها به مکان و انتقال داده می شود. چالش های خارجی نیز شامل بهینه سازی مدل شارژ می باشد به طوری که حداکثر درآمد را ایجاد می کند. چالش های داخلی مورد بحث در [3] شامل مسائل مربوط به داده های محلی نیز می باشد. ماهیت یک برنامه از لحاظ داشتن اطلاعات فشرده باید در هنگام قرار دادن VM و برنامه ریزی اتصالات مربوط به این برنامه بررسی شود. برای دستیابی به اهداف عملکرد و هزینه مذکور، ارائه دهندگان محاسبات ابری نیاز به سیستم تخصیص منابع جامعی دارند که هر دو منابع شبکه و محاسبات را مدیریت می کند. چنین سیستمی کارآمد می تواند تأثیر مهمی

sina-pub.ir

داشته باشد زیرا منابع اضافی به طور مستقیم به درآمد منتقل می شوند. بخش های زیر به شرح ذیل سازماندهی می شوند: بحث در مورد تلاش های تحقیقاتی مربوطه در بخش زیر معرفی می شود که منجر به مشارکت در این مقاله شده است. شرح مدل دقیق در بخش «شرح مدل» آمده است. بخش "فرمولاسیون ریاضی" فرمول مسئله ریاضی را ارائه می دهد. روش های اکتشافی در قسمت "راه حل اکتشافی" ارائه شده است. راه حل های غیربهینه در بخش "راه حل های غربهینه" ارائه شده است. نتایج در بخش "نتایج" نشان داده و تجزیه و تحلیل می شوند. در نهایت بخش "نتیجه گیری" مقاله را نتیجه گیری و کار آتی را معرفی می کند.

کار مرتبط

تلاش های قبلی برای بهینه سازی مجموعه ای متنوع از منابع ابر انجام شد. در [4]، دستجردی و بویا چارچوبی را برای ساده ساختن سرویس ابر ارائه می دهند. روش پیشنهادی آنها ترکیب سرویس بر اساس زمان استقرار، هزینه و قابلیت اطمینان ترجیح داده شده توسط کاربران بهینه می شود. نویسندگان از ترکیب الگوریتم های تکاملی و ترکیب بهینه منطق فازی با هدف به حداقل رساندن تلاش کاربران استفاده و همزمان ترجیحات خود را بیان می کنند. علیرغم داشتن طیف گسترده ای از الزامات کاربر در مدل سازی مسئله و ارائه فرمول بهینه همراه با یک منطق فازی اکتشافی، [4] مسئله را از آینده چشم انداز کاربر و نه ارائه دهنده مطرح می کند. هدف اصلی این است که بهترین ترکیب خدمات ممکن را ارائه می دهند که مشکل مسیر دلالی را به جای تمرکزبر روی عملکرد مرکز داده های ابری را به وجود می آورد . شرایط SLA داده تضمین شده توسط ارائه دهنده ابر بدون بررسی نحوه حاصل شدن آنها را بررسی می کنند.

وی و همکاران [5] کیفیت خدمات (QoS) شامل مشکل تخصیص منابع محدود برای خدمات محاسبات ابری را بررسی می کنند. آنها روش نظری بازی را برای یافتن راه حل تقریبی این مشکل ارائه می دهند. راه حل پیشنهادی خود را در دو مرحله اجرا می کند: (1) مرحله 1: حل بهینه سازی مستقل برای هر یک از شرکت کنندگان در نظریه بازی؛ مرحله 2: اصلاح استراتژی های چندگانه راه حل اولیه شرکت کنندگان مختلف مرحله 1 با توجه به بهینه سازی و انصاف بررسی می شود. مدل در [5] نشان دهنده مشکل رقابت برای منابع در محیط ابر است. هر سیستم /

sina-pub.ir

گره / دستگاه نشان دهنده منبعی است که دارای هزینه مربوطه و زمان اجرا برای هر کار است. جزئیات بیشتر از نظر درجه بندی چندگانه محاسبات و منابع شبکه در هنگام برنامه ریزی مورد نیاز است. حافظه، ذخیره سازی، توانایی محاسباتی و پهنای باند (حداقل) باید به صورت جداگانه در یک مدل ایده آل در نظر گرفته شود. علاوه بر این، تاثیر منابع شبکه به طور کامل در [5] بررسی نشده است. همچنین، هیچ بحث مفصل برای سناریوهای مجازی داده نشده است.

در [6]، بلاگ لازوف و همکاران چارچوب معماری را علاوه بر اصول تخصیص منابع برای محاسبه ابرکارآمد انرژی تعریف می کنند. آنها الگوریتم هایی را برای نقشه برداری انرژی VM ها گره فیزیکی مناسب توسعه می دهند. آنها الگوریتم های برنامه ریزی را پیشنهاد می کنند که انتظارات QoS و ویژگی های مصرف انرژی منابع مرکز داده را بررسی می کنند. این شامل، اول، تخصیص VM ها با استفاده از روش کاهش مناسب اصلاح شده و سپس بهینه سازی تخصیص VM فعلی با استفاده از جابه جایی VM است. با در نظر گرفتن چالش های جابه جایی ممکن است از نظر سقوط عملکرد ناشی از کپی ، انتقال تاخیر و چالش های برنامه ریزی همراه با آسیب پذیری ارائه دهندگان برای نقض[7] SLA است، راه حلی که نیاز به جابه جایی VM را به حداقل برساند، یک مورد ترجیحی است. علاوه بر این، هیچ مهلتی برای وظایف در این کار در نظر گرفته نشده است.

دوان و همکاران [8]مشکل برنامه ریزی برای برنامه های کاربردی جریان موازی در مقیاس بزرگ در ابرهای هیبریدی را به عنوان یک بازی مشارکت متوالی فرموله می کند. آنها یک الگوریتم چند منظوره ر ا برای ارتباطات و ذخیره سازی پیشنهاد می کنند که زمان اجرای و هزینه اقتصادی را بهینه می کند در حالیکه پهنای باند شبکه و محدودیت های مورد نیاز ذخیره سازی را برآورده می کند. در اینجا، زمان محاسبه به عنوان یک عملکرد مستقیم از محل سایت محاسبات و کارکرد به جای استفاده از یک واحد تک برای اندازه وظیفه مدل سازی می شود. حافظه به عنوان یک منبع استفاده نمی شود. مهلت های کاری بررسی نمی شوند. هدف این است که مجموعه ای از وظایفی که از برنامه خاص تشکیل شده است، تکمیل کنید. این مدل به اجرای شغل در شبکه نزدیک تر است به جای آنکه مدل رایج تر در ابر باشد که یک VM را با نیازهای خاصی از منابع رزرو می کند و سپس وظایف مربوط به آنها را

sina-pub.ir

انجام می دهد. علاوه بر این، فرضیه ارائه شده این است که درخواست های تبادل داده می توانند همزمان با محاسبات بدون هیچ گونه وابستگی اجرا شوند.

یکی دیگر از تغییرات را می توان در [9] مشاهده کرد که در آن دو الگوریتم زمانبندی، یعنی فهرست سبز، و نوبت گردشی مورد آزمایش قرار گرفتند.

تمرکز دوباره بر بازدهی انرژی بود، اما مدل ارائه شده شامل مدل سازی شبکه های دقیق بود همانطور که آن بر اساس شبیه ساز شبکه NS-2 بود.

درخواست کاربر به عنوان وظیفه مدل سازی می شود، وظایف به عنوان درخواست واحد مدل سازی می شود که شامل مشخصات منابع در قالب الزامات منابع محاسباتی (MIP، حافظه و ذخیره سازی) علاوه بر الزامات تبادل اطلاعات (از جمله مقادیر نشان دهنده پرونده های پردازشی که برای میزبان فرستاده می شود، وظایف قبل از اجرا برنامه ریزی می شود، داده های ارسال شده به سرور های دیگر در طول اجرا و اطلاعات خروجی بعد از اجرا ارسال می شود). هیچ مدل بهینه ای ارائه نمی شود. در [10]، یک زمانبند منابع انعطاف پذیر برای بازده انرژی برای مراکز شبکه های مهمی (NetFCs) پیشنهاد شده است. نقش زمانبند برای کمک به خدمات ابری زمان-واقعی مشتریان خودرو (VCs) برای مقابله با تاخیر و مسائل مربوط به تأخیر است. زمانبندن های مذکور در لبه شبکه وسیله نقلیه عمل می کنند و به VCs از طریق لینک های متحرک تک هاپ مبتنی بر TCP / IP زیر ساختار-به خودرو (I2V) خدمت می کنند.

هدف بهره برداری از وضعیت های محلی ارزیابی شده ارتباطاتTCP / IP ، به منظور به حداکثر رساندن ارتباطات کلی- به علاوه -محاسبات کارایی انرژی است در حالی که الزامات سخت افزاری QoS در مورد حداقل میزان انتقال، حداکثر تاخیر و تاخیر-تفاوت زمانی را برآورد می کند، زمانبند کارایی انرژی به طور مشترک انجام می شود : (i)کنترل پذیرش ترافیک ورودی توسط NetFC پردازش شود (ii)ارسال حداقل انرژی از ترافیک پذیرفته شده؛ (iii)پیکربندی سازگاری و تثبیت ماشین های مجازی (VMs) توسط NetFCs؛ میزبانی می شود و (iv) کنترل تطبیقی ترافیک تزریق شده به اتصالات تلفن همراهTCP / IP می باشد.

sina-pub.ir

در [11] یک زمانبند حداقل انرژی بهینه برای تخصیص مشترک پویا آنلاین به اندازه وظایف، میزان محاسبات، میزان ارتباطات و توان ارتباطی در مراکز داده شبکه مجازی (NetDCs) که تحت محدودیت های تاخیر در کار هر روزه عمل می کند، ارائه می شود. زیرساخت های NetDC ذکر شده از ماشین های مجازی فرکانس با فرکانس چندگانه (VMs) تشکیل شده است که با پهنای باند و شبکه محلی ناحیه (LAN) متغیر از نظر انرژی محدود متصل شده اند. یک روش دو مرحله ای برای تحلیل دقیق راه حل دقیق CCOP پیشنهاد شده است. زمانبند مطلوب به دست آمده متمایل به اجرای آنلاین مقیاس پذیر، توزیع شده و خصوصیات تحلیلی آن در قالب بسته می باشد. عملکرد واقعی تحت هر دو زمان متغیر تصادفی بطور ترکیبی ایجاد شده تست می شود و جهان واقعی طرح های حجم کاری را ارزیابی می کند.

برخی از آثار جدیدتر عبارتند از راه حل FUGE [12] . نویسندگان برنامه ریزی شغلی را ارائه می دهند که هدف آن اختصاص دادن شغل به مناسب ترین منابع، با توجه به ترجیحات و الزامات کاربر است. هدف FUGE انجام توازن بار مطلوب با توجه به زمان اجرا و هزینه می باشد. نویسندگان اصلاح الگوریتم ژنتیک استاندارد (SGA) و تئوری فازی را برای طراحی یک حالت پایدار مبتنی بر فازی به منظور بهبود عملکرد SGA از لحاظ متصل کردن اصلاح کردند. الگوریتم FUGE شغل را به منابع با توجه سرعت پردازش مجازی(VM) ، حافظهVM ، پهنای باند VMو طول کار اختصاص می دهد. یک اثبات ریاضی ارائه شده است که مشکل بهینه سازی با محاسبه شرایط معروف تحلیلی برجسته شده است (به ویژه شرایط کاروش کان تاکر)

در [13]، مسئله مدیریت صرفه جویی انرژی در هر دو مرکز داده و ارتباطات تلفن همراه حل شده است. یک زمانبند تخصیص منابع پویا و سازگار توزیع شده با هدف کمینه کردن محاسبات مصرف انرژی به علاوه ارتباطات در حالی که تضمین کیفیت خدمات (QoS) محدودیت های کاربر ارائه شده است. زمان بند برای معیارهای زیر ارزیابی می شود: زمان اجرا، استفاده از مکان مناسب و استفاده از پهنای باند.

