

موضوع: بهبود تخصیص منابع در محاسبات ابری سیار با استفاده از تقسیم وظایف

مبتنی بر DRL

۱- امین کریم الله دهکردی

(کارشناسی ارشد مهندسی کامپیوتر)

۲- دکتر گلناز آقایی

(دکتری تخصصی سیستمهای نرم افزاری و استادیار دانشگاه)

۱-۱- چکیده

مسئله اصلی در این پژوهش، بهبود تخصیص منابع و مدیریت توازن بار در محیطهای محاسبات ابری سیار است که با چالشهایی نظیر مصرف انرژی بالا، نوسانات در تقاضای منابع، و نیاز به توازن بار بین ماشینهای مجازی مواجه است. در این مقاله، به منظور حل این چالشها، رویکرد جدیدی مبتنی بر الگوریتم بهینهسازی کلونی مورچهها (ACO) ارائه شده است. الگوریتم پیشنهادی با بهره‌گیری از اصول هوش جمعی و قابلیت جستجوی کارآمد، به دنبال بهبود تخصیص منابع، کاهش مصرف انرژی، و افزایش توازن بار در سیستمهای ابری سیار است. برای ارزیابی عملکرد روش پیشنهادی، این الگوریتم با چندین روش مرجع از جمله الگوریتمهای توازن بار معمولی و الگوریتمهای بهینهسازی متداول مانند الگوریتم ژنتیک (GA) و الگوریتم بهینهسازی ازدحام ذرات (PSO) مقایسه شده است. معیارهای ارزیابی شامل زمان پاسخ، مصرف انرژی، و میزان توازن بار در سیستمهای ابری بوده‌اند. نتایج شبیه‌سازیها نشان داد که روش پیشنهادی بهبود قابل توجهی در تمامی معیارها ارائه می‌دهد. به طور خاص، الگوریتم پیشنهادی توانست مصرف انرژی را تا ۱۵٪ کاهش دهد، زمان پاسخ سیستم را تا ۲۰٪ بهبود بخشد، و توازن بار را در مقایسه با روشهای مرجع به طور چشمگیری ارتقاء دهد. در نهایت، این پایان‌نامه با ارائه یک رویکرد نوآورانه مبتنی بر ACO، راهکاری موثر برای بهبود مدیریت منابع و توازن بار در محیطهای محاسبات ابری سیار پیشنهاد می‌کند که می‌تواند منجر به افزایش کارایی و کاهش هزینه‌های عملیاتی در این سیستمها شود.

واژگان کلیدی: محاسبات ابری، تخصیص منابع، کاهش مصرف انرژی، الگوریتم کلونی مورچه، بهینه‌سازی.

۱-۲- مقدمه

اینترنت، از ابتدای آغاز کار خود تاکنون، دست خوش تحولات فراوانی شده است که بعضی از آنها موجب تغییر شیوه زندگی بشر در چند دهه اخیر گشته است. یکی از جدیدترین تغییرات در نحوه کارکرد اینترنت، با معرفی رایانش ابری صورت پذیرفته است. این فناوری جدید به دلیل ویژگی‌هایش به سرعت محبوب شده است چرا که در رایانش ابری همه نوع امکانات به کاربران، به عنوان یک سرویس ارائه شده است. طبیعتاً هر تغییر و مفهوم جدیدی در دنیای فناوری، مشکلات و پیچیدگی‌های خاص خود را دارد. بهره‌گیری از رایانش ابری نیز از این قاعده مستثنی نبوده و چالشهای فراوانی را پیش روی صاحب نظران این حوزه قرار داده است که از

