

مدل سازی و ارزیابی کارآیی یک مرکز سرویس پیام چندرسانه‌ای با در نظر گرفتن ترافیک پیام‌های ارسال مجدد شده

علی موفر
دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف
movaghar@sharif.edu

رضا متین‌نژاد
دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف
matinnejad@ce.sharif.edu

پیام، یک مدل ساده M/M/1 نمی‌تواند برای مدل‌سازی سیستم MMS به کار رود، مدلی از سرور/تقویت‌کننده سرویس پیام چند رسانه‌ای یا MMS-RS ارائه کرده است و به کمک یک روش چرخشی به تعیین مقدار بهینه‌ی طول بافر پرداخته است.

در سیستم پیام‌رسانی شبکه‌های GPRS/UMTS در صورتی که در فاصله‌ی انتظار پیام در صف، دریافت‌کننده‌ی پیام به محدوده‌ی ند سرویس‌دهنده‌ی شبکه‌ی رادیویی یا SGSN دیگری وارد شود، پیام موردنظر بلافاصله از صف SGSN فعلی خارج شده و در صف SGSN جدید قرار می‌گیرد. مرجع [9] با توجه به این کارکرد، مدل کامل‌تری از سیستم ارائه کرده است، که در آن نرخ ورود پیام‌های ارسال شده از SGSN های دیگر و نرخ ارسال پیام به دیگر SGSN ها نیز در نظر گرفته شده است و به کمک مدل ارائه شده روابطی برای زمان انتظار و احتمال از دست رفتن پیام‌ها، استخراج شده است.

در [7,8] مدلی به کار برده شده که تفاوت عمده‌ی آن با مدل‌های قبلی در این است که فرآیندهای تصادفی مربوط به تمامی پارامترهای سیستم، متغیر با زمان در نظر گرفته شده‌اند و فرضی در مورد نمایی یا پواسن بودن فرآیندها نشده است. با استفاده از این روش الگوریتمی برای تعیین مقدار بهینه‌ی طول بافر به صورت پویا، با توجه به تغییرات ترافیک سیستم، ارائه شده است.

گفته شد که هرگاه در فاصله‌ی انتظار پیام در صف، دریافت‌کننده‌ی پیام به محدوده‌ی SGSN دیگری وارد شود، پیام موردنظر بلافاصله از صف SGSN فعلی خارج شده و در صف SGSN جدید قرار می‌گیرد. در چنین شرایطی این گونه پیام‌ها ممکن است بیش از حد تحت تأثیر زمان طولانی انتظار در صف قرار بگیرند و زمان انقضای آنها فرا برسد. برای غلبه بر این مشکل روشی توسط [6] پیشنهاد شده است، که در آن با تخصیص بافری به طول N به پیام‌هایی که از سرورهای مجاور ارسال شده‌اند، احتمال انقضای آنها کاهش می‌یابد.

چکیده: در این مقاله، مدلی از یک مرکز سرویس پیام چندرسانه‌ای ارائه شده و مورد ارزیابی قرار گرفته است. یکی از بارزترین ویژگی‌های یک مرکز سرویس پیام چندرسانه‌ای، انجام وظیفه‌ی ذخیره و ارسال مجدد پیام‌های ناموفق در ارسال‌های قبلی است. در مدل‌هایی که تاکنون برای ارزیابی کارآیی مرکز سرویس پیام چندرسانه‌ای ارائه شده‌اند، از این ویژگی صرف‌نظر شده است. در این مقاله ترافیک پیام‌های ارسال مجدد شده نیز در مدل پیشنهادی در نظر گرفته شده است. با استفاده از نتایج تحلیل مدل، روشی برای به دست آوردن حدبالای متوسط زمان انتظار پیام‌ها در صف و حدبالای احتمال از دست رفتن پیام‌ها ارائه شده است. همچنین مدل پیشنهادی و دیگر نتایج به دست آمده، برای در بر گرفتن ترافیک پیام‌های ارسال مجددشده به تعداد مراتب دلخواه، گسترش داده شده‌اند.

واژه های کلیدی: سرویس پیام چندرسانه‌ای، مدل‌سازی، ارزیابی کارآیی، ارسال مجدد.

۱- مقدمه

در دهه‌ی اخیر شاهد افزایش روز افزون مشترکین تلفن‌های همراه بوده‌ایم که به نوبه‌ی خود باعث افزایش قابلیت دستگاه‌های تلفن همراه و بهبود سرویس‌های ارائه شده به مشترکین شده است. موفقیت سرویس‌هایی چون سرویس پیام کوتاه یا SMS و آهنگ‌ها و بازی‌های قابل دانلود، باعث ارائه‌ی سرویس‌های تواناتری هم‌چون سرویس پیام چندرسانه‌ای یا MMS بر پایه‌ی سرویس عمومی بسته‌های رادیویی یا GPRS و سیستم جهانی ارتباطات بی‌سیم یا UMTS، در حوزه‌ی پیام رسانی بی‌سیم شده است.