با نگاهی به راه حل های موجود درپیشینه، واضح است که هر آزمایش بر چند جنبه از چالش های تخصیص منابع مواجه شده در این ناحیه متمرکز است. ما سعی می کنیم جنبه های مختلف را در جدول 1 خلاصه کنیم.

sina-pub.ir

یک راه حل ایده آل، ویژگی / پارامترهای موجود در جدول 1 را برای ساخت یک راه حل کامل ترکیب می کند. این شامل فرمول بهینه سازی است که منابع محاسباتی و شبکه را در یک سطح گرانروی عملی پوشش می دهد. سروکارداشتن با پهنای باند به عنوان یک کالا ثابت، کافی نیست. جزئیات مسیریابی هر درخواست برای بازتاب نقاط داغ در شبکه مورد نیاز است. این برای منابع محاسباتی نیز هست. نیازهای CPU، حافظه و ذخیره سازی حداقل آنچه که باید بررسی شود را تشکیل می دهد. علاوه بر این، تعدادی از تلاش های قبلی متمرکز بر پردازش منابع است، در حالی که برخی بر منابع شبکه تمرکز دارند. سوال اینجاست که:

چگونه ما می توانیم درخواست های رزرو مشتری VM را با در نظر گرفتن نیازهای تبادل اطلاعات در ذهن خود حفظ و پردازش کنیم؟ رویکرد مشترک این است که قرار دادن VM و برنامه ریزی ارتباطات جداگانه یا در دو مرحله مختلف متوالی انجام شود. این شرایط QoS را به مخاطره می اندازد و باعث می شود که ارائه دهنده اقدامات کاهش را در هنگام تقاضای محاسبات و شبکه VM آغاز کند. اینها مراحل عبارتند از: ارائه مازاد به عنوان احتیاط یا مهاجرت VM و پیش شرط اتصال پس از مسائل مانند گره های شبکه شروع به افزایش می کند. حداقل کردن رویداد جابه جایی VM یک هدف عمده عملکرد است. جابه جایی آفلاین VM ، به هرحال سریع یا کارآمد می باشد، بدان معنی است که مدت زمان صرف شده برای مشتریان است. این واقعا با یک محیط مشتری خواسته نمی شود که در دسترس بودن پنج 9 (99.999٪ از زمان در دسترس بودن) انتظار می رود. همانطور که برای جابه جایی آنلاین، بار با کپی / افزونگی لازم مورد نیاز است. این چالش های مرتبط با مهاجرت VM باعث می شود که معماران راه حل های محاسبات ابری از هر راه حلی استقبال می کنند که اصلا شامل جابه جایی نمی باشد.

این کمبود نیاز به یک راه حل تخصیص منابع دارد که هر دو تقاضا را در زمان مشابه بررسی می کند. این راه حل، تقاضای ارتباطات آینده VM همراه با نیازهای محاسباتی قبل از قرار دادن VM را بررسی می کند. در این مورد، نیازهای شبکه شامل نه تنها نیازهای پهنای باند به عنوان یک شماره ثابت و یا متغیر، بلکه محل منبع / مقصد از اتصال درخواست می باشد. بدان معنی است که گره ها / VM ها (احتمالا) مبادله داده ها با VM هستند.

sina-pub.ir

همانطور که این VM ها بطور دقیق گره خورده نسبتا نزدیک به یکدیگر زمانبندی می شوند، استرس شبکه به حداقل می رسد و نیاز به بهینه سازی مکان VM به طور چشمگیری کاهش می یابد.

در این کار، ما قصد داریم تا مسئله تخصیص رزرو مشتری VM و درخواست زمانبندی اتصال به منابع مربوط به مرکز داده را حل کنیم در حالی که اهداف ارائه دهنده ابر را برآورده می کنیم. مشارکت ما شامل موارد زیر است:

1- مسئله تخصیص منابع برای مراکز داده ابر به منظور به دست آوردن راه حل بهینه فرمول بندی کنید. این فرمول بندی، نیازهای منابع محاسباتی با جزئیات دقیق را در زمان سناریو واقعی رایج در ابر را بررسی می کند. آن همچنین شرایط مطرح شده توسط درخواست اتصال (شرایط عمر درخواست / مهلت، نیازهای پهنای باند و مسیریابی) را در همان زمان بررسی می کند. یک مزیت مهم این رویکرد نسبت به رویکردهای مورد استفاده در تلاش های قبلی، بررسی هر دو مجموعه نیازهای منابع همزمان قبل از تصمیم گیری زمانبندی است. این فرمول از دیدگاه ارائه دهندگان مورد توجه قرار گرفته و هدف آن حداکثر سازی عملکرد است.

2- فرمول بندی عمومی را به شیوه ای انجام دهید که آن خود را محدود به محیط یک شبکه داخلی نمی کند. درخواست های دریافتی ارتباطی می تواند از بسیاری از ابرهای خصوصی یا عمومی توزیع شده حاصل شود. افزون بر این، زمانبند انعطاف پذیری را برای قرار دادن VM ها در هر یک از مراکز داده ارائه دهنده ابر فراهم می کند که در چندین شهر قرار دارند. این مراکز داده (ابرها) نشان دهنده گره های ارتباطی شبکه هستند. مشکل کامل با استفاده از کتابخانه بهینه سازی IBM ILOG CPLEX حل می شود[14].

3. معرفی چند روش اکتشافی برای انجام دو مرحله از روند برنامه ریزی. سه روش برای مرحله رزرو VM و دو روش برای ارتباطات برنامه ریزی تست شد. عملکرد این روش ها بررسی و سپس با برخی از روش های موجود در ابتدا ذکر شده مقایسه شد.

4- یک روش غیربهینه برای حل مسئله مشابه برای مورد مقیاس بزرگ را معرفی کنید. این روش مبتنی بر تکنیک تجزیه مشکل اصلی به دو مشکالت فرعی جداگانه است. اولین نفر به عنوان مشکل اصلی معرفی می شود که تخصیص VM را برای سرورهای مرکز داده بر اساس یک تابع رابطه گره VM انجام می دهد. دومین به عنوان

sina-pub.ir

زیرمجموعه نامیده می شود، برنامه ریزی درخواست های ارتباطی تایید شده توسط مسئله اصلی انجام می دهد . این روش غیر بهینه نتایج بهتر را از روش های اکتشافی به دست می آورد در حالی که این نتایج را در دوره های زمانی قابل قبولی در مقایسه با فرمولاسیون مطلوب دریافت می کنند.

شرح مدل

ما مدلی را برای مقابله با تخصیص مشکل منابع برای گروهی از درخواست های کاربر ابر معرفی می کنیم. این شامل فراهم آوردن هر دو محاسبات و منابع مراکز داده شبکه است. این مدل شامل یک شبکه گره های سرویس دهنده (ابرهای عمومی) و مشتری گره ها (ابرهای خصوصی) است. این گره ها در انواع مختلف شهرهای یا نقاط جغرافیایی مانند شکل 2 قرار دارند

جدول 1: مقایسه تلاش های تخصیص منابع ابری

رفرنس/ویژگی	4	5	6	7	8	راه حل پیشنهادی
مدل بهینه ارائه شده	بله	بله	نه	بله	نه	بله
دیدگاه	کاربر/دلال	ارائه دهنده	ارائه دهنده	ارائه دهنده	ارائه دهنده	ارائه دهنده
منابع محاسباتی	انواع VM (عمومی)	گره محاسباتی	CPU، حافظه و ذخیره سازی	گره محاسباتی	CPU، حافظه و ذخیره سازی	CPU، حافظه و ذخیره سازی
منابع شبکه	مقدار داده	بر زمان اجرا تاثیری ندارد	نه	BW و مقدار داده	BW، منبع و مقصد	BW، منبع و مقصد
برنامه ریزی هر دو منابع محاسباتی و شبکه را بررسی می کند	نه	نه	نه	بله	نه	بله
مهلت	نه	بله	نه	نه	بله	بله

sina-pub.ir

درخواست/طول عمر						
مدلسازی VM	کاربرد VMS و VM پیشنهاد شده	نه	جایگاه VM و جابه جایی آن بررسی شده	نه	جایگاه VM و جابه جایی آن بررسی شده	جایگاه VM بررسی شده

آنها با استفاده از شبکه لینک های دو طرفه متصل هستند. هر پیوند در این شبکه به تعداد خطوط برابر (جریان) تقسیم می شود. فرض بر این است که این عامل دانه بندی پیوندها می تواند کنترل شود. ما همچنین فرض می کنیم که هر مرکز داده شامل تعدادی از سرورهای متصل از طریق ارتباطات اترنت می باشد. هر سرور یک مقدار ثابت از حافظه، واحد محاسبات و فضای ذخیره سازی خواهد داشت. در گام اول، زمانی که مشتریان نیاز به میزبانی ابری دارند، آنها درخواست برای رزرو تعدادی از VM ها ارسال می کنند. همه این VM ها می تواند همان نوع یا انواع مختلف باشد. هر ارائه دهنده ابر انواع مختلف VM ها را برای مشتریان خود ارائه می دهد تا فرم را انتخاب کنند. این نوع در مشخصات هر منابع محاسباتی مانند حافظه، واحد CPU و ذخیره سازی متفاوت است. ما از این سه نوع منابع در آزمایش خود استفاده خواهیم کرد. در نتیجه، هر یک از VM های درخواست شده به یک سرور در یکی از مراکز داده اختصاص داده شده است. همچنین مشتری چندین درخواست برای رزرو یک ارتباط ارسال می کند. دو نوع اتصال [15، 16] وجود دارد:

1- درخواستی برای ارتباط یک VM به VM دیگر وجود دارد که هر دو VM ها قبلا فضای بر روی یک سرور در یک مکان از مراکز داده اختصاص دادند (ابرهای عمومی). 2- درخواست برای اتصال VM به گره مشتری است. در اینجا، VM واقع در گره مرکز داده متصل به ستاد مرکزی مشتری یا ابر خصوصی می باشد. شبکه سرویس گیرنده ابر در شکل 2 ارائه شده است. برای هر درخواست، مشتری منبع، مقصد، زمان شروع و مدت اتصال را تعریف می کند. بنابراین، هدف به حداقل رساندن میانگین تاخیر برای کلیه درخواست های ارتباطی می باشد. یک نمونه از درخواست ها مشتری در جدول 2 نشان داده شده است. درخواست هایی با برچسب "Res" درخواست رزرو

sina-pub.ir

VM هستند . درخواست های با برچسب "Req con" عبارتند از: درخواست اتصال بین یک VM و یک گره مشتری یا بین 2 VMs می باشد. یک مثال از پیکربندی VM در جدول 3 [17، 18] نشان داده شده است.