آن جمله می‌توان به مواردی نظیر: توازن بار، امنیت، قابلیت اطمینان، مالکیت، پشتیبان گیری از داده‌ها و قابلیت حمل داده‌ها اشاره کرد. الگوریتم‌های زمان بندی بسیاری برای توازن بار روی ابر پیشنهاد شده اند که می‌توان آنها را به الگوریتم‌های تکاملی و غیر تکاملی تقسیم بندی کرد. روش‌های غیرتکاملی پیچیدگی‌های پیاده سازی کمتری نسبت به روش‌های تکاملی دارند که از آنها می‌توان به الگوریتم‌های نوبت گردش و روش دانلود دو سویه یاد کرد. کلونی مورچه‌ها، بهینه سازی الگوریتم ازدحام ذرات، روش‌های بهینه سازی الگوریتم ژنتیک از نمونه الگوریتم‌های تکاملی می‌باشند که با توجه به ماهیت چند هدفی روش‌های زمانبندی با در نظر گرفتن NP- Complete بودن مسئله توازن بار در ابرها در عمل روش‌های تکاملی کیفیت بهتری دارند. الگوریتم‌های تکاملی مسئله را با راه حل‌های مختلف حل می‌کنند و هر مجموعه از راه حل‌ها که نسبت به بقیه پاسخ بهتری داشته باشند به عنوان مجموعه جواب بهینه در نظر گرفته می‌شود.

۳-۱- بیان مسئله

با رشد فزاینده فناوری‌های اطلاعات و ارتباطات، محاسبات ابری به یکی از ارکان اساسی زیرساخت‌های فناوری تبدیل شده است. این فناوری به کاربران و سازمان‌ها اجازه می‌دهد تا به منابع محاسباتی و ذخیره‌سازی داده‌ها به صورت پویا و مقیاس‌پذیر دسترسی پیدا کنند، بدون اینکه نیاز به سرمایه‌گذاری سنگین در سخت‌افزار داشته باشند. در عین حال، با توجه به افزایش استفاده از دستگاه‌های سیار و نیاز مبرم به دسترسی به خدمات ابری از طریق این دستگاه‌ها، محاسبات ابری سیار (Mobile Cloud Computing) به عنوان یک حوزه نوظهور با پتانسیل‌های بزرگ، مطرح شده است. این نوع از محاسبات امکان استفاده از منابع محاسباتی قدرتمند ابری را برای دستگاه‌های سیار با توان محاسباتی محدود فراهم می‌کند. با این حال، یکی از چالش‌های اساسی در این حوزه، مدیریت بهینه منابع و توازن بار بین ماشین‌های مجازی است که به دلیل ویژگی‌های خاص محیط‌های سیار، از جمله نوسانات تقاضا و محدودیت‌های انرژی، به مراتب پیچیده‌تر از محیط‌های ابری سنتی می‌باشد. در این مقاله، ما به بررسی و تحلیل دقیق این چالش‌ها پرداخته و به دنبال ارائه یک رویکرد نوآورانه و کارآمد برای مدیریت منابع و توازن بار در محیط‌های محاسبات ابری سیار هستیم. به‌طور مشخص، هدف این پژوهش، توسعه و ارزیابی یک الگوریتم بهینه‌سازی جدید مبتنی بر هوش جمعی است که قادر باشد با در نظر گرفتن نوسانات تقاضا، بهینه‌سازی مصرف انرژی، و کاهش زمان پاسخ، عملکرد سیستم‌های ابری سیار را بهبود بخشد. این الگوریتم با بهره‌گیری از قابلیت‌های کلونی مورچه‌ها (ACO) طراحی شده است که با توجه به انعطاف‌پذیری و کارایی بالا، پتانسیل زیادی برای حل این مسئله دارد.

۴-۱- پیشینه پژوهش

الینا پسینی و همکارانش در سال ۲۰۱۵ با استفاده از یک الگوریتم ACO سعی بر تعادل سازی بین توان و زمان پاسخ در یک بستر laas داشتند که برای رسیدن به هدفشان یک وزن برای تعدادی از کاربران سرویس داده شده (weight SU) و یک وزن برای تعدادی از ماشین‌های مجازی ایجاد شده (weight VMs) در نظر گرفتند و از دو سناریوی HTC و HPC با اختصاص دادن وزن‌های ترکیبی به ترتیب (۱،۰) و (۰،۱) و یک سناریوی مخلوطی HTC - HPC با اختصاص دادن وزن (۰،۵) و (۰،۵) استفاده کردند که کارشان باعث افزایش بهبودی در تعداد ماشین‌های مجازی، ایجاد تعادل بین تعداد کاربران سرویس داده شده و تعداد ماشین مجازی ایجاد شده در هنگام استفاده از سناریوی ترکیبی HTC-HPC شد ولی هنگام استفاده از دو سناریوی خالص HTC و HPC در رده دوم

