روش نوین توزیع بار ترافیک مبتنی بر BGP در مراکز داده توزیع شده جغرافیایی

امیر اسمعیلی، بهادر بخشی سراسکانرود

چکیده: **مسائل مربوط به مراکز داده به عنوان زیرساخت ارایه سرویس‌های فناوری اطلاعات همواره حائز اهمیت بوده است. از سوی دیگر، در سالهای اخیر، بنابه دلایل متعددی از جمله تفاوت هزینه انرژی در نقاط مختلف، امنیت بالاتر و تامین کیفیت سرویس، مراکز داده توزیع شده جغرافیایی مورد توجه قرار گرفته است که در آن مرکز داده متشکل از چندین خوشه در نقاط مختلف است. مساله اصلی در این نوع مراکز داده، توزیع کارای بار ترافیکی بین خوشه‌ها است. با وجود اینکه روشهای متعددی برای این منظور پیشنهاد شده است ولی به دلیل متمرکز بودن توزیع کننده، سرعت پایین واکنش به تغییرات بار ترافیکی و یا محدود بودن به یک تکنولوژی خاص، راه حل کارا، عملی و جامع این مساله به شمار نمی‌روند. در این مقاله معماری جدیدی پیشنهاد شده که در آن، برای توزیع بار ترافیکی، مفهوم مسیریابی anycast با تنظیمات مناسب BGP و در نظر گرفتن ملاحظات عملی پیاده‌سازی می‌شود. نتایج شبیه‌سازی بهبود قابل توجه این روش نسبت به روشهای موجود را نشان می‌دهد.**

کلید واژه: **توزیع بار ترافیکی، مراکز داده توزیع شده، BGP، anycast، شبکه‌های نرم‌افزار محور**

۱- مقدمه

امروزه با گسترش روز افزون محاسبات ابری و سرویس‌های مبتنی بر آن، دسترس‌پذیری، امنیت و کارایی بالای این سرویس‌ها از درجه اهمیت بسیار بالایی برخوردار است. برای تامین آنها از منابع زیرساخت محاسباتی و ذخیره‌سازی مراکز داده استفاده می‌شود که به دو دسته کلی متمرکز و توزیع شده تقسیم می‌شوند. با توجه به اینکه مراکز داده توزیع شده جغرافیایی از چندین خوشه در نقاط مختلف جهان تشکیل شده‌اند[۱]، علاوه بر توزیع بار بین سرورهای یک خوشه (که پیش از این روش‌های متعددی برای آن پیشنهاد شده است) می‌بایست روش موثری برای توزیع بار بین خوشه‌ها وجود داشته تا همه خوشه‌ها به صورت موثر قادر به ارایه سرویس باشند، این جنبه از توزیع بار چندان مورد توجه قرار نگرفته است.

مساله توزیع بار بین سرورها (در یک مرکز داده متمرکز)، عموما از جنس اختصاص منابع در سطح سیستم عامل به ماشین‌های مجازی به منظور بهینه‌سازی اهدافی مانند نرخ پذیرش یا مصرف انرژی [۱] است. توزیع بار در این سطح هدف این مقاله نیست.

در مراکز داده توزیع شده جغرافیایی، مساله توزیع بار به تقسیم بار ترافیکی در سطح کاربران و درخواست‌ها بین خوشه‌ها می‌پردازد که مساله مد نظر در این مقاله نیز است. کارهای موجود در این حوزه سه نوع روش تقسیم بار را پیاده‌سازی نموده‌اند که عبارتند از تقسیم بار متمرکز[۲] [۲]، سیستم‌های مبتنی بر سرویس دهنده DNS[3] و روش‌های مبتنی بر مسیریابی. هر یک از این روش‌ها دارای معایب خاص خود می‌باشد. در روش‌های متمرکز، وجود یک نقطه تقسیم کننده ترافیک علاوه بر ایجاد تک نقطه اشکال[۴]، مفهوم مرکز داده توزیع شده جغرافیایی را زیر سوال می‌برد. در روش مبتنی بر DNS نیز سرعت کم انتشار رکوردها و دقت کم این روش بازدهی و سرعت واکنش به تغییرات بار را با کاهش روبرو می‌نماید. با وجود اینکه روش‌های مبتنی بر مسیریابی با استفاده از آدرس‌های ویرایش ششم anycast دارای سرعت عملکرد بالایی هستند ولی مشکلات این روش‌ها عدم پایداری ارتباطات[۵] در هنگام تغییر خوشه، قابل استفاده نبودن مسیریابی داخلی[۶] در سطح مسیریاب‌های اینترنتی و همچنین محدودیت‌های آدرس دهی anycast می‌باشد.

در این مقاله معماری جدیدی جهت توزیع بار ترافیکی در مراکز داده توزیع شده جغرافیایی پیشنهاد شده است. در این روش با استفاده از مفهوم مسیریابی anycast، پروتکل BGP و شبکه‌های نرم‌افزار محور، یک معماری ترکیبی ارائه شده است. مشکلات مربوط به پایداری ارتباطات در بخش نرم‌افزار محور شبکه و با استفاده از کنترلر مرکزی حل شده است. استفاده از BGP باعث انعطاف‌پذیری بیشتر شده و قابلیت پیاده‌سازی عملی معماری پیشنهادی در اینترنت را

امیر اسمعیلی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران (email: aesmaeili@aut.ac.ir)

بهادر بخشی سراسکانرود، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران (email: bbakhshi@aut.ac.ir)

[۱] در هریک از این خوشه‌ها چندین سرویس دهنده وجود دارد.

[2] Central Load Balancer
[3] Domain Name System
[4] Single Point of Failure (SPF)
[5] Connection Establishment
[6] Interior Gateway Routing Protocols (IGP)

فراهم می‌کند؛ همچنین سرعت انتشار تغییرات در این روش از سایر روش‌های موجود بیشتر می‌باشد.

ساختار بخش‌های آتی این مقاله به این شرح است: در بخش ۲ روش‌های موجود در زمینه توزیع بار در مراکز داده مورد بررسی قرار خواهد گرفت. در بخش ۳ معماری پیشنهادی ارائه و توضیح داده خواهد شد. شبیه‌سازی، آنالیز و مقایسه روش پیشنهادی با سایر روش‌ها در بخش ۴ انجام می‌گیرد. بخش ۵ نیز به نتیجه‌گیری و ارائه پیشنهادات در زمینه کارهای آتی اختصاص دارد.

