



سیروکومولوس^۱: بررسی و ارائه مدلی کارا به منظور کاهش مصرف و هزینه انرژی مراکز داده با پراکندگی جغرافیایی

امیر حسین مجرد*

دانشجوی کارشناسی ارشد، مدیریت سیستم های اطلاعاتی، دانشگاه شیراز واحد پردیس بین الملل

ahmjrd@live.com

دکتر غلامحسین دستغیبی فرد

دانشیار بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر دانشگاه شیراز، شیراز - ایران

چکیده: امروزه به کارگیری مراکز داده با پراکندگی جغرافیایی به منظور تقسیم بار و افزایش تحمل خرابی بسیار مورد توجه ارائه دهندگان خدمات ابری می باشد. یکی از چالش های ارائه دهندگان خدمات ابری کاهش زمان پاسخ و هزینه های انرژی می باشد که در این راستا، با توجه به نرخ متغیر هزینه های انرژی در کشور های مختلف و زمان پاسخ به سرویس گیرندگان، ماشین های مجازی را در بین کشور های مختلف میزبانی می نمایند. روش های ارائه شده عمدتاً از هزینه های برق، میزان انتشار کربن ناشی از برق مصرفی مراکز داده و مدت زمان پاسخ، به عنوان پارامترهای تاثیر گذار در الگوریتم های خود استفاده نموده اند. در این مقاله، الگوریتم پیشنهادی (سیروکومولوس) علاوه بر موارد فوق، هزینه های انتقال ماشین های مجازی، را به عنوان یکی دیگر از پارامترها، در نظر گرفته است. نتایج ارزیابی الگوریتم سیروکومولوس، برای سه مرکز داده از سری مراکز داده آمازون، در مقایسه با الگوریتم های پیشین نشان میدهد، هزینه های انرژی تا حداکثر ۲،۵۳٪ و مدت زمان پاسخ تا حداکثر ۲،۱۲٪ بهبود پیدا کرده است.

کلمات کلیدی: محاسبات ابری، جابه جایی ماشینهای مجازی، کاهش مصرف انرژی مراکز داده، تقسیم بار

^۱نوعی ابر که پیام آور هوای صاف و آرام است. ساختمان آن‌ها اغلب متشکل از قطعات سفید رنگ بوده و معمولاً پیش از ابرهای سیروس در آسمان ظاهر می‌شوند.



۱. مقدمه:

در محاسبات ابری منابع سخت افزاری، پردازشی، نرم افزار، خدمات، پلت فرم و فضای ذخیره سازی، از طریق بستر اینترنت در دسترس سرویس گیرندگان قرار خواهد گرفت. هدف از بهره گیری این دسته از خدمات، حذف هزینه های هنگفت نگهداری تجهیزات رایانه ای برای شرکت ها و موسسات می باشد. به طور کلی محاسبات ابری به معنای اشتراک گذاری خدمات میزبانی شده از طریق اینترنت می باشد. پایه ریزی این تکنولوژی بر اساس منطبق به اشتراک گذاشتن پردازنده ها، فضاهای ذخیره سازی، شبکه و کاربری هایی که توسط سوم شخص ارائه می شود، بنا نهاده شده است [1]. تقاضای سیری ناپذیر استفاده از منابع ابری توسط مشاغل و اشخاص، باعث گردیده است که مدیریت منابع^۲، بعنوان یکی از مهمترین دغدغه های تصمیم گیری مدیران خدمات ابری محسوب شود. امروزه یکی از محبوب ترین خدمات ابری، زیرساخت به عنوان سرویس^۳ نامدارد، که درحقیقت اجاره دادن منابع سخت افزاری سطح پایین ماشین های فیزیکی^۴ مانند پردازنده، فضای حافظه و فضای ذخیره سازی، در قالب ماشین های مجازی می باشد. این دسته از خدمات معمولا به صورت، پرداخت با توجه به نیاز مشتری^۵ ارائه می گردد که باعث کاهش هزینه های جاری و راه اندازی کاربران می گردد [2]. خدمات بسیاری در قالب سرویس های ابری قابل ارائه می باشد. در این راستا، ارائه دهندگان سرویس های ابری، انباره عظیمی از منابع مورد نیاز این خدمات را فراهم نموده اند. این منابع عظیم می توانند در قالب یک مرکز داده و یا مراکز داده با پراکندگی جغرافیای ارائه شوند.

در اکثر قریب به اتفاق شبکه های مبتنی بر پروتکل IP، بهینه بودن توان عملیاتی و به کارگیری مناسب از لینک های ارتباطی^۶، از اصلی ترین مواردی است که در طراحی آن شبکه ها مد نظر گرفته می شود. استفاده از پهنای باند بالا، به خودی خود، نیاز های کنونی شبکه های امروزی را مرتفع نمی سازد. وجود پهنای باند بالا معمولا برای انتقال حجم بزرگی از داده و یا کاربری های که عنصر زمان در آنها دخیل نمی باشند، الزامی می باشد. شبکه های کنونی علاوه بر پهنای باند مناسب، نیازمند مدت زمان پاسخ متعادل و معقولی نیز می باشند. برخی از کاربری هایی که نیازمند انتقال حجم اندکی داده در بازه زمانی مشخص می باشند همانند انتقال صدا و تصویر و یا کاربری هایی که نیازمند دریافت پاسخ های بلادرنگ می باشند، مانند شبکه سنسورها و یا بازی های برخط، نمونه هایی از این ارتباطات می باشند. در نتیجه تاخیر در ارتباطات، نکته حیاتی و پررنگی در شبکه های کنونی بوده که نادیده گرفتن آنها، باعث تخریب کاربری آن در شبکه می باشد [3]. گسترده نمودن جغرافیایی مراکز داده، در صورت پیاده سازی مناسب تقسیم بار، دارای مزایای بسیاری می باشد. کاهش تاخیر^۷ در ارائه خدمات و افزایش نرخ انتقال داده، از مزایای نزدیکی جغرافیایی سرویس گیرنده به مرکز داده می باشد. وجود تاخیر در ارتباط، در کاربری های همچون انتقال صدا بر بستر شبکه^۸ و نیز بازی های آنلاین، باعث افت چشم گیر کیفیت می گردد و به منظور ارتقا کیفیت خدمات^۹ می بایست تاخیرهای موجود در ارتباطات به حداقل ممکن برسد. مقالات بسیاری در خصوص نحوه بهینه سازی انرژی در زمان تقسیم بار [6] [5] [4] به منظور افزایش بهره وری سیستم و کاهش هزینه های انرژی صورت پذیرفته است. این تقسیم بار از طریق مکانیزم های DNS پویا، تغییر مسیر درخواست های HTTP^{۱۰} و یا انتقال درخواست های

^۲ Resource management^۳ infrastructure as a service (IaaS)^۴ physical machines (PMs)^۵ pay-as-you-go^۶ Link utilization^۷ latency^۸ Voice-over-IP (VoIP)^۹ Quality of Service (QoS).^{۱۰} HTTP redirection



HTTP^{۱۱} امکانپذیر می باشد که این موارد در منابع [9] [8] [7] توضیح داده شده اند. خدمات تحت وب در مقیاس های اینترنتی بزرگ اغلب توسط چندین مرکز داده با پراکندگی جغرافیایی در حال ارائه خدمات می باشند، به طوری که چندین هزار سرور در هر مرکز داده میزبانی می شوند. به منظور ارائه این خدمات نیازمند هزاران مگاوات برق بوده که معادل ده میلیون دلار هزینه سالانه می باشد [4].

با توجه به تقاضای رو به رشد مصرف انرژی مراکز داده، انگیزه ای قوی برای محققان در سراسر دنیا به وجود آمده است که از دو جنبه اقتصادی و زیست محیطی به مطالعه و تحقیق در راستای بهینه سازی مصرف انرژی در مراکز داده بپردازند. بهینه سازی مهندسی پردازنده های چند هسته ای، سیستم های سرمایه گذاری اختصاصی مراکز داده، سیستم های برق رسانی مراکز داده، مجازی سازی و ... همگی به منظور بهبود بخشیدن بهینه سازی مصرف انرژی صورت پذیرفته است [11] [10] [13] [12]، تمرکز این مقالات در راستای کاهش هزینه های برق مراکز داده و سامانه های مصرف کننده برق مانند سیستم های سرمایه گذاری می باشد [14].

