

Reducing Energy and Carbon Cost of Cloud Data Centers through Optimal Virtual Machine Placement algorithms and an Automatic Servers Scale Controller Model

Ehsan Khodayarsesht¹, Alireza Shameli-Sendi^{2*}

1- Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran.

2*- Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran.

¹e.khodayarsesht@mail.sbu.ac.ir, ^{2*}a_shameli@sbu.ac.ir

Corresponding author's address: Alireza Shameli-Sendi, Faculty of Computer Science and Engineering, Shahid Beheshti University, Daneshjou Blvd, Velenjak, Tehran, Iran.

Abstract- Nowadays, one of the most significant challenges in cloud computing is the massive amount of costs that are being paid by cloud service providers for energy consumption and carbon tax. Accordingly, many efforts, such as increasing the utilization of servers, have been done by cloud providers to reduce these costs. However, many studies have focused only on the reduction of energy consumption on a single data center. Nevertheless, in recent years, there have been introduced lots of novel models that their final goal is to minimize the energy and carbon costs of a cloud provider through geographically distributed cloud data centers. However, one of the most significant defects visible in many models is about keeping up all data centers' servers in the ready state even when they are in the idle state. In this work, we intend to extend an optimal virtual placement machine (VM) placement algorithm via a useful and straightforward model based on an automated server scale controller mechanism and overall CPU utilization of data centers. Finally, we simulated our proposed model with the CloudSimPlus simulator. The results indicate that our model could significantly reduce the energy and carbon costs compared with the previous models.

Keywords- Cloud Computing, Energy Consumption, Carbon Footprint, VM Placement.

کاهش هزینه انرژی و کربن مراکز داده ابری با استفاده از الگوریتم‌های بهینه جانمایی ماشین‌های مجازی و یک مدل خودکار کنترل کننده مقیاس سرورها

احسان خدایارسرشت^۱، علیرضا شاملی سندی^{۲*}

۱- دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران.

۲- دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران.

^۱ e.khodayarsesht@mail.sbu.ac.ir, ^{۲*} a_shameli@sbu.ac.ir

* نشانی نویسنده مسئول: علیرضا شاملی سندی، تهران، ولنجک، بلوار دانشجو، دانشگاه شهید بهشتی، دانشکده مهندسی و علوم کامپیوتر.

چکیده- امروزه یکی از مهم‌ترین چالش‌ها در رایانش ابری، هزینه‌های سرسام‌آوری می‌باشد که توسط ارائه‌دهندگان خدمات ابری بابت مصرف انرژی و مالیات کربن پرداخت می‌شود. براین اساس، تلاش‌های فراوانی توسط ارائه‌دهندگان خدمات ابری برای کاهش این دسته از هزینه‌ها صورت گرفته است که از جمله آن‌ها می‌توان به افزایش بهره‌وری سرورهای مراکز داده اشاره کرد. گرچه اکثر پژوهش‌های انجام‌شده تنها بر روی یک مرکز داده واحد متمرکز هستند، اما در سال‌های گذشته، مدل‌های متعددی در خصوص به حداقل رساندن هزینه‌های انرژی مصرفی و انتشار کربن ارائه شده است که سعی در بکارگیری چندین مرکز داده توزیع شده متعلق به یک ارائه دهنده دارند. یکی از نواقص مهمی که در خیلی از روش‌ها مشاهده می‌گردد، نگهداری تمامی سرورهای مراکز داده در حالت آماده باش می‌باشد، حتی زمانی که بیکار هستند. ایده اصلی این مقاله جانمایی بهینه ماشین‌های مجازی در مراکز داده توزیع شده با استفاده از افزودن قابلیت افزایش و کاهش خودکار مقیاس سرورهای یک مرکز داده بر اساس بهره‌وری جاری پردازنده‌های آن مرکز می‌باشد. مدل ارائه شده با داده‌های واقعی در شبیه ساز CloudSimPlus پیاده سازی گردید و نتایج آن را با چند نمونه از الگوریتم‌های رقیب مورد مقایسه قرار دادیم. نتایج شبیه‌سازی نشان می‌دهد که مدل پیشنهادی در مقایسه با مدل‌های قبلی توانسته است به‌طور قابل توجهی هزینه‌های مربوط به انرژی و انتشار کربن را در مراکز داده کاهش دهد.

واژه‌های کلیدی: رایانش ابری، مصرف انرژی، انتشار کربن، جانمایی ماشین مجازی.

۱- مقدمه

زیادی میزبان^۱ یا سرور هستند. از این رو، مصرف زیاد انرژی یکی از بزرگترین مشکلاتی است که مراکز داده ابری با آن روبه‌رو هستند [۲، ۳]. علاوه بر این، مصرف زیاد انرژی توسط مراکز داده ابری باعث انتشار حجم زیادی کربن در جو کره زمین می‌شود، به‌طوری که براساس [۴] مشخص شده است که صنعت فناوری اطلاعات مسئول تولید ۲٪ از CO₂ جهانی است. همچنین در بسیاری از کشورهای جهان بابت انتشار کربن^۱ توسط مراکز داده ابری از ارائه‌دهندگان خدمات ابری نسبت به حجم کربن تولید شده، مالیات اخذ می‌شود [۵]. این امر بار سنگین دیگری را بر دوش ارائه

امروزه یکی از مهم‌ترین دستاوردهای صنعت فناوری اطلاعات و ارتباطات^۱، رایانش ابری^۲ است. ابر امکان دسترسی به انواع مختلف خدمات ابری را می‌دهد که بر اساس [۱] مهم‌ترین دسته‌بندی شامل زیرساخت به عنوان سرویس^۳، پلتفرم به عنوان سرویس^۴ و نرم افزار به عنوان سرویس^۵ است. علاوه بر این تمامی خدمات ارائه شده توسط ابر به صورت فراگیر^۶ بوده و در سرتاسر جهان قابل دسترس است. تمامی خدمات ارائه شده توسط ارائه‌دهندگان خدمات ابری^۷ در اصل توسط مراکز داده‌ای^۸ ارائه می‌شوند که هر یک شامل تعداد

مجازی در بین مراکز داده توزیع شده یک ارائه دهنده خدمات ابری

- روشن کردن درصدی از سرورهای موجود در هر مرکز داده با شروع کار آن‌ها و افزایش و کاهش خودکار تعداد سرورهای فعال هر مرکز داده در صورت نیاز
- بکارگیری بارکاری واقعی مراکز داده و ارزیابی همه جانبه مدل پیشنهادی در شبیه‌ساز کلودسیم

ادامه مقاله به این صورت خواهد بود: در بخش ۲ به بررسی کارهای مرتبط می‌پردازیم. در بخش ۳ در مورد مدل سیستم پیشنهادی، پارامترها و تابع هدف^{۱۹} مقاله بحث خواهیم کرد. در بخش ۴ الگوریتم‌های پیشنهادی را معرفی می‌کنیم. در بخش ۵ به شبیه‌سازی و تجزیه تحلیل نتایج بدست آمده خواهیم پرداخت. در نهایت نتیجه‌گیری مقاله و همچنین کارهای آینده در بخش ۶ ارائه شده است.

۲- کارهای مرتبط

تا به امروز مقالات علمی مختلفی در زمینه رایانش سبز^{۲۰} منتشر شده است که اغلب آن‌ها بر روی مدیریت انرژی درون یک مرکز داده تمرکز دارند و به تازگی کارهایی در حوزه کاهش مصرف انرژی به کمک فدرال‌سازی ابرها نیز انجام شده است. در [۱۰] هدف اصلی کاهش هزینه مصرف انرژی از طریق مهاجرت ماشین مجازی در یک محیط ابری فدرال شده می‌باشد. در این مقاله کاهش انتشار کربن و تأخیر در نظر گرفته نشده است. در یک کاری دیگر توسط نویسندگان همین مقاله سعی شده است تا با استفاده از مهاجرت ماشین مجازی در یک محیط فدرال، تنها میزان انتشار کربن توسط مراکز داده ابری را کاهش دهد [۱۱]. در کاری دیگر توسط [۱۲] یک معماری رزرو منابع مبتنی بر هزینه انرژی ماشین مجازی در سه وضعیت در حال اجرا، بیکار و در حال مهاجرت، ارائه شده است تا بتوان پیش بینی بهتری از درخواست‌های ورودی توسط کاربران داشت. همچنین در مقالاتی دیگر سعی شده است تا با استفاده از تکنیک اندازه‌گیری پویا^{۲۱} بتوان تعداد سرورهای مورد نیاز جهت سرویس دهی به بارهای کاری ورودی به مرکز داده را پیش بینی کرد و از آنجاکه سرورهای بیکار^{۲۲} بیش از نصف انرژی را در اوج مصرف سرورها استفاده می‌کنند [۱۳]، این تکنیک می‌تواند مصرف انرژی را به طور قابل توجهی کاهش دهد [۱۴، ۱۵]. بزرگ‌ترین معایب مقالات [۱۲، ۱۴، ۱۵] این است که تمامی آن‌ها تنها در حوزه یک مرکز داده متمرکز هستند و هدف اصلی آن‌ها فقط کاهش مصرف

دهندگان خدمات ابری قرار داده و هزینه‌های مصرف انرژی را دو چندان می‌کند. در نتیجه تحت نظر قرار دادن کلیه مسائل مربوط به مصرف انرژی مراکز داده ابری، نقش بسیار مهمی در کاهش هزینه‌ها و افزایش سود آوری، کاهش انتشار گازهای گل‌خانه‌ای و در نهایت کمک به ارتقاء سلامت محیط زیست دارد. به‌طور کلی پارامترهای تأثیر گذار زیادی در کنترل مصرف انرژی مراکز داده ابری دخیل هستند. همانطور که در [۶] اشاره شده است، ایده برتری رایانش ابری این است که یک سرویس ابری باید دارای قابلیت دسترسی مداوم^{۱۱}، قابلیت ارتجاعی^{۱۲} و مقیاس پذیر^{۱۳} باشد تا بتواند نیازهای مشتری را برآورده کند. این مسأله یکی از موارد تأثیر گذار در مصرف انرژی مراکز داده ابری است زیرا یک ارائه دهنده خدمات ابری باید همیشه از در دسترس بودن منابع کافی جهت پاسخ‌گویی مناسب به نیاز کاربران اطمینان حاصل کند. اما اینکه چه میزان منابع کافی می‌باشد، خود یک سوال بزرگ است زیرا الگوهای بار کاری^{۱۴} می‌توانند به‌طور غیرمنتظره‌ای تغییر کنند، در نتیجه ارائه دهندگان ابری نیاز به تدارک منابع مازاد^{۱۵} داشته تا در صورت افزایش ناگهانی بار کاری، بتوانند پاسخ‌دهی مناسب را انجام دهند [۷].

