



## بهینه سازی تخصیص ویدئو سرورها در محیط وب با استفاده از برنامه ریزی خطی optimize allocation of video servers on the internet with linear programming

(۱) مهدی حسنی گودرزی

دانشگاه آزاد اسلامی - واحد تهران جنوب  
Email:mehdigoodarzi@azad.ac.ir

(۲) آرش زعیم

دانشگاه آزاد اسلامی - سازمان مرکزی  
Email:arashzaeim@yahoo.com

چکیده - امروزه توزیع بار ترافیک سرویس ویدئویی در شبکه های گسترده به کاربران از موضوعات مهم در حوزه شبکه و نرم افزار می باشد. در مدل پیشنهادی اساس انتخاب بر پایه تاخیر بسته های ارسالی و پیام تصدیق آنهایی باشد. ولی مانند روشهای حریصانه و اکتشافی هوشمند نیست. در اغلب مدل های ارائه شده از کنترل پذیرش جهت حفظ کیفیت سرویس استفاده شده است مدل ارائه شده TP نام دارد و سعی بر این است که راه حل مناسبی جهت انتخاب سرور مناسب، با در نظر گرفتن شرایط کلاینت های نزدیک سرور، توزیع بار شبکه روی چندین منبع، انطباق ظرفیت بهینه سرور با درخواستها کلاینت ها ارائه گردد.

کلید واژه - Transportation Problem (TP)، ارسال ویدئوی در خواست شده Video-On-Demand (VoD)، برنامه ریزی خطی Linear Programming (LP)، کیفیت سرویس (QoS)، کنترل پذیرش (AC)، Admission Control، کارایی انسداد Round Trip Time (RTT)، Bloking Performance (BP)، Centralized Control Protocol (CCP)، کنترل توزیع شده Distributed Control Protocol (DCP)

### ۱- مقدمه

کلاینت X به سمت سرور Y می باشد. در مدل پیشنهادی می خواهیم با سنجش فاصله بین کلاینت و سرور هزینه سرویس را مینیمم کنیم. در این مدل سعی بر ایجاد دسته های سرور و تنظیم ظرفیت این گروه هاست، در یک گروه سرور روشی برای مدیریت درخواستها ارائه شده که نتیجه آن میزان ظرفیت شبکه را متوازن می کند. آستانه کنترل پذیرش برای سرورهایی که بطور مجزا درون گروه ها قرار دارند پیاده سازی شده است، همچنین گروه ها می توانند بشکل عادلانه اطلاعات را برای درخواستهای کاربران به اشتراک بگذارند. در روش کارایی انسداد، هنگامیکه نرخ انسداد بهینه باشد، می تواند بعنوان عاملی برای اختصاص سرور بصورت بهینه استفاده شود. دو پروتکل برای گروه بندی سرورها پیشنهاد شده است که وظیفه انتخاب مسیرهای گوناگون برای تحویل اطلاعات از چندین منبع به سرویس گیرندگان را بعهده دارد. این مدل روشی بهینه برای دستیابی ساکنان یک منطقه به سرورها ایجاد می کند، بدین شکل که سرورهای مجزا را به گره های منطقه ای بشکل بهینه تقسیم بندی می کند. اینگونه دسته بندی مزایای ذیل را بدنبال خواهد داشت.

استفاده از جریان اطلاعات ویدئویی مانند مشاهده و VoD در سالهای اخیر به شدت بالا رفته و می توان گفت حدوداً به میزان ۵۸ درصد افزایش یافته است. از یک سو کاربرد خانگی آن نیز تقریباً دوبرابر شده است، از این دسته سرویسها می توان:

- Microsoft's Windows Media
- Macromedia Streaming Shockwave .service
- Streaming Apple's Quicktime
- YouTube Player

را نام برد. در این راستا [5,1] Akamai راه حل مناسبی برای ارسال محتوای درخواستی ارائه نموده است. عملکرد این روش بدین شکل است که محتوای داده ها به سمت کلاینت فرستاده می شود. در این روش محتوای درخواستها به انتهای شبکه منتقل شده و با استفاده از AC میزان QoS را تضمین می نماید. VoD، نرم افزارهای تحت وبی هستند که بر خلاف وب سرورهای ارائه کننده مولتی مدیا نیاز مبرمی به QoS دارند. در این مدل روشی به نام TP با استفاده از برنامه ریزی خطی ارائه شده است. تابع هدف در این مدل، هزینه سرویس به درخواستهای ویدئویی ارسالی از

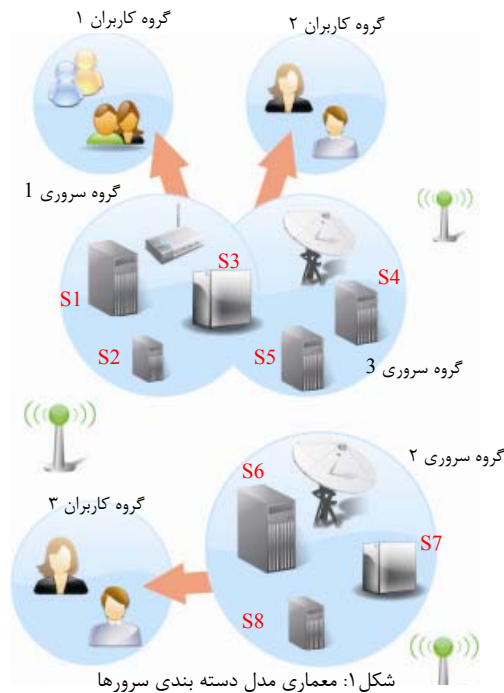
وب جهت جریان داده ای ویدئویی ارائه نموده اند ولی آگاهی چندانی از روشهای ارائه شده توسط این کمپانی ها در دست نیست و نمی توان این مدل را برای آنها نیز مورد آزمایش قرار داد. در صورتی که ممکن است نرخ داده ای پائین تری داشته باشند. در این مدل VoD بعلت پایداری و گستردگی مورد بررسی قرار گرفته است. کنترل پذیرش و کارایی انسداد بخش اساسی در طراحی اینگونه نرم افزارهاست. زمانیکه رد سرویس درخواستی راه حل جلوگیری از سر بار شبکه جهت نگاهداری QoS باشد، آنگاه تخصیص منابع بسیار مهم به نظر می رسد. همچنین مطابقت منابع سرورها با میزان درخواستهای کاربران از این طریق نشان داده شده است. این مدل با سایر روشهای بهینه سازی ظرفیت تفاوتی دارد. همچنین از مکانیزم تولید سلسله مراتب توپولوژی که توسط Doar ارائه شده، استفاده خواهیم کرد [20]. در [35] نویسنده دو مدل مولد توپولوژی اینترنت یعنی ساختار بر پایه درجه و کاربرد سلسله مراتبی آن در اینترنت را مورد بررسی قرار داده است. [21,27] به این سرویس وب به حالت گسترده پرداخته است. مرجع [21] به تکنولوژی سنجش ترافیک در شبکه اینترنت با استفاده از ابزارهایی از جمله Tstat می پردازد، که ما نیز از این روش برای شبیه سازی مدل استفاده خواهیم نمود. [28] روشی برای آشکار سازی ترافیک شبکه با استفاده از توزیع (M/Pareto) و نیز برای ورود و خروج از توزیع پواسن استفاده شده است. [29] روشی بهینه برای جستجوی ویدئوی مورد نظر ارائه نموده است. [6] اغلب انتخاب سرور مناسب کاری است که معطوف به خود وب سرور می باشد، این مدل که سیستم وب نگاشت نامیده شده است، روشی جهت اختصاص کلاینتها و نگاشت آنها به سرور ها در سیستمهای توزیع شده بیان و اثبات شده است. سایر مراجع که در زمینه انتخاب سرور و توازن بار ارائه شده است، تمرکزشان بر روی انتخاب از میان تعداد محدودی سرور تصویر شده و دانلود داده ها بصورت موازی می باشد که در [7,12] به آنها پرداخته شده است. [13] تحقیقات مقایسه ای برای سنجش و انتشار مسیر سرور و شبکه با سر بار کم در اینترنت ارائه می نماید. ما