در این مقاله سعی بر آن شده که چندین فاکتور تاثیرگذار بر راندمان مراکز داده و نیز هزینه های انرژی مورد ارزیابی قرارگیرد. این فاکتورها شامل، هزینه های انرژی برق سرورها، هزینه های جابه جایی ماشین های مجازی بین مراکز داده، بررسی امکانپذیر بودن جابه جایی ماشین مجازی با توجه به زمان انتقال و قطعی خدمات، هزینه های برق سیستم های سرمایه گذاری، ترافیک شبکه بین مراکز داده و کاربران، میزان نرخ خواندن و نوشتن اطلاعات^{۱۲} از/ بر روی اطلاعات، توافق نامه سطح خدمات^{۱۳} می باشند. همیشه دستیابی به بهترین حالت برای کلیه آیتم های مطرح شده به صورت یکجا امکانپذیر نمی باشد و گاه برخی از المان ها با یکدیگر در تضاد و تلاقی می باشند. با توجه به پراکندگی جغرافیایی مراکز داده و نیز اختلاف زمانی بین مکان های جغرافیایی، هزینه های برق مصرفی، ساعات اوج مصرف برق، پهنای باند و ترافیک های متفاوت به هر مرکز داده و نوع و هزینه سیستم های سرمایه گذاری در هر مرکز داده متفاوت می باشد. از آنجا که مشخصات ماشین های مجازی و فیزیکی موجود در مراکز داده نیز یکسایان نمی باشد، طبیعتاً چالش های دیگری مانند ظرفیت آزاد اتاق های سرور، تقاضای های متفاوت برای منابع، ترافیک شبکه، توده انبوهه منبع درخواست های شبکه و ... نیز گریبانگیر تصمیم گیری به منظور انتقال ماشین های مجازی می باشند. در روش پیشنهادی سیروکومولوس، که یک روش مبتنی بر گراف می باشد، به بررسی و کنترل این پارامترهای تاثیر گذار در محیط ابری، پرداخته ایم. با بهره گیری از دیاگرام ورونوی، ارجح ترین مرکز داده برای میزبانی ماشین های مجازی مورد نیاز را در بازه های مشخص زمانی، برآورد و طبقه بندی می کنیم و در نهایت میزان صرفه جویی در هزینه و میزان مصرف برق محاسبه می گردد.

۲. تحقیقات مرتبط

بیشتر فعالیت های انجام شده و تحقیقات موجود به منظور بهینه سازی مصرف انرژی و هزینه های برق، در یک مرکز داده صورت پذیرفته اند که بهترین ماشین فیزیکی برای میزبانی ماشین های مجازی را از بین ظرفیت های آزاد ماشین های فیزیکی مستقر در آن مرکز داده محاسبه می نماید و با این جایگزینی، حداکثر استفاده از ظرفیت آن مرکز داده صورت پذیرفته و ماشین های فیزیکی بلا استفاده، خاموش شده و یا در حالت خواب قرار میگیرند. در این صورت علاوه بر صرفه جویی در برق مصرفی ماشین های فیزیکی در میزان مصرف برق تجهیزات شبکه موجود در آن مرکز داده و نیز در مصرف برق سیستم سرمایه گذاری اتاق سرور نیز صرفه جویی می گردد.

^{۱۱} Forwarding HTTP requests

^{۱۲} Input / Output per second IOPS

^{۱۳} Service Level Agreement (SLA)



در [15] الگوریتمی ارائه گردیده است، که تعداد سرور های روشن یک مرکز داده را به حداقل مورد نیاز کاهش می دهد، در کنار کاهش این تعداد، آستانه اطمینانی نیز در نظر گرفته شده است که در صورت نیاز به افزایش ناگهانی منابع، با کندی سیستم ها مواجه نباشیم و همیشه سرور هایی روشن و در حالت آماده باش در چرخه سرویس دهی باشند. در [16] به بررسی میزان مصرف برق تجهیزات شبکه مانند سویچ ها و روتر های شبکه، پرداخته شده است و روش هایی به منظور کاهش و بهینه سازی مصرف انرژی تجهیزات سویچینگ ارائه گردیده است و میزان مصرف برق به ازای پهنای باند های متفاوت پورت های تجهیزات آنالیز و اندازه گیری شده اند، لیکن از آنجا که لینک های فیبر نوری، با توان مصرفی برق ثابتی عملیاتی می باشند، لذا این مقاله صرفاً برای یک مرکز داده قابل پیاده سازی می باشد.

در [6] از الگوریتم های توزیع شده ای^{۱۴} استفاده می نمایند که با کمک تکنیک بهینه سازی، تجسم مشبک^{۱۵}، مجموع هزینه های انرژی و نگهداری مراکز داده را کاهش می نمایند. علاوه بر آن، مقاله مذکور، هزینه تاثیرات اجتماعی یک مرکز داده را نیز مورد ارزیابی و بررسی قرار می دهند و سعی در به حداقل رساندن مضرات آن نموده اند. هزینه های تاثیرات اجتماعی، معیاری برای سنجش میزان اثر یک مرکز داده بر محیط زیست پیرامون می باشد. با بررسی در دسترس بودن انرژی های تجدید پذیر و انتقال بار محاسباتی و پردازشی ماشین های مجازی به مراکز داده هایی مجهز به انرژی های سبز، سعی بر کاستن تاثیرات منفی زیست محیطی مراکز داده، نموده اند.

علاوه بر موارد مطرح شده، تحقیق ها و آنالیز هایی بر الگوی میزان مصرف انرژی در مراکز داده نیز صورت پذیرفته است. در [17] میزان مصرف برق برای کلیه المان های یک مرکز داده با توجه به نوع سرویس های میزبانی شده و سناریو های پیاده سازی آن خدمات، مورد ارزیابی قرار گرفته شده است. ضمناً پژوهش هایی در خصوص نحوه تعیین نمودن مقصد خدمات مورد تقاضا^{۱۶} با در نظر گرفتن میزان انتشار کربن مراکز داده، انجام پذیرفته است. در [18] با توسعه مدل ارائه شده در [6] با کسر هزینه های برق تولید شده، توسط ژنراتورها و مولد های برق مستقر در مراکز داده، این امکان را فراهم نموده که مراکز داده ها ای که از انرژی های سبز استفاده می نمایند، بار پردازشی و ذخیره سازی بیشتری را نسبت به سایر مراکز داده بر عهده گیرند. در [19] با استفاده از الگوریتم ژنتیک، محاسبات انتقال و جابه جایی ماشین های مجازی صورت پذیرفته و با کاهش تعداد سرورهای فیزیکی روشن، میزان انتشار کربن حاصل از فعالیت یک مرکز داده را تقلیل داده اند.

Doyle و همکاران، در [20] با استفاده از دیگرام ورونوی اقدام به پیدا کردن نزدیک ترین سایت مرکز داده به توده انبوه مشترکین نموده است و با تمرکز بر میزان تاثیرات انتشار کربن در هر مرکز داده و هزینه های مصرف برق، اقدام به چیدمان مجدد ماشین های مجازی در بین مراکز داده، نموده اند. در [20]، فاصله بین خدمات گیرندگان و مراکز داده از طریق مدت زمان پاسخ مسیر محاسبه می گردد و فقدان محاسبه میزان پهنای باند آزاد لینک ارتباطی، هزینه های مربوط به آن و هزینه و مدت زمان انتقال یک ماشین مجازی از یک مرکز داده به مرکز داده دیگر، در الگوریتم پیشنهادی مشاهده می گردد. در صورتی که کلیه اطلاعات فضای ذخیره سازی ماشین های مجازی در کلیه مراکز داده، همسان باشند در این صورت، انتقال ماشین های مجازی با حداقل زمان و هزینه ممکن صورت خواهد پذیرفت که صرفاً هزینه های انتقال فضای RAM ماشین های مجازی می باشد، در غیر اینصورت می بایست کلیه اطلاعات از فضای ذخیره شده از مرکز داده مبدا، به سمت مرکز داده مقصد انتقال یابد که با توجه به حجم فضای ذخیره سازی ماشین مجازی، فرآیندی زمانبر و گاه ناقص تفاهم نامه سطح خدمات می باشد.