۲- بررسی روش‌های موجود

مراکز داده توزیع شده جغرافیایی بر اساس استاندارد TIA/942 در بالاترین سطح درجه اطمینان قرار دارند [۳]، بنابراین مساله توزیع بار در آنها بسیار حائز اهمیت است. همانگونه که ذکر شد، هدف این مقاله مساله توزیع بار ترافیکی بین خوشه‌های مرکز داده توزیع شده است، بنابراین در ادامه مقالات مرتبط در این خصوص بررسی می‌گردند.

۲-۲ توزیع ترافیک ورودی مرکز داده

همانطوری که اشاره شد، روش‌های توزیع بار ترافیک کاربران بین مراکز داده شامل سه روش متمرکز، مبتنی بر DNS و مسیریابی است که در ادامه بررسی شده است.

۲-۲-۱ توزیع کننده متمرکز بار ترافیکی

بیشترین مقالات در زمینه توزیع بار مراکز داده مربوط به توزیع کننده‌های متمرکز می‌باشد. این توزیع کننده‌ها که بیشتر در شبکه‌های نرم‌افزار محور مورد استفاده قرار می‌گیرند بر اساس مدل‌های واکنشی، به توزیع بار در مرکز داده می‌پردازند. در مقاله [۴] با استفاده از شبکه‌های نرم‌افزار محور [۵] یک راه حل متمرکز ارایه شده است که در آن با استفاده از یک کنترلر متمرکز NOX [۶] ترافیک بین سرورهای مختلف تقسیم می‌گردد. این معماری قابلیت پیاده‌سازی الگوریتم‌های واکنشی (خلوت ترین سرور[1] و کمترین زمان پاسخ[2]) را با استفاده از اطلاعاتی که از سرورهای شبکه دریافت می‌شود دارد. یکی از مهمترین الگوریتم‌های توزیع بار متمرکز، الگوریتم Round Robin وزن‌دار است که در آن ترافیک تخصیصی به خوشه iام که با C_i نشان داده می‌شود، طبق رابطه زیر محاسبه می‌گردد

$$C_i = \frac{L_{in}}{\sum_{i=1}^{N} w_i} w_i$$

در این رابطه w_i وزن خوشه iام و L_{in} کل ترافیک ورودی به مرکز داده می‌باشد.

یکی از مشکلات اصلی توزیع کننده‌های متمرکز ایجاد تک نقطه اشکال[3] در شبکه است که برای حل این موضوع می‌بایست دستگاه به صورت افزونه دیده شود. علاوه بر این محل قرارگیری توزیع کننده مرکزی نیز مشکل دیگر این معماری محسوب می‌شود که مفهوم مرکز داده توزیع شده را زیر سوال خواهد برد. سایر مقالاتی که در زمینه توزیع کننده‌های متمرکز می‌باشند مانند [۷] و [۸] نیز این مشکلات را دارند.

۲-۲-۲ تقسیم بار ترافیکی با استفاده از DNS

دسته دوم مقالات توزیع بار ترافیک ورودی، روش‌های مبتنی بر سرویس دهنده DNS می‌باشند. این دسته از پژوهش‌ها از سرویس دهنده نام در شبکه برای تقسیم بار استفاده می‌نمایند، به این صورت که تمام خوشه‌ها دارای اسم یکسان ولی آدرس مجزا و منحصربفردی هستند. ترافیک کاربران در مرحله ترجمه اسم به آدرس، به آدرس‌های متفاوتی نگاشته می‌شوند که موجب توزیع بار می‌گردد. اکثر الگوریتم‌های مبتنی بر سرویس دهنده نام، به صورت واکنشی و با مانیتور نمودن وضعیت کنونی سرورها اقدام به توزیع بار می‌نمایند. مهمترین مقالات در این زمینه، راه حل شرکت akami [۹] و همچنین راه حل شرکت گوگل برای مدیریت مراکز داده خود می‌باشد [۱۲]. شرکت akami بیش از ۱۲۰۰۰ سرور خود که در نقاط مختلف جهان پراکنده می‌باشند را از طریق توزیع بار مبتنی بر سرویس دهنده نام مدیریت می‌نماید.

اصول کار سرویس دهنده‌های نام در شبکه، بر مبنای یک مدل سلسله مراتبی بوده که شامل زنجیره‌ای است که از ۱۳ سرورهای ریشه اینترنت[4] شروع شده و به کاربر اینترنتی ختم می‌شود. در این فرآیند هر گره نام سرویس دهنده را به سرور بالادستی داده و آدرس آن را دریافت می‌نماید، این آدرس تا مدت زمان خاصی معتبر می‌باشد و پس از آن نیاز به سوال مجدد می‌باشد. در راه حل شرکت akami، درخواست کاربران به سرور مدیریت نام شبکه مرکز داده ارسال و برحسب نوع درخواست، محل کاربر، ترافیک هر یک از سرورها و ... نام یک خوشه بعنوان خوشه سرویس دهنده به کاربر اعلام می‌شود. نقطه ضعف اصلی این روش عدم کنترل بر روی رکوردهای ارسالی DNS پس از اعلام به کاربر می‌باشد. با توجه به ماهیت سلسله مراتبی، بسته‌های ارسالی به شبکه احتمالا در سرورهای نام پایین دستی تا زمان پایان اعتبار آن‌ها نگهداری شده و به سرور اصلی مراجعه انجام نخواهد شد. این موضوع کنترل مدیر شبکه به کاربران ورودی را با چالش روبرو خواهد نمود چرا که در صورت تغییر حجم بار خوشه‌ها تا قبل از اتمام زمان اعتبار، امکان تغییر مسیر درخواست کاربران وجود ندارد.

---

[1] Least Connection
[2] Lowest Response Time
[3] Single Point of Failure
[4] Root Hints

۳-۲-۲ تقسیم بار ترافیکی بر اساس مسیریابی

در روش‌های مبتنی بر مسیریابی، الگوریتم‌های لایه شبکه مسیر ترافیک کاربران به طرف خوشه‌های مرکز داده را تغییر می‌دهند. در اکثر مقالاتی که با استفاده از این روش به توزیع بار در سطح خوشه‌های مرکز داده پرداخته اند از آدرس‌های ویرایش ششم IP و دامنه آدرس‌دهی anycast استفاده شده است [۱۱]. مفهوم anycast در مقابل مفاهیم همه پخشی[1] و چند پخشی[2] ارائه شده و به معنی ارسال به یکی از اعضای گروه (معمولا نزدیکترین) می‌باشد. روش‌هایی مانند [۱۳] و [۱۴] که با استفاده از این دامنه آدرس‌دهی و مسیریابی داخلی به توزیع بار در شبکه پرداخته‌اند از سرعت واکنش بالاتری به نسبت سایر روش‌ها برخوردار می‌باشد و نتایج شبیه‌سازی آن‌ها افزایش ۱۰ درصدی در بازدهی شبکه از لحاظ تعداد ارتباطات همزمان را نشان می‌دهد.