- مجموعه سرورهایی که با کلاینت ها ارتباط ویدئویی دارند کارایی بهتری در ارسال داده ها و کاستن پکت های گم شده یا خراب شده خواهند داشت.
- وجود سرورهای جداگانه بطور فیزیکی داده ها را به اجبار به چند شبکه می فرستد و این خود باعث پائین آمدن تراکم در شبکه خواهد شد.
- تکه تکه شدن داده ها کاهش یافته و به توازن بار و تحمل شکست خواهیم رسید.
- توجه داشته باشیم مدل شبکه مورد بررسی مبتنی بر IP و اتصال گرا می باشد. این مدل روش بسیار مناسبی برای تخصیص بهینه سرورهاست، هدف اصلی این روش تخصیص ظرفیت سرورها به درخواستهای کاربران با توجه به هزینه ارائه سرویس از سرور  $i$  به کاربر درخواست کننده  $Z$  می باشد، در این مدل از RTT بعنوان مقیاس هزینه این مدل در نظر گرفته شده است. موضوع اصلی مطرح شده در این مدل تهیه یک معماری دسته بندی سرورها بر اساس تخصیص بهینه آنها و نیز تهیه پروتکل های مؤثر برای بکارگیری این سرورهاست. ضمناً مشکل آدرس دهی پویا به سرورها در نظر گرفته شده و راه حلی برای طراحی میزان انتشار شبکه ارائه خواهد شد. توجه داشته باشیم بعلت سر بار پردازشی زیاد نمی توان از الگوریتمهای حریمانه یا اکتشافی در حل مشکل ظرفیت داده برای کاربردهای مولتی مدیا استفاده نمود. جهت اثبات مدل و نیز بررسی ترافیک شبکه از نرم افزار Tstat جهت شبیه سازی ترافیک شبکه است استفاده شده است.
- قسمت های دیگر مدل ارائه شده بدین شرح می باشند.
- بخش دوم: کارهای مشابه مرتبط
- بخش سوم: مدل پیشنهادی بیان و توصیف شده است.
- بخش چهارم: مدل ارائه شده شبیه سازی شده است.
- ضمیمه (۱): به بیان روش حل [32] مدل های TP اشاره ای خواهیم نمود.

## ۲- کارهای مشابه:

با وجود اینکه شرکت هایی همچون Akamai و Inktomi و RealNetworks، سرویس هایی مبتنی بر

TP روشی جهت مطابقت مجموع درخواست کاربران گروه و کل ظرفیت سرورها ارائه می‌کند.



شکل ۱: معماری مدل دسته بندی سرورها

تمامی سرورهایی که به گروه های کاربران مشابه سرویس می دهند دسته بندی شده و به درخواستهای گروه های سرویس گیرنده پاسخ می دهند. ممکن است که یک سرور در حالت های مختلف مابین چندین گروه کاربر به اشتراک گذاشته شود و هر گروه کاربر فقط به ظرفیت اختصاص یافته به آن در سرور مربوطه دسترسی دارد. در شکل (۱) نموداری برای سازماندهی سه دسته از سرورهای یک ناحیه شهری نشان داده شده است. در مدل پیشنهادی پاسخ درست TP راه حل مناسب سازماندهی سرورها می باشد درست پاسخ TP می باشد. گروه سروری ۱ شامل سرورهای ۱ و ۲ و ۳ می باشد و به گروه کاربران ۱ سرویس دهی می کند، گروه سروری ۲ شامل سرورهای ۳ و ۴ و ۵ می باشد و به گروه کاربران ۲ سرویس دهی می کند، باید توجه داشت که سرور شماره ۳ در هر دو گروه قرار دارد، سه سرور دیگر در مکان دیگری به گروه کاربران ۳ سرویس دهی می کند.

نیز از این مرجع RTT را بعنوان مقیاس سنجش انتخاب نموده ایم. برای این کار از ابزارهایی همچون IDMaps [14] و همچنین King [15] که RTT مابین کلاینت و سرور را تخمین می زند، استفاده می کند. ابزارهایی همچون IDMaps فاصله اطلاعات را در شبکه عمومی اینترنت سنجیده و گزارش می دهد. و به این طریق می توان فاصله مابین دو IP واقعی را مشخص نمود. [16] نویسنده از روش Multiple Description Coding (MDC) برای سرورهای ویدئویی استفاده نموده است. و نتایج گستردگی مسیرها هنگامی که با چندین منبع و MDC ترکیب شده اند را اثبات می نماید از مدل GT-ITM که در حد خود جای بحث دارد استفاده می کند. مرجع [17] پیکربندی مناسبی برای ارسال همزمان جریان ویدئو توسط چند فرستنده به یک کلاینت ارائه می کند. [18] از مجموعه سرورهای تصویر شده با استفاده از نرخ تخصیص و الگوریتم تقسیم بندی پکتها برای دستیابی به خروجی بالا استفاده می کند. برخی دیگر نیز در زمینه شبکه های رودرو ارائه شده است و از عمده آن می توان از CoopNet را نام برد [19].

### ۳- معماری جریان داده ای توزیع شده :

کمپانی هایی همچون Akamai محتوای ارسالی را در لبه های شبکه هایشان متمرکز می کنند و با استفاده از استراتژی قرار گیری سرور در مکانهای مناسب، کلاینت ها می توانند سرورها را با استفاده از کوتاهترین RTT پیدا کنند. برای تهیه چندین منبع قسمتهای مختلف ویدئو می توانند برای داشتن کارایی بهتر سرعت و پشت سر هم از منابع متعدد ارسال شوند، استفاده از چندین منبع جهت توازن بار و تحمل پذیری در برابر شکست به ما کمک می کند، بدین شکل که چندین مسیر متنوع در شبکه به وجود می آید. و نگاشتی مطابق ظرفیت سرورها با درخواستهای کاربران با کارایی انسداد بهینه حاصل خواهد شد. برای مثال یک محدوده شهری برای تعریف گروه های کاربران و جمعیت آنها بررسی شده و فرض شده سرورهای مجزا از لحاظ جغرافیایی در منطقه مورد نظر قرار دارند. مدل

### ۳-۱- تخصیص بهینه سرورها

با استفاده از مدل TP، سرورهای مجرد برای دیدن درخواستهای گروه‌های کاربر به آنها تخصیص داده شده و دسته بندی می‌شوند و هرگروه به تعدادی کلاینت سرویس می‌دهد. سه پارامتر کمیتی برای فرموله کردن ماتریس انتقال استفاده شده است.

- ظرفیت سرور
  - میزان درخواست کاربران
  - هزینه ارسال پاسخ به درخواست گروه کاربری  $j$  توسط سرور  $i$
- میزان ظرفیت جریان سرور بر اساس میزان پهنای باند دیسک در پاسخ به درخواستهای کلاینت‌ها سنجیده می‌شود. در نهایت ماتریس هزینه بر اساس میزان نزدیکی سرورها به گروه‌های کاربر و با استفاده از مقیاس RTT سنجیده می‌شود.

جدول ۱: متغیرهای تصمیم مدل پیشنهادی

$m$	تعداد سرورها
$n$	تعداد گروه‌های کاربران
$s_i$	میزان ظرفیت جریان سرور $i$
$d_j$	میزان درخواست گروه کاربر $j$
$c_{ij}$	هزینه سرویس دادن سرور $i$ به کاربران گروه $j$

ظرفیت می‌باشد. در مدل TP ظرفیت سرور و درخواستهای کاربران با استفاده از واحد مگابیت بر ثانیه اندازه گیری شده است.  $c_{ij}$  میانگین هزینه RTT بین گروه کاربران و سرور می‌باشد. RTT تخمینی اساسی برای بدست آوردن تأخیر مابین دو میزبان اینترنتی می‌باشد و عموماً با استفاده از دستورات Ping و TraceRoute انجام می‌شود. در این حیطه ابزارهای فراوانی وجود دارد که برای نمونه از IDMaps برای این تخمین استفاده شده است [14]. در جدول (۲) میزان ماتریس میانگین RTT-Capacity-Demand استفاده شده در مدل TP نشان داده شده است. ماتریس نشان داده شده مسئله انتقال متوازی است که می‌خواهیم بیشترین ظرفیت را با کل درخواستها مساوی شود.