در [21] تقسیم بندی محیط با استفاده از دیگرام ورونوی صورت پذیرفته که به منظور دسترسی روبات ها و پوشش آنها در آن فضا به کار برده می شود. این مسئله را می توان به عنوان نسخه محدود شده ای از مسئله مکان یابی تسهیلات^{۱۷} ارائه نمود، که در [22] اثبات گردیده است که از مسائل NP-Hard می باشند و در [23] راه حل هایی برای حل کردن این

^{۱۴} Distributed Algorithms

^{۱۵} Gradient Projection

^{۱۶} Direct Service Request (DSR)

^{۱۷} the facility location problem



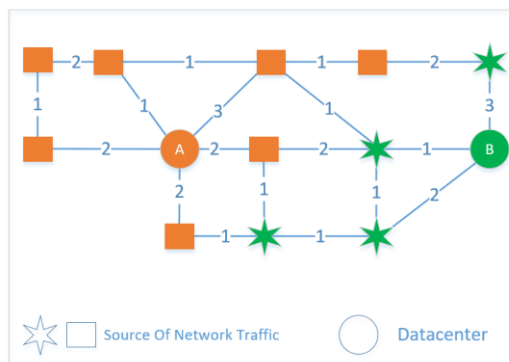
مسائل ارائه شده است. هر چند که روش های رایج حل این دسته از مسائل بار پردازشی سنگینی را در بر خواهد داشت، لیکن در صورتی که ورودی مسئله ایستا باشد، جواب بهینه به دست می آید. اشاره به این نکته الزامی است که در محیط ابری، تقاضا برای یک سرویس بسیار متغیر بوده و به کارگیری از سامانه ای که پاسخ گوی افزایش تغییرات باشد، ضروری می باشد.

۳. فرمول بندی مسئله

در این بخش به منظور روشن تر شدن موضوع به توضیح در مورد برخی از اصطلاحات این مقاله می پردازیم.

۳-۱. تقسیم بندی در محاسبات ابری

در این مقاله مراکز داده و منابع تولید ترافیک شبکه به عنوان نقاط دیاگرام ورونوی در نظر گرفته شده اند. یک سلول ورونوی نمایانگر آن می باشد که کدام مرکز داده می بایست در آن بازه زمانی مشخص، به مبدا ترافیک شبکه خدمات رسانی نماید و یا به عبارتی کدام ماشین مجازی می بایست در کدام مرکز داده میزبانی شود. در تصویر ۱، دو مرکز داده در نظر گرفته شده است که با دایره نمایش داده شده اند و هر مبدا ترافیک شبکه دارای حداقل دو مسیر به هر دو مرکز داده می باشد. زمانی منابع تولید ترافیک شبکه با یک مرکز داده در یک دسته بندی قرار میگیرند که مجموع وزن یال های موجود در مسیر آن منابع تا مرکز داده کمتر از سایر مسیر ها باشد. با جدا سازی رنگ بندی ها و نشانه منابع مشخص شده است که کدام منابع ترافیک شبکه با کدام مرکز داده در یک طبقه بندی قرار دارند، به گونه ای که منابع تولید ترافیکی که با نمایه مربع مشخص شده اند با مرکز داده A و منابعی که نمایه ستاره مشخص شده اند با مرکز داده B در یک سلول قرار گرفته اند.



تصویر ۱. نمونه ای از چگونگی طبقه بندی منابع تولید ترافیک شبکه و مراکز داده

۳-۲. بیان مسئله

فرض می کنیم $|D|$ مجموعه مراکز داده و $|S|$ مجموعه ای شامل منابع تولید ترافیک شبکه می باشد که با توجه به موقیت جغرافیایی و به صورت کشورهای تولید کننده ترافیک، دسته بندی می شوند. فرض می کنیم که $|N|$ مجموعه ای از نقاط باشد که شامل مراکز داده و منابع تولید ترافیک شبکه باشد، به گونه ای که توسط $|E|$ یال بدون جهت به یکدیگر متصل می باشند که وزن هر یال $|w|$ می باشد.

جهت محاسبه وزن از فرمول (۸) استفاده می شود:

$$w_i = f(T_i, P_i, M_i) = T_i + I_1(P_i) + I_2(M_i) \quad \forall i \in |w| \quad (8)$$

که در آن T_i مدت زمان مورد نیاز برای سرویس رسانی به درخواست می باشد و P_i هزینه های مصرف برق و M_i هزینه های انتقال ماشین های مجازی می باشد. ضمناً I_1 و I_2 به منظور مشخص نمودن ضریب اهمیت متغیر های تعریف شده یعنی هزینه های برق و یا هزینه های جابه جایی ماشین مجازی به کار برده می شود. شایان ذکر است که هرچه هزینه انتقال ماشین مجازی بیشتر باشد، قطعی در سیستم و اختلال در تضمین سطح خدمات بیشتر می گردد.



مجموعه $|N|$ به تعداد D سلول تقسیم بندی می شود، که در نهایت نمایانگر آن می باشد که کدام مرکز داده، عهده دار سرویس رسانی به کدام ماشین مجازی می باشد می باشد. این نکته باید در نظر گرفته شود که هیچ یک از سلول های تعریف شده نباید تهی باشد، بدین معنا که هیچ مرکز داده ای نباید به صورت بلا استفاده باقی بماند. با توجه به ماهیت دیاگرام ورونوی، طبیعتاً هیچ نقطه ای در دو سلول عضو نمی باشد، در این صورت حتماً یک مرکز داده در بازه زمانی مشخص شده به سرویس دهی به کشور مورد نظر مشغول می باشد. با استفاده از دیاگرام ورونوی، هزینه های جابه جایی ماشین مجازی، هزینه های برق مصرفی و نیز مدت زمان پاسخ، به حداقل ممکن خواهند رسید.

بنابراین برای هر مرکز داده $i \in |D|$ دیاگرام ورونوی V_i مجموعه ای از نقاط می باشد که فاصله کمتری تا مرکز داده i دارند. به منظور محاسبه فاصله بین منبع تولید ترافیک شبکه و مرکز داده، $d(x, y)$ یک گراف وزن دار می باشد، که می بایست مجموع وزن کلیه مسیرهای ممکن تا مرکز داده، به صورت مجزا محاسبه شود و کمینه مقدار به دست آمده، به عنوان مسیر برگزیده و مرکز داده منتخب آن ماشین مجازی محسوب می گردد. لازم به ذکر است که در صورتی که یک نقطه با دو یا چند مرکز داده دارای فاصله یکسانی باشد، مرکز داده ای انتخاب می شود که دارای نقاط کمتری در سلول خود باشد. شبه کد زیر به منظور انتخاب ماشین مجازی و قراردادن در سلول یک مرکز داده به کار برده می شود

```

1   $U := P_i(t) \cup P_j(t)$ 
2  For  $x \in U$ 
3     $W_i := \{x \in U : d(x, i) \leq d(x, j)\}$ 
4     $W_j := \{x \in U : d(x, i) > d(x, j)\}$ 
5  End
6   $V_i(t+1) := W_i$ 
7   $V_j(t+1) := W_j$ 

```

۳-۳. هزینه های انتقال برخط ماشین های مجازی^{۱۸}

انتقال برخط ماشین های مجازی به فرآیند کپی کردن یک ماشین مجازی از یک سرور فیزیکی به سرور دیگری، در حالتی که ماشین مجازی روشن باشد، گفته می شود [24]. در هر حال انتقال ماشین مجازی، باعث افت کارایی و راندامان سیستم در زمان انتقال می شود [25]. ضمناً نکته ای که هرگز نباید آن را نادیده گرفت، هزینه های تحمیل شده برق جانبی این فرآیند می باشد که در مراکز داده عظیم، عدد قابل ملاحظه ای است [26]. در حال حاضر انتقال برخط ماشین های مجازی، در خصوص ماشین های مجازی ای که بر روی حافظه ذخیره سازی محلی^{۱۹} سرورها میزبانی می شوند، قابل پیاده سازی نمی باشد و حتماً نیازمند بهره گیری از تکنولوژی هایی همانند دستگاه های ذخیره سازی پیوست شده به شبکه^{۲۰} و یا شبکه ذخیره سازی^{۲۱} می باشد که به صورت متمرکز اطلاعات کلیه ماشین های مجازی در یک یا چند دستگاه ذخیره سازی مجتمع بوده و در دسترس کلیه سرور های فیزیکی مستقر در آن مرکز داده می باشد. از این رو فرآیند انتقال برخط ماشین های مجازی به کپی کردن محتوای موجود در RAM و CPU بین مجازی سازها^{۲۲} محدود می باشد. به منظور انجام این فرآیند از تکنیک پیش-کپی^{۲۳} استفاده می شود [24].