با توجه به این روند، نیاز زیادی به ارائه راهکارهایی برای بهبود عملکرد این سرویس‌ها وجود دارد [3]. در همین راستا کارهای ارزشمندی در میان مقالات علمی معتبر به چشم می‌خورد. مرجع [5] با تأکید بر این نکته که با توجه به زمان محدود انتظار پیام‌ها یا همان زمان انقضای

دلایل عدم موفقیت در ارسال یک پیام را می‌توان در سه دسته کلی قرار داد [1,12]:

۱. غیر قابل ارسال: در مورد پیام‌هایی که در دسته غیر قابل ارسال قرار می‌گیرند، سیستم تلاشی برای ارسال مجدد آنها انجام نخواهد داد. به عنوان نمونه در شرایطی که مقصد قابلیت دریافت پیام را نداشته باشد و یا اینکه، پیام به اشتباه آدرس‌دهی شده باشد، پیام در این دسته قرار می‌گیرد.

۲. غایب: در مورد پیام‌هایی که در این دسته قرار می‌گیرند، سیستم بر اساس تریگری که از مقصد دریافت می‌کند اقدام به ارسال مجدد این پیام‌ها می‌کند. به عنوان نمونه در شرایطی که دستگاه تلفن همراه مقصد خاموش باشد و یا خارج از برد رادیویی باشد، پیام در این دسته قرار می‌گیرد. در این شرایط پیام در انتظار ارسال مجدد باقی می‌ماند، تا اینکه دستگاه تلفن همراه، مجدداً خود را در شبکه ثبت کند، به این معنی که تلفن همراه سیگنال‌های حضور خود را به شبکه بی‌سیم ارسال کند.

۳. عدم موفقیت موقت: در مورد پیام‌هایی که در دسته عدم موفقیت موقت قرار می‌گیرند، سیستم ارسال مجدد را پس از زمان تأخیری که وابسته به پیکربندی شبکه است، انجام می‌دهد. به عنوان نمونه در شرایطی که حافظه‌ی دستگاه تلفن همراه مقصد پر باشد و یا یک خطای سیگنالینگ در شبکه بی‌سیم رخ داده باشد، پیام در این دسته قرار می‌گیرد.

هر پیام چندرسانه‌ای که به یکی از دلایل فوق در ارسال به مقصد ناموفق باشد، در حافظه‌ی دائمی MMS-RS ذخیره می‌شود. پس از سپری شدن زمان تأخیر مناسب و با فراهم شدن شرایط ارسال مجدد پیام، MMS-RS پیام مورد نظر را مجدداً در بافر ارسال قرار می‌دهد. در صورتی که پیام در ارسال مجدد، باز هم با موفقیت به مقصد نرسد، سناریوی ارسال مجدد آن دومرتبه تکرار می‌شود. پس از چندین تلاش مجدد، در نهایت یا پیام با موفقیت به مقصد می‌رسد، و یا اینکه زمان انقضای آن فرا می‌رسد و از ارسال آن پیام صرف‌نظر می‌شود.

در ادامه و پس از تعیین پارامترهای ترافیک پیام‌های ارسال مجدد شده، مدلی از MMS-RS با در نظر گرفتن ترافیک این پیام‌ها ارائه می‌کنیم.

۱-۲ بررسی و تعیین پارامترهای ترافیک پیام‌های

ارسال مجدد شده

در ابتدای این بخش تعریفی از مرتبه‌ی یک پیام ارسال مجدد شده، مطابق رابطه‌ی ۱، ارائه می‌دهیم.

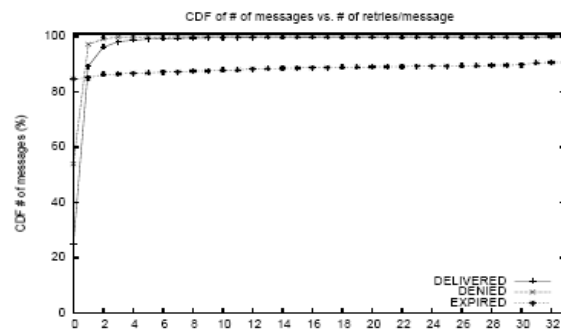
$$(1) \quad \text{مرتبه پیام} = \text{تعداد تلاش‌های قبلی برای ارسال پیام}$$

مطابق این تعریف پیام‌های مرتبه صفر، همان پیام‌های جدید ورودی به MMS-RS هستند. پیام‌های مرتبه یک، همان پیام‌های ناموفق در ارسال اول هستند که باید برای دومین بار ارسال شوند و پیام‌های مرتبه دو، پیام‌های ناموفق در ارسال‌های اول و دوم هستند، که باید برای سومین بار ارسال شوند.