در این معماری وضعیت کنونی شبکه مانیتور شده و بر این اساس، ترافیک ورودی کاربران در لایه مسیریابی به خوشه خلوت‌تر ارسال می‌شود. یکی از مشکلات موجود در این زمینه، حفظ نشست‌های برقرار شده فعلی با خوشه سرویس دهنده می‌باشد؛ در صورتی که کاربران قدیمی یک خوشه به دلیل ازدحام به خوشه دیگری مسیریابی شوند، می‌بایست این ترافیک در خوشه جدید شناسایی و به خوشه اصلی بازگردانده شود. در [۱۳] تغییرات در لایه کاربرد برای حل این مشکل پیشنهاد شده است و با قرار دادن آدرس سرویس دهنده اصلی ترافیک در لایه کاربرد تشخیص داده می‌شود. در [۱۴] علاوه بر این روش، قراردادن آدرس سرویس دهنده اصلی در لایه شبکه و بخش Source Route Option پیشنهاد شده است. با وجود اینکه این روش‌ها برای حل مشکل پایداری امکان‌پذیر است ولی محدودیت‌های خاصی را نیز دارا می‌باشد. اول اینکه در لایه کاربرد احتمالا در همه شرایط مدیر شبکه قادر به ایجاد تغییرات نمی‌باشد و دوم اینکه به دلیل اینکه Source Route توسط گره‌های شبکه قابل تغییر می‌باشد، در این روش آسیب پذیری حمله مردی در میان[3] وجود دارد و امروزه به‌صورت پیش فرض در مسیریاب‌های اینترنتی اعمال نمی‌شود. نکته بعدی محدودیت استفاده از دامنه آدرس دهی anycast می‌باشد که مدیران مرکز داده را مجبور به ثبت و استفاده از این دامنه آدرس‌ها می‌نماید.

مساله مهم دیگر در خصوص روش‌های مبتنی بر مسیریابی، نوع الگوریتم مسیریابی مورد استفاده می‌باشد. بطور کلی الگوریتم‌های مسیریابی داخلی دارای محدودیت‌هایی چون تعداد مسیریاب‌های موجود در شبکه، نیاز به دسترسی به کلیه مسیریاب‌های موجود در شبکه برای تغییر معیارهای مسیریابی و مسیریابی نقطه به نقطه می‌باشند که این امر با ماهیت کلان اینترنت در تناقض بوده و تنها قابل پیاده‌سازی در محیط‌های آزمایشگاهی و کوچک می‌باشد. از این رو برای یک مرکز داده که دارای خوشه‌های مختلفی در سراسر جهان می‌باشد نیاز به استفاده از الگوریتم مسیریابی خارجی سطح کلان می‌باشد که در کارهای پیشین به آن توجه نشده است.

با توجه به بررسی‌های انجام شده در این بخش، روش‌های موجود یا به صورت متمرکز عمل میکنند که با مفهوم مراکز داده توزیع شده در تناقض است یا از DNS استفاده می‌کنند که کنترل دقیق و سریع بر روی ترافیک ندارد و یا از تکنولوژی‌هایی مانند مسیریابی داخلی و دامنه آدرس خاص استفاده می‌کنند که به صورت عملی قابل پیاده‌سازی در سطح جهانی نیست. بنابراین راهکار عملی برای حل این موضوع به عنوان یک مساله باز همچنان مطرح است.

۳- معماری روش پیشنهادی

در این بخش به ارائه معماری روش پیشنهادی پرداخته خواهد شد. برای این ابتدا نیازمندی‌های مد نظر بیان شده و اجزای معماری پوشش دهنده این موارد شرح داده می‌شود.

۳-۱ نیازمندهای معماری توزیع بار ترافیکی

اولین مساله، پروتکل مسیریابی است. این پروتکل می‌بایست بدون نیاز به دسترسی به همه مسیریاب‌های شبکه و تنها با تغییر معیارهای مسیریابی در دروازه مرکز داده، مسیر بسته‌های ورودی را تغییر دهد. به عبارت دیگر، برای این منظور باید از پروتکل‌های مسیریابی خارجی[4] که ترافیک را در سطح کلان و شبکه به شبکه کنترل می‌کنند استفاده گردد. مساله دوم، نحوه آدرس‌دهی خوشه‌های مراکز داده است. در کارهای پیشین [۱۳] و [۱۴] از دامنه آدرس‌های anycast ویرایش ششم برای مسیریابی استفاده شده است در حالی که در اکثر شبکه‌ها همچنان از IPv4 استفاده می‌شود. بنابراین لازم است تا از مفهوم anycast استفاده شود ولی پیاده‌سازی آن محدود به یک تکنولوژی خاص نباشد. سومین نیازمندی حل مشکل پایداری ارتباطات در هنگام تغییر مسیریابی از یک خوشه به خوشه دیگر می‌باشد. اگرچه در کارهای پیشین [۱۳] و [۱۴] از روش‌های مختلف در لایه کاربرد و شبکه برای حل این مشکل استفاده شده است، اما هریک از این روش‌ها بنا به دلایل محدودیت‌های لایه نرم‌افزار و موارد امنیتی دارای مشکلاتی می‌باشند که پیاده‌سازی را با چالش مواجه می‌نماید. جهت حل این مشکل لازم است کاربرانی که هنگام بروز ازدحام به خوشه اشتباهی مسیریابی می‌شوند تشخیص داده شده و به خوشه درست هدایت شوند.

۳-۲ اصول طراحی معماری

در معماری پیشنهادی برای مرتفع نمودن این نیازمندی‌ها از یک معماری ترکیبی از شبکه‌های نرم‌افزار محور و شبکه‌های سنتی

[1] Broadcast
[2] Multicast
[3] Man In the Middle attack (MIME)
[4] Exterior Routing Protocol (EGP)

استفاده شده است. برای حل مشکل مسیریابی از پروتکل مسیریابی خارجی BGP استفاده شده است. در این پروتکل، برای هر شبکه یک هویت مستقل در نظر گرفته شده است که AS[1] نامیده می‌شود. الگوریتم انتخاب بهترین مسیر در این پروتکل از ۱۲ مرحله تشکیل شده است که ۸ مرحله اصلی آن در جدول ۱ آمده است. مسیرها به ترتیب مطرح شده در این جدول مقایسه شده و با پیدا شدن اولین مسیر بهتر، جستجو خاتمه می‌یابد.