^{۱۸} Virtual machine Live Migration

^{۱۹} Local Hard Disk

^{۲۰} Network Attached Storage (NAS)

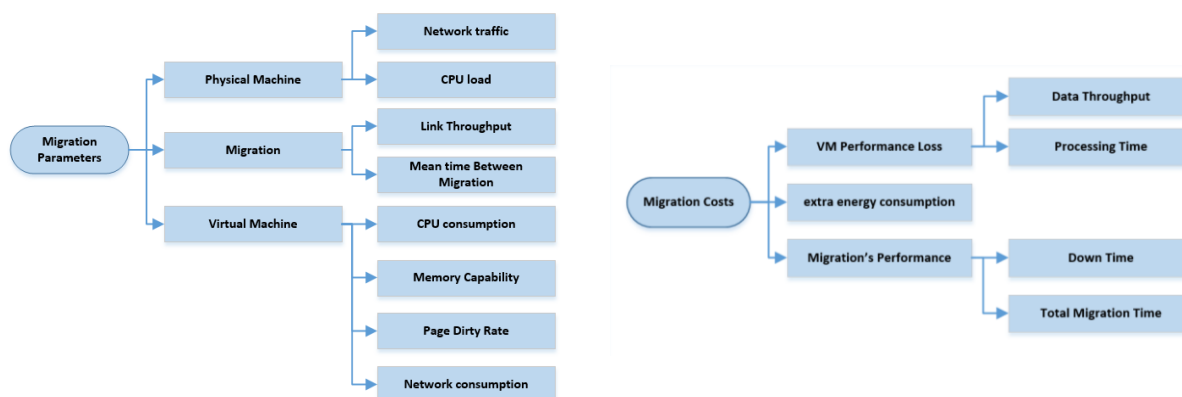
^{۲۱} Storage Area Network (SAN)

^{۲۲} hypervisors

^{۲۳} pre-copy



انتقال بر خط ماشین های مجازی پیش تر در مقالات متعددی مورد بحث و بررسی قرار گرفته است. همانگونه که در تصویر ۲- راست نمایش داده است، ما هزینه های انتقال ماشین های مجازی را به دو دسته اصلی طبقه بندی میکنیم: راندمان و انرژی مصرفی. راندمان برخط ماشین های مجازی یکی از مباحث مهم در تحقیقات کنونی می باشد. در خصوص راندمان ماشین های مجازی در زمان انتقال برخط، دو عامل همیشه مورد توجه قرار میگیرد: مدت زمان انتقال کامل ماشین مجازی و مدت زمان قطعی سیستم. مدت زمان انتقال کامل ماشین مجازی، به مدت زمانی که از آغاز فرآیند انتقال ماشین مجازی تا لحظه ای که ماشین مجازی در سمت مقصد، به فعالیت خود ادامه دهد، گفته می شود. از طرفی مدت زمان قطعی سیستم، به زمانی اطلاق شده که ماشین مجازی از چرخه سرویس دهی خارج شده که در اصطلاح در فاز توقف-و-کپی^{۲۴} می باشد. در زمان انتقال ماشین مجازی، کاهش راندمان و کارایی سیستم، دومین پارامتر مهم می باشد. به منظور تشخیص و مدیریت صفحات کثیف^{۲۵} فضای حافظه، مجازی ساز مبدا می بایست حافظه ماشین مجازی را تحت نظارت کامل داشته باشد که این فرآیند باعث کند شدن چرخه سرویس رسانی می گردد. علاوه بر نظارت مستمر بر رخ دادهای حافظه داخلی ماشین مجازی، پردازش و اجرای دستورات درون آن ماشین نیز می بایست تحت نظارت کامل قرار گیرد. به منظور عملیاتی شدن فازهای پیش-کپی و توقف-و-کپی مجازی ساز نیازمند منابع سخت افزاری بیشتری می باشد. این منابع شامل بار پردازشی بیشتر و ترافیک شبکه بالاتر می باشد. لذا در صورتی که منابع مورد نیاز در سمت مجازی ساز موجود نباشد، افت محسوسی بر روی کارایی ماشین مجازی مورد انتقال و نیز سایر ماشین های مجازی میزبانی شده بر روی آن مجازی ساز، مشاهده می گردد [27].



تصویر ۲ - طبقه بندی هزینه های انتقال ماشین مجازی (راست) [24] و طبقه بندی پارامترهای جابه جایی ماشین های مجازی (چپ)

شایان ذکر است که هزینه های سربار انرژی حاصل از بار تحمیل شده، در راستای انتقال ماشین های مجازی، قابل چشم پوشی نیست [28] لذا با توجه به اینکه، هدف اصلی از انتقال ماشین های مجازی، کاهش هزینه های برق می باشد، بنابراین هزینه های وارد شده به سیستم در زمان انتقال ماشین های مجازی می بایست در محاسبات در نظر گرفته شود. در این مقاله فرض بر آن بوده است که تمامی مراکز داده دارای ظرفیت آزاد چه از نظر منابع سخت افزاری مانند فضای حافظه، فضای ذخیره سازی و پردازنده در سمت گیرنده و فرستنده و چه از نظر فیزیکی و سرمایشی، برای میزبانی ماشین های

^{۲۴} stop-and-copy

^{۲۵} dirty pages



مجازی انتقال یافته را دارند. بنابر این فرض قابلیت انتقال ماشین های مجازی در هر شرایطی موجود خواهد بود. در تصویر ۲- چپ پارامتر هایی که در هزینه جابه جایی ماشین های مجازی دخیل هستند طبقه بندی شده اند [29]. بسیاری از محققین معتقدند که مدت زمان کلی انتقال ماشین مجازی که t_{mig} می باشد، وابستگی فراوانی به موارد ذیل دارد:

الف) مجموع کلی فضای حافظه داخلی v_{mig} که می بایست به ماشین فیزیکی دیگری انتقال یابد.
ب) متوسط سرعت شبکه (متوسط پهنای باند شبکه) بین هر دو ماشین فیزیکی ([30], [28], [31], [27], [32]) که از رابطه (۹) قابل محاسبه می باشد:

$$t_{mig} = \frac{v_{mig}}{b} \quad (9)$$

در [18] جهت محاسبه حجم حافظه انتقال یافته در دور i ام انتقال مطابق (۱۰) می باشد:

$$v_{mig,i} = \begin{cases} v_{mem} & \text{if } i = 0 \\ d * l * t_{i-1} & \text{else} \end{cases} \quad (10)$$

در صورتی که فرآیند انتقال ماشین های مجازی دارای n مرحله باشد و یک مرحله نهایی توقف-و-کپی، آنگاه مجموع اندازه فضای حافظه v_{mig} و مدت زمان قطعی که t_{down} می باشد از رابطه های (۱۱) و (۱۲) به دست می آید:

$$\begin{aligned} v_{mig} &= \sum_{i=1}^{n+1} v_{mig,i} \\ t_{down} &= \frac{d * i * t_n}{b} \end{aligned} \quad (11)$$

با توجه به (۹)، (۱۰)، (۱۱)، و (۱۲)، حداقل و حداکثر مقدار زمان انتقال و زمان قطعی به صورت (۱۳) و (۱۴) می باشد:

$$\frac{v_{mem}}{b} \leq t_{mig} \leq \frac{(m_{ch}+1) * v_{mem}}{b} \quad (13)$$

$$0 \leq t_{down} \leq \frac{v_{mem}}{b} \quad (14)$$

۳-۴. هزینه های سیستم های سرمایه‌ی

بیش از نیمی از هزینه های برق یک مرکز داده، مربوط به سیستم های سرمایه‌ی اتاق سرور^{۲۶} می باشد. ضریب محاسبه سیستم های سرمایه‌ی و هوادهی در یک مرکز داده به نسبت ۱،۸ الی ۲،۵ برابر برق مصرفی تجهیزات پردازشی، ذخیره سازی و ارتباطی موجود در اتاق سرور می باشد. عوامل تاثیر گذار در این ضرایب، میزان بازده سیستم سرمایه‌ی، طراحی و چیدمان سیستم سرمایه‌ی در آن مرکز داده می باشد [33].

در یک مرکز داده هزینه های سیستم های سرمایه‌ی وابستگی شدیدی به نحوه طراحی آن و شرایط آب و هوایی دارد. یکی از محبوب ترین طراحی های سیستم های سرمایه‌ی راهرو مهاری^{۲۷} می باشد [34] که با ممانعت از ادغام هوای سرد تولید شده توسط سیستم های سرمایه‌ی با هوای گرم خروجی از تجهیزات پردازشی، راندمان سیستم سرمایه‌ی را افزایش می دهد. یکی دیگر از انواع طراحی های رایج سیستم های سرمایه‌ی مراکز داده، سیستم هوای آزاد^{۲۸} می باشد. در این روش با توجه به ورودت دمای محیط بیرون، از هوای سرد محیط خارجی برای سرمایه‌ی اتاق سرور استفاده می کنند [33].

به منظور محاسبه هزینه های سرمایه‌ی مراکز داده، جامعه آماری متشکل از سه مرکز داده با ظرفیت ۸۰ رک در هر یک مد نظر گرفته شده است. مدل رک های انتخاب شده با ارتفاع ۴۲ یونیت می باشد. ضمناً فرض گردیده است که کلیه مراکز داده از سرورهای HP مدل DL-380-G9 مطابق مشخصات جدول ۲ استفاده می نمایند.