نگهداری همیشگی حجم زیادی از منابع ابر در حالت آماده باش جهت مقابله با افزایش ناگهانی بار کاری، منجر به افزایش مصرف انرژی و هزینه‌های عملیاتی خواهد شد. در این مقاله، قصد داریم تا با استخراج پارامترهای تأثیر گذار، روشی مبتنی بر فعال نگه‌داشتن حداقل سرورهای مورد نیاز ارائه دهیم که علاوه بر حفظ کیفیت سرویس^{۱۶} و جلوگیری از نقض قرارداد سطح خدمات^{۱۷}، در نهایت منجر به کاهش مصرف انرژی و انتشار کربن توسط مراکز داده ابری شود. یکی از پارامترهایی که ما قصد داریم جهت مدیریت میزان سرورهای فعال مورد نیاز مراکز داده ابری مورد استفاده قرار دهیم، میزان بهره‌وری مرکز داده ابری^{۱۸} است. موارد مختلفی را می‌توان به عنوان بهره‌وری یک مرکز داده ابری در نظر گرفت، مثل: میزان بهره‌وری مصرف برق مرکز داده و یا میزان بهره‌وری از لحاظ منابع فعال مرکز داده. اما چیزی که قرار است ما به‌طور خاص تر بر روی آن تمرکز کنیم، میزان بهره‌وری مرکز داده ابری از لحاظ بهره‌وری پردازنده‌های سرورهای فعال آن مرکز داده می‌باشد. بر این اساس، در این مقاله از مؤلفه میزان بهره‌وری مرکز داده ابری جهت به حداقل رساندن سرورهای فعال مورد نیاز استفاده شده است به گونه‌ای که نقض قرارداد سطح خدمات نیز در نظر گرفته شود. نوآوری‌های اصلی این مقاله به شرح ذیل می‌باشد:

- ارائه یک مدل تحلیلی جامع جهت تصمیم‌گیری در خصوص انتخاب بهترین مکان برای ایجاد یک ماشین

جدید که از رویکرد یادگیری ماشین چند عامله غیر متمرکز استفاده می‌کند، معرفی کردند. علاوه بر این در مقاله فوق از یک الگوریتم اکتشافی متمرکز برای مهاجرت ماشین‌های مجازی در جهت به حداقل رساندن مصرف انرژی در مراکز داده ابری استفاده شده است. در [۲۰] یک تکنیک ادغام ماشین مجازی مبتنی بر پیشبینی ارائه شده است. نویسندگان این مقاله از ترکیب مدل پیشبینی $GM(1,1)$ و زنجیره مارکوف^{۳۵} استفاده می‌کند تا بتواند میزان بهره‌وری پردازنده سرورهای کم بهره‌ور^{۳۶} و بیش از حد بهره‌ور^{۳۷} را در آینده نزدیک براساس تاریخچه داده‌های آن‌ها پیشبینی کند. در مقاله [۲۱] نیز هر دو مسئله کاهش مصرف انرژی و همچنین کاهش تعداد نقض قرارداد سطح خدمات در نظر گرفته می‌شود.

هدف نهایی این مقاله کاهش هزینه مصرف انرژی و مالیات کربن و همچنین کاهش انتشار کربن توسط مراکز داده ابری با استفاده از ایده تابع هدف و همچنین مدل افزایش و کاهش خودکار مقیاس سرور به صورت افقی در یک مرکز داده می‌باشد. مدل ارائه شده با داده‌های واقعی در شبیه ساز CloudSimPlus پیاده سازی گردید و همچنین نتایج آنرا با بهترین مدل‌های جاری مورد مقایسه قرار دادیم. نتایج شبیه‌سازی نشان می‌دهد که الگوریتم پیشنهادی توانسته به طور میانگین موجب کاهش حداکثری ۴۷٪ در مصرف انرژی برق شهری (سوخت‌های فسیلی) و همچنین کاهش حداکثری ۵۴٪ در انتشار کربن نسبت به الگوریتم‌های قبلی شود.

۳- مدل سیستم پیشنهادی

در این بخش قصد داریم در ابتدا به معرفی معماری سیستم مورد نظر و اجزای آن بپردازیم. سپس به معرفی پارامترهای تاثیر گذار در کنترل خودکار مقیاس سرورها و همچنین پارامترهای مهم در حوزه جانمایی ماشین مجازی با در نظر گرفتن مصرف انرژی و انتشار کربن خواهیم پرداخت. در نهایت به معرفی تابع هدف و محدودیت‌های^{۳۸} مدل سیستم پیشنهادی می‌پردازیم. لیست تمامی نمادهای بکار رفته در این مقاله در جدول ۱ نشان داده شده است

جدول ۱: توضیح نمادها

نماد	توضیح	نماد	توضیح
S_d^{set}	مجموعه‌ای از سرورهای موجود در مرکز داده d	DC^{set}	مجموعه‌ای از مراکز داده
$S_d^{inactive}$	مجموعه‌ای از سرورهای غیر فعال (خواب) در مرکز داده d	S_d^{active}	مجموعه‌ای از سرورهای فعال (روشن) در مرکز داده d
v^h	مدت زمان نگهداری ماشین مجازی v	VM	مجموعه‌ای از VM ها
DC_d^{cat}	تعداد دسته‌بندی‌های سرور در مرکز داده d	DC_{dt}^{temp}	دمای محیط خارجی مرکز داده d در زمان t

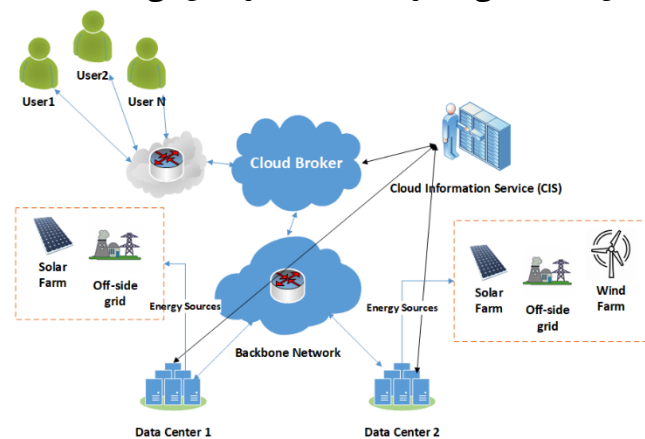
انرژی است که این امر لزوماً منجر به کاهش انتشارات کربن نخواهد شد. در [۱۶] یک رویکرد زمانبندی^{۳۳} جدید بنام PreAntPolicy ارائه شده است که شامل یک مدل پیش‌بینی مبتنی بر ریاضیات هندسه و زمانبندی بر اساس حالت توسعه یافته الگوریتم کلونی مورچه‌ها^{۳۴} است. در این روش مدل پیش بینی ارائه شده تعیین می‌کند که آیا می‌توان اجرای برنامه زمانبند را با توجه به پیش‌بینی روند بار کاری انجام داد، و همچنین برنامه زمان بند مسئولیت زمان بندی منابع در حالی که مصرف انرژی را به حداقل می‌رساند، با فرض در نظر گرفتن کیفیت سرویس^{۳۵}، بر عهده دارد. در [۱۷] یک سیستم زمان بندی پویا بلا درنگ جهت اجرای کارآمد برنامه‌های کاربردی مبتنی بر وظیفه^{۳۶} بر روی پلتفرم‌های محاسباتی توزیع شده^{۳۷} با هدف به حداقل رساندن مصرف انرژی، ارائه می‌شود. همانطور که می‌دانیم، زمان بندی وظایف بر روی چندین پردازنده یک مسأله NP-hard شناخته شده و پیدا کردن جواب کاملاً بهینه برای این دسته از مسائل غیر ممکن است. در مقاله فوق یک الگوریتم در زمان چند جمله‌ای^{۳۸} معرفی می‌شود که مجموعه‌ای از قوانین مکاشفه‌ای^{۳۹} و تکنیک‌های تخصیص منابع^{۴۰} را در جهت دستیابی به یک راه حل مناسب در یک بازه زمانی قابل قبول، با هم ترکیب می‌کند. الگوریتم پیشنهاد شده، در جهت به حداقل رساندن یک تابع چند-هدفه^{۴۱} تلاش می‌کند که در آن به ترکیب مصرف انرژی و مدت زمان اجرا بر اساس فاکتور کارایی انرژی که توسط سرویس دهنده و یا کاربر تعریف شده است بپردازد. در مقاله [۱۸] با در نظر گرفتن چندین مرکز داده توزیع شده یک ارائه دهنده خدمات ابری، یک الگوریتم جدید جانمایی ماشین مجازی در جهت افزایش پایداری زیست محیطی^{۴۲} و کاهش انتشار کربن ارائه شده است. در مقاله [۸] سعی شده است تا با پیدا کردن تأثیر گذارترین پارامترها بر روی هزینه انرژی و انتشار کربن، یک رویکرد جانمایی ماشین مجازی کارآمدتری را ارائه کند. علاوه بر این در مقاله [۸] از معیار PUE^{33} به شکل پویا استفاده شده است که از ترکیب پارامتر میزان بهره‌وری مصرف برق مرکز داده و دمای خارجی مرکز داده استفاده می‌شود. حقیقتاً و دوستان [۱۹] یک روش ادغام ماشین مجازی^{۴۴}

$DC_{d,thresh}^{util.cpu}$	میزان حد آستانه بهره‌وری پردازنده در مرکز داده d	$DC_{d,t}^{util.cpu}$	میزان بهره‌وری فعلی پردازنده‌های مرکز داده d در زمان t
$S_{s,t}^{util.cpu}$	میزان بهره‌وری پردازنده سرور s در زمان t	$DC_{d,t}^{util.p}$	میزان بهره‌وری مصرف برق تجهیزات IT مرکز داده d در زمان t
V_{ram}	میزان حافظه مورد نیاز ماشین مجازی V	V_{cpu}	تعداد پردازنده‌های مورد نیاز ماشین مجازی V
S_{ram}	ظرفیت حافظه سرور S	S_{cpu}	تعداد پردازنده‌های سرور S
S_{need}	تعداد سرورهای مورد نیاز جهت فعال‌سازی در یک مرکز داده	pct_s^{active}	درصد اولیه سرورهای فعال هر مرکز داده
$S_{need_{cat}}$	تعداد سرور مورد نیاز جهت فعال‌سازی از هر دسته‌بندی	$S_{deadline}$	حداقل زمان مجاز بیکار ماندن یک سرور
$P_{s, idle}$	میزان مصرف انرژی سرور در بهره‌وری حداقل (0%)	$P_{s, peak}$	میزان مصرف انرژی سرور در بهره‌وری حداکثر (100%)
$P_{s,t}^O$	میزان مصرف انرژی سرور S در زمان t	$P_{s,t}^U$	میزان مصرف انرژی سرور S در زمان t و بهره‌وری پردازنده U
R_E^O	نرخ انتشار کربن توسط انرژی off-side brown	$P_{s,t}^{U+}(v)$	میزان مصرف انرژی سرور S پس از اجرای ماشین مجازی v با بهره‌وری $U+$
C_E^O	قیمت انرژی off-side brown	T_E^O	میزان مالیات انتشار کربن توسط انرژی off-side brown
$C_S^F(v)$	هزینه انتشار کربن توسط سرور S جهت اجرای ماشین مجازی v	$C_S^E(v)$	هزینه انرژی سرور S جهت اجرای ماشین مجازی v
$E_S^{IT}(v)$	انرژی مورد نیاز تجهیزات IT سرور S جهت اجرای ماشین مجازی v	$C_S^T(v)$	هزینه کل (انرژی + کربن) سرور S جهت اجرای ماشین مجازی v
$E_S^T(v)$	کل انرژی مورد نیاز $(IT + Overhead)$ سرور S جهت اجرای ماشین مجازی v	$E_S^O(v)$	انرژی سرور مورد نیاز سرور S جهت اجرای ماشین مجازی v

ارائه دهنده ابری شامل تعدادی مرکز داده می‌باشد که تمامی آن‌ها توسط یک ستون فقرات^{۴۱} یکسان (فرض این مقاله) به یکدیگر متصل شده و از لحاظ جغرافیایی توزیع شده می‌باشد. همچنین هر مرکز داده جهت تأمین انرژی مورد نیاز سرورها، تجهیزات شبکه و سایر دستگاه‌های خود، دارای دو نوع منبع انرژی مختلف است: منبع انرژی خورشیدی^{۴۲} (به اختصار G) و منبع انرژی توری^{۴۳} (به اختصار O) که به عنوان سوخت فسیلی در نظر گرفته می‌شود. انرژی خورشید توسط پنل‌های خورشیدی^{۴۴} هر مرکز داده تولید شده و انرژی توری توسط شبکه برق شهری تأمین می‌شود. علاوه بر این، هر مرکز داده ابری شامل مجموعه‌ای از سرورها با تنظیمات مختلف است. مؤلفه مهم دیگر واسط ابر است و وظیفه اصلی آن دریافت درخواست ایجاد ماشین مجازی از کاربران و پیدا کردن میزبان^{۴۵} یا سرور مناسب در یکی از مراکز داده ابری جهت جانمایی است. به عبارت دیگر واسط ابر به عنوان رابط بین کاربران و مراکز داده ابری عمل می‌کند. CIS مولفهای است که شامل اطلاعاتی در مورد وضعیت فعلی تمامی مراکز داده یک ارائه دهنده ابری مانند منابع فیزیکی موجود و پارامترهای مرتبط با انرژی (مثل PUE مرکز داده، منابع انرژی، نرخ انتشار کربن، میزان بهره‌وری جاری سرورها و

۳-۱- معماری سیستم پیشنهادی

شکل ۱ نمای کلی معماری استفاده شده را نشان می‌دهد.