جدول ۲: ماتریس انتقال

ماتریس ظرفیت	گروه سوم	گروه دوم	گروه اول	
S1	۳۹۹	۳۵۰	۳۷۶	۹۵۵
S2	۴۷۵	۴۲۶	۲۶۵	۷۰
S3	۳۸۳	۲۷۴	۳۲۹	۴۰
S4	۵۱۱	۲۹۸	۴۵۷	۴۰
S5	۳۰۱	۴۲۲	۴۱۷	۱۵۵
Demand	۱۰۰	۱۰۰	۱۰۰	۳۰۰

مجموع ظرفیت سرور را ۳۰۰ درخواست همزمان در نظر گرفته ایم. درخواست کاربر نیز در جدول (۲) آورده شده است و برای تمامی گروه‌های کاربر فرض شده است و ظرفیت سرورها را نیز ماکزیمم فرض می‌کنیم. درخواستهای گروه کاربران می‌تواند به نسبت تولید شود. که در قسمتهای بعدی در زمان شبیه سازی نشان داده خواهد شد. در جدول (۳) تخصیص بهینه سرورها به گروه‌های کاربر نشان داده شده است که پس از حل و اعمال مدل ساده TP بدست می‌آید. راه حل این مدل در ضمیمه (۱) آورده شده است [32]. دسته بندی سرورها پس از TP بدین صورت خواهد شد.

$$Z = \text{Minimize} \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

$$\text{Subject to} \sum_{j=1}^n x_{ij} = s_i \text{ for } i = 1, 2, \dots, m$$

$$\sum_{i=1}^m x_{ij} = d_j \text{ for } j = 1, 2, \dots, n$$

$$x_{ij} \geq 0 \text{ for all } i \& j$$

فرمول ۱: تابع هدف و محدودیتهای مدل

$Z$  میزان هزینه تخصیص را مشخص نموده و  $x_{ij}$  میزان ظرفیت جریان داده ای تخصیص یافته از سرور  $i$  به گروه کاربران  $j$  می‌باشد. تابع هدف این مدل خطی به شکل ذیل می‌باشد. در این مدل ظرفیت سرورها ( $s_i$ ) و درخواست گروه‌های کاربران ( $d_j$ ) می‌باشد. برای گروه‌های کاربران تعداد درخواستهای همزمان بر اساس یک میانگین نرخ داده و برای سرور پهنای باند مبنای اصلی

جدول ۳: ماکزیمم تخصیص

	گروه اول	گروه دوم	گروه سوم
S1	۳۰	۲۰	۴۵
S2	۷۰	۰	۰
S3	۰	۴۰	۰
S4	۰	۴۰	۰
S5	۰	۰	۵۵

کنترل پذیرش که توسط سرور  $i$  انجام می شود بدین شکل خواهد بود:

درخواست وارد شده توسط گروه کاربر  $j$  در صورتی برای سرور  $i$  مجاز است که این دو خاصیت را داشته باشد.

۱. براساس آستانه :

$$N_{current_j} + 1 \leq N_{allocated_j}$$

۲. براساس ظرفیت:

$$\sum_{j=1}^n N_{current_j} + 1 \leq N_{capacity}$$

•  $N_{current_j}$ : تعداد درخواستهای جاری گروه کاربر  $j$  از سرور  $i$

•  $N_{allocated_j}$ : تعداد درخواستهای تخصیص سرور  $i$  توسط گروه کاربر  $j$

•  $N_{capacity}$ : ماکزیمم ظرفیت سرور  $i$

مدل ارائه شده بر پایه آستانه تحمل سرور می باشد، که در اینجا ظرفیت تخصیص سرور را تا سرحد آن در نظر گرفته شده است. قانون ظرفیت گرا جهت اطمینان از نپذیرفتن سربار اضافی توسط سرور بکار برده می شود.

### ۳-۳-۳- سازماندهی پویای مجدد برای گروه های سرور

برای اشتراک داده ها و همچنین تعیین کیفیت کارایی انسدادهای می توان از کنترل پذیرش استفاده نمود. سپس به نرخ انسدادهای گروه های کاربر و اختصاص بهینه سرور به گروه های کاربر نظارت داشت. می توان کارایی انسدادهای را بعنوان یک پارامتر قوی جهت تغییر درخواست کاربر و نیز تعیین کننده وضعیت شبکه استفاده نمود و نیز می توان از آن بعنوان راه اندازی برای سازماندهی مجدد دسته های سرور استفاده کرد. پس از آن با تغییر درخواستهای کاربر و یافتن یک تخصیص بهینه، مجدداً الگوریتم TP را در زمان سازماندهی مجدد اجرا کرد. این تخصیص جدید یک نتیجه پویا را به وسیله تغییر در سیاست کنترل پذیرش در هر سرور به وجود می آورد. موضوع مهمی که در سازماندهی مجدد می توان مشاهده نمود، اداره درخواستهایی است که در حال دریافت سرویس هستند، که قبل و بعد از سازماندهی مجدد از لحاظ تخصیص یک سرور خاص می توان

$S_1$  و  $S_2$  به گروه کاربران شماره ۱ سرویس دهی می کنند،  $S_1$  و  $S_3$  و  $S_4$  به گروه کاربران شماره ۲ سرویس دهی می کنند و  $S_1$  و  $S_5$  به گروه کاربران شماره ۳ سرویس دهی می کنند. این اختصاص سرورها بعداً در شبیه سازی تحلیل کارایی انسدادهای استفاده خواهد شد.

### ۳-۲- کنترل پذیرش سرور

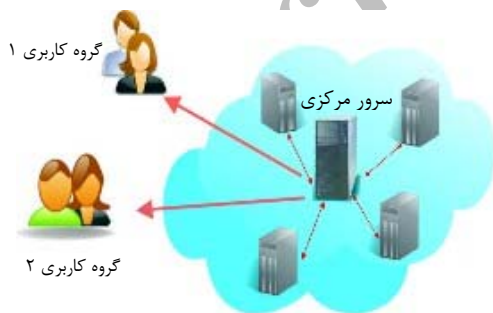
اغلب در برنامه های مولتی مدیا حالتی وجود دارد که هر گروه سروری برای جلوگیری از سربار می خواهد به تنهایی میزان کنترل پذیرش را کنترل کند. اساس کنترل پذیرش عموماً رد سرویس درخواستی در حالتی است که سرور منابع کافی برای سرویس دهی بر اساس آستانه معینی را ندارد. لکن با اشتراک سرورها میزان بار قابل تحمل توسط گروه افزایش می یابد. در [33] احتمال بستن درخواست پذیرش آورده شده است. در این معماری ممکن است یک سرور برای دو یا چند گروه کاربر به اشتراک گذاشته شده باشد، علاوه بر کنترل پذیرش به سیاست اشتراک گذاری منابع هم نیاز خواهیم داشت. بطور مثال سرور  $S_1$  در جدول (۳) درون سه گروه به اشتراک گذاشته شده است، که به هر کدام از درخواستهای کاربران گروه های ۱ و ۲ و ۳ پاسخ می هد. درخواستهای هر گروه به میزان بهینه تخصیص سرور به آن گروه محدود می شود. مثلاً اگر میزان آستانه تحمل گروه ۱ مقدار ۳۰ باشد هنگامی که ظرفیت سرور به آن آستانه برسد بسته خواهد شد. هدف این است که این ظرفیت برای پاسخ به درخواستهای گروه ۲ و ۳ نیز وجود داشته باشد تا اینکه بتوانیم برای آنها منابع را بطور عادلانه به اشتراک بگذاریم. بطور کلی

### ۳-۴- پروتکل‌های خوشه بندی

در این بخش دو شیوه برای پاسخ به درخواستهای کاربران در گروه‌های سرور بررسی شده است.