جدول ۱: فرآیند انتخاب بهترین مسیر در BGP [۱۵]

Seq.	Name	Preference
1	Weight	High
2	Local Preference	High
3	Locally Generated	True
4	AS-Path	Lower
5	Origin	IGP>EGP>Unknown
6	MED	Lower
7	External	eBGP>iBGP
8	IGP Cost	Lower

در معماری پیشنهادی از مرحله چهارم انتخاب بهترین مسیر استفاده شده است. بطور کلی AS-Path یک دنباله از شماره ASهای عبور دهنده ترافیک می‌باشد که هر مسیریاب نام AS خود را به دنباله اضافه می‌نماید. در فرآیند انتخاب بهترین مسیر، مسیری که کوتاهترین AS-Path را دارد به عنوان مسیر بهتر انتخاب می‌شود. یکی از تکنیک‌های تغییر مسیریابی در BGP اعمال تغییر در تعداد ASهای اضافه شده در دنباله توسط یک مسیریاب می‌باشد[2]. در معماری پیشنهادی نیز کلیه خوشه‌های مرکز داده با یک آدرس و AS-Pathها مختلف به شبکه اعلام شده و با تغییر در AS-Path ها ترافیک به خوشه‌های مختلف باز مسیریابی می‌شود.

مشکل استفاده از دامنه آدرس‌های anycast نیز به این ترتیب حل خواهد شد؛ با توجه به اینکه در استاندارد BGP هر یک از آدرس‌های اعلام شده توسط مسیریاب‌های مرزی یک AS می‌توانند با AS-Pathهای متفاوت در شبکه جهانی اعلام شوند و مسیریاب‌های مختلف با توجه به کوتاهترین دنباله ASها مسیر بهینه را انتخاب خواهند نمود (این ویژگی به خاصیت Multi-homing معروف می‌باشد و به منظور انتقال ترافیک بین شبکه‌های مختلف از مسیرهای متفاوت از آن استفاده می‌شود). با معماری پیشنهادی که از چند دروازه مختلف و یک هویت AS تشکیل شده است مفهوم anycast بدون استفاده از دامنه آدرس‌های مورد نظر پیاده‌سازی شده است. نوآوری آخر کار حل مشکل پایدار نگه داشتن ارتباطات کاربران در هنگام بازمسیریابی اشتباه است. برای حل این مشکل در معماری پیشنهادی از یک بخش نرم‌افزار محور استفاده شده است. در این بخش از شبکه سرورهای مربوط به هر خوشه مرکز داده به سوییچ‌های پشتیبانی کننده Open Flow 1.4 متصل بوده و توسط یک کنترلر مرکزی [۱۶] و [۱۷] که همه سوییچ‌های خوشه‌های مختلف به آن متصل می‌باشند مدیریت می‌شوند[3].

ذکر این نکته حائز اهمیت است که روش پیشنهادی، علاوه بر پوشش مشکلات روش‌های قبل، قادر به تغییر مسیر ترافیک حملات توزیع شده عدم سرویس[4] از یک خوشه شلوغ به سایر خوشه‌های خلوت می‌باشد که این امر به نوبه خود باعث آرام‌سازی[5] حملات (با توجه به هدف حمله کنندگان در این روش) و انتقال ترافیک حمله از یک خوشه به کلیه خوشه‌ها می‌شود.

۳-۳ جزییات طراحی معماری پیشنهادی

همانگونه که ذکر شد، معماری پیشنهادی، یک معماری ترکیبی از شبکه‌های سنتی و نرم‌افزار محور است. برای توصیف معماری و جزییات الگوریتم از نمادهای جدول ۲ استفاده می‌شود.

جدول ۲: نمادهای مورد استفاده در معماری پیشنهادی

ردیف	پارامتر	نماد
۱	تعداد خوشه ها	α
۲	AS اینترنت	β
۳	AS مرکز داده	μ
۴	تعداد سرورهای خوشه	S
۵	حد ازدحام	C_{exc}
۶	حد بحران	$C_{\max}$
۷	حد خلوتی خوشه	C_{min}
۸	ظرفیت خوشه	C

۳-۳-۱ بخش سنتی مرکز داده

بخش سنتی مرکز داده از تعدادی دروازه اینترنتی تشکیل شده است که تحت هویت μ با شبکه جهانی که AS آن β نامیده می‌شود، در هر خوشه همسایگی BGP دارند. مطابق شکل ۱، دروازه‌های هر خوشه با ارسال دنباله AS-Path به همسایه‌های خارجی خود ترافیک ورودی به خوشه را کنترل می‌نمایند. در بخش سنتی مرکز داده نیز ارتباط میان سوییچ‌های بخش نرم‌افزار محور با دروازه‌ها از طریق لینک‌های با پهنای‌باند بالا انجام می‌شود.

۳-۳-۲ بخش نرم‌افزار محور مرکز داده

در بخش نرم‌افزار محور معماری پیشنهادی، در هر خوشه از یک سوییچ پشتیبانی کننده OF v1.4 استفاده شده است که سرورهای خوشه به این سوییچ متصل می‌باشند.

---

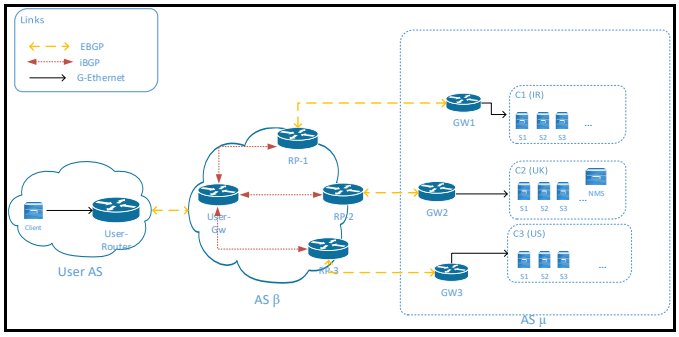
[1] Autonomous System (AS)

[2] As-Path Prepending

[3] با توجه به ماهیت توزیع شده خوشه‌های مرکز داده در معماری پیشنهادی می‌توان از کنترلرهای توزیع شده [۱۸]- [۲۰] استفاده نمود.

[4] Distributed Denial Of Service (DDOS)

[5] Mitigation



شکل ۱: بخش سنتی مرکز داده با α خوشه

بین سوییچ‌های خوشه‌های مختلف یک ارتباط حلقه وجود دارد تا در زمان نیاز به اصلاح مسیریابی کاربران به خوشه اصلی، از این لینک‌ها استفاده شود. در هر خوشه از یک کنترلر SDN استفاده شده است که وظیفه مدیریت سوییچ‌های هر خوشه را بر عهده دارند. هنگام مسیریابی اشتباه یک کاربر قدیمی (به دلیل تغییر مسیریابی قبل از اتمام نشست آن) کنترلر با تشخیص این موضوع از طریق لینک‌های بین خوشه‌ها کاربر را به خوشه درست بازمسیریابی می‌نماید.