قرار گرفت و نسبت به GA و Random عملکرد بدتری از خود نشان داد به همین دلیل باید بر روی میزان مصرف منابع شبکه و پیام‌های ارسالی تحقیقات بیشتری انجام گیرد.

سید ابراهیم دشتی و همکارانش در سال ۲۰۱۵ یک معماری سلسه مراتبی برای برآوردن اهداف ارائه دهنده و مصرف کننده ابر ارائه دادند در واقع یک مدل جهت قرار دادن VMS به صورت پویا جهت بهبود بهره‌وری انرژی توسط الگوریتم PSO در محاسبات ابری طراحی کردند و برای تضمین کیفیت خدمات از وظایف کاربران و کاهش بهره‌وری انرژی یک الگوریتم PSO تغییر یافته برای تخصیص VMS به میزبان و مهاجرت بیش از حد VMS پیشنهاد داده اند و با سه الگوریتم PABFD و FF و BF مقایسه شد که باعث بهبود زمان و بهره‌وری انرژی شد ولی با الگوریتم‌های کارآمد و بهتر از خودش مقایسه نشد.

دینش کومار و همکارانش در سال ۲۰۱۵ یک مدل بهینه از PSO بر اساس استراتژی برنامه ریزی VM برای به حداقل رساندن اتلاف کل منابع و سرور می‌باشد. با سه الگوریتم BF و FF و WF مقایسه شد ولی اگر با الگوریتم‌های جدید تر و کارآمدتر مقایسه شده بود نتیجه متفاوت تری نشان می‌داد.

۵-۱- روش شناسی پژوهش

در روش پیشنهادی ما از دو قسمت استفاده کرده‌ایم، در قسمت اول با استفاده از الگوریتم کلونی مورچه بهترین کار انتخاب می‌شود و در قسمت دوم به بهترین ماشین اختصاص داده می‌شود.

در روش پایه الگوریتم کلونی مورچه برای انتخاب کارها از یک رابطه ایستا استفاده می‌شود به طوری که در ابتدا برای تمام کارها مقدار فرمون محاسبه می‌شود و بر اساس آن تا انتها کارها انتخاب می‌شوند ولی در روش پیشنهادی ما از یک رابطه پویا استفاده کرده‌ایم که در فواصل زمانی مشخص مقدار فرمون کارها به روز رسانی می‌شود و مقدار فرمون بر اساس شرایط جدید سیستم محاسبه می‌شود.

در روش پایه برای محاسبه فرمون از زمان ورود کار و اندازه کار استفاده می‌شود ولی در روش پیشنهادی علاوه بر زمان ورود و اندازه زمان مرگ هم به عنوان عامل جدید در نظر گرفته شده است.

در روش پایه الگوریتم کلونی مورچه هزینه‌ای که کار برای ماشین ایجاد می‌کند در نظر گرفته نشده است ولی در روش پیشنهادی هزینه اجرای کار هم به عنوان یک عامل مهم در نظر گرفته شده است.

در این قسمت برای اجرای کارها به صورت سراسری از الگوریتم مورچه استفاده خواهیم کرد، در این روش هر مورچه را یک جواب برای مسئله در نظری می‌گیریم و مورچه‌ها به صورت تصادفی توزیع شده‌اند.

تعدادی خوشه وجود دارد و تعداد مورچه‌های هر خوشه برابر می‌باشد

قصد داریم که وظایف را به ماشین‌های مجازی اختصاص دهیم و برای این کار ما سه معیار اصلی زیر را در نظر می‌گیریم

• اندازه کار

• توان پردازشی ماشین

• معیار makespan

وظایف ورودی را به صورت $TaskList = \{t_1, t_2, \dots, t_n\}$ و ماشین‌های مجازی را به صورت $VM = \{vm_1, vm_2, \dots, vm_m\}$ نشان می‌دهیم. برای نگاشت وظایف ب ماشین‌های مجازی محلی از رویکرد تکاملی کلونی مورچه استفاده می‌شود.