فرمول ریاضی

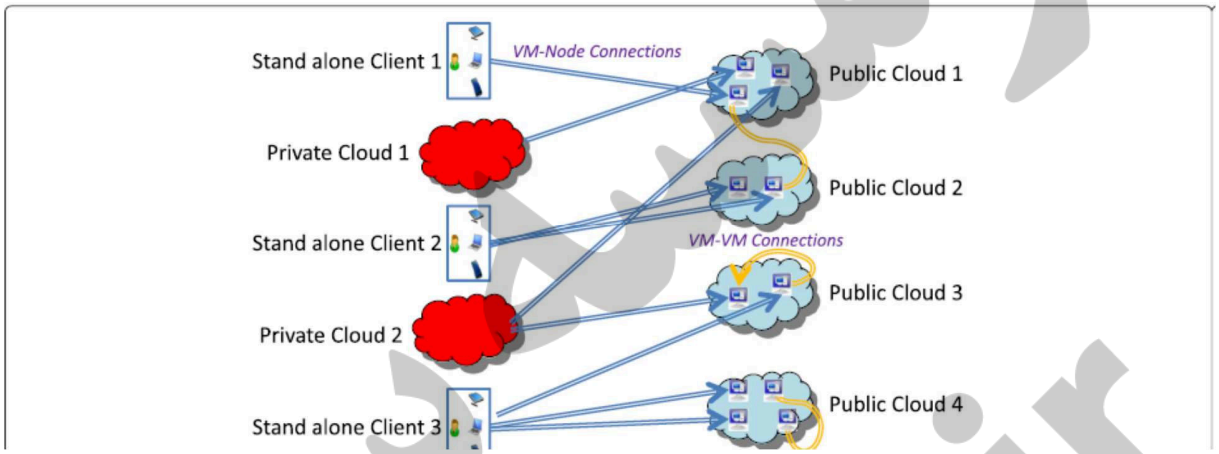
برای حل مسئله برنامه ریزی منابع در محاسبات ابری محیط، ما یک مدل تحلیلی معرفی می کنیم که ما مسئله را به صورت یک مسئله خطی عدد صحیح مختلط فرمول بندی می کنیم. ما بهینه سازی مشکل حداقل میانگین تاخیر را برای کلیه درخواست های ارتباطی رزرو مدل می کنیم در حالی که نیازهای درخواست اتصال مجازی مشتریان مختلف را برآورد می کند. این مدل با استفاده از نرم افزار IBM ILOG CPLEX برای مجموعه ای کوچک از درخواست ها حل شده است.

نشانه ها

پارامترهای محیط و شبکه به شرح ذیل بیان می شوند. مجموعه VM ها و مجموعه ای از سرورها توسط VM و Q به ترتیب نشان داده شده است. Mqm نشان دهنده مقدار منابع (به عنوان مثال حافظه) در دسترس در سرور است که در آن $q \in Q$ و $m \in \{$حافظه(mem) ، واحد پردازنده(cu) ، ذخیره سازی (sg)$\}$ چنین است که Mqm = 30نشان می دهد که حافظه در سرور q، 30 گیگابایت در دسترس است و فرض می شود که m دلالت بر نوع خاصی از منابع مورد نیاز دارد، یعنی، حافظه در یک سرور. Kvm برای نشان دادن مقدار منابع مورد نیاز برای استفاده هر VM درخواستی مورد استفاده قرار می گیرد که KVM = 7 نشان می دهد $VM\ v \in VM$نیاز به 7 گیگابایت حافظه فرض می شود که m نشان دهنده حافظه منابع بر روی یک سرور است. مجموعه ای از مسیرهای شبکه و مجموعه ای از لینک ها توسط P و L به ترتیب نشان داده شده است. alpیک پارامتر باینری است به طوری که alp = 1اگر پیوند $l \in L$ در مسیر $p \in P$ باشد؛ در غیر این صورت 0 است که در فرمول بندی ما، روش مسیریابی متناوب ثابت با مجموعه ای ثابت از مسیرهای موجود بین یک گره و هر گره دیگر استفاده شده است. این مسیرها،

sina-pub.ir

مسیرهای متناوب ثابت نشان می دهند که درخواست می تواند برنامه ریزی شود زمانی که ازیک سرور باقی مانده در گره اول به سرور متصل در گره دیگر حرکت می کند. bqcpیک پارامتر باینری است به طوری که1 = bqcp است.



شکل 2: مثالی از شبکه ارائه دهنده مشتری ابری.: مشتریان می توانند از ابرهای خصوصی اشان، ساختمان مرکزی یا از دستگاه های منحصر به فرد در اینترنت ارتباط برقرار کنند. مرکز ارائه دهندگان اطلاعات ابرهای عمومی را معرفی می کنند.

جدول 2 : نمونه ای از مجموعه ای از درخواست های تخصیص منابع

مشتری	درخواست	نوع	شروع	مدت	منبع	مقصد
C-1	Res VM1	High-CPU	T=10	125	-	-
C-2	Res VM2	High-Storage	T=15	400	-	-
C-1	Res VM3	Standard	T=20	150	-	-
C-2	Res VM4	High-Memory	T=10	70	-	-
C-1	Req con	VM-VM	T=15	10	VM1	VM3
C-1	Req con	VM-C	T=18	20	VM3	C1
C-2	Req con	VM-VM	T=25	8	VM4	VM2
C-2	Req con	VM-C	T=30	30	VM4	C2

sina-pub.ir

اگر مسیر $p \in P$ یکی از مسیرهای متناوب از سرور باشد، $q \in Q$ به سرور، $c \in Q$؛ در غیر این صورت صفر است، I مجموعه ای از درخواست ارتباطی را نشان می دهد. هر درخواست اتصال، $i \in I$ توسط یک منبع(si) ، مقصد (di)، زمان شروع درخواست شده (ri) و مدت اتصال TARD .(ti) را مشخص می کند. $TARD$ نشان دهنده تاخیر مجاز (تاخیر پذیرفته شده) برای هر درخواست ارتباطی می باشد. فرمول بندی سناریوهای را تحت پوشش قرار می دهد که شبکه ها می توانند لینک را به سهام یا جریان تقسیم کنند که به انعطاف پذیری بیشتر با فرمول بندی و پوشش مجموعه ای از موقعیت ها اجازه می دهد. مجموعه سهام (طول موج در مورد شبکه نوری) می تواند شامل هر عدد از طول موجها براساس خود مشکل باشد. مجموعه λ مجموعه ای از تمام طول موج های موجود در شبکه می باشد. این پارامتر h مورد استفاده در محدودیت 6 نشان دهنده تعداد زیادی است که به تضمین راه حل کمک می کند که با توجه به شرایط در محدودیت مشتق شده است. علاوه بر این، پارامتر باینری Wij نشان می دهد اگر درخواست i قبل از درخواست j برنامه ریزی شده باشد. با استفاده از این پارامتر محدودیت 6 را برای هر جفت درخواست فقط یک بار تست شده تضمین می کند.

متغیرهای تصمیم گیری

Fiیک متغیر تصمیم عدد صحیح است که نشان دهنده زمان شروع برنامه ریزی برای درخواست اتصال، $i \in I$.

Xvqیک متغیر تصمیم دوتایی است به طوری که $Xvq = 1$ اگر $v \in VM$ در سرور$q \in Q$ برنامه ریزی شده باشد. Yipw یک متغیر تصمیم باینری است به طوری که $Yipw = 1$ اگردرخواست، $i \in I$در مسیر، $p \in P$و طول موج، $w \in \lambda$. برنامه ریزی شده باشد.

تابع هدف

این مسئله به عنوان مسئله برنامه ریزی خطی اعداد صحیح (MILP) فرمول بندی شده است. هدف ازMILP کاهش

sina-pub.ir

جدول 3 : پیکربندی VM برای نمونه 3 انواع (VM) مورد استفاده در آزمایشی که توسط آمازون EC2 ارائه شده است [17]

نوع نمونه	استاندارد فوق العاده بزرگ (SXL).	حافظه فوق العاده بالا (MXL)	اضافی بزرگ CPU ، (CXL)
حافظه	15 GB	17 GB	7 GB
CPU (EC2 units)	8	6.5	20
ذخیره سازی	1690 GB	490 GB	1690 GB

به حداقل رساندن میانگین تاخیر ارتباط درخواست مشتری به و از VM ها است. استقامت در اینجا به عنوان تفاوت بین شروع زمان درخواست شده توسط مشتری (توسط ri معرفی شده) و زمان شروع برنامه ریزی شده توسط ارائه دهنده (توسط Fi معرفی شده) محاسبه شده است. حل کننده به دنبال راه حلی است که بهترین راه حل را برای مشتریان به ارمغان می آورد درحالیکه به مشتریان دیگر آسیب نمی رساند. راه حل تحت این فرض کار می کند که درخواست های مشتریان وزن / اهمیت مشابه ای به ارائه دهنده دارند. هدف عملکرد این مشکل به شرح زیر است:

$$MIN \sum_{i} (F_i - r_i) \qquad i \in I, \qquad (1)$$

محدودیت ها

تابع هدف منوط به محدودیت های زیر می شود:

sina-pub.ir

$$\sum_{q\in Q} X_{vq} = 1, \qquad v \in VM, \tag{2}$$

$$\sum_{p\in P}\sum_{w\in\lambda} Y_{ipw} = 1, \qquad i \in I, \tag{3}$$

$$\sum_{v\in VM} X_{vq} \times K_{vm} <= M_{qm}, \quad q \in Q, m \in \{m, c, s\}, \tag{4}$$

$$Y_{ipw} + (X_{s_i q} + X_{d_{ic}} - 3b_{qcp}) <= 2, \quad i \in I, q \in Q, c \in Q, p \in P, w \in \lambda, \tag{5}$$

$$\sum_{p\in P}[(t_i \times a_{lp} \times Y_{ipw}) + (h \times a_{lp} \times Y_{ipw}) + (h \times a_{lp} \times Y_{jpw})] + F_i - F_j + h \times W_{ij} <= 3h, \quad i, j \in I, l \in L, w \in \lambda, \tag{6}$$

$$W_{ij} + W_{ji} = 1, \qquad i, j \in I, \tag{7}$$

$$X_{vq}, Y_{ipw}, W_{ij} \in \{0, 1\}, \tag{8}$$

$$F_i - r_i >= 0, \qquad i \in I, \tag{9}$$

$$F_i - r_i <= TARD, \qquad i \in I, \tag{10}$$

$$F_i, r_i >= 0, \qquad i \in I. \tag{11}$$

در معادله (2) ما اطمینان می دهیم که VM دقیقا برای یک سرور مشخص می شود. در معادله (3)، ما اطمینان می دهیم که درخواست ارتباط دقیقا بر روی یک مسیر فیزیکی و یک طول موج قرار خواهد گرفت (جریان / سهم یک لینک). در (4) ما تضمین می کنیم این VM در سرورهایی با ظرفیت کافی از منابع محاسباتی مورد نیاز VM ها اختصاص داده خواهد شد. در (5)، ما اطمینان می دهیم که یک ارتباط فقط بر روی یکی از مسیرهای قانونی جایگزین بین VM و شریک ارتباط ایجاد خواهد شد (یکی دیگر از VM یا گره مشتری). در (6) ما اطمینان می دهیم که حداکثر یک درخواست را می توان در یک لینک خاص در یک زمان در هر طول موج برنامه ریزی کرد و هیچ درخواست دیگری در همان لینک برنامه ریزی نخواهد شد و طول موج تمام می شود. محدودیت (7) تضمین می کند محدودیت 6 تنها برای یک بار برای هر جفت درخواست تست می شود. این نشان می دهد که درخواست i قبل از درخواست j شروع خواهد شد. در معادله (9) و (10)، ما اطمینان داریم که زمان برنامه ریزی شده برای درخواست در پنجره مجاز تاخیر در این آزمایش است.

راه حل اکتشافی

مدل اکتشافی

sina-pub.ir

مدل پیشنهادی در این مقاله به چالش تخصیص منابع در زمان ارائه منابع محاسباتی (پردازنده، حافظه و ذخیره سازی) اشاره دارد و منابع شبکه با آن مواجه می شوند. یک کنترل کننده مرکزی درخواست های مذکور را با هدف به حداقل رساندن میانگین تاخیر و درخواست مسدود کردن مدیریت می کند. هدف راه حل حل کرد ن چالش هزینه ارائه دهنده و موضوع عملکرد برنامه های ابری می باشد.