#### ۳-۴-۱- کنترل مرکزی (CCP) :

در این پروتکل از تقسیم بار متمرکز استفاده شده است. تمامی درخواستهای کلاینتها توسط یک سرور مرکزی دریافت شده و هریک از این درخواستها با استفاده از اطلاعات ظرفیت هر سرور بطور یکنواخت میان سرورهای گروه تقسیم می‌شود. در این مدل سرور مرکزی اطلاعات عمومی از ظرفیت سرورهای درون گروه بدست آورده و این چرخه را برای آگاهی همزمان از وضعیت هرکدام از ویدئو سرورها تکرار می‌کند. صرفنظر از پروتکل جریان داده یکی از پروتکل‌های UDP یا TCP برای جلوگیری از گم شدن پیامها مابین سرور مرکزی و ویدئو سرورها و بالعکس استفاده شده است. در این مدل هر درخواست ارسال شده برای یک منبع ویدئویی توسط سرور مرکزی به آخرین سرور در حال سرویس ارسال می‌شود، لذا این پروتکل می‌تواند به صورت یک جریان داده یک سرور بیان شود، به شکلی که یک سرور فقط به یک کلاینت سرویس دهی کند، در مقایسه با DCP که در بخش بعد به آن پرداخته می‌شود. در شکل (2-1) این پروتکل نشان داده شده است.



شکل 2-1: پروتکل مدل کنترل مرکزی

اگر ویدئو سرور موفق به ارسال پیام اطلاعات وضعیت خودش به سرور مرکزی نشود، فرض بر این است که این سرور فعال نیست و هیچ درخواستی از سمت کلاینتها به آن ارسال نخواهد شد تا اینکه پیام فعال شدن خود را به سرور مرکزی بفرستد و آنرا تکرار نماید.

مشاهده نمود.  $N_{old}$  و  $N_{New}$  تعداد درخواستهای قبل و بعد از سازماندهی می‌باشد:

تخصیص قدیمی کوچکتر یا مساوی تخصیص جدید است  $N_{old} \leq N_{new}$  یا تخصیص قدیمی بزرگتر

از تخصیص جدید است  $N_{new} > N_{old}$ .

قدرمطلق  $|N_{old} - N_{new}|$  درخواست، میزان ماکزیمم

تعداد درخواستهای دریافتی قبلی سرور را نشان می‌دهد. میزان پیشرفت آنها در زمان سازماندهی مجدد در

بخش اول صفر است و مورد دوم بین صفر و  $|N_{old} - N_{new}|$  می‌باشد. در مورد اول انتقال از

تخصیص قدیمی به تخصیص جدید بسیار یکنواخت می‌باشد و می‌تواند به صورت بلادرنگ فرض شود زیرا

هرگز ظرفیت سرور بیشتر از حد توانش نیست. در مورد

دوم همه یا تعدادی از  $|N_{old} - N_{new}|$  درخواستها

ممکن است در حال دریافت سرویس درون سرور باقی بمانند که یک تغییر وضعیت آنی و رسیدن به

سازماندهی مجدد برای درخواستهای تخصیص سرویس را قطع می‌کند. فاز انتقال برای اجازه دادن به

درخواستهای زمانبندی شده قبلی جهت تکمیل ضروری می‌باشد. در این فاز سرور بررسی می‌شود که از حد

ماکزیمم یعنی  $|N_{old} - N_{new}|$  برای پاسخ به

درخواستهای بسته شده بیشتر نباشد و سپس مجدداً راه اندازی می‌گردد. پس از طی زمان انسداد و بعلت

سازماندهی مجدد توان مصرفی کم خواهد شد. در حقیقت وقتی تأثیر مناسب آنرا می‌بینیم که

$|N_{old} - N_{new}|$  تکمیل شود. وقتی که در فاز انتقال،

نوسان وجود داشته باشد سازماندهی مجدد انجام نمی‌شود. یک دوره تناوب به صورتی در نظر گرفته شده

که درخواستهای قدیمی به سمت درخواستهای جدید ارسال شوند. بنابراین فرض می‌کنیم در طراحی مدل،

این موضوع انتخابی نیست و می‌دانیم که سربار اضافی و احتمال قطع ارتباط، موجب نقص و خسارت به QoS

در مدیریت درخواستهای در حال پیشرفت خواهد شد.

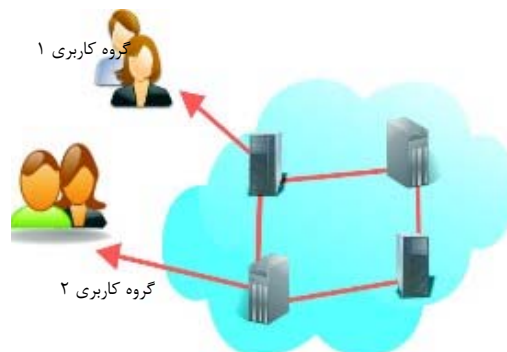
آنرا به سرور سوم تحویل می‌دهد و به همین ترتیب. هنگامیکه درخواست ارسال شد، شماره سریال مربوطه و لیست سرور نیز به همراه آن ارسال می‌شود، شماره سریال ارسالی، سگمندی را نشان می‌دهد که در سرور بعدی در حال پردازش است و لیست سرور شامل آدرس سرورهای است که در حال پردازش سگمنتها هستند. در حالی یک سرور جدید انتخاب می‌شود که درون لیست سرور ها نباشد. در صورت ارسال درخواست هر سرور شناسه خود را درون این لیست اضافه می‌کند. این پروتکل به شکل چند سرور یک جریان داده توصیف شده است. هنگامی که یک سرور جدید به درخواست پاسخ نداد سایر سرورها آنرا درون لیست سروری خود نشان نمی‌دهند. در این مدل سرورهایی که درون سرور لیست نیستند مشکل دارند. این پروتکل با دو رخداد توصیف شده اند. فایده پروتکل DCP این است که شبکه توازن بیشتری دارد و گم شدن پکت بسیار کم می‌شود. که در شبیه سازی مدل به آن پرداخته شده است. این مدل در صورتی خوب کار می‌کند که تمامی سرورها ظرفیت یکسانی داشته باشند، بطوریکه یک درخواست وارد شده بتواند بطور موازی مابین آنها توزیع شود. ولی هنگامی که سرورهای مجزای درون گروه متفاوت باشند استفاده هیبرید از دو پروتکل DCP و CCP بهینه می‌باشد.

#### ۴- ارزیابی کارایی مدل:

در این بخش به شبیه سازی و ارزیابی مدل با دو مقیاس انسداد و گم شدن یا تأخیر پکتها خواهیم پرداخت. مدل این معماری برای ۵ ویدئو سرور  $(S_1 \dots S_5)$  در شبکه اینترنت تعبیه شده که طبق ساختار سلسله مراتبی LANs, MANs, and WANs. جریان داده ای با پروتکل UDP سرویس داده می‌شود ولی بعلت اهمیت پیام برای ارسال TCP استفاده شده است، بر روی سرورها TP را اعمال می‌کنیم، همانطور که در جدول (3)، (2) گروه بندی شده اند، درون هر گروه پروتکل‌های DCP و CCP بطور مناسب استفاده شده است. کنترل پذیرش بر

#### ۳-۴-۲- کنترل توزیع شده (DCP) :

اکثر روشهای کنترل مرکزی ایجاد سربار و بروز خطا در خود سرور مرکزی را دارند. منابع محدود سرور مرکزی نیز می‌تواند بعنوان یک سربار در نظر گرفته شود. ولی پیاده سازی آن ساده است.