مرجع [4] با ارائه تکنیکی مبنی بر تخصیص پویای پهنای باند به کاربران شبکه بی‌سیم، به بهبود کارایی این شبکه پرداخته است و با ارائه یک مدل زنجیره‌ی مارکوف دوبعدی به تحلیل کارایی این روش پرداخته است.

یکی از بارزترین ویژگی‌های شبکه سرویس پیام چندرسانه‌ای، ویژگی ذخیره و ارسال است. پیام‌هایی که به هر دلیل در ارسال به مقصد ناموفق هستند، در MMS-RS ذخیره می‌شوند تا مجدداً ارسال گردند. این پیام‌ها یا در نهایت با موفقیت به مقصد ارسال می‌شوند و یا اینکه، پس از چندین بار تلاش برای ارسال آنها، زمان انقضای آنها فرا می‌رسد و به دور انداخته می‌شوند [1,12].

از آنجا که ترافیک حاصل از ارسال مجدد پیام‌ها، در شبکه‌هایی همچون شبکه‌های TCP/IP غالباً ناچیز است، صرف‌نظر کردن از ترافیک این پیام‌ها، یکی از فرضیات متداول در مدل صف این شبکه‌ها است. با این وجود، همانگونه که نمودار شکل ۱ نشان می‌دهد، درصد قابل توجهی از پیام‌هایی که وارد MMS-RS می‌شوند، در اولین ارسال خود با موفقیت به مقصد نمی‌رسند. با این وجود در تمامی کارهای بررسی شده، از ترافیک پیام‌های ارسال مجددشده در مدل‌سازی MMS-RS صرف‌نظر شده است.



شکل (۱): توزیع تجمعی تعداد پیام‌ها بر حسب تعداد تلاش‌های مجدد برای ارسال [3]

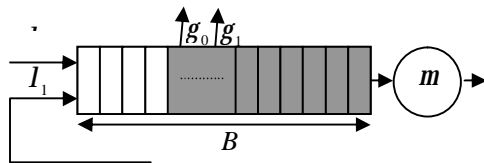
در ادامه‌ی این مقاله ابتدا به بررسی دقیق‌تر سناریوی ارسال مجدد یک پیام پرداخته شده است. سپس مدلی از MMS-RS، با در نظر گرفتن ترافیک پیام‌های ناموفق در ارسال اول، ارائه شده است و معادلات تعادلی محلی برای مدل استخراج شده است. به کمک مدل پیشنهادی، روشی برای به دست آوردن حدبالای متوسط زمان انتظار پیام‌ها در صف و حدبالای احتمال از دست رفتن پیام‌ها ارائه شده است. سپس، مدل مربوطه و دیگر نتایج به دست آمده، در مورد ترافیک پیام‌های ارسال مجدد شده تا مرتبه N ، تعمیم داده شده است. در انتها صحت نتایج به دست آمده به کمک شبیه‌سازی مورد تحقیق قرار گرفته است.

۲- مدل‌سازی پیام‌های ارسال مجدد شده

است، در نتیجه زمان انقضای پیام‌های مرتبه d را نیز به خوبی می‌توان بر اساس یک متغیر نمایی با پارامتر g_d مدل کرد. باید گفت که این فرض به مراتب از فرض مشابهی که در مورد پیام‌های مرتبه صفر در دیگر منابع انجام شده است به واقعیت نزدیکتر است. چرا که وقتی پیام‌ها برای اولین بار به MMS-RS وارد می‌شوند، کمتر پیامی ممکن است زمان انقضای کوچکی داشته باشد، ولی در توزیع نمایی احتمال تولید چنین مقادیری هم وجود دارد. در عین حال در مورد پیام‌های مرتبه d واضح است که ممکن است تعدادی از پیام‌ها در همان بدو ورود به صف دارای مقادیر انقضای زمانی کمی باشند.

۲-۲. آرایه مدلی از MMS-RS با در نظر گرفتن ترافیک پیام‌های ارسال مجدد شده

در شکل ۴ مدل اولیه‌ای از یک مرکز سرویس پیام چندرسانه‌ای ارائه شده است که در آن فقط ترافیک پیام‌های مرتبه صفر و یک در نظر گرفته شده است. در ادامه و پس از تحلیل این مدل، آن را برای دربرگرفتن پیام‌های مرتبه‌ی بالاتر گسترش می‌دهیم.