شکل ۱ نمای کلی معماری استفاده شده

همانطور که مشخص است این معماری از ۴ مؤلفه اصلی ارائه دهنده ابری، واسط ابر^{۴۶}، خدمات اطلاعات ابر و کاربران ابر^{۴۰} تشکیل شده است. لازم به ذکر است که در این مدل فرض شده است که هر درخواست ورودی از سمت کاربر، یک درخواست ماشین مجازی است که باید بر روی یکی از سرورها جانمایی شود. در این مدل هر

غیره)، P^{CS} انرژی مصرفی توسط تجهیزات سرمایه‌اشی (مانند چیلرها) و در نهایت P^{Net} که شامل انرژی مصرفی توسط تجهیزات شبکه‌ای (مانند روترها) می‌شود. در این مقاله جهت محاسبه PUE به صورت پویا از ایده [۸] استفاده کرده‌ایم. بطور کلی در محاسبه پویای PUE دو مورد دمای محیط خارجی مرکز داده ($DC_{d,t}^{temp}$) و میزان بهره‌وری مصرف برق مرکز داده ($DC_{d,t}^{util,p}$) تاثیر گذار هستند:

$$PUE(DC_{d,t}^{util,p}, DC_{d,t}^{temp}) = \frac{0.2 + 0.01DC_{d,t}^{util,p} + 0.01DC_{d,t}^{temp}}{DC_{d,t}^{util,p}} \quad (2)$$

۳-۲-۲- مدل مصرف سرور

نحوه مصرف انرژی توسط یک سرور به شکل زیر محاسبه می‌شود:

$$P_{s,t}^U = P_{s,Idle} + (P_{s,Peak} - P_{s,Idle}) * S_{s,t}^{util,cpu} \quad (3)$$

۳-۲-۳- منابع انرژی تجدید پذیر

ارائه دهندگان خدمات ابری بزرگ از انرژی تجدید پذیر جهت کاهش وابستگی خود از برق شبکه توری (برق شهری) که هم پرهزینه بوده و هم برای محیط زیست مضر است (چون ممکن است توسط سوخت‌های فسیلی تأمین شود) استفاده می‌کنند [۲۲، ۲۳]. انرژی خورشیدی به عنوان تنها منبع انرژی تجدیدپذیر برای هر مرکز داده به صورت مجزا در نظر گرفته شده و تنها تغییرات روز و شب برای این منبع انرژی لحاظ می‌شود.

علاوه بر این، بالاترین میزان اولویت منبع انرژی برای هر مرکز داده، منبع انرژی تجدید پذیر در نظر گرفته شده است. فرآیند کار بدین صورت می‌باشد که هر مرکز تا جایی که می‌تواند از انرژی تجدید پذیر استفاده خواهد کرد تا باعث به حداکثر رساندن میزان بهره‌وری مصرف انرژی تجدید پذیر در هر مرکز داده شود.

۳-۲-۴- قیمت انرژی

قیمت انرژی تجدید پذیر به علت ناچیز بودن آن در طول زمان، صفر در نظر گرفته می‌شود. در نتیجه در مقاله تنها قیمت انرژی برای منبع انرژی توری (C_E^0)، بر اساس شهرهایی که مراکز داده در آن مستقر هستند (cents/KWh)، در نظر گرفته شده است.

۳-۲-۵- نرخ انتشار کربن و مالیات کربن

بسته به نوع منبع انرژی، نرخ انتشار کربن به طور قابل توجهی می‌تواند تغییر کند. میزان انتشار کربن برای منبع انرژی تجدید پذیر صفر در نظر گرفته شده است. در نتیجه نرخ انتشار کربن به تفکیک

میزان مصرف انرژی) می‌باشد. همانطور که قبلاً اشاره شد، اطلاعات موجود در CIS توسط واسط ابر جهت جانمایی ماشین مجازی استفاده می‌شود.

مؤلفه آخر کاربران ابر هستند که درخواست‌های ایجاد ماشین مجازی خود را برای واسط ابر ارسال می‌کنند. در این مدل هر درخواست دارای دو ویژگی اصلی است. ویژگی اول نوع ماشین مجازی^{۴۶} درخواستی را نشان می‌دهد و ویژگی دوم مربوط به مدت زمان نگهداری^{۴۷} (۱/۱) ماشین مجازی می‌باشد. علاوه بر این، کاربران ابر نیاز به داشتن تضمینی در رابطه با QoS هستند که باید توسط ارائه دهنده ابر تأمین شود. در این مقاله QoS بر اساس تعداد درخواست‌های ماشین مجازی که به دلیل نبود منابع کافی رد می‌شوند، تعیین شده است. به این مفهوم که هرچه میزان رد درخواست‌ها پایین‌تر باشد، QoS بالاتر خواهد بود.

۳-۲- پارامترها

در این قسمت به معرفی پارامترهای تاثیر گذار بر روی مصرف انرژی، انتشار کربن و هزینه‌های مرتبط با آن خواهیم پرداخت. به‌طور کلی در پژوهش‌های این حوزه شش پارامتر راندمان انرژی مرکز داده^{۴۸}، مدل مصرف سرور^{۴۹}، منابع انرژی تجدید پذیر^{۵۰}، قیمت انرژی، نرخ انتشار کربن^{۵۱} و میزان مالیات انتشار کربن^{۵۲} به عنوان پارامترهای اصلی در نظر گرفته شده است. پارامترهای دیگری که ما در مقاله خود اضافه نموده ایم عبارت است از: (۱) درصد سرورهای فعال اولیه، (۲) میزان افزایش سرورهای فعال، (۳) حد آستانه بهره‌وری مرکز داده، (۴) حداکثر زمان مجاز بیکار ماندن یک سرور، (۵) میزان بهره‌وری پردازنده‌های مرکز داده و (۶) تعداد سرور مورد نیاز از هر دسته‌بندی جهت فعال سازی.

۳-۲-۱- راندمان انرژی مرکز داده

میزان اثربخشی انرژی یک مرکز داده به معیار PUE آن بستگی دارد. PUE معیاری جهت اندازه‌گیری انرژی سربار است. مانند انرژی که بابت سیستم سرمایه‌اشی، روشنایی و UPS^{۵۳} ها در جهت پشتیبانی از بار کاری ورودی به سیستم مصرف می‌شود. جهت محاسبه PUE تنها کفایت کل انرژی مصرفی توسط مرکز داده را بر انرژی مصرفی تجهیزات IT (سرورها) تقسیم کنیم. انرژی مصرفی کل شامل انرژی مصرفی سربار و انرژی مصرفی تجهیزات IT است:

$$PUE = \frac{P^{IT} + P^{EE} + P^{CS} + P^{Net}}{P^{IT}} \quad (1)$$

که عبارت است از انرژی مصرفی توسط تجهیزات IT (سرورها)، P^{EE} انرژی مصرفی توسط تجهیزات برقی (مانند باتری‌ها، UPS و

۳-۲-۱۰- میزان بهره‌وری پردازنده‌های مرکز داده

$$DC_{d,thresh}^{util.cpu} = \frac{\sum_{s=0}^{S_d^{active}} S_{s,t}^{util.cpu}}{|S_d^{active}|} \quad (7)$$

که $S_{s,t}^{util.cpu}$ نشان‌دهنده میزان بهره‌وری سرور s در زمان t و $|S_d^{active}|$ نشان‌دهنده تعداد سرورهای فعال می‌باشد.

۳-۲-۱۱- تعداد سرورهای مورد نیاز جهت فعال‌سازی از هر دسته‌بندی

زمانی که مرکز داده نیاز به افزایش تعداد سرورهای فعال خود دارد، تعداد سرورهای انتخاب شده جهت فعال‌سازی باید به گونه‌ای باشد که تا حد امکان انواع مختلف سرورهای مرکز داده (DC_d^{cat}) پوشش داده شوند:

$$S_{cat}^{need} = \frac{S^{need}}{|DC_d^{cat}|} \quad (8)$$

در بالا صورت کسر، S^{need} تعداد سرورهای مورد نیاز جهت فعال‌سازی را نشان می‌دهد (δ) تا بتوان میزان بهره‌وری کل مرکز داده را مجدداً به زیر حد آستانه بهره‌وری مرکز داده ($DC_{d,thresh}^{util.cpu}$) باز گرداند. همچنین تقسیم آن بر $|DC_d^{cat}|$ نشان‌دهنده تعداد سرورهای مورد نیاز از هر دسته‌بندی است.

۳-۳- تابع هدف

تابع هدف ارائه شده، با هدف کاهش هزینه اجرای بار کاری بر روی سیستم است که بر اساس کاهش مصرف انرژی و انتشار کربن عمل می‌کند. ما در این مقاله قصد داریم علاوه بر روش پیشنهادی در پژوهش‌های قبل، میزان حداقل انرژی مصرفی توسط سرورها را تا حد امکان کاهش دهیم. هزینه انرژی و کربن سرورهای مراکز داده ابری به صورت زیر محاسبه می‌شود.

$$C_T = C_E + C_F \quad (9)$$

$$C_E = \sum_{d \in DC^{set}} \sum_{s \in S_d^{set}} \sum_{v \in VM} C_s^E(v) x_{sv} \quad (10)$$

$$C_F = \sum_{d \in DC^{set}} \sum_{s \in S_d^{set}} \sum_{v \in VM} C_s^F(v) x_{sv} \quad (11)$$

C_E نشان‌دهنده هزینه انرژی مصرفی تمامی سرورها است و C_F نیز نشان‌دهنده هزینه انتشار کربن توسط کلیه سرورها است. همچنین $x_{s,v}$ نشان‌دهنده وجود یا عدم وجود یک ماشین مجازی بر روی یک سرور است. به عبارت دیگر، اگر $x_{s,v}$ برابر یک باشد، یعنی ماشین مجازی v بر روی سرور s قرار دارد و اگر صفر باشد یعنی ماشین مجازی v بر روی سرور s قرار ندارد. نماد ماشین مجازی نیز نشان‌دهنده مجموعه‌ای از $VM = \{v_1, v_2, \dots, v_n\}$ می‌باشد.

شهرهای مختلف تنها برای منبع انرژی توری (R_E^0) در نظر گرفته شده است (Tons/MWh). به همین ترتیب تنها برای منبع انرژی توری (T_E^0) مالیات کربن در نظر گرفته شده است (Dollar/Ton).

۳-۲-۶- درصد سرورهای فعال اولیه

درصد سرورهای فعال اولیه پارامتری است که نشان می‌دهد زمانی که یک مرکز داده ابری شروع به کار می‌کند، چه تعداد از سرورهای خود را باید در حالت آماده باش نگه دارد. حداقل مقدار این پارامتر باید بزرگتر از صفر و حداکثر آن برابر یک (کلیه سرورها) باشد:

$$pct_s^{active} = 0 < pct_s^{active} \leq 1, s \in S_d^{set} \quad (4)$$

۳-۲-۷- افزایش تعداد سرورهای فعال

از این پارامتر زمانی استفاده می‌شود که میزان بهره‌وری مرکز داده از مقدار تعیین شده در حد آستانه بیشتر شود. در این صورت مرکز داده باید به صورت خودکار به میزان فعلی سرورهای فعال خود، اضافه کند:

$$S^{need} = \left\lceil \frac{\sum_{i=0}^{S_d^{set}} S_{s,t}^{util.cpu}}{DC_{d,threshold}^{util.cpu}} - |S_d^{active}| \right\rceil \quad (5)$$

پارامتر S^{need} نشان‌دهنده تعداد سرورهای مورد نیاز جهت فعال‌سازی است تا بتوان بهره‌وری کلی مرکز داده از نظر پردازنده را کاهش داد.