برای هر درخواست، مشتری منبع، مقصد، زمان شروع و مدت زمان ارتباط را تعریف می کند. بنابراین، این مشکل در زیر دسته پیشرفته رزرو مسائل قرار می گیرد.

کنترل کننده مرکزی (می تواند برای مثال یک کنترل کننده شبکه تعریف شده نرم افزار [19, 20] (SDN) باشد)، جداول داده مسیرهای شبکه در دسترس، منابع سرور موجود و زمان انقضای ارتباط به منظور کنترل درخواست های جدید حفظ می شود. کنترل کننده سپس درخواست های VM ها را بر روی سرورها با توجه به روش یا سیاست مورد استفاده تخصیص می دهد. آن دسترسی منابع را با توجه جداول به روز می کند. پس از آن، کنترل کننده درخواست های ارتباطی را برای برآوردن نیاز مشتری برنامه ریزی و مسیریابی می کند. جداول دسترسی مسیر شبکه نیز به روز می شود. همانند هدف اولیه، کنترل کننده با هدف به حداقل رساندن میانگین تاخیرکلیه درخواست های ارتباطی رزروی را پیشرفته می باشد. همچنین هدف دوم، به حداقل رساندن تعداد درخواست های مسدود شده می باشد. به این هدف نائل می شوید بدون در نظر گرفتن آنچه که مسیر استفاده می شود. هدف سیاستهای اکتشافی/ تکنیکهای پیشنهادی بهتر شدن است اگرچه از نظر ریاضی مقادیر متریک عملکرد مطلوب نیست، در حالی که ارائه این راه حل قابل قبول در مقدار زمان قابل قبول است

تکنیک های اکتشافی برای به حداقل رساندن تاخیرروند تخصیص به دو مرحله متوالی تقسیم می شود:

1- اختصاص VM ها در سرورهای مرکز داده. اینجا همه درخواست های رزرو VM بر اساس دسترسی منابع سرور قبل از هرگونه درخواست ارتباط ارائه می شود.

2- زمانبندی درخواستهای ارتباط در مسیرهای شبکه در دسترس. این پس از آنکه همه VM ها به منابع اختصاص داده شد، رخ می دهد و عملیات در سرورها شروع می شود.

sina-pub.ir

برای اولین مسئله فرعی، سه تکنیک اکتشافی ارزیابی شد. برای مرحله دوم (زیرمجموعه)، دو تکنیک های اکتشافی مورد آزمایش قرار گرفتند. برای آزمایش کامل یک اکتشاف برای هر مسئله فرعی استفاده شد. اکتشافات مذکور به شرح زیر تقسیم می شوند.

تکنیکهای اکتشافی رزرو VM

الف) تکنیک توزیع زمان برابر(ED) : در این اکتشاف، TMiزمان کل رزرو شده توسط درخواست ارتباط از دستگاه مجازیVMi (مجموع مدت زمان ارتباط) می باشد. بعد، سهم یک سرور با تقسیم واحدهای کل زمان محاسبه می شود، تمام VM ها را از طریق تعدادی از سرورها درخواست شده اند. آن بر اساس این فرض است که همه سرورها (برای منابع محاسباتی و شبکه) ظرفیت مشابهی دارند. سپس، برای هر سرور، VMها منابع محاسباتی به سرورهای متناظر یک به یک اختصاص داده می شوند. وقتی سرور تعدادی از VM ها را اختصاص می دهد که سهم سرور محاسبه شده را پوشش دهد/ مصرف کند، VMبعد اختصاص یافته منابع در سرور زیر و مراحل قبلی تکرار می شوند. الگوریتم در شبه کد در شکل 3 توضیح داده شده است.

ب) تکنیک فاصله گره (ND)

اول، میانگین فاصله بین هر دو گره محاسبه شد. دو گره دورتر از یکدیگر (با حداکثر فاصله) انتخاب می شوند. سپس حداکثر تعداد VM ها بر روی سرور این دو گره مذکور اختصاص داده می شود. بعد، گره باقی مانده ارزیابی می شود، گره با حداکثر فاصله متوسط به دو گره قبلی انتخاب می شود. همان فرایند تکرار می شود تا زمانی که تمام سیستم های VMS برنامه ریزی شوند.

sina-pub.ir

```
Input: Virtual machine set VM, Server
       set Q, connection request set R
Output: Allocation of VMs on servers,
        TM_i has the total connection time
        requested by VM_i
for TM_i ∈ TM do
    TM_i = 0
end for
for VM_i ∈ VM do
  for R_j ∈ R do
    if R_j.source = VM_i or R_j.dest = VM_i then
      TM_i = TM_i + R_j.duration
    end if
  end for
end for
TM_total = ∑_i TM_i
ServerShare = TM_total /|Q|
i = 0
for S_j ∈ Q do
  ThisServerShare = ServerShare
  while S_j isNotFull and ThisServerShare > TM_i
  do
    Schedule VM_i on S_j
    ThisServerShare = ThisServerShare − TM_i
    i = i + 1
  end while
end for
```

شکل 3 روش اکتشافی توزیع زمان برابر

الگوریتم در شبه کد در شکل 4 توصیف می شود.

fillNode تابعی است که اساسا سعی می کند بسیاری از VM ها احتمالی در گره نامیده شده را برنامه ریزی کند تا زمانی که منابع گره خسته می شوند. fillNode در شکل 5 نشان داده شده است.

ج) تکنیک توزیع مبتنی بر منابع (RB):

در این اکتشافی، انتخاب سرور بر اساس نوع VM درخواست شده می باشد. همانطور که در جدول 3 نشان داده شده است، سه نوع VM ها در آزمایش استفاده می شود: i) حافظه بالا فوق العاده بزرگ (MXL) پیکربندی حافظه بالایی دارند؛ 2) CPU فوق العاده بزرگ (CXL) دارای یک قدرت محاسباتی بالا هستند

sina-pub.ir

iii) استاندارد فوق العاده بزرگ

(SXL) برای برنامه های معمولی بیشتر مناسب است که فضای ذخیره سازی زیادی نیاز دارند. بسته به نوع درخواست VM توسط مشتری، اکتشافی سرور را با بالاترین مقدار منابع مربوطه در دسترس انتخاب می کند. سپس VM منابعی را به سرور اختصاص می دهد. این باعث می شود که توزیع متعادل تر شود.

تکنیک های اکتشافی رزرو ارتباطی

a) تکنیک اولویت مدت (DP):

```
Input: Virtual machine set VM, Server
       set Q, Node set N, Path set P,
       where P_ijk is path k between nodes
       i and j, NP is a fixed Number of paths
       between node i and node j
Output: Allocation of VMs on servers
for N_i ∈ N do
  for N_j ∈ N do
    A[i][j] = Σ_k^NP P_ijk.Length/NP
  end for
end for
Pick 2 nodes x, y with max A[x][y]
U = {x, y}
RemVMs = |VM|
RemVMs = fillNode(x, RemVMs)
RemVMs = fillNode(y, RemVMs)
while U ≠ N and RemVMs > 0 do
  maxDist = 0
  for N_i ∈ N and N_i ∉ U do
    avgDist = 0
    for B_j ∈ U do
      avgDist = avgDist + A[B_j][N_i]
    end for
    if avgDist > maxDist then
      maxDist = avgDist
      NextNode = N_i
    end if
  end for
  RemVMs = fillNode(NextNode, RemVMs)
  U = U ∪ {NextNode}
end while
```

شکل 4 تکنیک اکتشافی فاصله گره

sina-pub.ir

```
1: Function :fillNode
2: Input: Virtual machine set VM, Node x,
3:        RequestedVMs, server set Q
4: Output: servers in node x filled with max
5:         VMs possible
6: i = |VM| − RemVMs
7: for S_j ∈ Q and S_j residing in Node x do
8:   while S_j isNotFull and i < |VM| do
9:     Schedule VM_i on S_j
10:    i = i + 1
11:  end while
12: end for
13: return i
```

شکل 5 تابع: fillNode

در این اکتشافی، ارتباط با کوتاه ترین مدت زمان اولویت داده می شود. اول، درخواست ارتباط بر اساس مدت زمان درخواست شده مرتب می شوند. مرحله بعدی انتخاب ارتباط با کوتاه ترین مدت و برنامه ریزی آن را در کوتاه ترین مدت مسیر در دسترس است. این مرحله تکرار می شود تا زمانیکه کلیه درخواست ارتباطی ارائه شود. الگوریتم در شبه کد در شکل 6 توضیح داده می شود.

ب) الگوریتم حریص (GA):

در این اکتشاف، که در شکل 7 نشان داده شده است، برنامه ریزی بر اساس ارتباط شروع زمان درخواست شده (RST) می باشد. درخواست ارتباط با RST قبلی در مسیر اول در دسترس بدون در نظر گرفتن طول مسیر برنامه ریزی شده باشد.

تجزیه و تحلیل پیچیدگی راه حل های اکتشافی

مشکل تخصیص منابع در مرکز داده ابر تنوع از مشکل حلقه شناخته شده است. این مسئله کوله پشتی دو شکل دارد. در فرم تصمیم گیری–

sina-pub.ir

```
Input: Path set P where P_xyk is path k
       between nodes x and y,
       and connection request set R
Output: Scheduling of network connection
        requests on network paths
Sort R in descending order based on R_i.duration
for R_i ∈ R do
  for t = R_i.RST to MaxTimeUnits do
    Pick shortest path P_xyk where R_i.source =
    x and R_i.destination = y
    if P_xyk isAvailable(t, R_i.duration) then
      Schedule R_i on P_xyk at time unit t
      Move to next request
    end if
  end for
end for
```

شکل 6 تکنیک اکتشافی اولویت مدت زمان

```
Input: Path set P
       where P_xyk is path k between nodes
       x and y is Server set Q, connection
       request set R, NP is Number of paths
       between node x and node y
Output: Scheduling of network connection
        requests on network paths
Sort R in descending order based on R_i.RST
(requested start time)
for R_i ∈ R do
  for t = R_i.RST to MaxTimeUnits do
    x = R_i.source
    y = R_i.destination
    for k = 0 to NP do
      if P_xykisAvailable(t, R_i.duration) then
        Schedule R_i on P_xyk at time unit t
        Move to next request
      end if
    end for
  end for
end for
```

شکل 7 : تکنیک های اکتشافی حریصی

موسسه سینا sina-pub.ir

sina-pub.ir

که کمتر دشوار است - همانطور که سوال در مورد -NP -مکمل این است: آیا ارزش عینی حداقلK بدون بیش از وزن مخصوص W به دست می آید؟ فرم بهینه مسئله - فرمی است که ما سعی می کنیم در این کار حل کنیم - تلاش می کنیم تا احتمال ارزش عینی را بهینه سازی کنیم. فرم بهینه سازی NP-Hard است. این بدان معنی است که حداقل سختی تمام مشکلات NP است. آنجا راه حل فعلی در زمان چندجمله ای برای این فرم وجود ندارد. این باعث معرفی الگوریتم های اکتشافی شد. این ممکن است برای خواننده مورد توجه قرار گیرد تا پیچیدگی الگوریتم های معرفی شده اکتشافی را بازدید کند.