همانند سایر الگوریتم‌های تکاملی، ایجاد جمعیت اولیه در این روش نیز به صورت تصادفی است. در روش پیشنهادی ما هر ماشین مجازی در غالب یک مورچه در نظر گرفته می‌شود که قرار کارها را اجرا کنند.

تابع برازندگی، ویژگی کلیدی الگوریتم‌های بهینه سازی است که مقدار برازندگی هر راه حل را مشخص می‌کند. در هر تکرار از الگوریتم، تلاش برای زمان بندی K وظیفه به M ماشین مجازی است. یکی از عوامل مؤثر بر زمان بندی بهینه‌ی وظایف به ماشین‌های مجازی، ظرفیت پردازشی ماشین مجازی است قبل از این که مورچه‌ها به خوشه‌ها اختصاص داده شوند باید مقدار تابع برازندگی هر مورچه محاسبه شود، برای محاسبه تابه برازندگی هر مورچه از رابطه زیر استفاده می‌کنیم:

$$F = (\text{Processing_power}) / \text{makespan}$$

این تابع برازندگی بر اساس مقدار پردازنده ماشین و makespan محاسبه شده است، از آنجا که هر چقدر makespan کمتر باشد ماشین دز وضعیت بهتری قرار دارد رابطه فوق باعث می‌شود که پر قدرت‌ترین ماشین‌ها بالاترین مقدار تابع برازندگی را داشته باشند.

بعد از محاسبه تابع برازندگی برای کل جمعیت مورچه‌ها آن‌ها را به صورت نزولی مرتب می‌کنیم و لیستی از خوشه‌ها خالی هم داریم کل جمعیت با مورچه‌ها، به M خوشه تقسیم می‌شوند. این تقسیم به نحوی انجام می‌شود که هر خوشه شامل N مورچه باشد. روند تقسیم بندی مورچه‌ها بدین صورت است که مورچه‌ی اول به خوشه اول، مورچه‌ی دوم به خوشه دوم و مورچه‌ی M ام به خوشه M ام و مورچه‌ی M + ۱ به خوشه اول تعلق می‌گیرند. این روند به صورت مشابه تا مورچه‌ی آخر تکرار می‌شود. هر M خوشه، شامل N مورچه است.

از آنجا که ما مورچه‌ها را بر اساس تابع فیتنس به صورت نزولی مرتب کردیم اولین مورچه‌ای که به خوشه اختصاص داده می‌شود بهترین راه حل و آخرین مورچه خوشه بدترین راه حل می‌باشد، بنابراین در هر خوشه ترتیب ورود مورچه در ابتدا مهم می‌باشد ما قصد داریم که وظایف را به ماشین‌های مجازی اختصاص دهیم تا اجرا شوند ولی اهداف زیر را دنبال می‌کنیم.

اجرای کارها تا جای امکان به صورت محلی

کاهش بار پردازشی ماشین‌ها و make span

بنابراین برای یافتن بهترین جواب توسط الگوریتم مورچه دو معیار محلی بودن و makespan را باید در نظر بگیریم که در ادامه توضیح داده‌ایم

ظرفیت پردازی هر ماشین مجازی بر اساس رابطه زیر محاسبه می شود

$$\text{Processing_power} = \text{power} \times \text{PCount} \quad (۲)$$

در این رابطه power قدرت پردازش پردازنده ماشین مجازی و Pcount تعداد پردازنده های خالی می باشد.

زمان اجرای هر کار در ماشین مجازی طبق رابطه زیر محاسبه می شود:

$$\text{ExeTime} = \text{TaskTime} * \text{Processing_power} \quad (۳)$$

در این رابطه TaskTime اندازه کار می باشد که قصد اجرای آن را داریم زمان اجرا برای هر ماشین مجازی متفاوت خواهد بود. هر چقدر زمان اجرای یک کار در ماشین مجازی کمتر باشد مقدار makespan در ماشین کمتر خواهد شد برای رسیدن به این هدف طبق الگوریتم زیر عمل خواهیم کرد.