شکل ۲-۲: پروتکل مدل کنترل توزیع شده

برای مقایسه مدل CCP و مدل جدید نشان می‌دهیم که معماری توزیع شده کارایی بیشتری دارد. در این پروتکل از یک روش عبور نشانه برای شکستن و توزیع مساوی درخواستهای رسیده مابین گروه های سروری بکار گرفته شده است. این توزیع باعث می‌شود که ترافیک در سراسر شبکه با چند مسیریابی متفاوت پخش شود. در شکل (۲-۲) پروتکل توزیع شده را ترسیم شده است. هر دو روش با توجه به مقیاسهای گم شدن پکت و تأخیر ارسال پکت مقایسه خواهد شد. این دو پروتکل در [31] ارائه شده است. هر سرور می‌خواهد در گروه بخشهای ویدئویی خودش را به کلاینتها اختصاص دهد، تعداد سرورهای درون گروه  $K$  فرض شده است. زمانی که درخواست جدید از یک کاربر توسط سرور دریافت گردد این درخواست به  $K$  سگمنت شکسته شده و هر یک از این سگمنت ها  $n$  ثانیه برای ارسال جریان نیاز داشته باشند، و به هر سگمنت یک شماره سریال  $i$  نسبت داده شود و  $(1 \leq i \leq k)$  می‌باشد. سروری که درخواست را دریافت نمود اولین سگمنت را به جریان می‌اندازد. گردش سگمنت اول تا اینکه کامل شود  $d$  ثانیه طول می‌کشد، و آنرا برای کنترل به سایر سرور ها ارسال می‌کند. این انتخاب کاملاً تصادفی می‌باشد. دومین سرور دومین سگمنت را پردازش نموده و پس از کنترل

جدول ۴: پارامترهای شبیه سازی

$N_w = 1$	تعداد شبکه های WAN
$N_m = 3$	تعداد شبکه های MANs/WAN
$N_l = 15$	تعداد شبکه های LANs/WAN
$S_w = 5$	تعداد نود های WAN
$S_m = 10$	تعداد نود های MAN
$S_l = 20$	تعداد نود های LAN
$R_w = 3, R_m = 2, R_l = 1$	افزودگی روی شبکه های داخلی
$R_{ww} = 3, R_{ml} = 1$	افزودگی روی شبکه های داخلی
10 Mbps	سرعت شبکه
5	تعداد سرورها
15 min	میانگین زمان درخواست ویدئو
100 kbps CBR	ترافیک ارسال اطلاعات
300-5-- h	زمان شبیه سازی

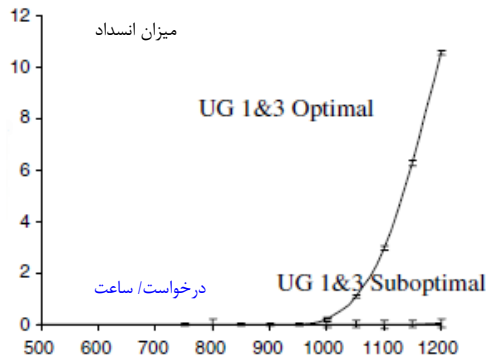
#### ۴-۱- نتایج شبیه سازی مدل طراحی انسداد

برای شبیه سازی این مدل سه آزمایش کارایی انسداد با معماری توزیع شده انجام داده ایم. مقیاس استفاده شده در این آزمایش درصد درخواستهای بلوک شده از هر گروه کاربر و مجموع آنها در سیستم می باشد. برای اولین آزمایش از ماتریس انتقال که در جدول (۲) نشان داده شده است استفاده شده است. ظرفیت جریان داده ای ترکیب شده ۵ سرور ۳۰۰ درخواست می باشد. ما فرض می کنیم هر گروه کاربر درخواستهای به تناسب ۱:۱:۱ را تولید نمایند، هدف از این آزمایش تخصیص بهینه منابع از طریق مدل TP و بیان کارایی انسداد می باشد. مقادیر بدست آمده میزان نزدیکی نودهای کلاینت به سرور را مشخص می کند. اما مقادیر RTT با در نظر گرفتن میزان تراکم بیشتر خواهد شد. در آزمایش اول کارایی انسداد را بر اساس تخصیص بهینه که در جدول (۳) بدست آمد بررسی کردیم. در شکل (۳-۱) کارایی انسداد برای تمامی سیستمها نشان داده شده است. تعداد درخواستها بر ساعت ۶۰۰ تا ۱۲۰۰ می باشد و میانگین طول زمان هر درخواست ۱۵ دقیقه است که بصورت توزیع نمایی تولید شده است. در آزمایش دوم از ماتریس انتقال جدول (۵) استفاده کرده ایم،

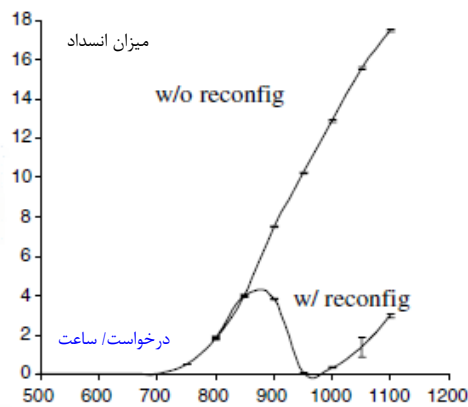
اساس آستانه، سرورها را مجبور به اشتراک حقیقی اطلاعات برای کاربران می کند، دو مدل ترافیک واقعی جریان مالتی مدیا و ترافیک پشت آن در حالت واقعی بررسی شده است. مقیاسهای برآورد آنها یکی درصد بستن درخواستهای نودهای سرور نسبت به گروههای کاربران و دیگری میزان گم شدن و تأخیر پکت می باشد. در مرحله اول نتایج تحلیل کارایی انسداد بدون سازماندهی مجدد سرورها بررسی شده است. در بخش (2-4) به تحلیل کارایی دو مدل CCP و DCP خواهیم پرداخت. همه نتایج براساس میانگین ۵ آزمایش بدست آمده است، هر آزمایش برای ۳۰۰ الی ۵۰۰ ساعت شبیه سازی شده است و میزان متوسط درخواستها مابین ۵۰۰ الی ۱۲۰۰ درخواست در نظر گرفته شده است. که ترکیب این ۵ آزمایش میانگین آماری مناسبی دارد. برای شبیه سازی این مدل از ورژن 2.26 Ns استفاده شده است، در [35] استفاده از این نرم افزار برای ایجاد توپولوژی شبکه بیان شده است، این مقادیر واقعی برای انواع مختلف شبکه (nodes/LAN, ) (LANs/MAN, MANs/WAN) استفاده شده است، بخش اعظمی از ترافیک اینترنت مربوط به TCP می باشد و به تناسب UDP ترافیک کمتری را به خود اختصاص می دهد، در TCP با TelNet و FTP از الگوی پواسن استفاده شده است. Session های مربوط به TelNet و FTP می توانند با یک توزیع پارتو (در حقیقت بیانگر توزیع درآمد است که بر اساس آن رابطه بین درآمد شخصی و جمعیت بررسی می شود) بررسی شوند. ترافیک پشت صحنه می تواند با تناوب روشن/خاموش نمودن توزیع پارتو مدل شود [28]. در جدول ۴ پارامترهای شبیه سازی نشان داده شده است.



در جدول (۲) تمامی درخواستها برای تمامی گروههای کاربر اجرا شده است و انتظار می رود که کارایی انسداد بطور مشابه منحنی بهینه نشان داده شده در شکل (۳و۲) باشد.