۳-۳- بخش مدیریت ترافیک مرکز داده

در معماری پیشنهادی، با توجه به ماهیت واکنشی الگوریتم توزیع بار، از یک سامانه مانیتورینگ[1] برای بررسی وضعیت شبکه استفاده شده است. این سامانه از طریق استانداردهای WMI و SNMP به جمع‌آوری اطلاعات مربوط به وضعیت کنونی سرورها می‌پردازد، از طرفی این سامانه در لایه کاربرد شبکه نرم‌افزار محور [۵] با کنترلر شبکه در ارتباط بوده تا در صورت نیاز در کنترلر دستورات لازم را اعمال نماید. از طرفی برای مدیریت ترافیک ورودی کاربران در شبکه نیاز است تا سامانه با دروازه‌های شبکه در ارتباط باشد که این مهم از طریق استاندارد CLI[2] برقرار شده است. شکل ۲ طراحی داخلی هر خوشه را نمایش می‌دهد.

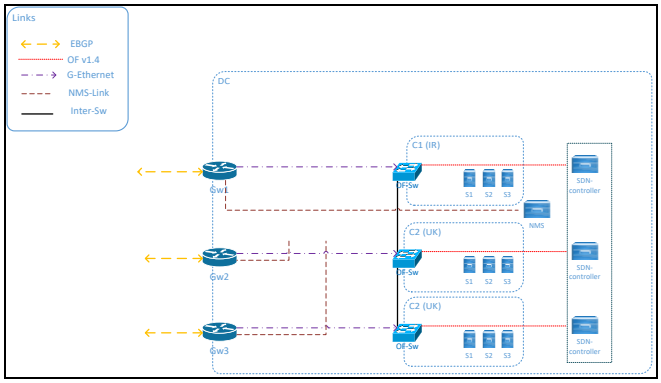
۳-۴- الگوریتم‌های معماری پیشنهادی

در معماری پیشنهادی، برای توزیع بار از چهار الگوریتم استفاده می‌شود که دو مورد از آنها با ورود هر بسته و دو مورد دیگر به صورت دوره‌ای انجام می‌شوند.

۳-۴-۱ الگوریتم توزیع بار بین خوشه‌ای

این الگوریتم در سطح کلان و بین خوشه‌های مرکز داده ترافیک را توزیع می‌نماید. این الگوریتم با ورود هر بسته و در سامانه مانیتورینگ اجرا و نتیجه آن از طریق کنترلر و دروازه‌های شبکه به ترتیب در بخش‌های نرم‌افزار محور و سنتی شبکه اعمال می‌شود. شبه کد این الگوریتم در شکل ۳ آمده است.

[1] Network Monitoring Server (NMS)
[2] Command Interface Line (CLI)



شکل ۲: طراحی داخلی هر خوشه و نحوه ارتباط آنها با یکدیگر

```
1 for each input packet
2   if packet = SYN
3      Usr++
4       if  Usr > C then Reject Packet and Usr--
5       if  Usr > C_max then AS+=2
6      if   Usr < C_max and Usr > C_exc then AS++
7      Intra-Cluster Load Balancer()
8   else if packet is for Current Cluster
9       route to S(Packet)
10 else   route to Correct Cluster
```

شکل ۳: الگوریتم توزیع بار بین خوشه‌ای

این الگوریتم به شرح زیر ترافیک ورودی به مرکز داده را از طریق افزایش دنباله AS-Path که توسط دروازه‌های خوشه اعلام خواهد شد مدیریت می‌نماید. اگر این بسته، شروع یک نشست جدید باشد (خط ۲)، تعداد نشستهای فعال (Usr) با سه حد آستانه بررسی می‌شود. در صورتی که تعداد کاربران (نشست‌های فعال) خوشه از ظرفیت خوشه (C) بیشتر باشد، این درخواست رد می‌شود (خط ۴). اگر تعداد کاربران بیش از حد بحرانی (C_{max}) باشد به دنباله AS-Path دو بار (خط ۵) و در صورتی که خوشه در حالت ازدحام باشد (خط ۶) به دنباله AS-Path فقط یک بار AS این مرکز داده (µ) اضافه می‌شود. با پذیرفته شدن این درخواست الگوریتم توزیع بار درون خوشه‌ای فراخوانی می‌شود تا درخواست به سرور مناسب ارسال گردد.

۳-۴-۲ الگوریتم توزیع بار درون خوشه‌ای

این الگوریتم در هر خوشه و به منظور توزیع بار بین سرورهای خوشه با ورود هر بسته اجرا می‌شود. اگر این بسته شروع نشست جدید باشد به خلوت ترین سرویس دهنده[3] ارسال می‌شود (خط ۳ و ۴). اگر این بسته درخواست جدید نبوده و به اشتباه به این ارسال شده باشد با تنظیم قواعد جدید مسیریابی در سوییچ، به خوشه صحیح بازمسیریابی می‌شود (خط ۶).

[3] Least Connection

```
1 for each input packet
2  if packet = SYN
3      LeastConnection(S)
4      set wildCard to flow table
5  else if packet isn't for Current Cluster
6      set MicroFlow to Correct OF switch
7  else route to host
```

شکل ۴: الگوریتم توزیع بار درون خوشه‌ای

۳-۴-۳ الگوریتم تشخیص خلوت شدن خوشه

پس از تشخیص ازدحام در خوشه به دلیل افزایش AS-Path توسط الگوریتم توزیع بار بین خوشه‌ای ممکن است برای مدت زمانی هیچ ترافیکی به خوشه وارد نشده و درخواستهای فعلی نیز به اتمام برسند. این وضعیت، حالت خلوت[1] نامیده می‌شود. دلیل آن این است که به دلیل افزایش بیش از حد دنباله AS-Path هیچ یک از مسیریاب‌های شبکه جهانی ترافیک را به خوشه مورد نظر ارسال نمی‌نمایند. برای حل این مشکل، الگوریتم تشخیص خلوت شدن خوشه به صورت دوره‌ای برای همه خوشه‌ها اجرا شده و با توجه به تعداد فعلی کاربران خوشه نسبت به تغییر دنباله AS-Path آن خوشه اقدام می‌نماید. لازم به ذکر است در صورتی که کل کاربران مرکز داده از یک حد آستانه (خط ۲) کمتر باشد، نیازی به تغییر AS-Path نیست چرا که این حالت به معنی کم بودن بار کل مرکز داده است.

```
1 when Timer expires
2  if not C(all Cluster) < 3*C_min
3     for each Cluster
4        if C(Cluster) < C_min
5           AS--
```