بر اساس تابع برازندگی برای مورچه ها ما در هر خوشه در جستجوی محلی در هر خوشه، موقعیت بدترین مورچه ی هر خوشه بر اساس تابع برازندگی، با توجه به موقعیت بهترین جواب آن خوشه یا حتی بهترین جواب کل خوشه ها بهبود می یابد و از این طریق میانگین برازندگی مورچه ها افزایش پیدا می کند.

۱-۶- شبیه سازی و ارزیابی روش پیشنهادی

در سال های اخیر، زیرساخت های فناوری اطلاعات به دلیل وجود تقاضا برای توان محاسباتی مورد نیاز برنامه های کاربردی به سرعت در حال رشد هستند و مراکز داده پیشرفته در محاسبات ابر میزبان انواع مختلفی از برنامه های کاربردی می باشند. هزینه های بالای انرژی و همچنین مسئله گازهای گلخانه ای حاصل از این مراکز داده بزرگ یکی از مسائل بااهمیت در طراحی و ساخت این مراکز می باشد. به همین دلیل ارائه روش هایی کارآمد برای کاهش مصرف انرژی توسط این مراکز داده بسیار مورد توجه محققان می باشد. مراکز داده پیشرفته مورد استفاده در محاسبات ابری میزبان انواع مختلفی از برنامه های کاربردی هستند. این برنامه ها ممکن است تنها برای چند ثانیه و یا برای دوره های طولانی تر از زمان، در مراکز داده اجرا شوند. در گذشته کارایی بالا به عنوان اصلی ترین معیار در طراحی مراکز داده مورد توجه بوده است و میزان مصرف انرژی این مراکز از اهمیت چندانی برخوردار نبوده است؛ اما امروزه با توجه به اهمیت منابع انرژی در سطح جهان میزان مصرف انرژی نیز به عنوان یکی از معیارهای مهم مورد توجه قرار گرفته است.

به دلیل اینکه یکی از مهم ترین مشکلات در خصوص مراکز داده مصرف انرژی بسیار زیاد آن ها و توزیع نامتوازن کارها است، لذا کاهش مصرف انرژی در این مراکز داده مستلزم استفاده بهینه از منابع پردازشی موجود از جمله ماشین های مجازی و میزبان های فیزیکی است. نحوه استقرار ماشین های مجازی در میزبان های موجود در مراکز داده از این جهت اهمیت دارد و قابل بحث می باشد که استفاده کمتر از میزبان ها درواقع معادل مصرف انرژی کمتر است و یکی از راه حل ها برای کاهش مصرف انرژی در مراکز داده کاهش بار پردازشی میزبان ها است؛ زیرا هرچقدر بار پردازشی میزبان ها کم شود مصرف انرژی آن ها هم کمتر می شود؛ اما حتی در حالتی که یک میزبان در حالت بیکار قرار بگیرد بازهم نسبت به حالتی که از تمام توان پردازشی کامل آن استفاده شود تنها ۳۰

درصد، مصرف انرژی کمتری خواهد داشت یعنی حدود ۷۰ درصد انرژی اوج توان خود را مصرف می کند . در فصل قبل یک روش جدید برای توازن بار و زمان بندی ارائه شد که در این فصل نتایج شبیه سازی آن بررسی شده است.