ابتدا متغیرهایی را که در این تجزیه و تحلیل در نظر گرفته شده را بازبینی می کنیم. VM نشان دهنده مجموعه VMاست، Nنشان دهنده مجموعه ای از گره ها است، S مجموعه ای از سرورها است، Rمجموعه درخواست های اتصال است، T تاخیر مجاز برای هر درخواست و D میانگین مدت زمان ارتباط است. این تجزیه و تحلیل با هدف تنها یک تقریب از پیچیدگی زمان برای نشان دادن ارائه شده است که این الگوریتم ها در داخل زمان چندجمله ای و به نوبه خود اجرا می شود - عملا توسط آن مقیاس گسترده ای از شبکه های ابر استفاده می شود. نگاه کنید به الگوریتم های یک به یک معرفی شده ، متوجه می شویم که توزیع زمان برابر پیچیدگی O (| VM || R | + | S ||دارد. فاصله گره الگوریتم ها در O (| N3 | + | S | + | V |) اجرا می شوند. توزیع مبتنی بر منابع در O (| V || S |) اجرا می شود که سریع ترین در میان الگوریتم جابه جایی VM 3 ما معرفی کردیم. مانند برای ارتباط الگوریتم های اکتشافی زمان بندی، مدت زمان اولویت در O (| R | |lg | R | + | R | .T.D) یا O (| R | + lg | R | + 1) T. | D) اجر می شود. در نهایت، الگوریتم زمان بندی اتصال حریص در O (R.T.D). اجرا می شود. بنابراین، تمام الگوریتم های ذکر شده در زمان چندجملهای اجرا می شود و می تواند نتیجه را برای مشکل مقیاس بزرگ در دوره های زمانی عملی تولید کند.

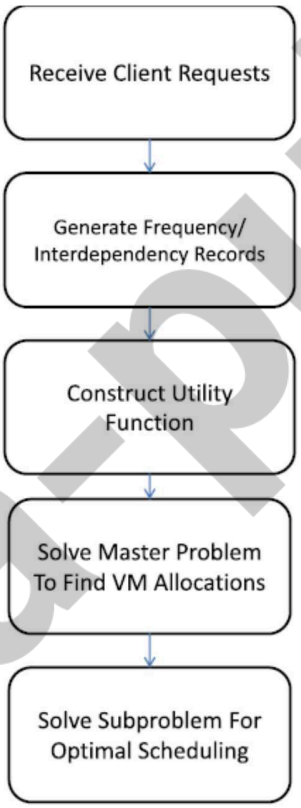
راه حل غیر بهینه

اگر چه راه حل بهینه می تواند با استفاده از فرمول بندی در بخش "فرمول ریاضی" حاصل شود، این تنها احتمال برای مقابله با مسئله در مقیاس کوچک است. حتی در زمان استفاده یک شبکه 5 گره با 4 سرور و 7 پیوند آنها را

sina-pub.ir

ارتباط می دهد، تعداد متغیرهای بهینه سازی می تواند به بزرگی 5000 متغیر باشد زمانی که برنامه زمانبندی 50 درخواست متعلق به 5 VM باشد. از سوی دیگر، روش های اکتشافی راه حل قابل قبول در زمان های نسبتا سریع حاصل می شود اما کیفیت راه حل قابل اثبات نیست. این ما را انگیزه دار می کند تا به مرحله بعدی حرکت کنیم که یافتن روش دستیابی به یک راه حل غیربهینه است. روش معرفی شده در اینجا بر اساس تکنیک تجزیه است. ما روش در شکل 8 نشان می دهیم. مراحل به شرح زیر است.

1- در مرحله 1، مجموعه ای از درخواست ارتباط شناخته شده برای ارزیابی وابستگی متقابل پیش پردازش می شود. این با محاسبه ارتباطات فرکانس بین هر دو نقطه در شبکه حل می شود. به طور خاصتر، فرکانس درخواست های ارتباط بین هر VMi و VMj محاسبه می شود و همچنین فرکانس درخواست ارتباطی بین VMi و گره k که یک ابر خصوصی را نشان می دهد. این به ما نشان می دهد کدام مسیر، بیشتر ارتباطات VM را دارد. این ارتباط با وابستگی این VM نزدیک است و باید به طور ایده آل برآنچه که برنامه ریزی شده تاثیر بگذارد.



شکل8: روش مرحله به مرحله غیر بهینه

sina-pub.ir

2- در مرحله دوم، یک تابع مفید بر اساس مقادیر فرکانس اتصال تولید شده در مرحله 1 ساخته شده است. تابع ابزار به عنوان تابع هدف مسئله اصلی عمل می کند که VM ها را در میزبان اختصاص می دهد.

3- بعدا یک مشکل اساسی که ما اختصاص VM ها به سرورها و ارتباطات مسیرهای خاص بدون برنامه ریزی آنها تولید می شوند را کنترل می کنیم. به عبارت دیگر، ما متغیر تصمیم گیریXvq با توجه به محدودیت های برنامه ریزی حل می کنیم. این یک وظیفه قابل اجرا برای VM ها ایجاد می کند که هدف آن برنامه ریزی VMهای نزدیک به یکدیگر هستند.

4- پس از گرفتن مکان وظایفVM ، یک مسئله فرعی که در آن تلاش می کنیم تا برنامه ریزی بهینه برای ارتباط داده تحت این وظیفه شرایط خاصVM پیدا کنیم. به عبارت دیگر، ما برای متغیر تصمیم گیری Yipw ، Fi در مسئله فرعی را حل می کنیم. حداقل تاخیر ایجاد شده از زیرمجموعه هدف ارزشمندی است که ما به دنبال آن هستیم. همانطور که در هر تجزیه مبتنی بر بهینه سازی، موفقیت روش تجزیه بستگی به راه حل مشکل اصلی انتخاب شده دارد . ما مسئله اصلی و مسئله فرعی را به شرح ذیل مطرح می کنیم.

مشکل اصلی فرمول سازی

ما ابتدا تابع فاصله را معرفی می کنیم. این نشان دهنده فاصله بین دو گره اندازه گیری شده توسط تعداد لینک در کوتاهترین مسیر بین آنها می باشد. تابع فرکانس بر اساس مدت ارتباط نیز اضافه شده است. این تابعی است که در آن زمان ارتباط به عنوان عامل غالب ترجیح داده می شود. تابع فرکانس مقداری است که وابستگی متقابل بین دو VMیا بین VM و یک ابر خصوصی (گره مشتری) را معرفی می کند. گزینه دیگر در اینجا بسته به تعداد ارتباطات مورد درخواست بین این دو نقطه به جای مقدار کل زمان ارتباط می باشد. هنگامی که ما مقدار تابع فرکانس را محاسبه می کنیم، عملکرد نرم افزار به شرح ذیل ساخته شده است:

$$MIN \sum_{v\in VM}\sum_{u\in VM}\sum_{s\in Q}\sum_{q\in Q}(Freq_{vu}\times Distance_{sq}\times X_{vs}\times X_{uq}), \quad (12)$$

منوط به

sina-pub.ir

(2), (4). (13)

مشکل اصلی، یافتن تخصیص VM است که ارزش نقطه به نقطه وابستگی متقابل را به حداکثر می رساند.

مسئله فرعی

همانطورکه مسئله فرعی بر برنامه ریزی تمرکز دارد، هدف عملکرد آن مشابه فرم مطلوب است، یعنی، به حداقل رساندن میانگین تاخیر در ارتباط می باشد. در این مورد، ارزش نهایی هدف آرام به طور مستقیم از راه حل مسئله فرعی نشئت می گیرد. تفاوت این است که مسئله فرعی در حال حاضر می داند که VM اختصاص داده شده است و ارتباطات بر طبق آن برنامه ریزی شده است. هدف مسئله فرعی به شرح زیر است.

$$MIN \sum_{i} (F_i - r_i) \qquad i \in I, \tag{14}$$

منوط به

(3), (5) − (11). (15)

نتایج

شبیه سازی محیط

مشکل با استفاده از یک رویداد گسسته مبتنی بر برنامه شبیه سازی، شبیه سازی شده است و در یک مقیاس عملی تر با استفاده از تکنیک های جستجوی اکتشافی که در گذشته مورد بحث قرار گرفته حل شده است. شبکه مورد استفاده برای آزمایش شبکه NSF است (در شکل 9). این شامل 14 گره است که 3 گره های مرکز داده هستند و بقیه گره های مشتری در نظر گرفته می شوند [21]. گره ها با استفاده از سرعت بالا شبکه با جزئیات گره پیوسته متصل می شوند که تا 3 برابر خطوط (جریان) در هر لینک افزایش می یابند. روش مسیریابی متناوب ثابت با 3 مسیر در دسترس بین یک گره و هر نوع گره دیگر استفاده می شود. پیکربندی سرور و درخواست پارامترهای داده در جدول 4 بطور مبسوط و مشروح بیان شده است. پیش شرط درخواست ارتباط در این آزمایش مجاز نیست.

اکتشافی

همانطور که در بخش های قبلی توضیح داده شد، هر آزمایش شامل دو مرحله و بنابراین دو اکتشافات مورد نیاز

sina-pub.ir

است:

یکی به برنامه VMs در سرور و دیگری برنامه ریزی درخواست ارتباط است. پنج تکنیک قبل از ایجاد 6 ترکیب احتمالی توضیح داده شد. به هرحال، ما انتخاب کردیم تا نتایج از بهترین 4 ترکیب (بهترین 4 راه حل کامل) را نمایش دهیم. این به خاطر محدودیت های فضایی است. 4 ترکیب انتخاب شده تمام 5 اکتشافات را پوشش می دهد.

سناریوهای شبیه سازی و اکتشافات ترکیبی برای دو مسئله فرعی به شرح زیر استفاده می شود:

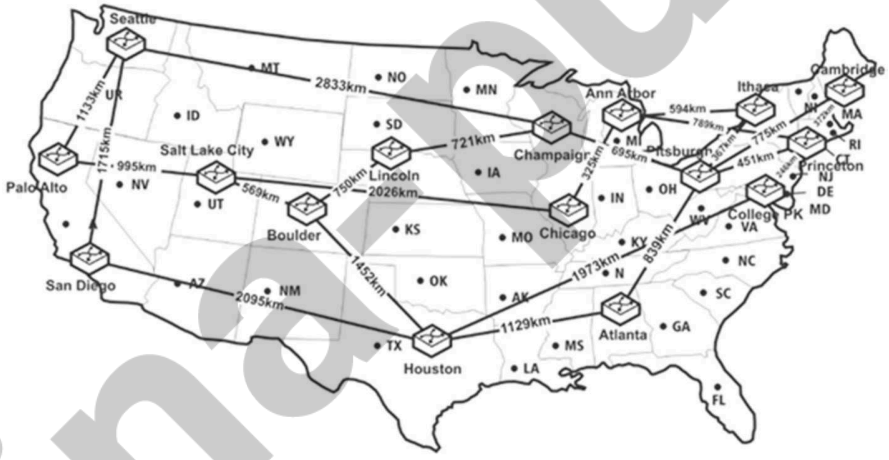
ED-GA-1: روش توزیع زمان برابر و الگوریتم حریص

RB-DP-2: روش توزیع مبتنی بر منابع و روش اولویت دوام.

ED-DP-3: روش توزیع زمان برابر و مدت زمان تکنیک اصلی

ND-DP-4: تکنیک گره دور و الگوریتم حرص

در شکل ها 10 و 11، عملکرد 4 روش بر اساس درصد مسدود شده مقایسه می شوند. این به عنوان افزایش بار درخواست ارزیابی می شود. شکل 10 مقایسه درصد درخواست های مسدود شده را نشان می دهد (درخواست هایی که نمی توان برنامه ریزی کرد) جایی که مقدار پارامتر تاخیر اندکی مجاز است (1 واحد زمان). این بدان معنی است که این سناریو شبیه به الزامات درخواست تحمیل به زمان بندی زمان واقعی می باشد.