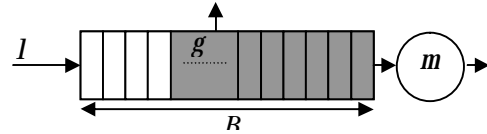
شکل (۴): مدل اولیه‌ی سرور پیام چندرسانه‌ای با در نظر گرفتن پیام‌های ارسال مجدد شده

باید دقت شود که در شکل ۴ نرخ انقضای g_0 در مورد پیام‌های مرتبه صفر و نرخ انقضای g_1 در مورد پیام‌های مرتبه یک است. با توجه به این که نرخ سرویس به پیام‌های مرتبه صفر و یک در مدل ارائه شده متفاوت است، یک زنجیره‌ی مارکوف دو بعدی را برای تحلیل این مدل به کار می‌بریم. در این مدل هر حالت سیستم با یک دو تایی (i, j) مشخص می‌شود که در آن i و j به ترتیب تعداد پیام‌های مرتبه صفر و تعداد پیام‌های مرتبه یک حاضر در سرور هستند. هرگاه ظرفیت سیستم برابر با B پیام باشد، تعداد حالات سیستم بر اساس رابطه‌ی ۲ محدود خواهد شد. بدیهی است که منظور از ظرفیت سیستم، مجموع ظرفیت صف و ظرفیت سرور است.

$$0 \leq i + j \leq B \quad (2)$$

به عنوان نمونه در شکل ۵ مدل زنجیره‌ی مارکوف سیستم، در حالتی که $B = 3$ باشد آورده شده است.

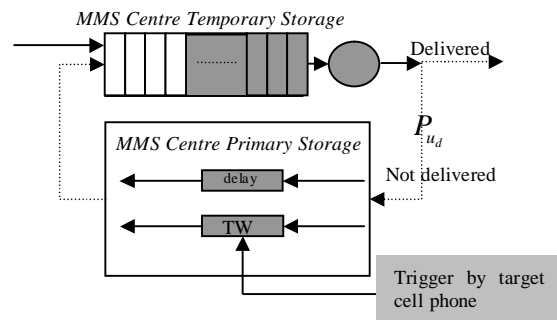
مدل ارائه شده در شکل ۲، در [5,6,9] برای مدل‌سازی ترافیک پیام‌های یک مرکز سرویس پیام چندرسانه‌ای به کار گرفته شده است.



شکل (۲): مدل اولیه‌ی یک مرکز سرویس پیام چندرسانه‌ای

فرضیاتی که در این مدل، در مورد ترافیک پیام‌های ورودی انجام گرفته شامل موارد زیر است: ترافیک ورودی پیام‌ها با یک فرآیند پواسون با نرخ I و زمان انقضای پیام‌ها با یک متغیر نمایی با پارامتر g مدل شده است. این دو فرض که به تناوب در منابع مختلف از جمله [10,11] استفاده شده‌اند، علیرغم اینکه ممکن است چندان با واقعیت سازگار نباشند، ولی برای تحلیل مدل در حالت متوسط بسیار قابل قبول به نظر می‌رسند.

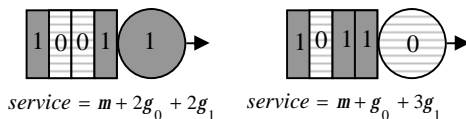
همان‌گونه که در شکل ۳ مشاهده می‌شود، پیام‌های ارسالی از MMS-RS در ارسال d ام با احتمال P_{u_d} ، با عدم موفقیت مواجه می‌شوند. مطابق آنچه که در بررسی سناریوی ارسال مجدد پیام‌ها مورد بحث قرار گرفت، ترافیک پیام‌های ارسال مجدد شده ورودی به MMS-RS، شامل دو منبع است. یک منبع شامل پیام‌هایی است که پس از ارسال ناموفق با اعمال یک زمان تأخیر، مجدداً به صف پیام‌ها وارد می‌شوند و دیگری پیام‌هایی که با تریگر از سوی تلفن همراه مقصد، مجدداً به صف پیام‌ها می‌پیوندند. در نتیجه تأخیر اعمالی به پیام‌ها قبل از ارسال مجدد آنها، در مورد پیام‌های مختلف، متفاوت خواهد بود. در این شرایط به خوبی می‌توان ترافیک پیام‌های مرتبه d را بر اساس یک فرآیند پواسون با نرخ I_d مدل کرد.