به منظور انجام محاسبات سیستم سرمایه‌ی، از شبیه ساز ANSYS Fluent استفاده گردیده است. در شبیه سازی صورت پذیرفته ۸ ردیف رک در هر مرکز داده به همراه ۸ سیستم سرمایه‌ی اتاق سرور با حداکثر ظرفیت سرمایه‌ی ۹۰ کیلووات بر

^{۲۶} Computer Room Air Colling (CRAC)

^{۲۷} Aisle containment

^{۲۸} Free Air Cooling(FAC)



ساعت، در نظر گرفته شده است. دمای خروجی سیستم سرمایشی ۱۶ درجه سانتیگراد و قدرت پرتاب هوا ۱۷,۰۰۰ مترمکعب در ساعت تنظیم شده است. در نهایت هزینه های سیستم سرمایش را می توان از رابطه (۱۵) محاسبه نمود:

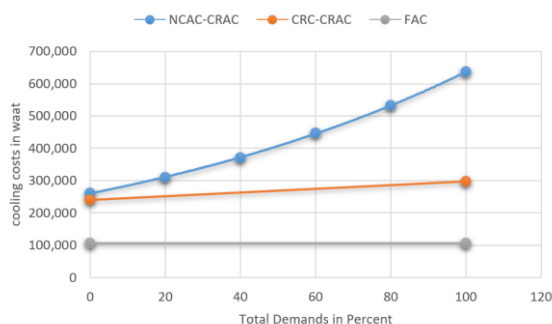
$$C = \frac{S}{COP(T_{sup} + (T_{safe} - T_{max}))} + P_{Fan} \quad (15)$$

که در آن C هزینه های سرمایش، S توان مصرفی سرورها، P_{Fan} توان مورد نیاز فن های موجود در دمنده های مستقر در اتاق سرور و COP ضریب عملکرد^{۲۹} سیستم سرمایشی می باشد. ضریب عملکرد سیستم سرمایشی نرخ گرمای خارج شده از محیط بر کار انجام شده جهت خروج آن گرما می باشد. همچنین در (۱۵) حداکثر دمای مجاز اتاق سرور می باشد که در این تحقیق مطابق استاندارد TIA942 معادل ۲۵ درجه سانتی گراد در نظر گرفته شده است و T_{max} دمای سرور ها می باشد.

جدول ۲- مشخصات سرور های استفاده شده در کلیه مراکز داده

	Specification	Quantity	Capacity
Processor	E5-2630Lv3	2	16 core
Memory	PC4-2133P	10	160 GB
Hard Disk	759212-B21	8	4.8 TB
Power supply	720478-B21	2	1000 w

در تصویر ۴ خروجی شبیه سازی صورت پذیرفته برای سه روش سرمایشی متفاوت در مرکز داده نمایش داده شده است. در CRC-CRAC^{۳۰} راهرو دربردارنده هوای سرد کاملا از راهرو گرم مجزا شده است. در NCAC-CRAC^{۳۱} راهروی گرم و سرد توسط هیچ مانعی از یکدیگر جدا نشده اند. در نهایت از آنجا که در سیستم های سرمایشی هوای آزاد، فقط فن دمنده ها فعالیت می نمایند، لذا میزان حرارت تولید شده در اتاق سرور تاثیری در میزان مصرف برق آن سیستم ها ندارد و یک رابطه خطی می باشد.



تصویر ۴. هزینه های برق مصرفی طراحی های مختلف سیستم های سرمایش با توجه به میزان بار موجود در مراکز داده.

۳-۵. هزینه های برق مصرفی

با توجه به تعرفه های متفاوت هزینه برق در کشورهای مختلف، این پتانسیل برای مالکان مراکز داده با پراکندگی جغرافیایی وجود دارد تا هزینه های برق مصرفی خود را کاهش دهند [7]. در این مقاله ما میزان صرفه جویی بالقوه در هزینه های انرژی

^{۲۹} coefficient of performance

^{۳۰} Cold Aisle Containment CRAC

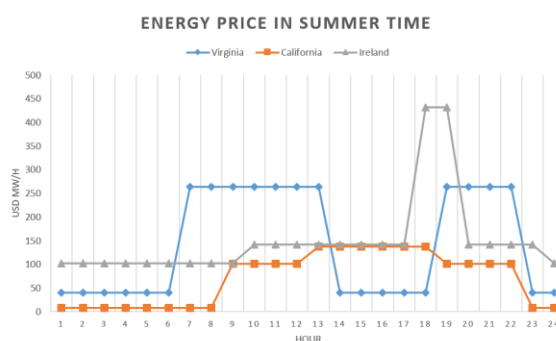
^{۳۱} None Cold Aisle Containment CRAC



را در بخشی از مراکز داده آمازون^{۲۲} مورد ارزیابی قرار داده ایم. این بخش شامل سه مرکز داده اصلی آمازون مستقر در کشور ایرلند، ایالت ویرجینیا و ایالت کالیفرنیا می باشد. اطلاعات جمع آوری شده شامل هزینه های برق مصرفی و نیز ساعات اوج مصرف و ساعات نیمه اوج مصرف می باشد.

حداکثر تعرفه برق استخراج شده مربوط به ایرلند می باشد که معادل ۴۳۲ دلار برای هر مگاوات ساعت مصرف برق اعلام شده است، بنابراین از دید نخست این فکر به ذهن خطور می نماید که ترافیک کمتری می بایست به سمت ایرلند هدایت شود تا در نهایت هزینه های برق کمتری برای مراکز داده داشته باشیم. لیکن ما در این تحقیق نشان داده ایم که با توجه به ترافیک شبکه و پهنای باند موجود، اختلاف ساعت بین مراکز داده و ساعات اوج مصرف انرژی، لزوماً کمترین ترافیک سرور ها متوجه مراکز داده مستقر در ایرلند نخواهد بود. در تصویر ۵ هزینه های مصرف برق برای سه منطقه ایرلند، ایالت کالیفرنیا و ایالت ویرجینیا در ۲۴ ساعت برای مصرف یک مگاوات ساعت، نمایش داده شده است. لازم به ذکر است که قیمت های معرفی شده برای قیمت های تابستانه می باشد. از آنجا که ارتفاع سرور های در نظر گرفته شده ۳ یونیت می باشد، لذا در هر مرکز داده ۱،۱۲۰ سرور فیزیکی میزبانی خواهد شد. با به کارگیری از نرم افزار HP Power Advisor، حداقل مصرف برق در سرورها زمانی که دارای صفر درصد بار می باشند ۱۲۴،۷۳ وات بر ساعت و حداکثر مصرف برق برای صد درصد بار بر روی سرورها، معادل ۲۷۲،۹۲ وات بر ساعت محاسبه گردیده است.

علاوه بر هزینه های برق سیستم های سرمایشی و سیستم های پردازشی موجود در مراکز داده، تجهیزات سویچینگ و مسیریابی مراکز داده نیز، در این مقاله مورد ارزیابی قرار گرفته اند. بدین صورت که مصرف برق، پورت های فعال و غیر فعال این تجهیزات نیز در محاسبات هزینه های انرژی دخیل شده اند و در صورتی که یک پورت سویچ بلا استفاده باشد، آنگاه آن پورت از حالت فعال به غیرفعال تغییر موقعیت خواهد داد. از آنجا که ما حداکثر ترافیک ساعتی هر ماشین فیزیکی را یک ترابیت در نظر گرفته ایم، این عدد معادل پهنای باند تقریبی ۲۹۱،۲ مگابیت بر ثانیه کارت شبکه سرور می باشد. کارت های شبکه سرور های معرفی شده در بخش ۴،۲ دارای توان عملیاتی یک گیگابیت بر ثانیه می باشند، بدین معنا که یک پورت شبکه سرور به تنهایی قادر به ارائه خدمات در حد مورد نیاز این مقاله می باشد، لیکن به علت در نظر گرفتن افزونگی در مراکز داده، دو پورت برای هر سرور فیزیکی، فعال در نظر گرفته شده که طبیعتاً دو پورت از سویچ را درگیر خود می نماید. طبق تحقیقات انجام شده در [35] میزان مصرف برق یک پورت فعال در سویچ های سیسکو سری ۳۷۵۰ برای پورت های دارای پهنای باند یک گیگابیت بر ثانیه، معادل ۰،۷ وات می باشد. در نتیجه به ازای هر جابه جایی ماشین فیزیکی، دو پورت سویچ از مراکز داده مبدا به حالت غیر فعال و دو پورت در مراکز داده مقصد به حالت فعال تغییر حالت خواهند داد.