شکل 9: شبکه NSF 14 گره

sina-pub.ir

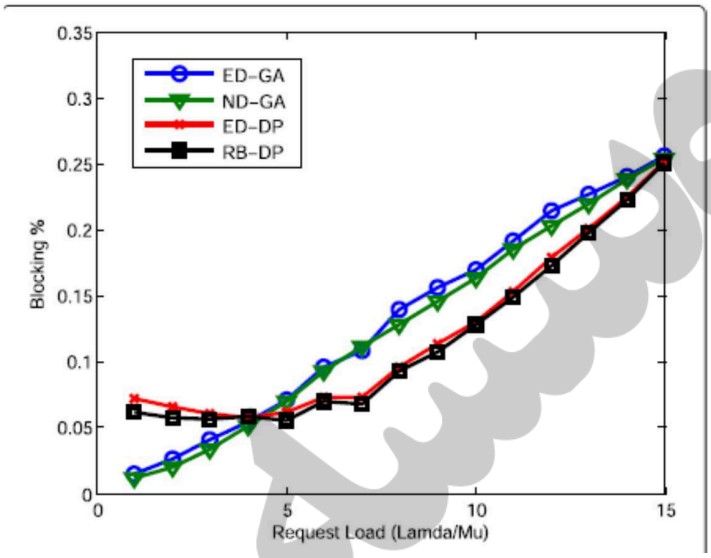
محور x نشان دهنده بار درخواستی است که توسط λ / μ اندازه گیری می شود. λ نشان دهنده میزان ورود و μ نشان دهنده میزان خدمات می باشد. شکل 11 همان مقایسه را نشان می دهد زمانی که مجاز بودن تاخیر در هر درخواست زیاد (30000 واحد زمانی) است. در هر دو سناریو متوجه شدیم که روش های ED-DP و RB-DP مزیت واضح با امتیاز دادن به طور مداوم درخواست های کم مسدود شده را نشان داده اند. عامل مشترک برای این 2 روش در حال استفاده DP برای برنامه ریزی اتصالات می باشد. بنابراین، این مزیت روش استفاده از DP بر GA هنگام برنامه ریزی درخواست اتصال در شرایط زمانی سخت یا واقعی را نشان می دهد. علاوه بر این، همانطور که در شکل 11 دیده می شود، RB-DP به خوبی مزیت بیش از ED-DP از نظر درصد مسدود کردن را نشان داده است.

جدول 4: پیکربندی پارامتر تجربی

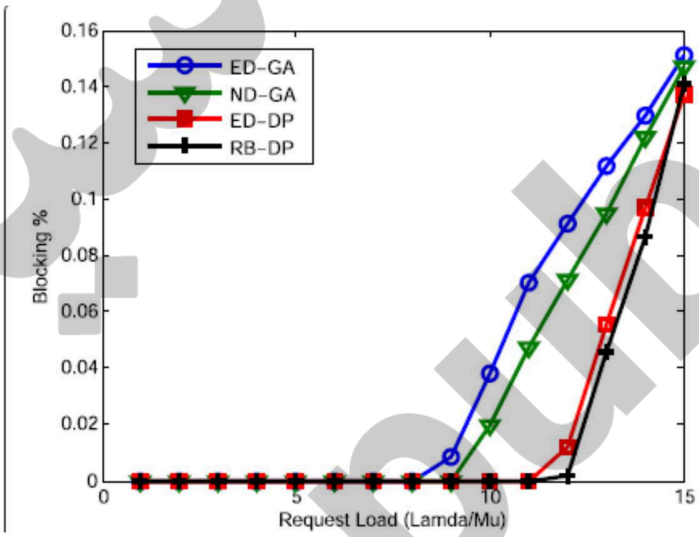
پارامتر	مقدار
تعداد کل سرورها	132
سرور / مرکز داده	44
درخواست رزرو VM	200
درخواست ارتباط	10000
توزیع RST	پواسون با لامبدا = ۱۰
توزیع مدت زمان ارتباط	نرمال با میانگین = ۲۰۰ واحد زمانی
توزیع منبع و مقصد	یکنواخت
مجاز بودن تاخیر برای هر درخواست	از دامنه ۱ تا ۵۰۰ واحد زمانی
زمان تجربی کل	۷۰،۰۰۰ واحد زمانی

با توجه به ماتریس عملکرد دیگر، میانگین تاخیر در هر درخواست، ارزیابی ها در شکل 12 نشان داده شده است. شکل نشان می دهد که مقادیر متوسط تاخیر در هر درخواست، در زمان استفاده از چهار روش تولید می شود. تاخیر مجاز در این آزمایش اندک است (25 واحد زمان). یک بار دیگر، روش های ED-DP و RB-DP نشان دهنده یک مزیت واضح با امتیاز دهی ثابت تاخیر کمتر در هر درخواست می باشد. همچنین، از شکل متوجه شدند که ED-DP نتایج کمی بهتر از RB-DP را تولید می کند (میانگین کمتر تاخیر).

sina-pub.ir



شکل 10: نتایج مسدود کردن درخواست برای روش های برنامه ریزی (تاخیر مجاز/درخواست= 1 واحد زمانی)



شکل 11: نتایج مسدود کردن درخواست برای روش های برنامه ریزی (تاخیر مجاز/درخواست= 30000 واحد زمانی)

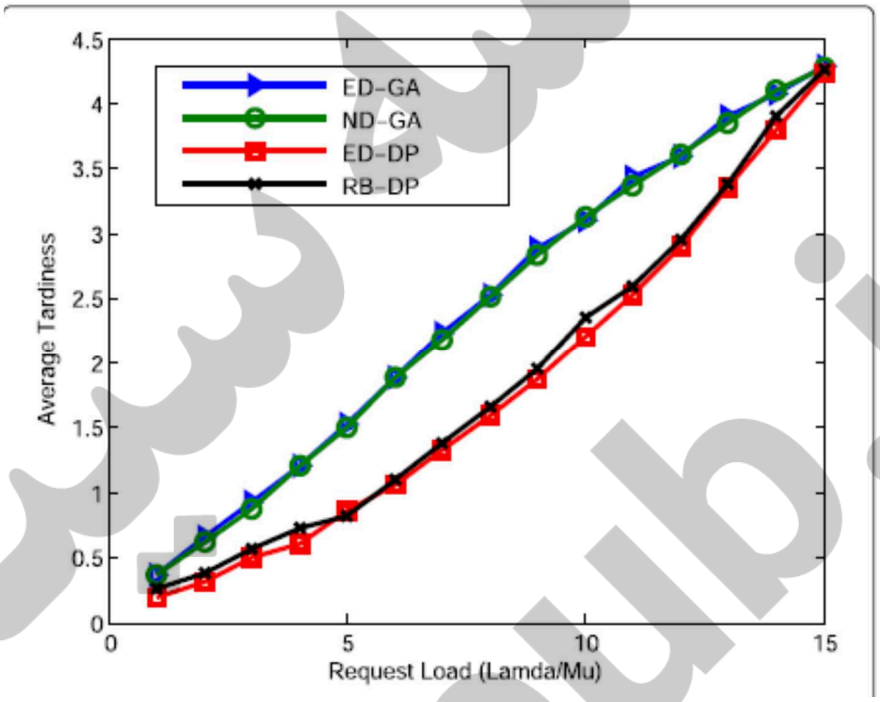
بنابراین، استفاده از روش RB-DP مناسب تراز سناریوهایی است که تاکید بر خدمت به بیشترین تعداد درخواست ها دارد. از سوی دیگر، استفاده از ED-DP برای سناریوهایی مناسب تر است که در آن عملکرد درخواست یا سطح خدمات فرد بیش از ارائه دادن درخواست های بیشتر اولویت بندی شده است.

نتایج راه حل ساده

sina-pub.ir

با توجه به شبکه ای که ما تست کردیم، از یک شبکه 5 گره در این آزمایش ها با 2 گره همانند گره های مرکز داده و بقیه به عنوان گره های مشتری (ابر خصوصی) استفاده کردیم. چهار سرور در آزمایش ها با 2 سرور در هر مرکز داده مورد استفاده قرار گرفت

برای اتصال گره ها در شبکه فیزیکی، 7 لینک استفاده و 20 مسیر مختلف تعریف شد. دو مسیر روتین جایگزین برای هر دو گره تعریف شد.



شکل 12 : درخواست میانگین تاخیر ناشی از روشهای زمانبندی انجام شده است (تاخیر مجاز / درخواست = 25 واحد زمان) داده حاوی اطلاعات مربوط به 5 نمونه VM است.

انتخاب این شبکه بخاطر دو عامل است. اول، متراکم کردن درخواست در معماری با منابع محدود که شبکه را تحت فشار زیادی قرار می دهد تا اثر ظرفیت شبکه را از بین ببرد که بر نتایج تاثیر خواهد داشت. این به کنترل بیشتر با حذف هر گونه عوامل مرتبط به طراحی شبکه یا توزیع گره اجازه خواهد داد که ممکن است فشار بر الگوریتم زمانبندی را کاهش دهد. این راه، اندازه مشکل فقط با استفاده از پارامترها کنترل می شود ما آزمایش می کنیم که

sina-pub.ir

کدام تعداد درخواست ها مشخصات و توزیع آنها هستند. دوم، آن را ساده تر می کند تا نتایج و زمان اجرای راه حل ساده نسبت به راه حل بهینه را مقایسه کند.

برای درخواست های ارتباط حاصل از مشتریان، همانند [22]، مقادیر ورودی آنها بر اساس روند پواسون تنظیم شده است. طول عمر درخواست ارتباط (مدت زمان) به طور معمول با میانگین 100 واحد زمانی توزیع شد و تعداد کل درخواست های ارتباطی به تدریج از 20 تا 3000 درخواست افزایش می یابد. هردرخواست ارتباطی با یک منبع، یک مقصد، زمان شروع درخواست و مدت زمان در ارتباط است. گره های منبع / VM به صورت یکنواخت توزیع شد.

برای ارزیابی راه حل های بهینه و ساده، ما از استودیوی بهینه سازی IBM ILOG CPLEX v12.4. استفاده کردیم. هر دو راه حل بهینه و ساده با استفاده از زبان برنامه نویسی بهینه (OPL) برنامه ریزی و چندگانه مراحل تست انجام شد. هر دو راه حل برای مقادیر متعدد بار شبکه نرمال تست شدند.

جدول 5 مقایسه بین مقادیر عینی بدست آمده با استفاده از طرح بهینه در مقایسه با مقادیر حاصل از طرح ساده (تجزیه شده) برای مقیاس اندک مسئله را نشان می دهد (حداکثر تا 200 درخواست). در حالی که راه حل بهینه قادر به زمانبندی تمام درخواستها بدون هیچ تاخیری (تداخلی)می باشد، راه حل ساده میانگین تاخیر قابل قبولی را در مقایسه بدست آورد.

همانطور که از جدول متوجه شدید، زمان اجرا برای طرح بهینه کمی برای مجموعه داده های کوچک بهتر است، اما همانطورکه تعدادی از درخواست ها رشد می کنند، تفاوت در زمان اجرا آشکار می شود. این تا زمانی ادامه پیدا می کند که راه حل بهینه غیر قابل انجام شود در حالی که راه حل ساده هنوز در دوره نسبتا کوتاه اجرا می شود. حداکثر تعداد درخواست ها راه حل بهینه بستگی به تعداد ماشین آلات مورد استفاده دارد و پارامترهای بارگذاری شبکه برای تولید داده های ورودی استفاده می شود.

در مورد مسائل بزرگ در مقیاس بزرگ، نتایج تجربی در جدول 6 نشان می دهد که راه حل ساده به میانگین تاخیر قابل قبول در مقایسه با راه حل مطلوب دست یافته است. اثر افزایش اندازه مشکل در مقدار متوسط تاخیردر هنگام

sina-pub.ir

استفاده راه حل ساده آشکار است. میانگین تاخیر به دست آمده کمتر از 10٪ از میانگین مدت زمان درخواست می باشد (طول عمر). این به خوبی در مجموعه محدود در [23] برای تاخیر قابل قبول ارتباط است که نصف طول عمر (50٪) یا مدت زمان درخواست ارتباط است.