شکل (۳): مکانیزم ارسال مجدد پیام‌ها در MMS-RS

برای تعیین توزیع زمان انقضای پیام‌های مرتبه d ، توجه می‌کنیم که پیام‌های مرتبه d ، قبل از اینکه برای $d+1$ امین بار وارد صف سرور شوند، d بار تأخیر انتظار در صف سرور، d بار تأخیر ارسال از طریق سرور و d بار تأخیر انتظار برای ارسال مجدد را گذرانده‌اند. از آنجا که مقادیر تمامی تأخیرهای ذکر شده در مورد پیام‌های مختلف متفاوت

که نرخ سرویس‌دهی سیستم در حالت اول برابر با $m + ig_0 + (j-1)g_1$ و در حالت دوم برابر با $m + (i-1)g_0 + jg_1$ است. این تفاوت به خوبی در شکل ۷، برای حالتی که $i=3$ و $j=2$ باشد به نمایش درآورده شده است. مشخص است که دیگر ترتیب‌های متفاوت قرارگرفتن پیام‌ها در صف، تأثیری در نرخ سرویس‌دهی سیستم در این دو حالت ندارد و آنچه تعیین‌کننده است، مرتبه‌ی پیامی است که در حال گرفتن سرویس است.



شکل (۷): مقایسه تفاوت نرخ سرویس‌دهی با توجه به مرتبه‌ی پیام

برای سهولت کار، ترتیبی را که در آن یکی از پیام‌های مرتبه صفر، در حالت (i, j) ، در حال سرویس گرفتن است را با (i_s, j_s) و ترتیبی که در آن یکی از پیام‌های مرتبه یک، در حالت (i, j) ، در حال سرویس گرفتن است را با (i_s, j_s) نمایش می‌دهیم. برای تعیین نرخ کل سرویس‌دهی سیستم در حالت (i, j) ، نیاز به تعیین احتمالات حضور در هر یک از دو حالت (i_s, j_s) و (i, j_s) داریم. در ادامه‌ی این بخش به دنبال به دست آوردن مقادیر p_{i_s, j_s} و p_{i, j_s} هستیم، که به ترتیب مطابق رابطه‌های ۳ و ۴ تعریف می‌شوند.

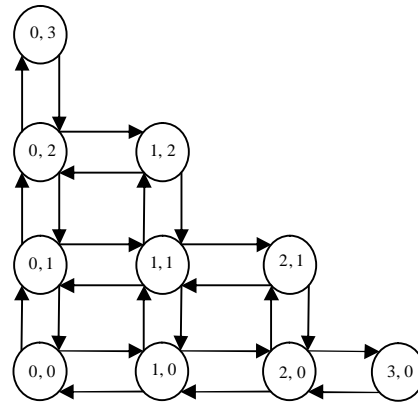
$$p_{i_s, j_s} = \text{احتمال اینکه یکی از } i \text{ پیام مرتبه صفر در حال گرفتن سرویس باشد، در شرایطی که سیستم در حالت } (i, j) \text{ حضور دارد.} \quad (3)$$

$$p_{i, j_s} = \text{احتمال اینکه یکی از } j \text{ پیام مرتبه یک در حال گرفتن سرویس باشد، در شرایطی که سیستم در حالت } (i, j) \text{ حضور دارد.} \quad (4)$$

مطابق این دو تعریف، بدیهی است که: $p_{i_s, j_s} + p_{i, j_s} = 1$. با توجه به اینکه مقادیر p_{i_s, j_s} و p_{i, j_s} دو احتمال مکمل می‌باشند، برای تعیین آنها کافی است نسبت این دو مقدار را، بر اساس رابطه ۵، به دست آوریم.

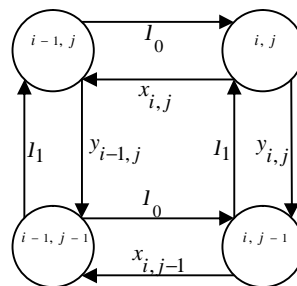
$$\frac{p_{i_s, j_s}}{p_{i, j_s}} = \frac{r_{i_s, j_s}}{r_{i, j_s}} = \frac{I_{i_s, j_s} \times m_{i_s, j_s}}{I_{i, j_s} \times m_{i, j_s}} \quad (5)$$

برای تعیین مقدار $I_{i_s, j_s} / I_{i, j_s}$ به این نکته توجه می‌کنیم که، کافی است نسبت تعداد ترتیب‌های مختلف ورود پیام‌های مرتبه صفر و پیام‌های مرتبه یک به صف سرور را، بین حالت‌های (i_s, j_s) و (i, j_s) تعیین کنیم. چرا که در شرایط بدون حافظه بودن ترافیک پیام‌ها و با توجه به فرض مستقل بودن توزیع ترافیک پیام‌های مرتبه صفر و ترافیک پیام‌های مرتبه یک، به خوبی می‌توان احتمال این ترتیب‌های