تصویر ۵. هزینه های برق تابستانه بر حسب دلار برای یک مگاوات ساعت مصرف، در ساعات مختلف شبانه روز در ایرلند، کالیفرنیا و ویرجینیا

^{۲۲} Amazon's EC2



۳-۶. مدت زمان اجرای کار

در بخش های قبل در خصوص مدت زمان انتقال ماشین های مجازی، هزینه های سرمایش مرکز داده و نیز هزینه های مصرف انرژی، صحبت شد ولی صرفاً در نظر گرفتن هزینه برق و کاهش هزینه های سرمایشی در انتخاب مرکز داده میزبان سرویس ها، می تواند باعث افزایش زمان پاسخ سرویس ها گردد. در اینصورت مالکین خدمات ابری می بایست در خصوص سناریو های تعادل بار در مراکز داده خود، تجدید نظر و یا بازبینی نمایند. در نتایج حاصله مشاهده شده است که تفاوت های بسیاری در مدت زمان پاسخ در ساعات مختلف شبانه روز بین نقاط مبدا و مقصد وجود دارد. یکی از معیار های بررسی زمان پاسخ، میزان تاخیر موجود بین مبدا درخواست گیرنده سرویس و سرویس دهنده می باشد. به منظور محاسبه این مدت زمان از خدمات شرکت Planet Lab برای پینگ کردن کلاینت ها در بازه زمانی ۱۰ دقیقه ای به مدت سه روز، بهره گرفته شده است. یکی از علل این اختلاف در مدت شبانه روز ازدحام^{۳۳} ترافیک وارده به سمت مراکز داده و یا ترافیک شبکه موجود سمت درخواست دهنده خدمات و اختلاف ساعت می باشد.

بنابراین به منظور کاهش زمان تاخیر در بازه های زمانی که ازدحام بالایی وجود دارد، انتقال سرویس دهندگان به یک مرکز داده دیگر می باشد. به منظور محاسبه مدت زمان پاسخ خدمات^{۳۴} می بایست زمان تاخیر و زمان پاسخ بین مبدا و مقصد معین گردد. طبیعتاً این احتمال نیز وجود دارد که در صورت جا به جایی ماشین های مجازی به یک مرکز داده دیگر، تفاوتی در مدت زمان پاسخ حاصل نشود زیرا که ترافیک موجود از سمت دروازه های خروجی سرویس گیرندگان باشد. لیکن کوتاه تر کردن مسافت بین سرویس گیرنده و سرویس دهنده از نظر فاصله جغرافیایی گاهاً و نه لزوماً باعث کاهش هزینه های برق تجهیزات ارتباطی بین مسیر می باشد.

۴. پیاده سازی و ارزیابی الگوریتم پیشنهادی

در این بخش، روش شبیه سازی و نحوه وزن دهی به یال های گراف معرفی می شوند. همانگونه که در ۴,۳ اشاره گردید، سه مرکز داده از مراکز داده آمازون، مورد ارزیابی و شبیه سازی قرار می گیرند. در این مقاله ۲۳۴ کشور به عنوان منابع تولید ترافیک شبکه در نظر گرفته شده اند. هر یک از این منابع تولید ترافیک شبکه از طریق یک یال به هر سه مرکز داده متصل می باشد. میزان تاخیر بین هر کشور تا سه مرکز داده به مدت سه روز و با بازه زمانی ده دقیقه ای محاسبه گردیده شده است [36]، [37] و [38]. در این مقاله فرض بر آن می باشد که ماشین های مجازی میزبانی شده بر روی هر ماشین فیزیکی دارای ترافیک شبکه مشابه ای می باشند. در این پروژه فرض شده است که اطلاعات هارد سرورها در مرکز داده مقصد نیز موجود می باشد و نیازی به انتقال اطلاعات هارد دیسک ماشین ها نمی باشد.

زبان برنامه نویسی انتخاب شده به منظور انجام محاسبات Matlab R2013b می باشد.

یکی از نیازهای اصلی و انکار ناپذیر مراکز داده با پراکندگی جغرافیایی بهره گیری از پهنای باند بالا، ایمن و بدون وقفه در بین مراکز داده می باشد. در این مقاله به منظور محاسبه هزینه های انتقال و مدت زمان جابه جایی ماشین ها مجازی نیازمند در دسترس داشتن مقدار پهنای باند بین مراکز داده می باشد. از آنجا که پهنای باند بین مراکز داده یکی از اسرار کسب و کار مالکان مراکز داده می باشد و از نظر امنیتی و روابط کسب و کار معمولاً در رده بسیار محرمانه تلقی می شود، اطلاعات دقیقی از میزان پهنای باند بین مراکز داده آمازون موجود نمی باشد و محاسبات متفاوتی در خصوص برآورد حدودی پهنای باند میان مراکز داده قابل اجرا می باشد [43]، [44]. در این مقاله ارتباط بین مراکز داده را به صورت یک توپولوژی حلقه در نظر گرفته ایم بدین معنا که هر مرکز داده به دو مرکز داده دیگر دارای یک ارتباط مستقیم نقطه به نقطه می باشد و پهنای باند بین

^{۳۳} congestion

^{۳۴} service request time



مراکز داده، چهار لینک ۱۰ گیگابیت بر ثانیه معادل ۴۰ گیگابیت بر ثانیه، فرض شده است که با توجه به فاصله تقریبی مراکز داده، میزان تاخیر لینک‌های ارتباطی محاسبه و پهنای باند مفید به دست آمده است.

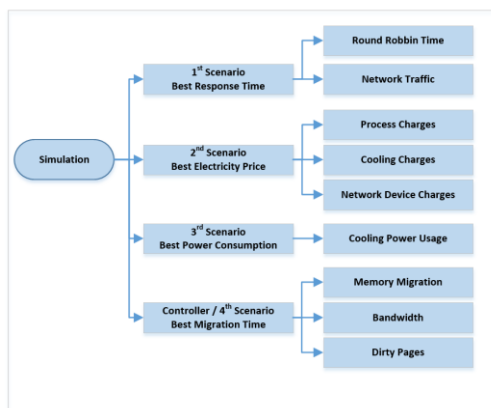
در شبیه‌سازی صورت پذیرفته چهار سناریو مختلف پیاده‌سازی شده است. در سناریوی نخست یا به عبارتی "بهترین زمان پاسخ"، زمان پاسخ تنها دغدغه سرویس‌دهندگان می‌باشد و هزینه‌های برق و یا مدت زمان انتقال به هیچ وجه مورد بررسی قرار نمی‌گیرند. در این صورت $I_1(P_i) = 0$ و $I_2(M_i) = 0$ می‌باشد در نتیجه وزن یال‌های گراف $w_i = f(T_i) = T_i$ می‌باشد. در سناریوی دوم یا به عبارتی "بهترین هزینه برق"، اولویت اصلی هزینه‌های انرژی می‌باشد، لذا ضریب هزینه‌های انرژی را ۱۰۰۰ برابر می‌کنیم تا وزن بیشتری نسبت به زمان داشته باشد. در این صورت زمان پاسخ و زمان انتقال بی‌ارزش می‌شوند، لذا خواهیم داشت $I_1(P_i) = 1000(P_i)$ و $I_2(M_i) = 0$ و در نهایت وزن یال‌های گراف به صورت رابطه (۱۶) محاسبه می‌گردد:

$$w_i = f(T_i, P_i) = T_i + 1000(P_i) \quad (16)$$

در سناریو سوم یا به عبارتی "بهترین مصرف انرژی"، میزان انرژی مصرف شده در مراکز داده مورد ارزیابی قرار می‌گیرد. لازم به ذکر است که همانگونه که در بخش ۳ اعلام شده، از آنجا که سیستم‌های سرمایه‌گذاری مراکز داده متفاوت می‌باشند، این احتمال وجود دارد که با صرف انرژی کمتری، هرچند گرانتر، اقدام به خنک‌سازی سرورها نماییم، لذا مصرف انرژی رابطه مستقیمی با هزینه انرژی نخواهد داشت. مدت زمان انتقال یک عامل کنترلی می‌باشد که به منظور قابلیت انتقال ماشین‌ها، در این مقاله گنجانده شده است. به عنوان مثال، سروری که میزبان سایت داندلود می‌باشد با چند ده ترابایت داده، هزینه و زمان بسیاری در انتقال، بر روی مرکز داده تحمیل می‌نماید. به منظور بررسی تاثیر این المان، در سناریوی چهارم، ضریب این عامل را به ۱۰۰۰ تغییر داده $I_1(P_i) = 0$ و $I_2(M_i) = 1000(M_i)$ و وزن یالها از رابطه (۱۷) محاسبه می‌گردد:

$$w_i = f(T_i, M_i) = T_i + 1000(M_i) \quad (17)$$

در تصویر ۶ این سناریوها و المانهای تاثیرگذار در پارامترهای محاسباتی آنها نمایش داده شده است. در هر کدام از سناریوهای معرفی شده مدت زمان انتقال، هزینه‌های انرژی و بار ترافیکی مرکز داده بعد از انتقال نیز جداگانه محاسبه شده و با یکدیگر مقایسه می‌شود.