۳-۲-۸- حد آستانه بهره‌وری مرکز داده

این پارامتر نشان‌دهنده حد آستانه بهره‌وری مرکز داده از لحاظ بهره‌وری پردازنده‌های کلیه سرورها است. زمانی که بهره‌وری مرکز داده از این حد آستانه بیشتر شود، مرکز داده موظف است در صورت امکان به تعداد سرورها فعال خود اضافه کند.

$$DC_{d,threshold}^{util.cpu} = 0 \leq DC_{d,threshold}^{util.cpu} \leq 1, d \in DC^{set} \quad (6)$$

۳-۲-۹- حداکثر مدت زمان مجاز بیکار ماندن یک سرور

زمانی که کار تمامی ماشین‌های موجود بر روی یک سرور به اتمام می‌رسد و دیگر هیچ ماشین مجازی بر روی آن قرار ندارد، سرور در وضعیت بیکار قرار خواهد گرفت.

در این پارامتر ($S_{idle}^{deadline}$) مشخص می‌شود که یک سرور چه مدت زمانی باید در حالت بیکار قرار داشته باشد تا بتوان پس از آن در صورت امکان در وضعیت خواب^{۵۴} قرار گیرد. مقدار این پارامتر بر اساس دقیقه بوده و می‌تواند بزرگتر مساوی صفر باشد.

۳-۴- محدودیت‌ها

تابع هدف که وظیفه به حداقل رساندن $C_T = C_s + C_v$ را دارد، دارای محدودیت‌های ذیل است:

کل منابع اختصاص داده شده به ماشین‌های مجازی موجود در یک سرور، نباید از حداکثر منابع موجود آن سرور بیشتر باشد:

$$\sum_{d \in DC^{set}} \sum_{s \in S_d^{active}} \sum_{v \in VM} v^{cpu} z_{sv} \leq S^{cpu}, \quad (12)$$

$$\sum_{d \in DC^{set}} \sum_{s \in S_d^{active}} \sum_{v \in VM} v^{ram} z_{sv} \leq S^{ram}$$

متغیر z_{sv} نشان می‌دهد که ماشین مجازی v بر روی سرور s قرار دارد یا خیر. در صورتی که z_{sv} برابر با یک باشد یعنی ماشین مجازی v بر روی سرور s قرار دارد و در صورتی که برابر صفر باشد یعنی ماشین مجازی v بر روی سرور s قرار ندارد.

مرکز داده تنها در صورتی می‌تواند شروع به روشن کردن سرورهای جدید کند که میزان بهره‌وری پردازنده‌های مرکز داده ($DC_{d,t}^{util.cpu}$) بزرگ‌تر از حد آستانه بهره‌وری مرکز داده ($DC_{d,thresh}^{util.cpu}$) شود:

$$DC_{d,t}^{util.cpu} > DC_{d,thresh}^{util.cpu} \quad (13)$$

همانطور که در (۵) نشان داده شد S^{need} تعداد سرورهای مورد نیاز جهت فعال‌سازی را نشان می‌دهد. S^{need} باید کوچکتر مساوی از تعداد سرورهای غیرفعال آن مرکز داده ($S_d^{inactive}$) باشد و در غیر این صورت تمامی سرورهای غیرفعال باید روشن شوند:

$$\begin{cases} activate S^{need} servers, & if S^{need} \leq |S_d^{inactive}| \\ activate |S_d^{inactive}| servers, & otherwise \end{cases} \quad (14)$$

در صورتی که نتیجه محاسبه تعداد سرورهای مورد نیاز از هر دسته‌بندی (S_{cat}^{need}) جهت فعال‌سازی کوچکتر از یک شود، باید به اندازه S^{need} تعدادی از سرورهای موجود در وضعیت خواب ($S_d^{inactive}$) را فعال کرد.

$$(15)$$

$$\begin{cases} Choose S_{cat}^{need} from each category S_d^{inactive}, & if S_{cat}^{need} \geq 1 \\ Activate S^{need} from S_d^{inactive}, & otherwise \end{cases}$$

سروری که به اندازه کافی ($S_{deadline}^{idle}$) در وضعیت بیکار قرار گرفته است، تنها در صورتی می‌تواند در وضعیت خواب قرار گیرد که میزان بهره‌وری پردازنده‌های مرکز داده ($DC_{d,t}^{util.cpu}$) بعد از به خواب رفتن آن سرور بیکار، بیشتر از حد آستانه بهره‌وری مرکز داده ($DC_{d,thresh}^{util.cpu}$) نشده و همچنین تعداد سرورهای فعالی که پس از به خواب رفتن سرور، بیکار باقی می‌مانند کمتر از درصد سرورهای فعال اولیه (pct_s^{active}) نشود:

$$\frac{1}{S_d^{active} - 1} * \sum_{s=0}^{S_d^{active} - 1} S_{s,t}^{util.cpu} \leq DC_{d,thresh}^{util.cpu}, \quad (16)$$

$$\frac{S_d^{active} - 1}{S_d^{set}} \geq pct_s^{active}$$

در آخر می‌توان مسئله بهینه‌سازی این مقاله را به صورت زیر خلاصه کرد:

$$\min C_T$$

$$s.t. (12) - (16)$$

۴- الگوریتم‌های پیشنهادی

در این بخش به معرفی الگوریتم افزایش و کاهش خودکار مقیاس میزبان و ترکیب آن با سه الگوریتم جانمایی پویا ماشین مجازی بنام CRA-DP و ERA-DP که توسط [۸] ارائه شده‌اند و همچنین الگوریتم رایج FFD [9] می‌پردازیم.

۴-۱- Cost and Renewable-Aware with Dynamic PUE (CRA-DP)

الگوریتم CRA-DP کامل‌ترین الگوریتم ارائه شده توسط [۸] است که هر دو مورد هزینه انرژی و انتشار کربن و همچنین مطلع بودن از منبع انرژی تجدید پذیر را به صورت همزمان در نظر می‌گیرد. در این الگوریتم زمانی که درخواست ایجاد یک ماشین مجازی از سمت کاربر برای broker ارسال می‌شود، در این مرحله چندین مرکز داده برای broker جهت جانمایی ماشین مجازی موجود است. الگوریتم CRA-DP ابتدا یک مرکز داده را انتخاب کرده و سپس یکی از سرورهای موجود در آن مرکز را انتخاب می‌کند. این الگوریتم با در نظر گرفتن منبع انرژی تجدید پذیر موجود در هر مرکز داده و PUE پویا آن مرکز، ابتدا مرکز داده‌ای را انتخاب می‌کند که کمترین میزان هزینه را برای ارائه دهنده ابری در پی دارد (به حداقل رساندن ΔC_T). سپس ΔC_T هر مرکز داده براساس هزینه انرژی و انتشار کربن افزوده شده به آن مرکز داده جهت اجرای ماشین مجازی فوق، به صورت صعودی مرتب می‌شود. این الگوریتم در قدم بعدی سعی در انتخاب سرور مناسب برای ماشین مجازی درخواست شده می‌کند. معیار انتخاب سرور در هر دو الگوریتم بر اساس کمترین میزان افزایش در مصرف انرژی سرور می‌باشد. به عبارت دیگر، پس از انتخاب مرکز داده مناسب جهت جانمایی ماشین مجازی، سروری از آن مرکز انتخاب می‌شود که دارای کمترین میزان افزایش در مصرف انرژی جهت اجرای ماشین مجازی فوق باشد. حال به مشکل اصلی این الگوریتم می‌پردازیم که در این مقاله سعی در بهبود عملکرد آن داریم. در تمامی الگوریتم‌های ارائه شده توسط [۸] فرض بر این است که از ابتدای شروع کار مراکز داده، تمامی سرورهای

الگوریتم از سه بخش تشکیل شده است. بخش اول تا سوم این الگوریتم در داخل هر مرکز داده شروع به کار کرده و روشن یا در وضعیت خواب نگه‌داشتن سرورها را کنترل می‌کند (الگوریتم ۱، ۲ و ۳). لازم به ذکر است که الگوریتم فوق با هر یک از ۳ الگوریتم CRA-DP، ERA-DP و FFD سازگار بوده و در بخش ۵ به هر یک از آن‌ها پیوست داده شده تا میزان افزایش بهبود در آن‌ها مورد بررسی قرار گیرد.

در الگوریتم ۱ و الگوریتم ۲ شرایط نیاز به سرورهای بیشتر در یک مرکز داده بررسی می‌شود. به این ترتیب اگر میزان بهره‌وری جاری مرکز داده ($DC_{d,t}^{util.cpu}$) بیشتر از حد آستانه بهره‌وری آن مرکز داده ($DC_{d,thresh}^{util.cpu}$) باشد و همچنین سروری در وضعیت خواب وجود داشته باشد (خط ۳ و ۴ الگوریتم ۱)، آنگاه الگوریتم ۲ جهت بررسی افزایش مقیاس سرورها فراخوانی شده و در صورت نیاز سرورهای جدید فعال می‌شوند.

الگوریتم ۱: Scale-up Controller

Input: DC^{set}
Output: S_d^{active}
1: while(true) do
2: foreach $d \in DC^{set}$
3: if $|S_d^{active}| \neq |S_d^{set}|$ **then**
4: if $DC_{d,t}^{util.cpu} > DC_{d,thresh}^{util.cpu}$ **then**
5: return $hostScaleOut(S_d^{set})$;
6: return S_d^{active} ;

در صورتی که الگوریتم ۲ فراخوانی شود ابتدا بر اساس S_{cat}^{need} محدودیت (۱۵) بررسی می‌شود (خط ۲). در صورت فراهم بودن شرایط، الگوریتم ۲ تلاش می‌کند تا از هر دسته‌بندی از سرورهای موجود در مرکز داده به اندازه S_{cat}^{need} سرور فعال کند (خطوط ۳ تا ۲۶). همچنین اگر محدودیت (۱۵) برقرار نشود، الگوریتم ۲ تلاش می‌کند به اندازه S_{cat}^{need} از سرورهای موجود در لیست $S_d^{inactive}$ سرورهای خواب را فعال نماید (خطوط ۲۸ تا ۳۴).

الگوریتم ۲: Host Scale-out

Input: S_d^{set}
Output: S_d^{active}
1: totalActivatedHosts = 0;
2: if $S_{cat}^{need} \geq 1$ **then**
3: loop1:
4: foreach $hostCategory \in DC_{cat}^{set}$

آن‌ها روشن بوده و در حالت آماده باش است. اما همانطور که قبلاً اشاره شد، هر سرور با حداقل میزان بهره‌وری (بهره‌وری صفر یا $P_{s,Idle}$) دارای یک حداقل میزان مصرف انرژی در واحد زمان است. در نتیجه این امکان وجود دارد که هر مرکز داده دارای تعداد زیادی سرور فعال باشد که بیکار بوده و هزینه انرژی و کربن زیادی را به آن مرکز داده تحمیل می‌کند. در این مقاله با معرفی الگوریتم Auto Scale-in and Scale-out Algorithm (AS) در بخش ۴-۴ ارائه شده است.

۲-۴ Energy and Renewable-Aware with Dynamic PUE (ERA-DP)

عملکرد این الگوریتم مانند الگوریتم CRA-DP-AS است با این تفاوت که تنها میزان انرژی مصرفی و منبع انرژی تجدید پذیر را جهت انتخاب مرکز داده و در نهایت جانمایی ماشین مجازی در نظر می‌گیرد.