شکل (۵): مدل تحلیلی سیستم در حالت B=3

برای تعیین نرخ انتقال از حالات مختلف سیستم به یکدیگر، قسمتی از مدل را که در شکل ۶ آورده شده است، در نظر می‌گیریم. انتقال از حالت $(i-1, j)$ به حالت (i, j) ، با ورود یک پیام مرتبه صفر و در نتیجه با نرخ I_0 انجام می‌گیرد. به طور مشابه انتقال از حالت $(i, j-1)$ به حالت (i, j) با ورود یک پیام مرتبه یک و در نتیجه با نرخ I_1 صورت می‌پذیرد.



شکل (۶): نرخ انتقال از حالت‌های سیستم به یکدیگر

در این مقاله برای تعیین مقادیر $x_{i, j}$ ، نرخ انتقال از حالت (i, j) به حالت $(i-1, j)$ و $y_{i, j}$ ، نرخ انتقال از حالت (i, j) به حالت $(i, j-1)$ ، یک روش ابتکاری به کار گرفته شده است. به این ترتیب که در بخش بعد، به تعیین نرخ کل سرویس‌دهی به پیام‌ها در حالت (i, j) می‌پردازیم و سپس به کمک معادلات تعادلی محلی سیستم، یک رابطه‌ی بازگشتی برای استخراج این مقادیر ارائه می‌دهیم.

۳-۲ تعیین نرخ کل سرویس‌دهی سیستم در حالت (i, j)

همان‌گونه که بیان شد در حالت (i, j) ، تعداد i پیام مرتبه صفر و j پیام مرتبه یک در سیستم حضور دارند. برای تعیین نرخ کل سرویس‌دهی سیستم به پیام‌ها در این حالت، آنچه که از بین ترتیب‌های مختلف سرویس‌دهی برای ما اهمیت دارد، این است که آیا یکی از i پیام مرتبه صفر در حال سرویس گرفتن است و یا یکی از j پیام مرتبه یک. چرا

اینجا برای تعیین نسبت مقادیر $x_{i,j}$ و $y_{i,j}$ ، معادلات تعادلی محلی شکل ۶، که در رابطه ۱۱ آورده شده است را به کار می‌بریم.

$$\begin{aligned} x_{i,j} p_{i,j} &= I_0 p_{i-1,j} \\ y_{i,j} p_{i,j} &= I_1 p_{i,j-1} \end{aligned} \quad (11)$$

$$y_{i-1,j} p_{i-1,j} = I_1 p_{i-1,j-1}$$

$$x_{i,j-1} p_{i,j-1} = I_0 p_{i-1,j-1}$$

رابطه ۱۲، نتیجه مستقیم چهار تساوی رابطه ۱۱ است.

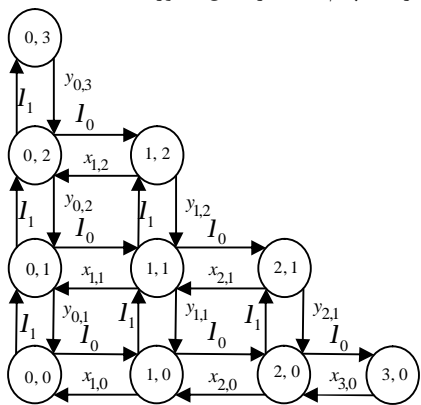
$$\frac{x_{i,j}}{y_{i,j}} = \frac{x_{i,j-1}}{y_{i-1,j}} \quad (12)$$

برای تعیین $x_{i,j}$ و $y_{i,j}$ ، کافی است توجه کنیم که مجموع این دو مقدار برابر با نرخ کل سرویس‌دهی در حالت (i, j) است [2]، به عبارت دیگر $m_{i,j} = x_{i,j} + y_{i,j}$ در نتیجه بر اساس رابطه ۱۲، رابطه‌های بازگشتی ۱۳ و ۱۴ به ترتیب برای تعیین مقادیر $x_{i,j}$ و $y_{i,j}$ به دست می‌آیند.

$$x_{i,j} = \frac{x_{i,j-1}}{x_{i,j-1} + y_{i-1,j}} m_{i,j} \quad (13)$$

$$y_{i,j} = \frac{y_{i-1,j}}{x_{i,j-1} + y_{i-1,j}} m_{i,j} \quad (14)$$

بر اساس این مقادیر مدل تکمیل شده شکل ۵، در حالتی که ظرفیت سیستم برابر با ۳ پیام باشد در شکل ۸ آورده شده است.