شکل ۳-۳: کارایی انسداد برای گروه کاربران ۱ و ۳



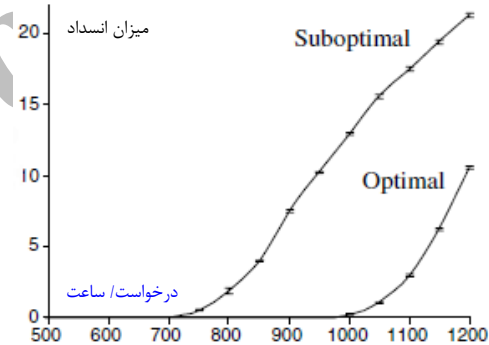
شکل ۳-۴: براساس پیکر بندی مجدد و بدون آن

توجه: منحنی کارایی که برچسب بهینه دارد نحوه دسته بندی مطلوب سرورها را نشان می دهد. منحنی کارایی که برچسب زیربهینه دارد گروه های سروری هستند که با آنها برای تغییر درخواست به ملایمت رفتار شده است. نتایج آزمایش دوم هنگامیکه درخواست کاربر تغییر نمود در منحنی زیربهینه در شکلهای (۳-۳و۱-۳) نشان داده شده است. کارایی نشان می دهد که سرورها برای یافتن شکل بهینه باید مجدداً پیکربندی شوند. به شکلی که در جدول (۵) نشان داده شده است درخواست گروه کاربر ۱ و ۳ از ۱۰۰ به ۸۰ کم شده است و درخواست گروه ۲ به میزان ۴۰ واحد بالا رفته است، در این شبیه سازی به تناسب ۱:۱,۷۵:۱ تولید شده است، گروه کاربر ۲ به میزان ۱,۷۵ زمان بیشتری از درخواستهای گروه

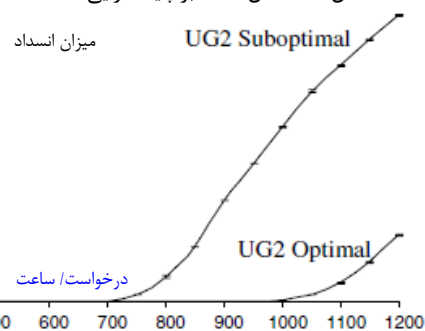
جدول ۵: ماتریس انتقال

مکانیزم ظرفیت	گروه سوم	گروه دوم	گروه اول	
S1	۳۹۹	۳۵۰	۲۷۶	
S2	۴۷۵	۴۲۶	۲۶۵	
S3	۳۸۳	۲۷۴	۳۲۹	
S4	۵۱۱	۲۹۸	۴۵۷	
S5	۳۰۱	۴۲۲	۴۱۷	
Demand	۱۰۰	۱۰۰	۱۰۰	۳۰۰

ظرفیت ترکیب شده جریان داده ۳۰۰ می باشد، هر چند که درهم سازی درخواستهای سه کاربر به نسبت ۱:۱,۷۵:۱ تغییر نمود. مقصود از این آزمایش این است که نشان دهیم که در تخصیص بهینه جدول (۳) بهینگی طولانی با استفاده از انسداد کلی نخواهیم داشت.



شکل ۳-۱: مدل کاملاً بر پایه کارایی انسداد



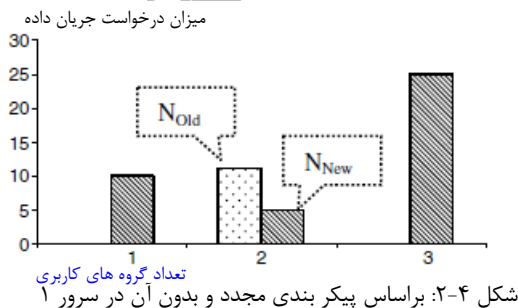
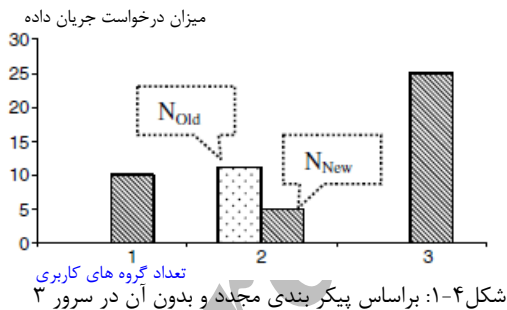
شکل ۳-۲: کارایی انسداد برای گروه کاربری ۲

سیستم یک انسداد بسیار آرام را نشان می دهد که طی ۱۱۰۰ درخواست پس از مدتی به طور متوسط بالا می رود. شکل (۳-۱) میزان کارایی انسداد را برای گروههای کاربر ۱ و ۳ نشان می دهد. دیده می شود که

در این فاز انسداد ادامه پیدا خواهد کرد و بالاخره شروع به کم شدن می‌کند. این رفتار در شکل (۳-۴) اثبات شده است. بدون پیکربندی مجدد همگی انسداد موجود در فضای نمونه  $1100 \text{ request/h}$  به مقدار  $5,17$  درصد است در صورتی که در همان حجم کار میزان آن با پیکربندی مجدد به میزان  $235$  درصد پائین می‌آید.

#### ۴-۱-۱- آنالیز فاز انتقال :

در شکل (۴-۱ و ۴-۲)، یک تصویر لحظه‌ای از درخواستهای در حال سرویس روی سرورهای ۱ و ۳ پس از  $300$  ساعت با  $850 \text{ stream/h}$  از زمان پیکربندی مجدد نشان داده شده است، سرور  $S3$  در تخصیص جدید درخواست کمتری از گروه کاربران ۲ دارد پس درخواستهای تخصیص داده شده در پیکربندی قبلی در حال پردازش باقی می‌مانند و آنگاه سرویس دهی به درخواستهای وارد شده جدید در این سرور ادامه می‌یابد. این توضیح مربوط به برآمدگی در منحنی شکل (۳-۴) می‌باشد.

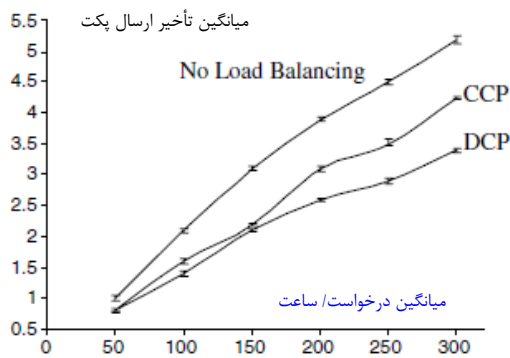


در فاز انتقال می‌توان درخواستها را از سرورهایی که ملزم به انجام آن می‌باشند به سرورهایی که به میزان

های ۱ و ۳ تولید می‌کند. و ماتریس هزینه RTT با مقادیر جدید بدست آمده بروز می‌شود که در شکل‌های (۳-۱ و ۳-۲) نشان داده شده است، تخصیص جدول (۳) دیگر بهینه نیست و موجب انسداد زیادی در گروه ۲ می‌شود. انسداد برای گروه‌های ۱ و ۳ در صورتی که سرورهای اختصاص داده شده به آنها به میزان کافی استفاده نشود نزدیک به صفر است. تمامی انسدادها همانطور که در شکل (۳-۱) نشان داده شد بالا هستند (روی آن زیربینه) نوشته شده است. و روی شکل (۳-۴) عبارت پیکربندی مجدد W/O نوشته شده است. سومین آزمایش شامل اجرای راه حل انتقال بر روی ماتریس جدید جدول (۵) می‌باشد. نتایج این تخصیص در جدول (۶) نمایش داده شده است و تخصیص قبلی درون پرانتز نشان داده شده است. در شکل (۳-۴) میزان کارایی انسداد با استفاده از پیکر بندی مجدد سرورها و بدون در نظر گرفتن آن نشان داده شده است. پیکر بندی مجدد وقتی شروع می‌شود که مقدار کارایی انسداد از میزان آستانه آن بیشتر شود. در این آزمایش  $3,5$  درصد برای  $850$ ، جریان داده بر ساعت بعنوان آستانه پیکربندی مجدد در نظر گرفته شده است. اگر میزان تخصیص بهینه باشد انسداد در این حجم کاری باید صفر شود که در شکل (۳-۱) آورده شده است. شکل (۳-۴) مقایسه کارایی مابین تخصیص قبلی و جدید را نشان می‌دهد. پیکربندی مجدد جهت بهینه سازی می‌باشد و باید میزان انسداد را کم کند. طی این فاز انتقال برخی از درخواستهای مربوط به تخصیص قبلی دایر هستند و قبل از اینکه تخصیص جدید صورت پذیرد نیاز دارند که تکمیل شوند (در بخش ۳-۳ به آن پرداختیم)