شکل ۵: الگوریتم تنظیم دوره‌ای AS-Path

۴-۴-۳ الگوریتم کنترل دنباله AS-Path

مسیریاب‌های اینترنتی برای جلوگیری از دستکاری[2] دنباله‌های AS-Path، دنباله‌هایی که بیش از ۲۵ عضو در آن‌ها قرار داشته باشد را به عنوان حمله در نظر گرفته و آنها را قبول نمی‌کنند. برای جلوگیری از رد شدن دنباله‌های ساخته شده، به صورت دوره‌ای دنباله‌ها بررسی شده و در صورت افزایش اعضای دنباله به بیش از ۲۰ عضو، از کل دنباله خوشه‌ها ۲ واحد کم می‌نماید. فرآیند کم شدن بر روی کلیه دنباله‌ها اعمال می‌شود تا اولویت ترافیک ورودی به مرکز داده در خوشه‌ها حفظ شود.

```
1 when Timer expires
2  for each Cluster
3    if AS-Path Cluster > 20
3       all Cluster AS -= 2
```

شکل ۶: الگوریتم تنظیم دوره‌ای AS-Path

با توجه به الگوریتمهای شرح داده شده، معماری پیشنهادی در دو سطح الف) توزیع بار در بین خوشه‌ها و ب) توزیع بار در هر خوشه بین سرورها به صورت واکنشی می‌پردازد. این الگوریتم با استفاده از ویژگی‌های مسیریابی در استاندارد BGP بدون نیاز به استفاده از دامنه آدرس‌های anycast مفهوم anycast را پیاده‌سازی نموده است. با توجه به اینکه مسیریاب‌های اینترنتی بین دو مسیر یکسان، مسیر با کوتاهترین AS-Path را انتخاب می‌نمایند، خوشه‌های مرکز داده در هنگام بروز ازدحام در شبکه با اضافه نمودن به دنباله AS-Path خوشه مسیریابی ترافیک را از آن خوشه دور می‌نمایند.

در خصوص استفاده از آدرس‌های یکسان برای همه خوشه‌های مرکز داده نیز با توجه به اینکه کنترلر نرم‌افزار محور شبکه با استفاده از مشخصات سوکت ارتباط شامل آدرس کاربر، آدرس مقصد، شماره پورت مبدا و شماره پورت مقصد ترافیک کاربران مختلف را به سرور اصلی هدایت می‌نماید، تداخلی در آدرس‌های داخل مرکز داده بوجود نمی آید (در داخل مرکز داده تصادم آدرس[3] بوجود نخواهد آمد).

استفاده از معماری ترکیبی نرم‌افزار محور و سنتی در کنار استفاده از امکانات BGP مدیران مرکز داده را قادر خواهد ساخت تا با کمترین تغییر در ساختار مرکز داده بیشترین بازدهی را در تعداد کاربران همزمان متصل شده به خوشه‌ها داشته باشند.

۴- شبیه‌سازی و ارزیابی کارایی

در این بخش به ارزیابی کارایی روش پیشنهادی پرداخته می‌شود. برای پیاده‌سازی معماری پیشنهادی، در بخش سنتی از شبیه‌ساز شبکه GNS3 استفاده شد که در آن، دروازه‌های هر خوشه با دریافت دستور از سامانه مدیریت شبکه، دنباله AS-Path مد نظر را با BGP به همسایگان خود اعلام می‌کنند و کاربران در شبکه اینترنت بطور اتوماتیک به خوشه مسیریابی می‌شوند. برای ایجاد تغییر در دنباله AS-Path هر خوشه از روش Route-map استفاده شده است. با توجه به اینکه بسته‌های ورودی به شبکه نرم افزار محور به صورت اتوماتیک به کنترلر ارسال می‌شوند و در صورت ورود بسته‌ها به دلیل تغییر AS-Path به خوشه نادرست، کنترلر با ایجاد روال موقتی[4] بسته‌ها را به خوشه درست بازمسیریابی می‌نماید، در شبیه‌سازی‌های انجام شده یک فرآیند بارگیری با تغییر خوشه ورودی ثابت مشاهده شد. در شبیه‌سازی دروازه های‌شبکه از مسیریاب‌های Cisco سری ۷۲۰۰ انتخاب شده‌اند و مسیریابی خارجی BGP بر روی آن‌ها پیاده‌سازی شده است. با توجه به پیاده‌سازی در محیط شبیه‌سازی رابطه میان ظرفیت خوشه و حد ازدحام و بحران در بهترین حالت به شرح زیر است.

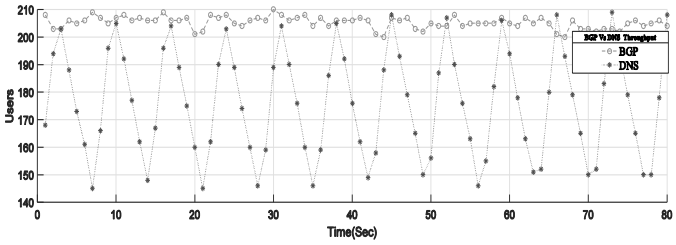
[1] Sparse State
[2] Route Manipulation
[3] IP Conflict
[4] Micro Flow

$$C_{exc} = 0.85C \ , C_{max} = 0.95C$$

برای شبیه‌سازی و ارزیابی کارایی روش پیشنهادی از مدل داده‌های آماری موجود در خصوص کاربران اینترنت استفاده شده است. سامانه مانیتورینگ که الگوریتم‌های بخش ۳-۴ را پیاده‌سازی می‌نماید، با استفاده از تکنولوژی NET. پیاده‌سازی شده است. شاخص‌هایی که برای مقایسه روش پیشنهادی با سایر روش‌ها انتخاب شده است شامل تعداد کاربران همزمان متصل شده به خوشه‌ها مرکز داده و همچنین تعداد درخواست‌های رد شده توسط مرکز داده می‌باشد. برای ارزیابی کارایی روش پیشنهادی، این روش با روش مبتنی بر DNS مقایسه شده است که برای توزیع بار ترافیک از روش Round Robin استفاده می‌نماید. لازم به ذکر است، علیرغم تشابه روش پیشنهادی به روش‌های مبتنی بر مسیریابی، مقایسه با این روش‌ها ارزش افزوده‌ای ندارد. در این روش‌ها با فرض‌های غیر واقعی که امکان پیاده‌سازی عملی ندارد (مانند استفاده از مسیریابی داخلی در سطح internet یا استفاده از source routing) معیارهای کارایی (مانند نرخ پذیرش) بهبود قابل توجهی خواهد داشت ولی در عمل نمی‌توان از آنها استفاده کرد.