۳-۴ First Fit Decreasing (FFD)

الگوریتم FFD یک الگوریتم حریصانه مکاشفه‌ای بسیار سریع و پرکاربرد است [۹]. نحوه عملکرد FFD به این صورت است که ابتدا لیست ماشین‌های مجازی درخواست شده را براساس منابع درخواستی آن‌ها (مانند پردازنده و یا حافظه) به صورت نزولی مرتب می‌کند. سپس براساس لیست سرورهای فعال مرکز داده، هر یک از ماشین‌های مجازی را جهت جانمایی در سرورهای موجود در لیست مورد بررسی قرار می‌دهد و به محض اینکه اولین سرور مناسب برای ماشین مجازی درخواستی یافت شود، آن سرور برای ماشین مجازی فعلی انتخاب شده و الگوریتم به سراغ ماشین مجازی بعدی موجود در لیست انتظار می‌رود. در این مقاله، الگوریتم FFD را جهت مقایسه با عملکرد الگوریتم‌های CRA-DP و ERA-DP انتخاب کردیم. لازم به ذکر است که به دلیل وجود چندین مرکز داده در مقاله فوق، نحوه عملکرد الگوریتم FFD به این صورت تعریف شده که لیست ماشین‌های مجازی درخواست شده به ترتیب تمامی مراکز داده ابری موجود در لیست D^{set} را بررسی می‌کند و به محض قابل اجرا بودن در هر یک از مراکز داده ابری، فرایند به اتمام می‌رسد و سایر مراکز داده بررسی نمی‌شوند. همچنین در صورتی که لیستی از ماشین‌های مجازی درخواست شده در یک مرکز داده قابل اجرا نباشند، مرکز داده بعدی موجود در لیست D^{set} انتخاب می‌شود.

۴-۴ الگوریتم Auto Scale-in and Scale-Out (AS)

وظیفه اصلی این الگوریتم به حداقل رساندن تعداد سرورهای فعال در هر مرکز داده با رعایت قرارداد سطح خدمات (SLA) است. این

```

3:  foreach host ∈ Sdactive
4:    if host.idleTime ≥ Sdeadlineidle then
5:      DCd,t+util.cpu =  $\frac{1}{S_d^{active} - 1} * \sum_{s=0}^{S_d^{active} - 1} S_{s,t}^{util.cpu}$ ;
6:      Sd,t+active =  $\frac{S_d^{active} - 1}{S_d^{set}}$ ;
7:      if DCd,t+util.cpu > DCd,threshutil.cpu and Sd,t+active ≥ pctsactive
          then
8:        host.setActive(false);
9:        Sdinactive.add(host);
10:       Sdactive.remove(host);
11:  return Sdactive

```

پیچیدگی زمانی الگوریتم ۱ و ۲ برابر است با $O(DC^{set} \times S_d^{inactive})$ با این فرض که تنها لیست سرورهای خواب ($S_d^{inactive}$) هر مرکز داده بررسی می‌شوند. همچنین پیچیدگی زمانی الگوریتم ۳ عبارت است از $O(DC^{set} \times S_d^{active})$.

۵- ارزیابی کارایی

در این قسمت به تجزیه و تحلیل نتایج بدست آمده از روش پیشنهادی پرداخته و تاثیر پارامترهای مختلف را بر روی مصرف انرژی و انتشار کربن بررسی می‌کنیم. در این مقاله الگوریتم AS معرفی شد که هدف اصلی آن به حداقل رساندن تعداد سرورهای فعال هر مرکز داده و در نهایت کاهش هزینه انرژی مصرفی و مالیات کربن توسط ارائه دهندگان خدمات ابری است. از این سو جهت افزایش کارایی و بهبود عملکرد الگوریتم‌های جانمایی ماشین‌های مجازی، الگوریتم پیشنهادی AS را با ۳ الگوریتم مکاشفه‌ای حریصانه جانمایی ماشین‌های مجازی ترکیب کردیم. این ۳ الگوریتم عبارت است از CRA-DP، ERA-DP و FFD. الگوریتم [۹] CRA-DP جهت جانمایی ماشین مجازی، علاوه بر در نظر گرفتن میزان موجود انرژی تجدید پذیر هر مرکز داده، دو مورد هزینه انرژی مصرفی و مالیات انتشار کربن توسط ماشین مجازی درخواست را در نظر می‌گیرد.

الگوریتم [۹] ERA-DP نیز مانند الگوریتم CRA-DP بوده با این تفاوت که تنها میزان مصرف انرژی ماشین مجازی درخواستی و میزان انرژی تجدید پذیر در مرکز داده را در نظر می‌گیرد. به عبارت دیگر، ERA-DP مانند CRA-DP عمل کرده با این تفاوت که هزینه انرژی مصرفی توسط ماشین مجازی درخواستی را در نظر نمی‌گیرد. الگوریتم [۹] FFD نیز یک الگوریتم مکاشفه‌ای حریصانه است که

```

5:  categoryActivatedHosts = 0;
6:  loop2:
7:  foreach host ∈ hostCategory
8:    if host.isActive() then
9:      host.setActive(true);
10:     Sdactive.add(host);
11:     Sdinactive.remove(host);
12:     totalActivatedHosts ++;
13:     categoryActivatedHosts ++;
14:     if categoryActivatedHosts == Scatneed then
15:       categoryActivatedHosts = 0;
16:       continue loop1;
17:     if totalActivatedHosts == Sneed then
18:       break loop1;
19:   if totalActivatedHosts < Sneed then
20:     foreach host ∈ Sdinactive
21:       host.setActive(true);
22:       Sdactive.add(host);
23:       Sdinactive.remove(host);
24:       totalActivatedHosts ++;
25:       if totalActivatedHosts == Sneed then
26:         break loop1;
27: else
28:   foreach host ∈ Sdinactive
29:     host.setActive(true);
30:     Sdactive.add(host);
31:     Sdinactive.remove(host);
32:     totalActivatedHosts ++;
33:     if totalActivatedHosts == Sneed then
34:       break;
35: return Sdactive;

```

در الگوریتم ۳ شرایط به خواب بردن سرورهای بیکار بررسی می‌شود. بر این اساس در صورتی که یک سرور به اندازه کافی ($S_{deadline}^{idle}$) در وضعیت بیکاری قرار داشته باشد (خط ۴)، ابتدا محدودیت (۱۶) بررسی می‌شود (خط ۷). سپس در صورت فراهم بودن شرایط، سرور بیکار از حالت روشن به حالت خواب تغییر وضعیت می‌دهد (خطوط ۸ تا ۱۰).

الگوریتم ۳: Scale-in Controller

```

Input: DCset
Output: Sdactive
1: while(true) do
2:   foreach d ∈ DCset

```

جدول ۲ پیکربندی سرورهای مرکز داده

مدل مصرف برق	پهنای باند (Gbps)	حافظه اصلی (GB)	سرعت هسته (MIPS ⁵⁸)	تعداد پردازنده	نوع سرور
نوع ۲	100	2	2500	1	Medium
نوع ۲		4		2	Large
نوع ۳		8		4	xLarge
نوع ۳		16	2700	8	2xLarge
نوع ۱		32		16	4xLarge

۵-۱-۲- ماشین‌های مجازی

در این مقاله از ماشین‌های مجازی که براساس نمونه‌های EC2^{۵۹} شرکت آمازون می‌باشند، استفاده کرده ایم. ماشین‌های مجازی موجود در این شبیه‌سازی به‌طور کلی به پنج دسته مختلف تقسیم می‌شوند. هر نمونه دارای سه ویژگی تعداد پردازنده، سرعت هسته (MIPS) و میزان حافظه اصلی (GB) مورد نیاز است. جدول ۳ نمونه‌های ماشین‌های مجازی را نشان می‌دهد.

جدول ۳ نمونه‌های ماشین‌های مجازی

نوع ماشین مجازی	تعداد پردازنده	سرعت هسته (MIPS)	حافظه اصلی (GB)
Medium	1	2500	2
Large	2		3.75
xLarge	4		7.5
2xLarge	8	2700	15
4xLarge	16		30

۵-۱-۳- مصرف برق سرورها

مصرف برق سرور با میزان بهره‌وری آن رابطه خطی دارد [۲۶]. مدل‌های برق مصرفی تعیین شده در این مقاله براساس بهره‌وری پردازنده سرور در جدول ۴ مدل مصرف برق سرورها نشان داده شده است. لازم به ذکر است که تمامی سرورها پس از تغییر وضعیت از حالت روشن به حالت خواب، نوع مدل برق مصرفی خود را به حالت Sleep Mode (نوع ۴) تغییر می‌دهند.

جدول ۴ مدل مصرف برق سرورها

نوع	نام مدل	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1	HP ProLiant G3	105	112	118	125	131	137	147	153	157	164	169
2	HP ProLiant G4	86	89.4	92.6	96	99.5	102	106	108	112	114	117
3	HP ProLiant G5	93.7	97	101	105	110	116	121	125	129	133	135
4	Sleep Mode	10										

در ابتدا لیست ماشین‌های مجازی درخواست شده را به‌صورت نزولی مرتب کرده و سپس سعی می‌کند هر یک از ماشین‌های مجازی را در اولین سروری که منابع کافی دارد ایجاد کند. در ادامه در بخش ۲-۵ به تحلیل نتایج شبیه‌سازی بدست آمده از هر یک از ۳ الگوریتم جانمایی ماشین‌های مجازی و ترکیب آن‌ها با الگوریتم AS خواهیم پرداخت.

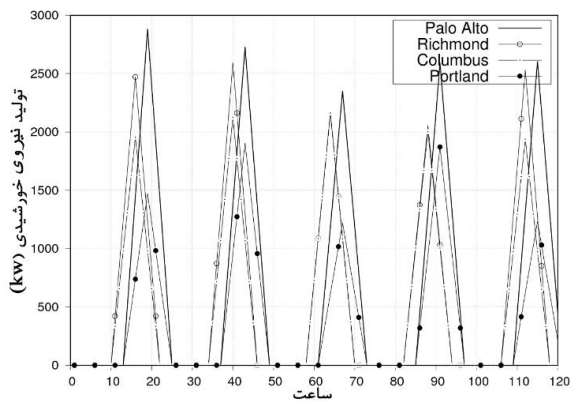
۵-۱-۱- تنظیمات مدل شبیه‌ساز

شبیه‌سازی در این مقاله، توسط ابزار شبیه‌سازی CloudSimPlus [۲۴] انجام شده است. همچنین تنظیمات مربوط به محیط ابر را براساس محیط رایانش ابری IaaS (زیرساخت به عنوان سرویس) انجام داده‌ایم. از آنجا که شبیه‌سازی CloudSimPlus بطور پیش‌فرض قابلیت پشتیبانی از شبیه‌سازی انرژی و انتشار کربن را ندارد، این قابلیت به این شبیه‌ساز اضافه شده است. علاوه بر این، قابلیت‌های دیگری اعم از هزینه انرژی مصرفی و کربن منتشر شده، دسترسی به منبع انرژی تجدید پذیر، مصرف انرژی سرور و محاسبه پویا PUE را به هسته اصلی ابزار شبیه‌ساز اضافه کردیم.

۵-۱-۱-۱- تنظیم مراکز داده

در این شبیه‌سازی چهار مرکز داده را بر اساس [۲۵] در چهار شهر مختلف ایالات متحده آمریکا که هر یک دارای منطقه زمانی^{۶۰} مختلفی هستند در نظر گرفته‌ایم. این چهار مرکز داده به ترتیب عبارت است از مرکز داده شهر Palo Alto در استان California، شهر Richmond در استان Virginia، شهر Columbus در استان Ohio و شهر Portland در استان Oregon.