جدول ۶: جدول تخصیص بهینه جدید

	گروه اول	گروه دوم	گروه سوم
S1	۰ (۳۰)	۹۵ (۲۰)	۰ (۴۵)
S2	۷۰	۰	۰
S3	۱۰ (۰)	۵ (۴۰)	۲۵ (۰)
S4	۰	۴۰	۰
S5	۰	۰	۵۵



شکل ۵-۲: میزان تأخیر دریافت پکت‌ها در پروتکل‌های ارائه شده شکل (۱-۵) میانگین گم شدن پکت را هنگامی که میزان جریان داده‌ها افزایش یافت را نشان می‌دهد. زمانیکه به بالاترین درجه توزیع توازن در کل شبکه دست می‌یابد پروتکل DCP بهتر انجام می‌شود. هر درخواست کلاینت توسط چندین مسیریابی متفاوت منتشر می‌شود. بنابراین مسیریابی در شبکه میزان بالایی از ترافیک را به خود اختصاص می‌دهد، به علت تراکم محلی گم شدن پکتها مینیمم شده است. زمانیکه وضعیت کلی سیستم بروز رسانی شد CCP بدتر انجام می‌پذیرد، فقط در این حالت توزیع بار بطور متناوب ایده آل نیست. در شکل (۵-۲) میانگین تأخیر پکتها در یک حجم مشابه که قبلاً نشان داده شده بود آمده است، زمانیکه هر جریان داده توسط تمامی سرورها سرویس دهی شود کارایی DCP از مدل CCP کمتر است (با توجه به دور و نزدیک بودن از کلاینت). هر دو پروتکل بهتر از روش پایه انجام می‌شوند. در [31] بطور کامل مورد بررسی قرار گرفته است. CCP و DCP پروتکل‌هایی هستند که کارایی متفاوتی دارند نمی‌توان گفت که هیچیک از دیگری بهتر است. پیاده سازی مدل CCP بسیار ساده تر است ولی تخصیص بهینه سرورها را تضمین نمی‌کند. مدل DCP زمانیکه گروه سرورها از لحاظ ظرفیت یکسان باشند بهتر کار می‌کند. و با توجه به این دو پروتکل کارایی انسداد ساده می‌شود.

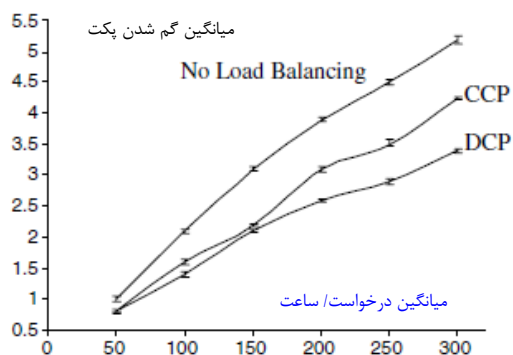
#### ۵-نتیجه گیری :

در این گزارش مدلی تحلیلی از تخصیص ظرفیت سرورها به درخواستهای کاربران بصورت پویا نشان داده شد، با ثبت RTT بعنوان مقیاس هزینه در مدل انتخاب سرورها به فرضیات نزدیک شدیم، در حال حاضر در CDNs توسط Real Networks و سایر ارائه کنندگان

کافی استفاده نشده مهاجرت داد. تصویری از سرور S1 در شکل (۳-۴) نشان می‌دهد که تخصیص جدید به گروه کاربران ۲ بسیار بیشتر از تخصیص قدیم به آنها می‌باشد. به هر حال تمامی درخواستهای در حال انجام باید تکمیل شوند تا تخصیص جدید شروع گردد (صفر برای گروه ۱ و ۹۵ برای گروه ۲ و ۰ برای گروه ۳). در فاز انتقال، آستانه جدید حدود ۰.۵ درصد از میزان قبلی که ۳.۵ درصد بود بیشتر شد، قبل از اینکه کاهش آنرا بدلیل پیکربندی مجدد بینیم (در شکل ۳-۴ این موضوع نشان داده شده است).

#### ۴-۲- نتایج تحلیل شبیه سازی پروتکلها :

در این بخش آنالیز کارایی دو مدل CCP و DCP دریک گروه بررسی شده است، مقیاس استفاده شده درصد پکت‌های گم شده یا خراب شده برحسب میلی ثانیه می‌باشد. یک گروه مشتمل بر ۴ ویدئو سرور یکسان استفاده شده است، میانگین تعداد درخواستهای برای ۱۰۰ کلاینت بین ۵۰ الی ۳۰۰ درخواست در ساعت متغیر است. کلاینتها درخواستهایشان را بر اساس توزیع پواسن تولید می‌کنند. پارامترهای مربوط به این شبکه در شکل (۴) نشان داده شده اند. این دو پروتکل با یکدیگر مقایسه شده اند و مبنای کارشان این است که ترافیک در سراسر شبکه توزیع شود. سروری که درخواست کاربر را دریافت کرد باید آنرا بطور کامل پردازش کند. اساس این سناریو وضعیتی را معین می‌کند که اینجا نیست، درحالیکه هر دو CCP و DCP در طراحی توازن بار بهتری را پیشنهاد نموده اند.



شکل ۵-۱: میزان گم شدن پکت‌ها در پروتکل‌های ارائه شده

		Destination					Supply	$u_i$
		1	2	3	4	5		
Source	1	16	20	16	13	22	17	50
	2	14	14	13	19	15		60
	3	19	19	20	23	30	10	50
	4(D)	M	0	M	0	0	50	50
Demand		30	20	70	30	60		$Z = 2,470 + 10M$

شکل ۶: روش گوشه شمال غرب

مرحله مقدار دهی اولیه: بدست آوردن راه حل شدنی از

طریق یکی از سه روش بالا

**تست بهینگی:** از توابع  $u_i$  و  $u_j$  مشتق گرفته و مشتق را مساوی صفر قرار دهیم. برای گرفتن سطری که بالاترین مقدار تخصیص را دارد و این مجموعه معادلات را برای هر  $i$  و  $j$  حل می کنیم:

$$C_{ij} = u_i + v_j$$

با توجه به اینکه  $x_{ij}$  متغیر اصلی می باشد. راه حل جاری وقتی بهینه است که

$$(C_{ij} - u_i - v_j) \geq 0$$

برای تمامی  $x_{ij}$  های غیر اساسی، اگر مقدار  $(C_{ij} - u_i - v_j) < 0$  آنگاه مدل را تکرار می کنیم.

#### مرحله تکرار:

متغیر غیر پایه  $x_{ij}$  را بعنوان بزرگترین مقدار منفی و بعنوان متغیر پایه وارد شده می گیریم.

مقدار متغیر پایه گرفته شده را کوچکترین مقدار زنجیره اعضا برای نگاهداری (شدنی بودن) تخمین می زنیم، این کار مجموعه ای از اعضای داده شده و گرفته شده را به ما می دهد، و بیان می کند کجا بترتیب تخصیص افزایش یافته و کجا کم می شود. تابع شدنی جدید را تخمین می زنیم با اضافه نمودن مقدار متغیر های گرفته شده برای تخصیص به اعضای گرفته شده و این مقدار را از مجموعه داده شده کم می کنیم. در نهایت راه حل شدنی بدست آمده را تست می کنیم.