برای شبیه سازی روش پیشنهادی در این مقاله دو مجموعه داده مربوط به ماشین ها و کارها در نظر گرفته شده است. در این سناریو در ابتدا تمام ماشین ها خالی هستند و هیچ کاری اجرا نشده است و فرآیند شبیه سازی در دو مد مختلف انجام می شود. در این سناریو دو آیتم تعداد ماشین و تعداد کار متغیر در نظر گرفته شده است. هر ماشین در شبیه سازی دارای ویژگی های زیر می باشد:

- اندازه کل ماشین
- مقدار فضای آزاد ماشین
- مدت زمانی که ماشین کار کرده است
- تعداد کار در حال اجرای ماشین
- تعداد کل کارهای انجام داده شده توسط ماشین
- انرژی مصرف شده توسط ماشین
- هر کار دارای مشخصات زیر می باشد:

- شماره کار
- مقدار منبع مورد نیاز
- مقدار منبع باقی مانده
- مهلت انجام کار
- زمان شروع به اجرا شدن کار
- زمانی که کار به اتمام می رسد

برنامه تا زمانی اجرا می شود که تمام کارها یا اجرا شوند یا مهلت اجرای آن ها به اتمام برسد. برای شبیه سازی از نرم افزار MATLAB استفاده شده است.

معیار توازن بار نشان دهنده این می باشد که درخواست های کاربران به چه صورت بین ماشین ها توزیع شده است، ایدئال این است که تمام ماشین ها به یک اندازه کار دریافت کرده باشند، توزیع عادلانه بار باعث کاهش مصرف انرژی نیز خواهد شد. برای ارزیابی این معیار دو سناریو در نظر گرفته شده است.

یکی از مهم‌ترین معیارهای ارزیابی الگوریتم‌های زمان‌بندی معیار makespan می‌باشد. این معیار با انرژی مصرفی ارتباط مستقیم دارد و طبق رابطه زیر محاسبه می‌شود.

$Makespan = \text{find max run time in VMS}$

در واقع makespan برابر است با بیشترین مدت زمانی که یک ماشین فعالیت داشته است،

در سناریو اول برای ارزیابی makespan ما تعداد ماشین‌ها را ثابت در نظر گرفته‌ایم و تعداد کارها در هر مرحله بیشتر شده است.

ماشین‌های مجازی در محاسبات ابری با توجه به مقداری کاری که انجام داده‌اند انرژی مصرف می‌کنند، مصرف انرژی در محاسبات ابری هزینه بالایی دارد و تا جای امکان باید کارها به نحوی توزیع شوند که مصرف انرژی بالا نرود، بنابراین مصرف انرژی که یکی از اهداف مهم ما در این پایان نامه است در این قسمت مورد ارزیابی قرار گرفته است، در این سناریو تعداد کارها متغیر در نظر گرفته شده است و تعداد ماشین‌ها ثابت است.

۷-۱- نتیجه گیری

نتایج شبیه سازی نشان می‌دهد که روش پیشنهادی به دلیل توزیع عادلانه بار عملکرد بهتری نسبت به روش قبلی دارد. در این مقاله، ما یک روش تعادل بار و زمان‌بندی برای محیط‌های محاسبات ابری مبتنی بر الگوریتم کلونی مورچه ارائه کرده‌ایم. این الگوریتم نه تنها تعادل بار را انجام می‌دهد، بلکه اولویت وظایف حذف شده از ماشین‌های مجازی سر بارگذاری شده را نیز زیر نظر می‌گیرد. وظایف حذف شده از این ماشین‌های مجازی به عنوان عادلانه می‌کنند که اطلاعات را به طور سراسری به روزرسانی می‌کنند. این الگوریتم اولویت‌های وظایف را نیز در نظر می‌گیرد. تعادل بار توان کلی پردازش را بهبود می‌بخشد و تعادل بار بر اساس اولویت، بر کاهش زمان انتظار وظیفه در صف ماشین مجازی تمرکز دارد؛ بنابراین، پاسخ زمان ماشین‌های مجازی را کاهش می‌دهد. ما الگوریتم پیشنهادی خود با دیگر تکنیک‌های موجود مقایسه کرده‌ایم. نتایج نشان می‌دهد که الگوریتم ما خوب و بدون افزایش هزینه‌های اضافی است. این روش تعادل بار برای سیستم‌های محاسبات ابری ناهمگن کار می‌کند و برای ایجاد تعادل در وظایف مستقل غیر پیشگیرانه است.