شکل (۸): مدل تحلیلی تکمیل شده سیستم در حالت $B=3$

در نهایت برای اینکه بتوانیم به کمک روابط بازگشتی ۱۳ و ۱۴، تمامی مقادیر $x_{i,j}$ و $y_{i,j}$ را تعیین کنیم، باید دقت کنیم که $y_{0,j} = m + (j-1)g_1$ [5,6,9] و $x_{i,0} = m + (i-1)g_0$.

۳- روشی برای استخراج حد بالای متوسط زمان تأخیر و حد بالای احتمال از دست رفتن پیام‌ها مدل سازی پیام-ها

در این بخش با استفاده از مدل پیشنهادی برای مرکز سرویس پیام چندرسانه‌ای و تکنیک‌های به کار گرفته شده در منابع دیگر، روشی

مختلف را یکسان در نظر گرفت. در حالت (i_s, j) ، یکی از i پیام مرتبه صفر در حال سرویس گرفتن است و در نتیجه ترتیب‌های مختلف ورود دیگر پیام‌ها، شامل $i-1$ پیام مرتبه صفر و j پیام مرتبه یک، برابر با $\binom{i+j-1}{i-1}$ است. توضیح این رابطه به این قرار است که از بین $i+j-1$ مکان موجود برای پیام‌ها در صف سرور، با انتخاب $i-1$ مکان، محل قرار گرفتن پیام‌های مرتبه صفر را مشخص کرده‌ایم و مکان‌های باقیمانده به پیام‌های مرتبه یک تخصیص می‌یابد.

به طور مشابه در حالت (i, j_s) که یکی از j پیام مرتبه یک در حال گرفتن سرویس است، ترتیب‌های مختلف ورود دیگر پیام‌ها، شامل i پیام مرتبه صفر و $j-1$ پیام مرتبه یک، برابر با $\binom{i+j-1}{i}$ است. با توجه به دو مقدار اخیر رابطه ۶ به دست می‌آید.

$$\frac{I_{i_s,j}}{I_{i,j_s}} = \frac{\binom{i+j-1}{i-1}}{\binom{i+j-1}{i}} = \frac{j}{i} \quad (6)$$

در مورد نسبت نرخ سرویس‌دهی در حالت‌های (i_s, j) و (i, j_s) ، تفاوت نرخ سرویس‌دهی مربوط به تفاوت نرخ انقضای پیام‌های مرتبه صفر و مرتبه یک می‌باشد و رابطه ۷ نتیجه مستقیم این مطلب است.

$$\frac{m_{i_s,j}}{m_{i,j_s}} = \frac{m + ig_0 + (j-1)g_1}{m + (i-1)g_0 + jg_1} \quad (7)$$

بر اساس روابط ۵، ۶ و ۷ نسبت مقادیر p_{i,j_s} و $p_{i_s,j}$ مشخص است و در نتیجه این دو مقدار به ترتیب از رابطه های ۸ و ۹ به دست می‌آیند.

$$p_{i_s,j} = \frac{i(m + ig_0 + (j-1)g_1)}{i(m + ig_0 + (j-1)g_1) + j(m + (i-1)g_0 + jg_1)} \quad (8)$$

$$p_{i,j_s} = \frac{j(m + (i-1)g_0 + jg_1)}{i(m + ig_0 + (j-1)g_1) + j(m + (i-1)g_0 + jg_1)} \quad (9)$$

در این جا آماده هستیم که با توجه به مقادیر احتمالات به دست آمده، به تعیین نرخ کل سرویس‌دهی سیستم در حالت (i, j) بپردازیم. با توجه به بحثی که انجام شد، نرخ سرویس‌دهی در حالت (i, j) ، با احتمال $p_{i,j}$ برابر با $m + (i-1)g_0 + jg_1$ و با احتمال $p_{i_s,j}$ برابر با $m + ig_0 + (j-1)g_1$ است. در نتیجه نرخ کل سرویس‌دهی در حالت (i, j) که آن را $m_{i,j}$ می‌نامیم، از رابطه ۱۰ به دست می‌آید.