تصویر ۶- سناریوهای مختلف شبیه‌سازی و عوامل تاثیرگذار در پارامترهای محاسباتی

تجمیع خروجی‌های حاصل از پیاده‌سازی این سناریوها، در جدول ۳ نمایش داده شده است. در کنار محاسبات صورت پذیرفته، بررسی هر یک از عناصر به صورت انفرادی نیز مورد مقایسه قرار گرفته است. بدین معنا که ضرایب سایر عناصر در محاسبات صفر شده و کارایی مرکز داده، با توجه به آن عنصر فعال محاسبه گردیده است.

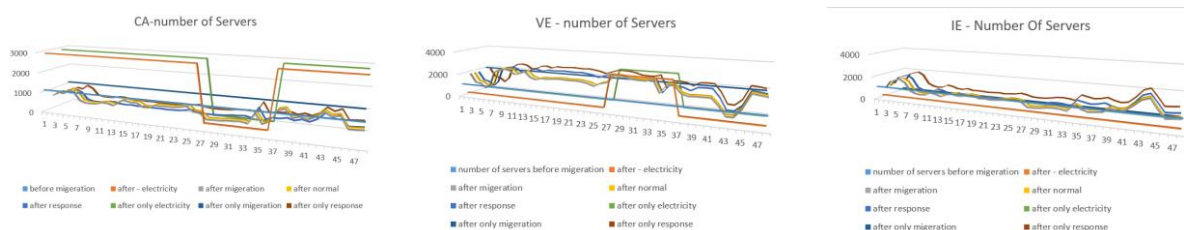
جدول ۳- نتایج به دست آمده از سناریوهای شبیه‌سازی شده

متوسط زمان پاسخ	متوسط هزینه برق	متوسط مصرف انرژی
-----------------	-----------------	------------------

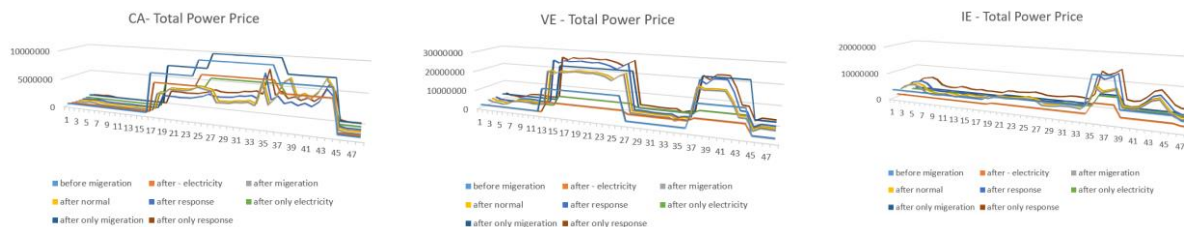


حالت اولیه	وات / ساعت	٪ صرفه جویی	دلار	٪ صرفه جویی	میلی ثانیه	٪ صرفه جویی
حالت اولیه	۷۱۲۸۶۲	۰	۳۳۱۲۵,۶	۰	۱۵۳	۰
سناریوی ۱ - بهترین زمان پاسخ	۵۶۲۸۷۵,۸۳	۲۱,۰۴	۱۵۵۲۹,۲	۵۳,۱۲	۷۹,۳۷	۴۸,۱۲
سناریوی ۲ - بهترین هزینه برق	۴۸۲۸۲۱	۳۲,۲۷	۱۲۰۸۰,۹	۶۳,۵۳	۱۳۶,۸۴	۱۰,۵۶
سناریوی ۳ - بهترین مصرف برق	۳۷۸۶۰۱	۴۶,۸۹	۱۳۳۳۹,۶۷	۵۹,۷۳	۱۲۵,۹۹	۱۷,۶۵
سناریوی ۴ - بهترین زمان انتقال	۵۷۳۴۲۶	۱۹,۵۶	۱۹۴۶۴,۶۰	۴۱,۲۴	۱۴۱,۲۹	۷,۶۵

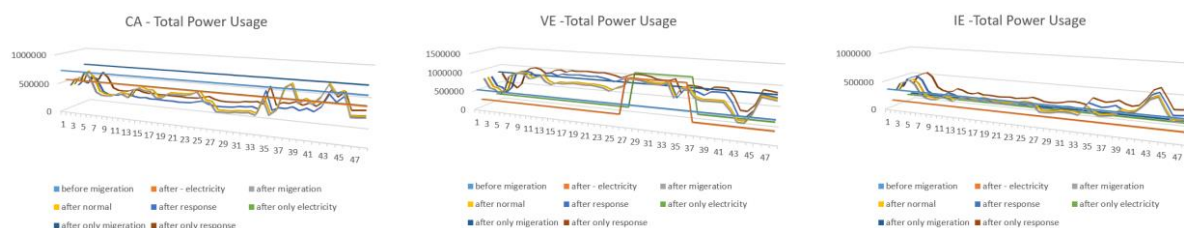
تعداد سرور های میزبانی شده فعال در هر مرکز داده، مطابق تصویر ۷ می باشد.



تصویر ۷- تعداد سرور های میزبانی شده در هر مرکز داده با توجه به سناریو های مختلف پیاده سازی شده. مقایسه هزینه های انرژی مصرفی و نیز میزان انرژی مصرف شده در مراکز داده نیز در تصویر ۸ و تصویر ۹ نمایش داده شده اند.



تصویر ۸- میزان هزینه های برق در مراکز داده با توجه به سناریو های مختلف شبیه سازی شده



تصویر ۹- میزان مصرف انرژی در مراکز داده با توجه به سناریو های مختلف شبیه سازی شده

در جدول ۴ توضیح طبقه بندی تصاویر ۷ الی ۹ قید شده است.

جدول ۴- توضیح بخش های مختلف تصاویر ۷ الی ۹

عنوان	توضیحات	عنوان	توضیحات
-------	---------	-------	---------



هزینه های برق دارای ضریب بیشتری می باشد	After Electricity	قبل از انجام محاسبات و تغییرات می باشد	Before migration
فقط هزینه برق مورد ارزیابی قرار گرفته	After Only Electricity	کلیه متغیر ها دارای ضریب یک می باشند	After normal
زمان پاسخ دارای ضریب بیشتری می باشد	After Response	هزینه جابه جایی دارای ضریب بالاتری می باشد	After Migration
فقط هزینه های زمان پاسخ مورد ارزیابی قرار گرفته	After Only Response	فقط هزینه های انتقال مورد ارزیابی قرار گرفته	After only migration

مقایسه روش سیروکومولوس و استراتوس [20] به شرح جدول ۵ اعلام می گردد. لازم به ذکر است که در استراتوس، میزان انتشار کربن در محاسبات قید شده، که رابطه مستقیم با میزان مصرف انرژی دارد.

جدول ۵- مقایسه الگوریتم های استراتوس و سیروکومولوس

متوسط زمان پاسخ میلی ثانیه	متوسط هزینه برق دلار	متوسط مصرف انرژی کیلو وات	
٪۴۷	٪۶۱	٪۲۱	استراتوس
٪۴۸,۱۲	٪۶۳,۵۳	٪۲۱,۰۴	سیروکومولوس
٪۱,۱۲	٪۲,۵۳	٪۰,۰۴	درصد بهبود



مراجع:

- [1] Anupama Sahu, Suwendu Kumar Jena, "Efficient load Balancing in Cloud Computing using Fuzzy Logic," *ISSN: 2250-3021 Volume 2, Issue 7*, pp. 65-71, July 2012.
- [2] Lena Mashayekhy, Mahyar Movahed Nejad, Daniel Grosu, "Physical Machine Resource Management in Clouds: A Mechanism Design Approach," *IEEE Transactions on Cloud Computing*, vol. 03, no. 03, pp. 247-260, July 2015.
- [3] Bob Briscoe, Anna Brunstrom, Andreas Petlund, David Hayes, David Ros, Ing-Jyh Tsang, Stein Gjessing, Gorrry Fairhurst, Carsten Griwodz and Michael Welzl, "Reducing Internet Latency: A Survey of Techniques and their Merit," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2149 - 2196, thirdquarter 2016.
- [4] A. Qureshi, J. Guttag, R. Weber, B. Maggs, and H. Balakrishna, "Cutting the electric bill for internet-scale systems," *Proceedings of ACM SIGCOMM, Barcelona*, vol. 17, no. 21, p. 123-134, August 2009.
- [5] R. Stanojević and R. Shorten, "Distributed dynamic speed scaling," *IEEE INFOCOM, San Diego*, vol. 14, no. 19, pp. 1-5, March 2010.
- [6] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, "Greening geographical load balancing," *SIGMETRICS, San Jose*, pp. 233-244, June 2011.
- [7] M. Conti, E. Gregori, and F. Panzieri, "Load distribution among replicated Web servers: a QoS-based approach," *SIGMETRICS Performance Evaluation Review*, vol. 27, no. 4, pp. 12-19, 2000.
- [8] Z. M. Mao, C. D. Cranor, F. Bouglis, M. Rabinovich, O. Spatscheck, and J. Wang, "A precise and Efficient Evaluation of the Proximity between Web Clients and their Local DNS Servers," *USENIX, Monterey*, vol. 10, no. 15, pp. 229-242, June 2002.
- [9] M. Pathan, C. Vecchiola, and R. Buyya, "Load and Proximity Aware Request-Redirection for Dynamic Load Distribution in Peering CDNs," *On the Move to Meaningful Internet Systems: OTM, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg*, vol. 5331, pp. 62-81, 2008.
- [10] S. Irani and K. R. Pruhs, "Algorithmic problems in power management," *SIGACT News*, vol. 36, no. 2, p. 63-76, 2005.
- [11] A. Beloglazov, R. Buyya, Y. C. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," Technical Report, 2010.
- [12] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg and Ivona 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, p. 599-616, June 2009.
- [13] J. Dong, X. Jin, H. Wang, Y. Li, P. Zhang, S. Cheng, , "Energy-Saving Virtual Machine Placement in Cloud Data Centers," *proceeding of 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, vol. 13, pp. 618 - 624, 2013.
- [14] Zhenhua Liu, Minghong Lin and Adam Wierman, "Greening Geographical Load Balancing," *IEEE/ACM Transactions on Networking*, vol. 23, no. 2, pp. 657 - 671, April 2015.
- [15] V. Mathew, R. K. Sitaraman, and P. Shenoy, "Energy-aware load balancing in content delivery networks," *EEE INFOCOM, Orlando*, vol. 25, p. 954-962, 30 March 2012.
- [16] P. Mahadven, S. Banerjee and P. Sharma, "Energy Proportionality of an Enterprise Network,"



Green Networking '10 Proceedings of the first ACM SIGCOMM workshop on Green networking, pp. 53-60, 2010.

- [17] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, "Green cloud computing: Balancing energy in processing, storage, and transport," *IEEE*, vol. 99, no. 1, p. 149–167, 2011.
- [18] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, "Geographical load balancing with renewables," *GreenMETRICS, San Jose*, vol. 7, no. 11, pp. 1-5, 2011.
- [19] F.F.Moghaddam, M. Cheriet, and K.K.Nguyen, "LowCarbon Virtual Private Clouds," in *Proceedings of IEEE International Conference on Cloud Computing, Washington DC*, vol. 4, p. 259–266, July 2011.
- [20] J. Doyle, R. shorten and D.O'mahony, "stratus: load balancing the cloud for carbon emissions control.," *Cloud Computing, IEEE Transactions*, vol. 1, no. 1, Jan 2013.
- [21] J. W. Durham, R. Carli, P. Frasca, and F. Bullo, "Discrete Partitioning and Coverage Control for Gossiping Robots," *EEE Transactions on Robots*, vol. 28, no. 2, p. 364–378, 2012.
- [22] R. J. Fowler, M. S. Paterson, and S. L. Tanimoto, "Optimal packing and covering in the plane are NP-complete," *Information processing letters*, vol. 12, p. 133–137, 1981.
- [23] R. Z. Hwang, R. C. T. Lee, and R. C. Chang, "The slab dividing approach to solve the Euclidean p-center problem," *Algorithmica*.
- [24] C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live migration of virtual machines," *2nd conference on Symposium on Networked Systems Design & Implementation*, vol. 2, no. 5, pp. 273-286, May 2005.
- [25] T. Imada, M. Sato and H. Kimura., "Power and QoS performance characteristics of virtualized servers," *10th IEEE/ACM International Conference on Grid Computing, Banff, Alberta, Canada*, pp. 232-240, 13-15 Oct. 2009.
- [26] A.-C. Orgerie, L. Lefevre and J.-P. Gelas, "Demystifying energy consumption in Grids and Clouds," *International Green Computing Conference, Chicago, IL, USA*, pp. 335-342, 15-18 Aug. 2010.
- [27] A. Verma, G. Kumar and R. Koller, "The cost of reconfiguration in a cloud," *The 11th International Middleware Conference Industrial track (Middleware Industrial Track '10). Bangalore, India*, pp. 11-16, December, 2010.
- [28] Y. Kuno, K. Nii and S. Yamaguchi., "A Study on Performance of Processes in Migrating Virtual Machines," *10th International Symposium on Autonomous Decentralized Systems (ISADS)*, pp. 567-572, 23-27 March 2011.
- [29] A. Strunk, "Costs of Virtual Machine Live Migration: A Survey," *2012 IEEE Eighth World Congress on Services*, 2012.
- [30] S. Akoush, R. Sohan, A. Rice, A.W. Moore and A. Hopper, "Predicting the Performance of Virtual Machine Migration," *The 18th Annual Meeting of the IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2010), Miami Beach, Florida*, pp. 37-46, 17-19 Aug. 2010.
- [31] F. Salfner, P. Tröger and A. Polze, "Downtime Analysis of Virtual Machine Live Migration," *The Fourth International Conference on Dependability (DEPEND 2011), French Riviera, France*, pp. 100-105, August 21-27, 2011.
- [32] S. Srikantaiah, A. Kansal and F. Zhao., "Energy Aware Consolidation for Cloud Computing.," *Proceedings of Workshop on Power Aware Computing and Systems at OSD (USENIX HotPower'08), San Diego, California, USA*, 7 December, 2008.
- [33] A.Jain , S.Taneja, "Money saving advantage Analysis of Datacenter Consolidation Using Global Virtual Datacenter Online Calculator and VMware ROI TCO Calculato," *International Journal of Progressive Sciences and Technologies (IJPSAT)*, vol. 3, no. 1, pp. 26-35, June 2016.



- [34] L. A. Barroso and U. Holzle, "The datacenter as a computer: An introduction to the design of warehouse-scale machines," Synthesis Lectures on Computer Architecture, 2009. [Online].
- [35] "Ethernet Power Consumption for Cisco®Switches," Cisco Systems, 2008. [Online]. Available: http://www.atrac.com/documents/Cisco/Enterprise/brochures/Cisco_BROCHURE_Ethernet%20Power%20Consumption%20for%20Switches_2.14.12.pdf. [Accessed 12 1 2016].
- [36] "Amazon Webservice Network test," 2016. [Online]. Available: <https://cloudharmony.com/speedtest-for-aws>. [Accessed 12 March 2016].
- [37] "Ping Maps," Oxford, 2016. [Online]. Available: <http://folk.uio.no/sigurdkn/pingmap/>. [Accessed 12 march 2016].
- [38] "Global Ping Statistics," 2016. [Online]. Available: <https://wondernetwork.com/pings>. [Accessed 12 March 2016].
- [39] ISPreview.co.uk, "European v USA Broadband Internet Usage and Surfing Trends," 19 August 2012. [Online]. Available: <http://www.ispreview.co.uk/story/2009/08/19/european-v-usa-broadband-internet-usage-and-surfing-trends.html>. [Accessed 2 may 2016].
- [40] N. M. S. University, "NMSU internet Usage," 2015. [Online]. Available: http://noc.nmsu.edu/internet_usage.cgi. [Accessed 3 may 2016].
- [41] Statista, "Amazon's audience reach in selected global markets as of 2nd quarter 2015," 2016. [Online]. Available: <http://www.statista.com/statistics/274711/worldwide-amazon-audience-composition/>. [Accessed 16 may 2016].
- [42] "Traffic share by country," 8 may 2016. [Online]. Available: https://pro.similarweb.com/#/website/audience-geography/amazon.com/*/999/3m. [Accessed 8 may 2016].
- [43] S. Anthony, "How big is the cloud?," extremetech, 23 may 2012. [Online]. Available: How big is the cloud?. [Accessed 12 march 2016].
- [44] "AWS Global Infrastructure," Amazon Co, 2016. [Online]. Available: <https://aws.amazon.com/about-aws/global-infrastructure/>. [Accessed 12 march 2016].