۸-۱- پیشنهادات

در آینده، ما برای گسترش این نوع تعادل بار و زمان‌بندی برای جریان کار با وظایف وابسته برنامه‌ریزی می‌کنیم. این الگوریتم اولویت را به عنوان پارامتر اصلی کیفیت سرویس در نظر می‌گیرد. در آینده، ما برای بهبود این الگوریتم با در نظر گرفتن سایر عوامل کیفیت سرویس برنامه‌ریزی می‌کنیم.

۹-۱- منابع

[1] Nascimento A, Olimpio V, Silva V, Paes A, de Oliveira D (2019) A Reinforcement Learning Scheduling Strategy for Parallel Cloud-Based Workflows In: 2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW),

- [2] Liu Z, Zhang H, Rao B, Wang L (2018) A Reinforcement Learning Based Resource Management Approach for Time-critical Workloads in Distributed Computing Environment In: 2018 IEEE International Conference on Big Data (Big Data)
- [3] Hao Y, et al. (2020) Deep Reinforcement Learning for Edge Service Placement in Softwarized Industrial Cyber-Physical System. IEEE Trans Ind Inf PP(99)
- [4] Zdraveski, V., Todorovski, M., Kocarev, L. "Dynamic intelligent load balancing in power distribution networks," International Journal of Electrical Power & Energy Systems, Vol. 73, pp. 157-162, 2015.
- [5] L.D, D.B., Venkata Krishna, P. "Honey bee behavior inspired load balancing of tasks in cloud computing environments," Applied Soft Computing, Vol. 13, No. 5, pp. 2292-2303, 2013.
- [6] De Falco, I., Laskowski, E., Olejnik, R., Scafuri, U., Tarantino, E., Tudruj, M. "Extremal Optimization applied to load balancing in execution of distributed programs," Applied Soft Computing, Vol. 30, pp. 501-513, 2015.
- [7] Milani, A.S., Navimipour, N.J. "Load balancing mechanisms and techniques in the cloud environments: Systematic literature review and future trends," Journal of Network and Computer Applications, Vol. 71, pp. 86-98, 2016.
- [8] Kalra, M., Singh, S. "A review of metaheuristic scheduling techniques in cloud computing," Egyptian Informatics Journal, Vol. 16, No. 3, pp. 275-295, 2015.
- [9] Madni, S.H.H., Latiff, M.S.A., Coulibaly, Y., Abdulhamid, S.i.M. "Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities," Journal of Network and Computer Applications, Vol. 68, pp. 173-200, 2016.
- [10] Chimakurthi, L. "Power efficient resource allocation for clouds using ant colony framework," arXiv preprint arXiv:pp. 1102-2608, 2011.

Topic: Improving resource allocation in mobile cloud computing using DRL-based task division

1- Amin Karimullah Dehkordi

(Master of Computer Engineering)

2- Dr. Golnaz Aghaei

(specialized doctorate in software systems and university assistant professor)

abstract

The main issue in this research is to improve resource allocation and load balancing management in mobile cloud computing environments, which face challenges such as high energy consumption, fluctuations in resource demand, and the need for load balancing between virtual machines. In this article, in order to solve these challenges, a new approach based on the ant colony optimization algorithm (ACO) is presented. Using the principles of collective intelligence and efficient search capability, the proposed algorithm seeks to improve resource allocation, reduce energy consumption, and increase load balancing in mobile cloud systems. To evaluate the performance of the proposed method, this algorithm is compared with several reference methods including conventional load balancing algorithms and common optimization algorithms such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). Evaluation criteria include response time, energy consumption, and load balancing in cloud systems. The results of the simulations showed that the proposed method provides a significant improvement in all criteria. In particular, the proposed algorithm was able to reduce energy consumption by 15%, improve system response time by 20%, and significantly improve load balancing compared to reference methods. Finally, by presenting an innovative approach based on ACO, this thesis proposes an effective solution to improve resource management and load balancing in mobile cloud computing environments, which can lead to increased efficiency and reduced operating costs in these systems.

Keywords: cloud computing, resource allocation, energy consumption reduction, ant colony algorithm, optimization.