جدول 5: مقادیر راه حل بهینه در برابر ساده و زمان اجرا

تعداد درخواست ها	بارشبکه	راه حل بهینه		راه حل ساده	
		مقدار میانگین تاخیر	زمان اجرا	مقدار میانگین تاخیر	زمان اجرا
30	0.86	0	3 s	5.73	8.14 s
50	0.86	0	7 s	10.12	9 s
200	0.86	0	2 min 24 s	10.785	1 min 2 s

همچنین بهبود قابل ملاحظه ای نسبت به عملکرد راه حل اکتشافی است که در همان جدول نشان داده شده است (مقدار میانگین تاخیر حدود 20 درصد از عمر درخواستی). جدول همچنین افزایش در میانگین تاخیر را در زمان افزایش تعداد درخواست ها نشان می دهد (اندازه مشکل). این به خاطر این واقعیت است که تاخیر به عنوان اولویت داده شده به درخواست قبل از آن تجمع می یابد. بر اساس زمان اجرا، همانطورکه تعداد درخواست ها رشد می کند، تفاوت در زمان اجرای بین راه حل های مطلوب و ساده آشکار می شود. راه حل بهینه غیر قابل انجام می شود در حالی که راه حل ساده هنوز در برنامه نسبتا کوتاه مدت 3000 درخواست در یک دوره بین 8-11 دقیقه بسته به بار شبکه اجرا می شود.

برای نشان دادن تاثیر پارامتر تاخیر مجاز بر اساس نسبت پذیرش درخواست، نتایج در جدول 7 ارائه شده اند. با استفاده از راه حل اکتشافی با ترکیب RB-DP، جدول افزایش در نسبت پذیرش را نشان می دهد همانطورکه ما تاخیر مجاز در هر درخواست را برای بارگذاری شبکه خاص افزایش می دهیم. برای اندازه گیری نسبت پذیرش، ما حداکثر پارامتر دوره انتظار را معرفی کردیم (یا تاخیر مجاز همانطورکه در بخش های قبلی مورد بحث قرار گرفت). این پارامتر نشان دهنده مدت زمانی است که یک درخواست اتصال انتظار می رود خدمت پیش از آن مسدود شده است.

sina-pub.ir

برای آن یک ارزش ایده آل همان ارزش استفاده شده در [23]، یعنی، نیمی از درخواست طول عمر می باشد. به عبارت دیگر، اگر ارتباط برای بیش از 50٪ مدت زمان خود صبر کرد و برنامه ریزی نشده بود سپس مسدود شده یا ارائه نمیشود. جدول 6 نسبت پذیرش و میانگین تاخیر برای درخواست با یک مدت زمان متوسط 100 واحد زمانی را نشان می دهد با توجه به این سناریو که در آن درخواست ها با خستگی بالا مسدود شده است تجارت بین میانگین تاخیر ارتباط و درصد (یا تعداد) ارتباطات مسدود شده را ارائه می دهد.

جدول 6: زمان اجرا و متوسط تا خیر برای درخواست ارتباط برای مشکلات بزرگ در هنگام افزایش متوسط مدت زمان اتصال به 100 واحد زمان

بار شبکه (RB-DP)	راه حل اکتشافی		راه حل ساده	راه حل ساده
زمان / طول عمر	میانگین تاخیر (درصد مدت		میانگین تاخیر	زمان اجرا
0.86	19.81%		2.88%	8 min 21 s
0.93	21.18%		6.36%	8 min 54 s
1	22.54%		9.08%	11 min 31 s

متوجه شده اید که میانگین تاخیرکاهش می یابد همانطورکه ما درخواستها را با تاخیر بالا حذف می کنیم و آنها را مسدود در نظر می گیریم. میانگین تاخیر کمتر از 2٪ از درخواست طول عمر می تواند تضمین شود اگر ما مایل به قربانی کردن 13 درصد از درخواست های مسدود شده باشیم. تصمیم به استفاده از این سناریو برای آب و هوای مناسب معماران راه حل ابر نیست. این بستگی به حساسیت مشتری به دقت / کیفیت در مقابل سرعت دستیابی به نتایج دارد.

مقایسه با راه حل های قبلی

هنگام برنامه ریزی مقایسه بین راه حل پیشنهادی و راه حل های موجود در پیشینه تحقیق، ما با چالش مواجه می شویم. همانطور که در جزئیات بخش کار "مرتبط" بحث شد "، راه حل های در دسترس پارامترهای مورد نظر و ابعاد پوشیده شده از مشکل تخصیص منابع ابر متنوع هستند. این محدودیت تعدادی از راه حل هایی دارد که به طور واقعی می تواند برای حل این عطر و طعم خاص مشکل مورد استفاده قرار گیرد. به هر حال، ما قادر به استفاده

sina-pub.ir

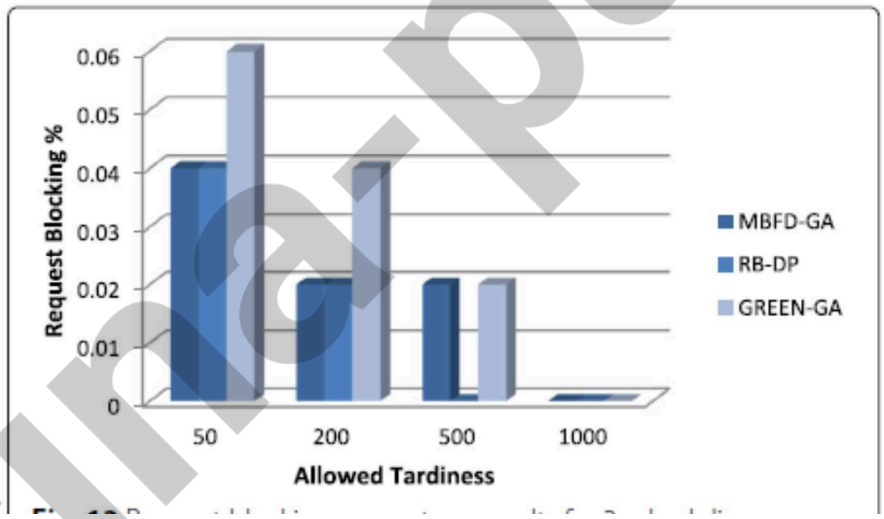
از الگوریتم های اجرا شده در [6] (بهترین روش مناسب کاهش اصلاح شده) و [24، 25] (برنامه ریزی سبز) برای حل مشکل مشابه و مقایسه عملکرد آنها به روشی که ما توسعه دادیم. تمرکز این بود: ظرفیت شبکه (حداقل درصد مسدود کردن) و عملکرد (زمانی که مسدود کردن مسئله نیست، به حداقل رساندن میانگین تاخیر در هر درخواست ارائه شده یک اولویت است). این مقایسه برای اولین شبکه کوچکتر به منظور کشف اثر استرس بر روی یک شبکه ابر انجام شد. سپس، همان مقایسه برای سناریوی شبکه واقع گرایانه تر بزرگتر انجام شد. همانند آزمایش های قبلی، تست ها برای اندازه های مسائل مختلف و سطوح مختلف مجاز هر درخواست انجام شد.

نتایج شبکه کوچک

شکل 13 عملکرد سه الگوریتم را بر اساس درخواست درصد مسدود برای سناریوهای متفاوت مقایسه می کند همانطورکه تاخیر سطح مجاز افزایش می یابد.

جدول 7 میزان پذیرش درخواست ارتباط برای بارهای شبکه مختلف با استفاده از راه حل های ساده

تاخیر مجاز (درصد درخواست طول عمر یا مدت زمان)	50%	200%	1000%
میزان پذیرش	86%	87%	100%
میانگین تاخیر برای درخواست های پذیرفته شده	1.98%	16.72%	219.767%



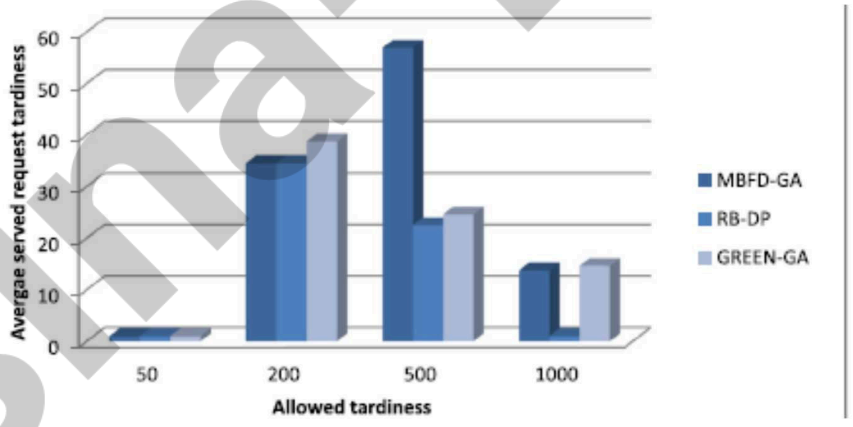
sina-pub.ir

شکل 13 درصد مسدود کردن درخواست نتایج برای 3 برنامه ریزی الگوریتم های اندازه گیری شده برای 4 **موارد مختلف تاخیر مجاز**

این شکل نشان می دهد که تکنیک ما (RB-DP) به طور مداوم بهتر (درصد مسدود کردن کمتر) از الگوریتم زمانبندی سبز عمل می کند. آن همچنین نشان می دهد که RB-DP در همان سطح MBFD برای سطح تاخیر مجاز قبل از نشان دادن یک مزیت برای تاخیر مجاز بالا انجام می شود. شکل 14 نتایجی را برای متریک دیگر، میانگین درخواست تاخیر ارائه می دهد. شکل 14 نشان می دهد که RB-DP با انجام همان سطح از دو الگوریتم های دیگر شروع می شود و در حالی که ما سطح تاخیر مجاز برای درخواست ها را افزایش می دهیم، RB-DP نشان دهنده مزیت قابل توجه است (همانطور که در آخرین مورد تاخیر مجاز = 1000 واحد زمانی مشاهده شد). تأثیر افزایش تاخیر مجاز، اساسا از بین بردن نیاز برای برنامه ریزی هر درخواست به محض اینکه آن (برای جلوگیری از درخواست مسدود) حاصل شود. در عوض، آن بر آزمایش برای نشان دادن الگوریتم تمرکز می کند که می تواند درخواست را در کارآمد ترین شیوه برنامه ریزی کند / ارائه دهد و این، به نوبه خود، تاخیر میانگین در هر درخواست را کاهش می دهد.