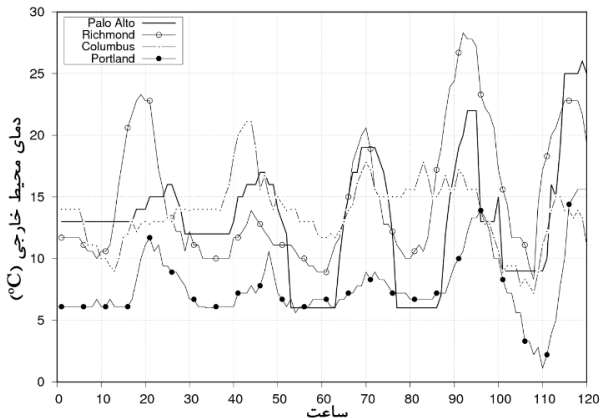
علاوه بر این، هر مرکز داده شامل ۱۵۰ عدد سرور فیزیکی ناهمگون با پنج پیکربندی مختلف است. جدول ۲ پیکربندی سرورهای مرکز داده را نشان می‌دهد. همچنین فرض شده است که تمامی سرورها به NAS^{۵۷} متصل بوده و در نتیجه از نظر حافظه ثانویه هیچگونه محدودیتی ندارند.



شکل ۲ تولید نیروی خورشیدی به مدت ۵ روز

۵-۱-۷- دمای خارجی

جهت تعیین میزان دما خارجی هر مرکز داده از اطلاعات ارائه شده توسط [۳۱] استفاده کردیم. شکل ۳ دمای خارجی هر مرکز داده را از تاریخ ۱۳ مارچ تا ۱۷ مارچ ۲۰۱۶ را به مدت ۵ روز متوالی نشان می‌دهد.



شکل ۳ دمای محیط خارجی مراکز داده به مدت ۵ روز

۵-۱-۸- تولید بارکاری

جهت تولید بارکاری (درخواست های ماشین مجازی) که به صورت پویا در طول فرایند شبیه‌سازی برای broker ارسال می‌شود، از دیتاست واقعی MetaCentrum 2 که توسط [۳۲] ارائه شده است استفاده می‌کنیم. همچنین فرض می‌کنیم هر یک از درخواست‌های ارسالی، درخواست ایجاد یک ماشین مجازی مستقل است که قرار است با بهره‌وری ۱۰۰٪ (از نظر پردازنده و حافظه اصلی) کار کند. علاوه بر این، فرض می‌کنیم که هر یک از درخواست دارای طول $MI^{62} 50,000,000$ است. به عنوان مثال فرض کنیم که یک بار کاری با طول ذکر شده بر روی یک ماشین مجازی با ۲ پردازنده که هر یک از آن‌ها دارای سرعت پردازش MIPS ۲۵۰۰ هستند شروع به اجرا می‌کند. اگر بار کاری فوق از بهره‌وری ۱۰۰٪ پردازنده‌های ماشین مجازی خود استفاده کند،

۵-۱-۴- نرخ انتشار کربن و مالیات کربن

نرخ انتشار کربن ($Tons/MWh$) و میزان مالیات آن (Dollars/Ton) به ترتیب بر اساس [۲۷, ۲۸] انتخاب شده است (جدول ۵).

۵-۱-۵- قیمت انرژی

هزینه انرژی را در دو بازه زمانی اوج مصرف^۶ (بین ۸ صبح تا ۱۰ شب) و خارج از اوج^۷ (بین ۱۰ شب تا ۱۰ صبح) برای هر یک از مراکز داده ابری، بر اساس [۲۹] به صورت جداگانه در نظر می‌گیریم. هزینه انرژی هر مرکز داده در بازه زمانی خارج از اوج به صورت نیم بها محاسبه می‌شود (جدول ۵).

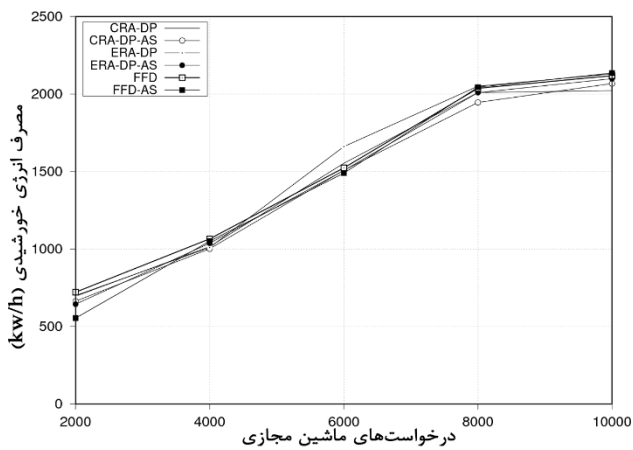
جدول ۵ ویژگی‌های مراکز داده

ویژگی‌های مراکز داده	Palo Alto	Richmond	Columbus	Portland
نرخ انتشار کربن ($Tons/MWh$)	0.2060	0.3715	0.6692	0.1393
مالیات کربن (Dollars/Ton)	15.3	10.77	18	48
قیمت انرژی (cents/KWh)	12.37	6.59	5.45	5.87
منطقه زمانی	-7 GMT	-4 GMT	-4 GMT	-7 GMT

۵-۱-۶- انرژی خورشیدی

جهت تولید انرژی خورشید از دیتاست ارائه شده توسط [۳۰] استفاده کرده ایم. در این مقاله مدت ۵ روز متوالی را برای شبیه‌سازی در نظر گرفتیم. علاوه بر این، کل مساحت مورد نیاز برای صفحات مسطح جاذب پروتوهای خورشیدی را ۵۰۰ متر مربع (m^2) در نظر گرفتیم.

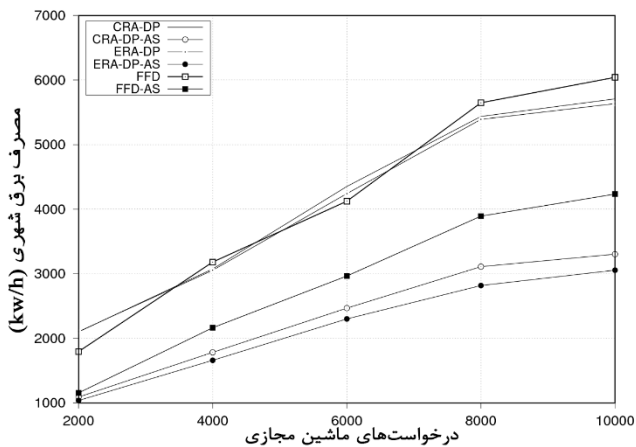
همچنین بازه زمانی تولید انرژی خورشیدی از ساعت ۶ صبح هر روز شروع شده و تا ساعت ۱۲ ظهر به اوج خود می‌رسد، سپس از ساعت ۱۲ تا ساعت ۶ بعد از ظهر میزان تولید انرژی خورشیدی شروع به کمتر شدن می‌کند تا اینکه از ساعت ۶ بعد از ظهر تا ساعت ۶ صبح روز بعد میزان تولید انرژی خورشیدی صفر خواهد بود. شکل ۲ تولید انرژی خورشیدی را به مدت ۵ روز متوالی برای ۴ مرکز داده مختلف نشان می‌دهد.



شکل ۴ نمودار مصرف انرژی خورشیدی

۵-۲-۲- مصرف برق شهری

از این بخش به بعد مقایسه روش پیشنهادی این مقاله (الگوریتم AS) و ترکیب آن با سه الگوریتم جانمایی CRA-DP، ERA-DP و FFD به‌طور چشم‌گیری نمایان می‌شود. همانطور که نمودار شکل ۵ نشان می‌دهد، میزان مصرف برق شهری که سوخت فسیلی محسوب می‌شود، در هر سه الگوریتم CRA-DP-AS، ERA-DP-AS و FFD-AS در مقایسه با CRA-DP، ERA-DP و FFD به‌طور قابل توجهی کاهش یافته است.



شکل ۵ نمودار مصرف برق شهری

این میزان کاهش به‌طور میانگین در هر سه الگوریتم ERA-DP-AS، CRA-DP-AS و FFD-AS نسبت به الگوریتم هم‌نوع خود به ترتیب برابر است با ۴۷٪، ۴۴٪، ۳۱٪. بنابراین همانطور که از نمودار مشخص است الگوریتم ERA-DP-AS دارای کمترین میزان مصرف برق شهری از ابتدای شروع فرایند شبیه‌سازی بوده است. از سوی دیگر هر سه الگوریتم CRA-DP، ERA-DP و FFD به ترتیب دارای بیشترین مصرف برق شهری بوده و تقریباً دارای راندمان مصرف یکسانی هستند. به‌طور دقیق‌تر می‌توان گفت که الگوریتم CRA-DP به‌طور میانگین ۱٪ از الگوریتم

بار کاری قادر به اتمام کار خود در مدت زمان ۲۰.۰۰۰ ثانیه است. در این مقاله در مجموع ۱۰.۰۰۰ بار کاری در طول فرایند شبیه‌سازی برای broker ارسال می‌شود. لازم به ذکر است که در دیتاست اشاره شده به هر بار کاری براساس میزان تعداد پردازنده درخواستی آن، یکی از ماشین‌های مجازی اشاره شده در بخش ۲-۱-۵ تخصیص داده می‌شود و در صورتی که تعداد پردازنده درخواستی بیشتر از تعداد پردازنده‌های بزرگتری ماشین مجازی موجود در شبیه‌سازی باشد، بزرگترین ماشین مجازی موجود در شبیه‌سازی به آن بارکاری اختصاص داده می‌شود.

۵-۱-۹- مقدار دهی پارامترها

مقادیر در نظر گرفته شده برای هر یک از سه پارامتر ثابت الگوریتم AS در جدول ۶ نشان داده شده است.

جدول ۶ مقدار پارامترهای ثابت الگوریتم AS

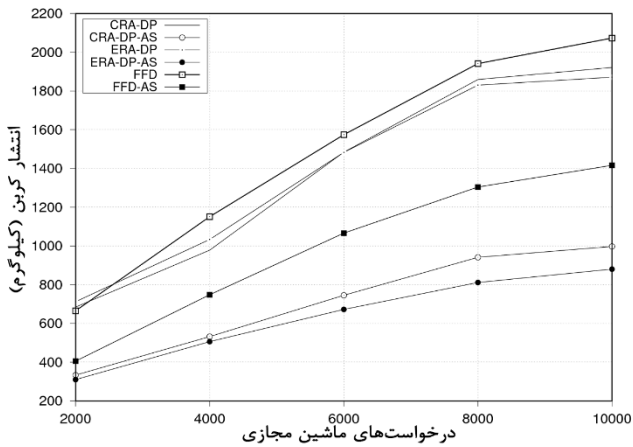
pct_s^{active}	$DC_{util,cpu,d,thresh}$	$S_{idle,deadline}$
40%	60%	15 Min

۵-۲- نتایج شبیه‌سازی

در این مقاله شبیه‌سازی را به مدت ۵ روز متوالی اجرا کردیم. همچنین برای اطمینان از نتایج حاصله، اجرای هر یک از الگوریتم‌ها را به مدت ۵ بار تکرار کرده و میانگین نتایج بدست آمده را گزارش نموده ایم. ما در شبیه‌سازی خود نتایج سه الگوریتم CRA-DP، ERA-DP، FFD و CRA-DP-AS، ERA-DP-AS، DP-AS، FFD-AS و ترکیبی از الگوریتم‌های جانمایی ماشین‌های مجازی و الگوریتم AS ارائه شده در این مقاله هستند، مورد مقایسه قرار می‌دهیم.