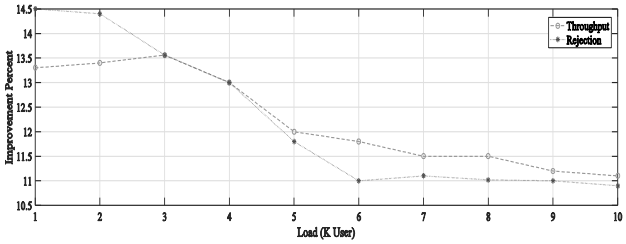
طبق این روش شبیه‌سازی، کاربران طبق مدل آماری مورد اشاره به مرکز داده مراجعه نموده و در صورت بروز ازدحام توسط یکی از روش‌های مورد بررسی کاربران به خوشه‌های دیگر باز مسیریابی می‌شوند برای ارزیابی دقیق تر تعداد کاربران متغیر در نظر گرفته شده است. تعداد کاربران از ۳۰۰۰۰ هزار کاربر تا ۱۲۰۰۰۰ کاربر مراجعه کننده به مرکز داده قرار داده شده است. هر کاربر بطور متوسط زمانی را در مرکز داده سرویس گرفته و پس از آن از مرکز داده خارج می‌شود. دو سناریو کاربران تک درخواستی و کاربران چند درخواستی مورد آزمایش قرار گرفته است.

در سناریو اول کاربران حجم ترافیک پایینی دارند، آنها با ورود به مرکز داده یک درخواست ارسال نموده و بین ۶۰۰ تا ۱۲۰۰ ثانیه در مرکز داده فعالیت نموده و پس از آن از مرکز داده قطع می‌شوند (طول عمر این درخواست‌ها یک مقدار تصادفی بین ۶۰۰ الی ۱۲۰۰ ثانیه است). تعداد کاربرانی که در این تست شرکت داده می‌شوند ۳۰۰۰۰ کاربر بوده و برای مقایسه با روش مبتنی بر DNS تعداد گام DNS برای توزیع بار ۱۰۰ کاربر در نظر گرفته شده است. کل مرکز داده دارای ۳ خوشه و حد رد کردن درخواست ۲۱۲ در نظر گرفته شده است. شکل ۶ تعداد کاربران هر خوشه در روش پیشنهادی و روش مبتنی بر DNS را نمایش می‌دهد. همانگونه که در این نمودار مشخص است، تعداد کاربران در روش مبتنی بر DNS به دلیل سرعت واکنش پایین این روش دایما در حال کم و زیاد شدن است در حالی که روش پیشنهادی با واکنش مناسب و سریع به صورت متوسط تعداد کاربران یکسانی را می‌پذیرد که این مساله در شکل‌های بعدی با جزییات بیشتری شرح داده شده است.



شکل ۷: تعداد کاربران هر خوشه در روش پیشنهادی (BGP) و روش DNS

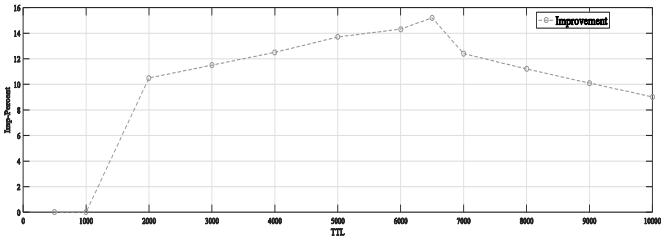
میزان بهبود در تعداد کاربران همزمان (نسبت تعداد متوسط کاربران مرکز داده در روش پیشنهادی نسبت به روش DNS) و همچنین بهبود در تعداد درخواست‌های رد شده (نسبت تعداد درخواست‌های رد شده توسط روش DNS نسبت به روش پیشنهادی) در شکل ۸ نمایش داده شده است.



شکل ۸: میزان بهبود در بسته‌های رد شده و بازدهی مرکز داده

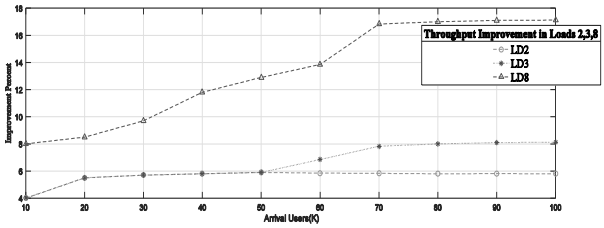
با توجه به نمودارهای شکل ۸ حداکثر میزان بهبود در تعداد کاربران همزمان ۱۳.۵ درصد و تعداد درخواست‌های رد شده ۱۴ درصد نسبت به حالت مبتنی بر DNS می‌باشد.

در کنار تعداد کاربران، پارامتر دیگری که بر روی بار مرکز داده تاثیرگذار است TTL درخواست‌ها می‌باشد. شکل ۹ نمایش دهنده میزان بهبود کارایی با افزایش TTL کاربران است. در این آزمایش، بیشترین میزان بهبود برای TTL برابر ۶۵۰۰ ثانیه است. برای TTLهای بسیار کمتر از این مقدار حجم بار ترافیکی مرکز داده چندان زیاد نیست بنابراین روش DNS هم کارایی نسبتا قابل قبولی دارد فلذا میزان بهبود کم است. برای TTLهای بسیار بزرگتر از این مقدار نیز مرکز داده کاملا در حالت اضافه بار است و منابع کافی برای توزیع بار وجود ندارد فلذا کارایی الگوریتم پیشنهادی خیلی بهتر از روش DNS نیست.

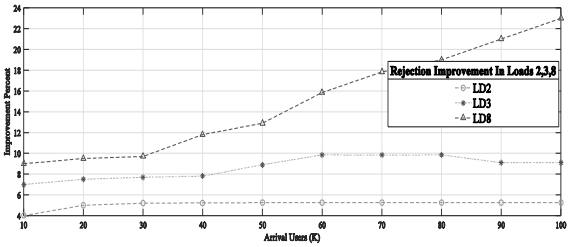


شکل ۸: مقایسه بهبود کارایی با افزایش TTL کاربران

در سناریو دوم، به ازای هر کاربر تعداد بیشتر از یک درخواست به خوشه ارسال می‌شود (سناریو کاربران با بار متغیر). در این سناریو، کاربران به صورت متغیر و با بار ۲ و ۳ و ۸ درخواست به خوشه مرکز داده وارد می‌شوند. نتایج مقایسه روش پیشنهادی با روش مبتنی بر DNS (با گام ۵۰) در اشکال ۱۰ و ۱۱ آمده است.