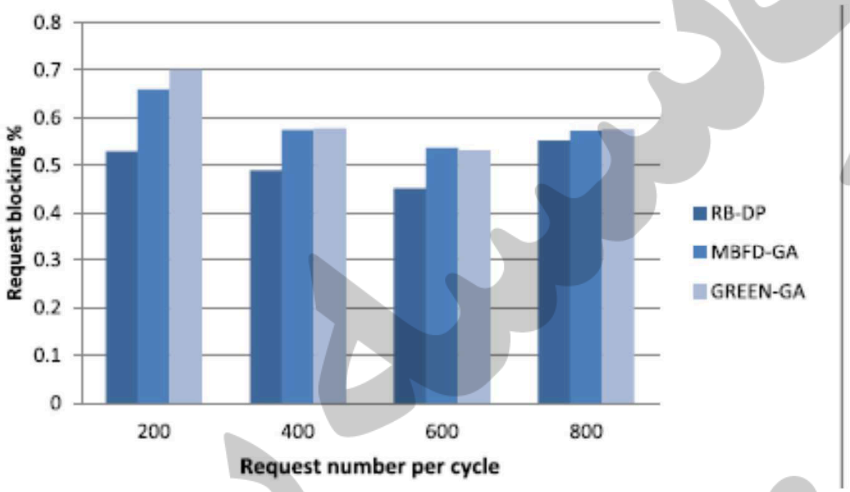
نتایج شبکه بزرگ

روند مشابهی در زمان آزمایش در شبکه های بزرگ انجام می شود (شبکه NSF). در شکل 15، درصد مسدود کردن برای سه الگوریتم برای اندازه های مسائل مختلف نشان داده شده است. اندازه مشکل در اینجا توسط تعدادی از درخواست های ارسال شده به کنترل مرکزی در هر چرخه نشان داده شده است.



sina-pub.ir

شکل 14: نتایج میانگین تاخیر درخواست برای 3 الگوریتم زمانبندی ارائه می دهد برای 4 موارد مختلف تاخیر مجاز ارزیابی شده است



شکل 15 نتایج درصد مسدود کردن درخواست برای روش های برنامه ریزی با تاخیر اجازه =5 واحد (بسیار کم) (نرخ ورود).

این نتایج برای سطح تاخیر مجاز = 5 واحد زمانی نشان داده شده است (سطح بسیار پایین) که فشار اضافی برای ارائه درخواست در یک دوره کوتاه از ورود آنها می افزاید و تمرکز الگوریتم های کار در ارائه بیشترین تعداد درخواست ها در سطوح تدریجی نیست. در شکل 16، نتایج مشابهی برای تعداد بیشتری از درخواست ها بین 3000 تا 10،000 در هر دوره نشان داده شده است. این ثابت می کند که نتایج تجربی ما سازگار است هنگامی که شبکه در معرض بار بالاتر که نزدیک است یا از ظرفیت آن فراتر می رود. در هر دو صورت، آنها نشان می دهند که تکنیک ما (RB-DP) همواره از دو الگوریتم دیگر تحت بارهای بالا بهتر عمل می کند.

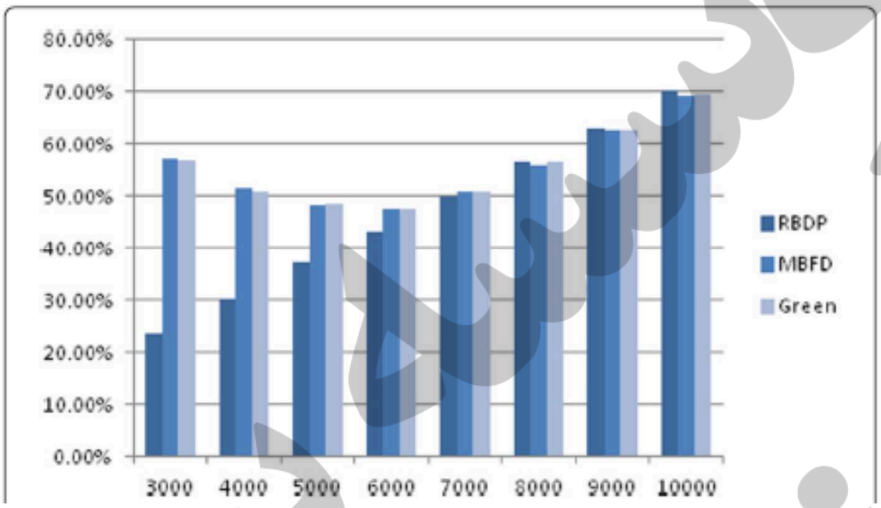
شکل 17 عملکرد الگوریتم ها را در مورد خاصی از سطوح تاخیر مجاز بالا بررسی می کند.

RB-DP مزایای واضح را از لحاظ مسدود کردن درصد متریک برای سطوح مختلف تاخیر مجاز ارائه می دهد.

حرکت به ماتریس دوم، شکل 18 عملکرد سه الگوریتم براساس درخواست میانگین تاخیر در زمان تغییر سطوح تاخیر مجاز (یا درخواست طول عمر) را نشان می دهد. RB-DP در یک سطح قابل مقایسه به دو الگوریتم دیگر

sina-pub.ir

برای تاخیر سطوح مجاز کوچک و سپس فراتراز عملکرد MBFD سطوح متوسط طول عمر درخواست شروع می شود و سپس به وضوح بیش از هر دو الگوریتم با سطوح بالاتر 400 واحد زمانی آغاز می شود.



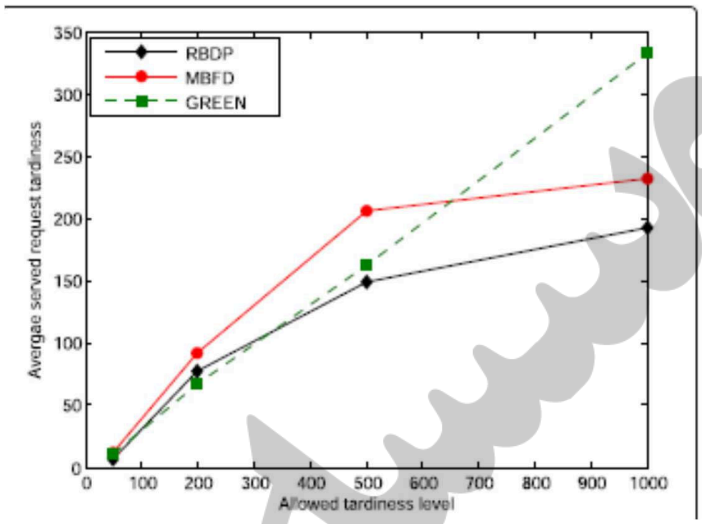
شکل 16 نتایج درصد درخواست مسدود کردن برای 3 الگوریتم برنامه ریزی برای تعداد زیادی از درخواست ها در هر چرخه (1-10K).

این نتایج پتانسیل راه حل ما را براساس به دست آوردن عملکرد بهتر در هر دو درصد مسدود کردن (درخواست ارتباط بیشتری مورد قبول و احتقان شبکه کمتر است) و میانگین تاخیر (کیفیت بهتر شرایط سرویس برای کاربران ابر) ثابت می کند.

نتیجه گیری

ما یک راه حل جامع برای مقابله با آن مشکل تخصیص منابع در مرکز داده های ابر محاسباتی معرفی کردیم [26]. اول، مشکل به صورت مدل خطی عدد صحیح ترکیبی فرمول بندی شد. این فرمول با استفاده از یک کتابخانه بهینه سازی برای مجموعه داده های کوچک حل شد. به هر حال، پیدا کردن راه حل بهینه برای سناریوهای کارآمد بزرگتر با استفاده از فرمول ریاضی مطلوب، امکان پذیر نیست. بنابراین، ما 5 روش اکتشافی را برای مقابله با دو طرف مشکل، یعنی رزرو VM و برنامه ریزی ارتباط معرفی کردیم. عملکرد این تکنیک ها مورد تجزیه و تحلیل قرار گرفت و مقایسه شد.

sina-pub.ir



شکل 18 نتایج تاخیر میانگین درخواست برای روش های برنامه ریزی (تاخیر مجاز / درخواست = 25 واحد زمان)

اگر چه راه حل مسئله مقیاس حل شد، یک راه حل اکتشافی بصورت بهینه سازی تضمینی ارائه نمی کند. این انگیزه را برای معرفی یک راه حل غیرقطعی به وجود می آورد. راه حل شامل 4 مرحله می باشد که وابستگی VM را به عنوان عامل غالب در فرآیند تخصیص VM مورد سوء استفاده قرار می گیرد. این به ما اجازه می دهد مرحلهٔ زمانبندی را بطور بهینه در مرحله بعدی حل کنیم که باعث می شود راه حل به طور قابل توجهی بهبود یابد. این راه حل ساده نتایجی مطابق با پارامترها از پیش تعیین شده در پیشینه تحقیق برای میانگین تاخیر ارتباط ارائه می دهد. نتایج نیز برای سناریوی نشان داده شد که مسدود کردن درخواست مجاز است. نتایج بدون قربانی کردن امکان سنجی محاسباتی حاصل شد که روش ما یک راه حل معتبر برای رسیدن به سطوح تاخیر ارتباط قابل قبول را نشان می دهد. علاوه بر این، راه حل پیشنهادی با دو الگوریتم برجسته در پیشینه تحقیق مقایسه شد. راه حل پیشنهاد شده نشان داده شد تا بر اساس به حداقل رساندن هر دو میانگین تاخیر درخواست و مسدود کردن درصد فرمول سناریوهای شبکه ابر مفید باشد. این باعث داشتن نامزد قوی برای استفاده در سناریوهای ابر می باشد که در آن تمرکز بر متریک مانند درخواست های پذیرفته شده ارتباط بیشتر و شبکه کمتر احتقان یا میانگین تاخیر درخواست می باشد (کیفیت بهتر شرایط خدمات برای کاربران ابر).

در آینده، ما قصد داریم از طرح خود برای آزمایش با سایر اهداف مهم ارائه دهنده ابر استفاده کنیم[27، 28]. حفظ حریم خصوصی در هنگام پردازش و ارتباط داده ها از طریق منابع ابر یک چالش مهم است.

sina-pub.ir

حریم خصوصی یک نگرانی عمده برای کاربران در ابر یا برنامه ریزی برای حرکت به ابراست. بهبود حریم خصوصی داده ها معیارها نه تنها برای مشتریان مهم هستند بلکه برای انطباق با مقررات دولتی حیاتی هستند که سریع تحقق می یابد. این بدان معنی است که سیستم تخصیص منابع باید لیستی از اولویت های خود را به آن اضافه کند که شامل معیارهای خصوصی علاوه بر عملکرد معمول و معیارهای هزینه می باشد. محدودیت در کنترل داده ها، حرکت داده ها و مکان های برنامه ریزی باید دیده شود. این حریم خصوصی به اطلاعات مربوط به منابع مورد نیاز ابر مشتریان ما منتقل می شود برخی از چالش های موضوع را مورد بررسی قرار داده ایم [3]. گام بعدی ما گسترش مدل برای کشف این امکانات است. این ابعاد مختلف برای دادن مزیت رقابتی به ارائه دهندگان ابر اضافه خواهد کرد که سطح مورد انتظار از حریم خصوصی را به مشتریان آینده نگر ارائه می کنند.

تقدیرنامه ها

نویسندگان تمایل دارند تا از دکتر رجول چودری برای کمک وی به اجرای کد و بخش کار مربوطه تشکر کنند. نویسندگان نیز دوست دارند تا از دکتر دائه هون بان از سامسونگ برای اطلاعات دقیق خود تشکر کنند.

منابع مالی

این کار تا حدودی توسط شورای پژوهشی علوم طبیعی و مهندسی کانادا (447230NSERC-STPGP) و پاداش پژوهشی جهانی سامسونگ(GRO) حمایت می شود.

مشارکت نویسندگان

دکتر محمدعباب شکور وضعیت هنر این رشته را بررسی کرد تجزیه و تحلیل تکنیک های تخصیص منابع فعلی و محدودیت های آنها را انجام داد. او مسئله را در فرم های بهینه وغیربهینه فرمول بندی می کند، شبیه ساز ارائه شده در این مقاله، و انجام آزمایش و نیز تجزیه و تحلیل اجرا شده است. پروفسور عبدالله شامی این تحقیق را آغاز و تحت نظارت، هدایت و مشارکت علمی خود را تأیید کرد، ارائه داده کلی، بررسی مقاله و تاییدیه خود را برای نسخه نهایی صادر کرد. دکتر عبدالقادر اوودا داده های عمومی را ارائه و مقاله را بررسی کرد. همه نویسندگان نسخه نهایی مقاله را خواندند و تایید کردند.