۵-۲-۱- مصرف انرژی خورشیدی

در این بخش به مقایسه میزان انرژی خورشیدی مصرف شده توسط هر یک از شش الگوریتم اشاره شده می‌پردازیم. شکل ۴ نشان می‌دهد که هر شش الگوریتم در زمینه مصرف انرژی خورشیدی، عملکرد نسبتاً یکسانی دارند. به‌طور دقیق‌تر می‌توان گفت که الگوریتم ERA-DP که تنها میزان مصرف انرژی و همچنین میزان موجودیت انرژی تجدید پذیر را در هنگام جانمایی ماشین‌های مجازی در نظر می‌گیرد، در هنگام ارسال درخواست‌های ایجاد ماشین مجازی (از ۲۰۰۰ بارکاری تا ۱۰.۰۰۰ بارکاری) بطور میانگین ۳٪ بیشتر از سایر الگوریتم‌ها (پنج الگوریتم دیگر) انرژی خورشیدی استفاده کرده است.

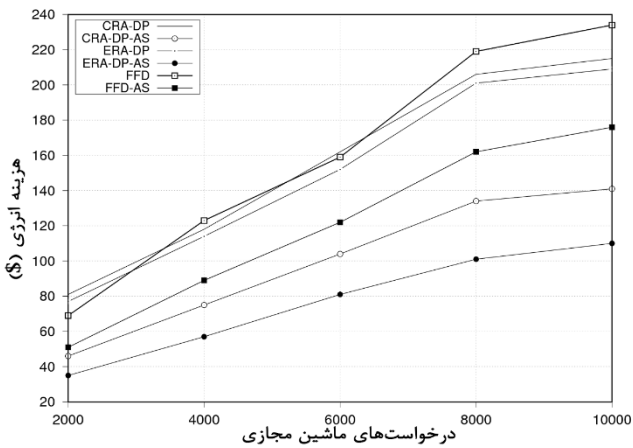


شکل ۷ نمودار میزان انتشار کربن

میانگین ۶٪ از CRA-DP و ۵٪ از ERA-DP کربن بیشتری منتشر کرده است.

۵-۲-۵- هزینه مصرف انرژی

بررسی نتایج مربوط به هزینه انرژی مصرفی در شکل ۸ نشان می‌دهد که الگوریتم‌های FFD، ERA-DP و CRA-DP هر یک به ترتیب شامل بیشترین میزان میانگین هزینه انرژی مصرفی نسبت به سایر الگوریتم‌ها هستند. از سوی دیگر الگوریتم‌های ERA-DP-AS، CRA-DP-AS و FFD-AS به‌طور میانگین میزان هزینه انرژی مصرفی را هر یک به ترتیب به میزان ۵۰٪، ۳۷٪ و ۲۵٪ نسبت به هم نوع خود کاهش داده‌اند.



شکل ۸ نمودار هزینه انرژی مصرفی

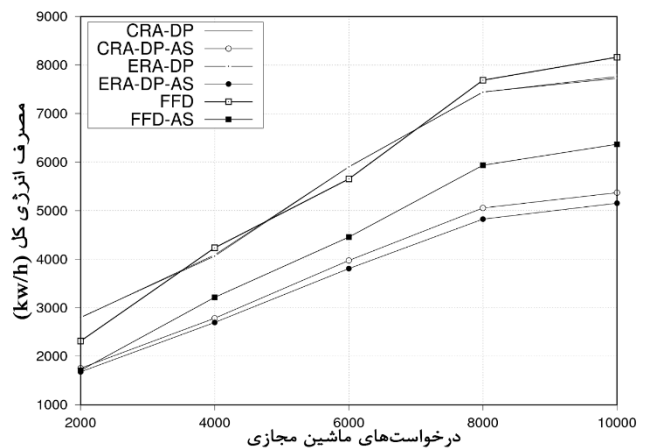
۵-۲-۶- هزینه انتشار کربن

تحلیل نتایج بدست آمده در مورد مالیات انتشار کربن در شکل ۹، بیانگر این است که در هر یک از الگوریتم‌های CRA-DP-AS، ERA-DP-AS و FFD-AS به‌طور میانگین شاهد کاهش ۵۱٪، ۴۷٪ و ۳۹٪ نسبت به الگوریتم‌های هم نوع هستیم. همچنین ۳ الگوریتم ERA-DP، CRA-DP و FFD به ترتیب دارای بیشترین میزان مالیات کربن هستند.

ERA-DP و ۲٪ از الگوریتم FFD برق شهری بیشتری مصرف می‌کند.

۵-۲-۳- مصرف انرژی کل

مصرف انرژی کل مراکز داده ابری (انرژی خورشیدی + برق شهری) در هر یک از شش الگوریتم مورد نظر در شکل ۶ نشان داده شده است. همانطور که مشخص هست نمودار مصرف انرژی کل دارای الگوی مشابهی با نمودار مصرف برق شهری در شکل ۵ است. براین اساس، الگوریتم‌های ERA-DP-AS، CRA-DP-AS و FFD-AS دارای کمترین میزان مصرف انرژی بوده و هر یک به ترتیب و به‌طور میانگین ۳۶٪، ۳۳٪ و ۲۳٪ انرژی کمتری نسبت به الگوریتم هم نوع خود مصرف می‌کنند. در اینجا نیز الگوریتم ERA-DP-AS دارای بهترین عملکرد است. همچنین هر سه الگوریتم CRA-DP، ERA-DP و FFD دارای بیشترین میزان مصرف انرژی بوده و هر دو الگوریتم CRA-DP و ERA-DP نسبت به الگوریتم FFD به‌طور میانگین ۳٪ انرژی بیشتری مصرف کرده‌اند.

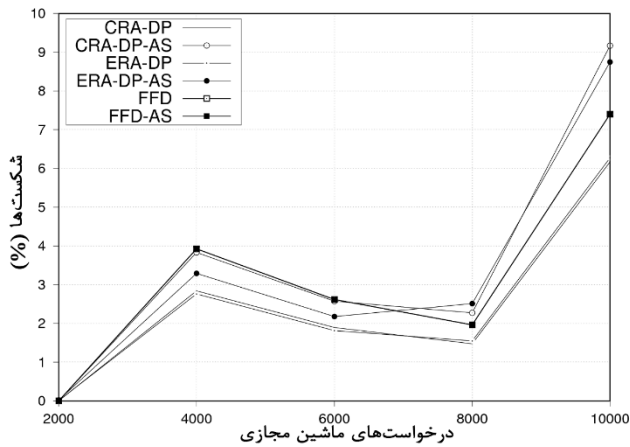


شکل ۶ نمودار مصرف انرژی کل (برق شهری + انرژی خورشیدی)

۵-۲-۴- میزان انتشار کربن

نتایج حاصله از شبیه‌سازی در مورد میزان انتشار کربن نشان می‌دهد که رفتار مشابهی با بخش ۵-۲-۳ اتفاق افتاده است. شکل ۷ نشان می‌دهد که الگوریتم‌های جانمایی ماشین‌های مجازی که با الگوریتم AS ترکیب شده‌اند و سعی کرده‌اند تعداد سرورهای فعال خود را به حداقل برسانند، دارای کمترین میزان انتشار کربن هستند.

از این رو الگوریتم‌های ERA-DP-AS، CRA-DP-AS و FFD-AS به‌طور میانگین میزان انتشار کربن را به ترتیب ۵۴٪، ۴۹٪ و ۳۴٪ نسبت به الگوریتم هم نوع خود کاهش داده‌اند. همچنین الگوریتم FFD دارای بیشترین میزان انتشار کربن بوده و به‌طور



شکل ۱۱ نمودار نقض قرارداد سطح خدمات

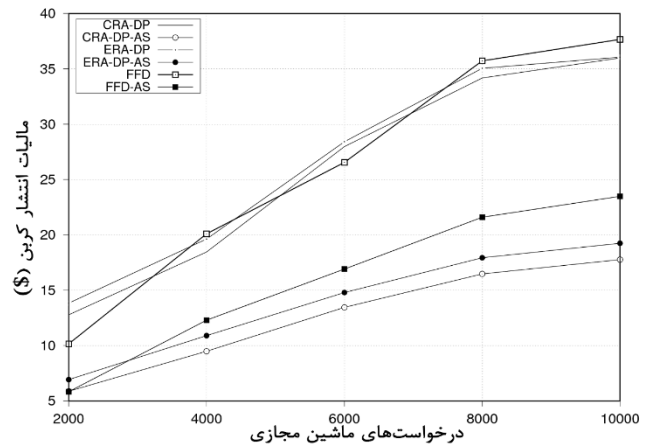
به عبارت دیگر الگوریتم FFD به‌طور میانگین ۰.۷٪ از ERA-DP و ۰.۳٪ از CRA-DP کمتر شامل مالیات انتشار کربن شده است. به عبارت دیگر الگوریتم‌های هم نوع خود، به‌طور میانگین دارای ۰.۱٪ نقض SLA بیشتری می‌باشند که این امر نشان می‌دهد الگوریتم AS به خوبی سطح کیفیت سرویس (QoS) را رعایت می‌کند. به عبارت دیگر شش الگوریتم CRA-DP-AS، ERA-DP-AS، FFD، FFD-AS، ERA-DP و CRA-DP به‌طور میانگین هریک به میزان ۰.۳٪، ۰.۳٪، ۰.۳٪، ۰.۲٪ و ۰.۲٪ متحمل نقض SLA شده‌اند.

نکته قابل توجه در نمودار شکل ۱۱ افزایش تعداد نقض SLA تا ۴۰۰۰ درخواست ماشین مجازی و کاهش مجدد آن تا ۸۰۰۰ درخواست ماشین مجازی می‌باشد، در حالی که تعداد درخواست‌ها در حال افزایش می‌باشد که البته این امر از ۸۰۰۰ درخواست تا ۱۰۰۰۰ افزایش یافته است.

این موضوع می‌تواند دلایل متعددی داشته باشد اما یکی مهم‌ترین آن‌ها می‌تواند مربوط به زمان رسیدن درخواست‌های ماشین مجازی باشد. به عبارت دیگر در دیتاست استفاده شده مربوط به بارهای کاری، فاصله زمانی بین درخواست‌های ارسال شده در بازه ۲۰۰۰ تا ۴۰۰۰ و همچنین ۸۰۰۰ تا ۱۰۰۰۰ متراکم‌تر است که این امر می‌تواند موجب افزایش نقض SLA شود.

۵-۲-۹- زمان اجرای فرایند شبیه‌سازی

شکل ۱۲ نمودار زمان اجرای فرایند شبیه‌سازی را نشان می‌دهد. همانطور که قبل‌تر اشاره شد، الگوریتم FFD به دلیل سرعت اجرای بالا دارای شهرت بسیار است. براین اساس، همانطور که در نمودار مشخص است، الگوریتم FFD با میانگین زمان اجرای 67.4 ثانیه در اجرای ۱۰۰۰۰ بارکاری، دارای کم‌ترین میزان زمان اجرا در بین کلیه الگوریتم‌ها است. همچنین سه الگوریتم ERA-DP-AS، CRA-DP-AS و ERA-DP با زمان‌های اجرای 153.4،

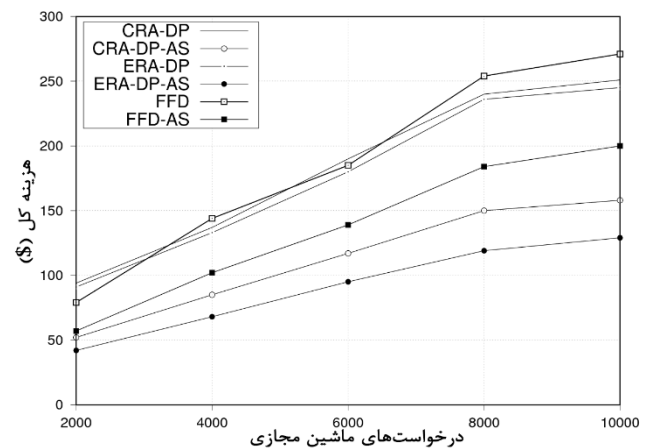


شکل ۹ مالیات انتشار کربن

به عبارت دیگر الگوریتم FFD به‌طور میانگین ۰.۷٪ از ERA-DP و ۰.۳٪ از CRA-DP کمتر شامل مالیات انتشار کربن شده است.