شکل ۱۰: میزان بهبود در تعداد کاربران همزمان در بارهای ۲ و ۳ و ۸



شکل ۱۱: میزان بهبود در تعداد درخواست‌های رد شده در بارهای ۲ و ۳ و ۸

با توجه به شکل ۱۰ بیشترین میزان بهبود در تعداد کاربران همزمان خوشه در بار ۸ و به اندازه ۱۷ درصد می‌باشد، همچنین بیشترین میزان بهبود در تعداد درخواست‌های رد شده نیز مربوط به بار ۸ و ۲۳ درصد می‌باشد. این نتایج نشان می‌دهد که روش پیشنهادی در حجم ترافیک بالا که توزیع بار اهمیت و ارزش بیشتری دارد، کارایی بالاتری نسبت به روش‌های موجود دارد.

روش پیشنهادی با توجه به اینکه دارای قابلیت پیاده‌سازی در شبکه جهانی می‌باشد و از مسیریابی BGP پشتیبانی می‌نماید، دارای مقیاس‌پذیری بیشتری نسبت به سایر روش‌های مبتنی بر مسیریابی می‌باشد. استفاده از این استاندارد باعث افزایش حجم جداول مسیریابی تا چند برابر روش‌های مبتنی بر مسیریابی داخلی و یا مسیریابی دستی می‌شود.

۵- تیجه گیری و پیشنهادات

در این مقاله یک معماری جدید برای توزیع بار ترافیکی در مراکز داده توزیع شده جغرافیایی ارائه شد که در مقایسه با سایر روش‌های موجود دارای انعطاف‌پذیری و میزان کارایی بیشتری نسبت به روش مبتنی بر Round Robin است. این روش برای مدیریت ترافیک ورودی به مرکز داده، مفهوم anycast را با استفاده از پروتکل BGP و از طریق تنظیم مناسب دنباله AP-Path پیاده‌سازی کرده است. از ایده این روش می‌توان برای تشخیص سریعتر و کم هزینه‌تر حملات توزیع شده DOS به خوشه‌های مراکز داده استفاده نمود.

مراجع

[1] H. Goudarzi and M. Pedram, "Force-directed geographical load balancing and scheduling for batch jobs in distributed datacenters," *IEEE International Conference* on *Cluster Computing (CLUSTER)*, Indianapolis, 2013.

[2] www.tia942.org/content/162/289/About_Data_Centers (last access: 11 May 2017).

[3] H. Goudarzi, S. Hatami and M. Pedram, "Demand-side load scheduling incentivized by dynamic energy prices," *International Conference on Smart Grid Communications (SmartGridComm)*, Brussels, 2011.

[4] Uppal, Hardeep, and Dane Brandon. "OpenFlow based load balancing." *Networking Project Report, University of Washington*, 2010.

[5] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," in *Proceedings of the IEEE*, vol. 103, no. 1, Jan. 2015, pp. 14-76.

[6] Gude, Natasha, et al. "NOX: towards an operating system for networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 3, July 2008, pp. 105-110.

[7] A. Serra, D. Gaiti, G. Barroso and J. Boudy, "Assuring QoS differentiation and load balancing on web servers clusters," *IEEE Conference on Control Applications (CCA)*, Toronto, 2005.

[8] Hailong Zhang and Xiao Guo, "SDN-based load balancing strategy for server cluster," *IEEE International Conference on Cloud Computing and Intelligence Systems*, Shenzhen, 2014.

[9] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman and B. Weihl, "Globally distributed content delivery," *IEEE Internet Computing*, vol. 6, no. 5, Sep/Oct 2002, pp. 50-58.

[10] F. Semchedine, L. Bouallouche-Medjkoune, O. Sayeh, S. Ayoub and D. Aïssani, "DNS-based load balancing with Cache for geographically distributed Web server systems," *Global Summit on Computer & Information Technology (GSCIT)*, Sousse, 2014.

[11] https://tools.ietf.org/html/rfc2526 (last access: 11 May 2017).

[12] Jain, Sushant, et al. "B4: Experience with a globally-deployed software defined WAN," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, Oct. 2013, pp. 3-14.

[13] H. Hashim and A. Manan, "Active anycast method for server load balancing," *Student Conference on Research and Development*, 2002.

[14] R. L. Pereira and T. Vazao, "Network Layer Implemented Anycast Load Balancing," *IEEE Symposium on Computers and Communications*, Aveiro, 2007.

[15] https://tools.ietf.org/html/rfc4271 (last access: 11 May 2017).

[16] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, S. Shenker, "NOX: towards an operating system

for networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 3, 2008, pp. 105-110.

[17] S, Kaur, S, Japinder, S. Navtej, "Network programmability using POX controller," *IEEE International Conference on Communication, Computing & Systems (ICCCS)*, 2014.

[18] A. A. Dixit, F. Hao, S. Mukherjee, T. V. Lakshman, R. Kompella, "Elasticon: an elastic distributed sdn controller," *ACM/IEEE symposium on Architectures for networking and communications systems*, 2014.

[19] A. Krishnamurthy, S. P. Chandrabose, A. Gember-Jacobson, "Pratyaastha: An efficient elastic distributed SDN control plane," *ACM workshop on Hot topics in software defined networking*, 2014.

[20] V. Yazıcı, M. O. Sunay, and A. Ö. Ercan, "Controlling a software defined network via distributed controllers," *NEM Summit, Implementing Future Media Internet Towards New Horizons*, 2012

امیراسمعیلی در سال ۱۳۸۷ مدرک کارشناسی مهندسی کامپیوتر خود را از دانشگاه شاهد تهران دریافت نمود و در سال ۱۳۹۵ مدرک کارشناسی ارشد مهندسی فناوری اطلاعات خود را از دانشگاه صنعتی امیرکبیر دریافت کرده است. نامبرده هم اکنون بعنوان رئیس اداره پشتیبانی زیرساخت فتا در سازمان تنظیم مقررات و ارتباطات رادیویی مشغول بکار می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارت از: مهندسی ترافیک، شبکه‌های نرم‌افزار محور و امنیت شبکه‌های کامپیوتری می‌باشد.

بهادر بخشی سراسکانرود در سال ۱۳۸۹ مدرک دکترای خود را در رشته مهندسی کامپیوتر از دانشگاه صنعتی امیرکبیر دریافت نمود. دکتر بخشی از سال ۱۳۹۱ تاکنون بعنوان عضو هیئت علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر می‌باشند. زمینه‌های تحقیقاتی مورد علاقه ایشان شبکه‌های نرم‌افزار محور و مجازی‌سازی کارکردهای شبکه می‌باشد.