#### مراجع:

- [1] Akamai Technologies, Akamai Streaming—When Performance Matters, White Paper, 2004.
- [2] RealNetworks, RealNetworks Production Guide for Release 8, September 2002.
- [3] Alexander Ferreira, Optimizing Microsoft Windows Media Services 9 Series, Technical

این وب سرویس از گروه های سروری استفاده می شود. اگر از راه حل های معماری گروه ها و آزمایشات بهینه برای جریان داده استفاده شود، کارایی بالاتری خواهیم داشت. یک مدل واقعی از اینترنت والگویی ترافیک آن در یک استراتژی شبیه سازی گسترده بکار برده شد. تحلیل و طراحی پروتکل های جریان داده و یک معماری جامع رودرو برای جریان داده ارائه نمودیم.

**ضمیمه ۱: TP** حالت خاصی از برنامه ریزی خطی می باشد که برای توزیع منابع به گیرندگان میباشد، بطور کلی منابع  $i$  با ظرفیت  $v_i$  واحد برای توزیع اطلاعات مابین گیرندگان و گیرندگان  $j$  به میزان  $d_j$  واحد می توانند از منابع درخواست نمایند. فرض اصلی در این مدل میزان هزینه توزیع اطلاعات میباشد،  $C_{ij}$  به طور مستقیم با تعداد توزیع اطلاعات مابین منبع  $i$  و گیرنده  $j$  متناسب است. می خواهیم هزینه را با توجه به  $RTT$  مینیمم کنیم. محاسبات این مدل در اندازه بزرگ بسیار پیچیده است. برای بدست آوردن یک راه حل شدنی میتوان از این سه راه حل استفاده نمود:

۱. قاعده گوشه شمال غرب
۲. روش تخمین ووگل
۳. روش تخمین راسل

راه حل شدنی اولیه به شکل تکراری برای بدست آوردن راه حل بهینه استفاده می شود که از هر دو روش ۲ و ۳ استفاده نموده است. روش قاعده گوشه شمال غرب: از گوشه شمال غربی جدول شروع به کار میکند یعنی از  $x_{11}$ ، اگر  $x_{ij}$  آخرین متغیر اصلی انتخاب شده باشد، برای انتخاب  $x_{ij+1}$  یک ستون به سمت راست منتقل می شود، در صورتی که منبع  $i$  باقیمانده موجودی داشته باشد ولی در غیر اینصورت به سطر پائین می رود. راه حل شدنی اولیه با این روش حاصل می شود ولی بدلیل تکرارهای زیاد نسبت به دو مدل بعدی خوب نیست.



- ACMSIGCOMMWorkshop on InternetMeasurement, 2002.
- [16] M. Guo, Q. Zhang, W. Zhu, Selecting path-diversified servers in content distribution networks, in: Proc. of IEEE GLOBECOM, 2003.
- [17] T. Nguyen, A. Zakhor, Multiple sender distributed video streaming, IEEE Transactions on Multimedia 6 (2) (2004).
- [18] T. Nguyen, A. Zakhor, Distributed video streaming over the Internet, in: Proc. of MMCN, 2002.
- [19] V.N. Padmanabhan, H. Wang, P. Chou, K. Sripanidkulchai, Distributing streaming media content using cooperative networking, in: Proc. NOSSDAV, 2002.
- [20] M. Doar, A better model for generating test networks, in: Proc. of GLOBECOM, 1996.
- [21] M. Mellia, R. Lo Cigno, F. Neri, Measuring IP, TCP behavior on the edge node, Planet IP and Nebula Joint Workshop, 2002.
- [22] H.J. Burch, F. Ercal, Mapping the Internet, Computer 32 (4) (1999).
- [23] A. Broido, K. Claffy, Internet topology: connectivity of IP graphs, CAIDA (July) (2001).
- [24] S. McCreary, K. Claffy, Trends in Wide Area IP Traffic Patterns. Available from: <<http://www.caida.org/outreach/papers/>>.
- [25] E. Zegura, K.L. Calvert, M.J. Donahoo, A quantitative comparison of graph-based models for Internet topology, IEEE/ACM Transactions on Networking 5 (6) (1997).
- [26] K.L. Calvert, M. Doar, E. Zegura, Modeling Internet topology, IEEE Communications Magazine (1997).
- [27] The Mercator Internet Mapping Project at <<http://www.isi.edu/scan/mercator/maps.htm>>.
- [28] T.D. Neame, M. Zukerman, Modeling broadband traffic streams, in: Proc. of GLOBECOM, 1999.
- [29] F. Fitzek, M. Reisslein, MPEG-4 and H.263 video traces for network performance evaluation, IEEE Network Magazine 15 (6) (2001).
- [30] P. Mundur, R. Simon, A. Sood, End-to-end analysis of distributed video on demand systems, IEEE Transactions on Multimedia (February) 6 (1) (2004).
- Brief, Microsoft New Media Platforms Division, February 2003.
- [4] QuickTime Streaming: End-to-end Solutions for Live Broadcasting and On-Demand Streaming of Digital Media, 3620 P. Mundur, P. Arankalle / Computer Networks 50 (2006) 3608–3621 Technology Brief, Mac OS X Server: QuickTime Streaming, 2004.
- [5] Akamai Technologies, Internet Bottlenecks: the Case for Edge Delivery Services, White Paper, 2000.
- [6] M. Andrews, B. Shepherd, A. Srinivasan, P. Winkler, F.Zane, Clustering and server selection using passive monitoring, in: Proc. IEEE INFOCOM, 2002.
- [7] A. Myers, P. Dinda, H. Zhang, Performance characteristics of mirror servers on the Internet, in: Proc. of IEEE INFOCOM, 1999.
- [8] M. Crovella, R. Carter, Dynamic server selection using bandwidth probing in wide-area networks, in: Proc. of IEEE INFOCOM, 1997.
- [9] M. Conti, E. Gregori, F. Panzieri, Load Distribution among Replicated Web Servers: A QoS-based Approach, WISP, ACM Press, 1999.
- [10] R.B. Bunt, D.L. Eager, Achieving load balance and effective caching in clustered web servers, in: Proc. of the 4th International Web Caching Workshop, San Diego, CA, 1999.
- [11] M. Colajanni, P.S. Yu, D.M. Dias, Analysis of task assignment policies in scalable distributed web-server systems, IEEE Transactions on Parallel and Distributed Systems 9 (6) (1998) 585–600.
- [12] P. Rodriguez, E.W. Biersack, Dynamic parallel access to replicated content in the Internet, IEEE Transactions on Networking 10 (4) (2002).
- [13] L. Amini, H. Schulzrinne, On probe strategies for dynamic multimedia server selection, in: Proc. of ICME, August 2002.
- [14] P. Francis, S. Jamin, C. Jin, V. Paxson, D. Raz, Y. Shavitt, L. Zhang, IDMaps: a global Internet host distance estimation service, in: Proc. IEEE INFOCOM, March 2000.
- [15] K.P. Gummadi, S. Saroiu, S.D. Gribble, King: estimating latency between arbitrary Internet hosts, in: Proc. of the 2nd



- [31] A. Matthur, P. Mundur, Dynamic load balancing across mirrored multimedia servers, in: Proc. of ICME, 2003.
- [32] F. Hillier, G. Lieberman, Introduction to Operations Research, McGraw-Hill Publishers, 1986.
- [33] P. Mundur, A. Sood, R. Simon, Class-based access policies for distributed video on demand systems, IEEE Transactions on Circuits and Systems for Video Technology 15 (7) (2005).
- [34] NS—Network Simulator at <http://www.isi.edu/nsnam/ns/>.
- [35] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, W. Willinger, Network topology generators: degree-based vs. structural, in: Proc. of SIGCOMM, 2002.

Archive of SID