۵-۲-۷- هزینه کل

همانطور که انتظار می‌رود با توجه به نتایج به دست آمده در بخش‌های قبلی، تحلیل هزینه کل (شکل ۱۰) که شامل هزینه انرژی مصرفی و همچنین مالیات انتشار کربن است، نشان می‌دهد که میزان هزینه کل در هر سه الگوریتم ERA-DP-AS، CRA-DP-AS و DP-AS به‌طور میانگین شامل کاهش ۰.۴۹٪، ۰.۳۹٪ و ۰.۲۷٪ نسبت به الگوریتم‌های هم نوع شده است. همچنین هر سه الگوریتم CRA-DP، FFD، ERA-DP به ترتیب دارای بیشترین میزان هزینه کل هستند.



شکل ۱۰ نمودار هزینه کل (مصرف انرژی + مالیات کربن)

۵-۲-۸- نقض قرارداد سطح خدمات

شکل ۱۱ نمودار قرارداد سطح خدمات را نشان می‌دهد. در این مقاله رخداد نقض SLA زمانی روی می‌دهد که هیچ یک از مراکز داده ارائه دهنده خدمات ابری به دلیل عدم وجود منابع کافی جهت سرویس‌دهی قادر به ایجاد ماشین مجازی درخواست شده از سوی کاربر نباشند. همانطور که مشخص است، دو الگوریتم

[7] N. Grozev and R. Buyya, "Inter-Cloud architectures and application brokering: taxonomy and survey," *Software: Practice and Experience*, vol. 44, no. 3, pp. 369-390, 2014.

[8] A. Khosravi, L. L. Andrew, and R. Buyya, "Dynamic vm placement method for minimizing energy and carbon cost in geographically distributed cloud data centers," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 2, pp. 183-196, 2017.

[9] Y. Ajiro and A. Tanaka, "Improving packing algorithms for server consolidation," in *int. CMG Conference*, 2007, vol. 253, pp. 399-406.

[10] M. Giacobbe, A. Celesti, M. Fazio, M. Villari, and A. Puliafito, "An approach to reduce energy costs through virtual machine migrations in cloud federation," in *2015 IEEE Symposium on Computers and Communication (ISCC)*, 2015, pp. 782-787: IEEE.

[11] M. Giacobbe, A. Celesti, M. Fazio, M. Villari, and A. Puliafito, "An approach to reduce carbon dioxide emissions through virtual machine migrations in a sustainable cloud federation," in *2015 Sustainable Internet and ICT for Sustainability (SustainIT)*, 2015, pp. 1-4: IEEE.

[12] L. Lefèvre and A.-C. Orgerie, "Designing and evaluating an energy efficient cloud," *The Journal of Supercomputing*, vol. 51, no. 3, pp. 352-373, 2010.

[13] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33-37, 2007.

[14] M. Lin, A. Wierman, L. L. Andrew, and E. Thereska, "Dynamic right-sizing for power-proportional data centers," *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1378-1391, 2012.

[15] D. Shen *et al.*, "Stochastic modeling of dynamic right-sizing for energy-efficiency in cloud data centers," *Future Generation Computer Systems*, vol. 48, pp. 82-95, 2015.

[16] H. Duan, C. Chen, G. Min, and Y. Wu, "Energy-aware scheduling of virtual machines in heterogeneous cloud computing systems," *Future Generation Computer Systems*, vol. 74, pp. 142-150, 2017.

[17] F. Juarez, J. Ejarque, and R. M. Badia, "Dynamic energy-aware scheduling for parallel task-based application in cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 257-271, 2018.

[18] A. Khosravi, S. K. Garg, and R. Buyya, "Energy and carbon-efficient placement of virtual machines in distributed cloud data centers," in *European Conference on Parallel Processing*, 2013, pp. 317-328: Springer.

[19] K. Haghshenas, A. Pahlevan, M. Zapater, S. Mohammadi, and D. Atienza, "Magnetic: Multi-agent machine learning-based approach for energy efficient dynamic consolidation in data centers," *IEEE Transactions on Services Computing*, 2019.

[20] S.-Y. Hsieh, C.-S. Liu, R. Buyya, and A. Y. Zomaya, "Utilization-prediction-aware virtual machine consolidation approach for energy-efficient cloud data centers," *Journal of Parallel and Distributed Computing*, vol. 139, pp. 99-109, 2020.

[21] R. Yadav, W. Zhang, O. Kaiwartya, P. R. Singh, I. A. Elgandy, and Y.-C. Tian, "Adaptive energy-aware algorithms for minimizing energy consumption and SLA violation in cloud computing," *IEEE Access*, vol. 6, pp. 55923-55936, 2018.

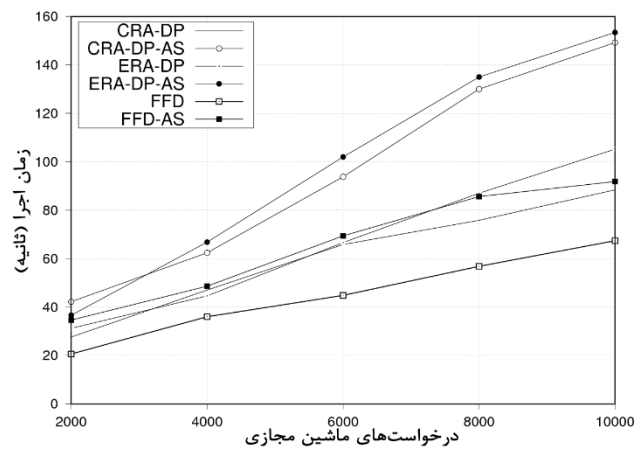
[22] J. Peterpaul, "Solar panel module and support therefor," ed: Google Patents, 1987.

[23] P. Thibodeau, "Wind power data center project planned in urban area," *ComputerWorld, Apr*, 2008.

[24] M. C. Silva Filho, R. L. Oliveira, C. C. Monteiro, P. R. Inácio, and M. M. Freire, "CloudSim plus: a cloud computing simulation framework pursuing software engineering principles for improved modularity, extensibility and correctness," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2017, pp. 400-406: IEEE.

[25] (2020, Jan. 29). *Data Center Map*. Available: <https://www.datacentermap.com/>

[26] S. Pelley, D. Meisner, T. F. Wenisch, and J. W. VanGilder, "Understanding and abstracting total data center power," in



شکل ۱۲ نمودار زمان اجرای فرایند شبیه‌سازی

۱۴۹.۲ و ۱۰۵.۲ ثانیه به ترتیب دارای طولانی‌ترین زمان اجرا هستند.

۶- نتیجه‌گیری و کارهای آینده

به‌طور خلاصه، ما در این مقاله به معرفی الگوریتم افزایش و کاهش خودکار مقیاس سرور (الگوریتم AS) پرداختیم. در این الگوریتم تمرکز اصلی در به حداقل رساندن تعداد سرورهای فعال جهت کاهش حداکثری انرژی مصرفی و انتشار کربن است. مدل ارائه شده در ابزار CloudSimPlus شبیه‌سازی گردید و نتایج حاصله را با بهترین کارهای انجام شده در این حوزه مقایسه کردیم. نتایج پیاده‌سازی نشان داد که الگوریتم پیشنهادی این مقاله می‌تواند به‌طور میانگین موجب کاهش حداکثری ۴۷٪ در مصرف انرژی برق شهری (سوخت‌های فسیلی) و همچنین کاهش حداکثری ۵۴٪ در انتشار کربن نسبت به الگوریتم‌های قبلی شود. در آینده قصد داریم تاثیر مهاجرت ماشین مجازی در کاهش انرژی مصرفی و انتشار کربن را مورد بررسی قرار دهیم.

مراجع

[1] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.

[2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation computer systems*, vol. 25, no. 6, pp. 599-616, 2009.

[3] A. Beloglazov, R. Buyya, Y. C. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," in *Advances in computers*, vol. 82: Elsevier, 2011, pp. 47-111.

[4] C. Pettey, "Gartner estimates ICT industry accounts for 2 percent of global CO2 emissions," *Dostupno na: https://www.gartner.com/newsroom/id/503867*, vol. 14, p. 2013, 2007.

[5] C. Hepburn, "Regulating by prices, quantities or both: an update and an overview," *Oxford Review of Economic Policy*, vol. 22, no. 2, pp. 226-247, 2006.

[6] G. Lewis, "Basics about cloud computing," *Software engineering institute carnegie mellon university, Pittsburgh*, 2010.

- [30] M. Dunlap, W. Marion, and S. Wilcox, "Solar radiation data manual for flat-plate and concentrating collectors," National Renewable Energy Lab., Golden, CO (United States)1994.
- [31] (2019, Oct. 21). *Travel weather averages (weatherbase)*. Available: <http://www.weatherbase.com>
- [32] (2020, Jan. 29). *The metacentrum 2 workload logs*. Available: https://www.cs.huji.ac.il/labs/parallel/workload/1_metacentrum2/index.html
- [27] (2015, Jan. 30). *US Department of Energy*. Available: http://cloud.agroclimate.org/tools/deprecated/carbonFootprint/references/Electricity_emission_factor.pdf.
- [28] (2015, Jan. 30). *Carbon Tax Center*. Available: <http://www.carbontax.org>
- [29] (2017, Apr. 7). *EIA-electricity data*. Available: <http://www.eia.gov/electricity/monthly/pdf/epm.pdf>

پاورقی‌ها:

- ³² Energy Sustainability
- ³³ Power Usage Efficiency (PUE)
- ³⁴ VM Consolidation
- ³⁵ Markov Chain
- ³⁶ Underutilized
- ³⁷ Over-utilized
- ³⁸ Constraints
- ³⁹ Cloud Broker
- ⁴⁰ Cloud Users
- ⁴¹ Backbone
- ⁴² Solar Energy
- ⁴³ Off-side Grid Energy
- ⁴⁴ Solar Panels
- ⁴⁵ Host
- ⁴⁶ VM Type
- ⁴⁷ Hold Time
- ⁴⁸ Data Center Power Efficiency
- ⁴⁹ Server Power Model
- ⁵⁰ Renewable Energy Sources
- ⁵¹ Carbon Footprint Rate
- ⁵² Carbon Tax
- ⁵³ Uninterruptible Power Supply (UPS)
- ⁵⁴ Sleep Mode
- ⁵⁵ Idle State
- ⁵⁶ Timezone
- ⁵⁷ Network Attached Storage (NAS)
- ⁵⁸ Million Instruction Per Second (MIPS)
- ⁵⁹ Instances
- ⁶⁰ On-peak
- ⁶¹ Off-peak
- ⁶² Million Instructions (MI)
- ¹ The Information and Communication Technology (ICT)
- ² Cloud Computing
- ³ Infrastructure as a Service (SaaS)
- ⁴ Platform as a Service (PaaS)
- ⁵ Software as a Service (SaaS)
- ⁶ Ubiquitous
- ⁷ Cloud Service Provider
- ⁸ Data Center
- ⁹ Host
- ¹⁰ Carbon Emission / Carbon Footprint
- ¹¹ Constant Availability
- ¹² Elasticity
- ¹³ Scalability
- ¹⁴ Workload
- ¹⁵ Overprovision Resources
- ¹⁶ Quality-of-Service (QoS)
- ¹⁷ Service Level Agreement (SLA)
- ¹⁸ Data Center Utilization
- ¹⁹ Objective Function
- ²⁰ Green Computing
- ²¹ Dynamic right-sizing technique
- ²² Idle Servers
- ²³ Scheduling
- ²⁴ Ant Colony Algorithm
- ²⁵ Quality of Service (QoS)
- ²⁶ Task-based Applications
- ²⁷ Distributed Computing Platforms
- ²⁸ Polynomial-time Algorithm
- ²⁹ Heuristic Rules
- ³⁰ Resource Allocation
- ³¹ Multi-objective Function