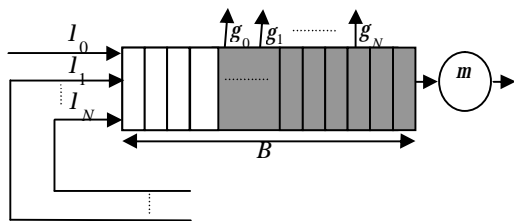
$$m_{i,j} = p_{i_s,j} [m + (i-1)g_0 + jg_1] + p_{i,j_s} [m + ig_0 + (j-1)g_1] \quad (10)$$

۴-۲ تکمیل مدل ارایه شده

در بخش ۲-۲ تا به آن جا پیش رفتیم که برای تعیین مقادیر $x_{i,j}$ ، نرخ انتقال از حالت (i, j) به حالت $(i-1, j)$ و $y_{i,j}$ ، نرخ انتقال از حالت (i, j) به حالت $(i, j-1)$ ، به نرخ کل سرویس‌دهی در حالت (i, j) نیاز داشتیم، که در بخش قبل به تعیین آن پرداختیم. در

۴- تعمیم نتایج به دست آمده برای پیام‌های ارسال مجدد شده مرتبه‌های بالاتر

از مهم‌ترین ویژگی‌های مدل ارائه شده و تحلیل‌های انجام شده در بخش‌های قبل، قابلیت گسترش آنها در مورد پیام‌های مرتبه‌ی بالاتر است. همان‌گونه که ذکر شد، یک مرکز سرویس پیام چندرسانه‌ای ارسال مجدد پیام‌های ناموفق در ارسال را بیشتر از یک بار انجام می‌دهد. در ادامه‌ی این بخش فرض می‌کنیم یک مرکز سرویس پیام چندرسانه‌ای، ارسال مجدد پیام‌های ناموفق در ارسال را، حداکثر تا مرتبه‌ی N انجام می‌دهد و نتایج به دست آمده در بخش‌های قبل را، برای تحلیل کارآیی آن تعمیم می‌دهیم. مدل کلی مرکز سرویس پیام چندرسانه‌ای در این شرایط، در شکل ۱۰ به نمایش درآمده است. با توجه به تعریف مرتبه پیام، بدیهی است که $I_0 > I_1 > \dots > I_N$ و $g_N > g_{N-1} > \dots > g_0$.



شکل (۱۰): مدل سرور با در نظر گرفتن ترافیک پیام‌های مرتبه N

مدل تحلیلی مرکز سرویس پیام چندرسانه‌ای، در این شرایط، یک زنجیره‌ی مارکوف $N+1$ بعدی خواهد بود. هر حالت سیستم با یک $N+1$ تایی (i_0, i_1, \dots, i_N) مشخص می‌شود که در آن i_d مشخص‌کننده‌ی تعداد پیام‌های مرتبه‌ی d است که در سیستم حضور دارند. معادلات تعادلی محلی حالت‌های سیستم از رابطه‌ی ۱۵ به دست می‌آیند.

$$I_d P_{i_0, \dots, i_{d-1}, \dots, i_N} = x_{d, i_0, \dots, i_N} P_{i_0, \dots, i_N} \quad (15)$$

$$\text{که در آن } 0 \leq \sum_{d'=0}^N i_{d'} \leq B \text{ و } 1 \leq i_d \leq B, 0 \leq d \leq N$$

مشابه آنچه که برای حالت $N=1$ بحث شد، $P_{i_0, \dots, i_{d-1}, \dots, i_N}$ یعنی احتمال اینکه در حالت (i_0, i_1, \dots, i_N) یکی از پیام مرتبه d در حال دریافت سرویس باشد، از رابطه‌ی ۱۶ به دست می‌آید.

$$P_{i_0, \dots, i_{d-1}, \dots, i_N} = \frac{i_d}{\sum_{d'=0}^N \frac{m + i_0 g_0 + \dots + (i_{d'} - 1) g_{d'} + \dots + i_N g_N}{i_{d'}}} \quad (16)$$

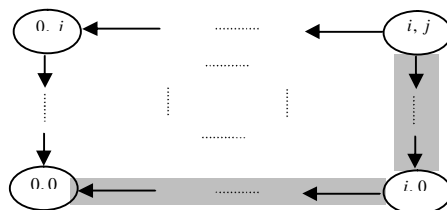
برای اینکه بتوانیم مقادیر احتمال حضور در حالت‌های مختلف سیستم را محاسبه کنیم، به محاسبه‌ی مقادیر $x_{d, i_0, i_1, \dots, i_N}$ که نرخ انتقال از

برای به دست آوردن حدبالای زمان تأخیر پیام‌ها و همچنین حدبالای احتمال از دست رفتن پیام‌ها ارایه شده است.

مشابه آنچه که در [5,6,9] عمل شده است، پیام $m_{i,j}$ را که در حالت (i, j) وارد سیستم شده است، در نظر می‌گیریم. ارسال پیام $m_{i,j}$ در شرایطی آغاز می‌شود که این i پیام مرتبه صفر و j پیام مرتبه یک، که جلوی $m_{i,j}$ قرار دارند، از سیستم خارج شوند. به این ترتیب که یا ارسال گردند و یا زمان انقضای آنها فرا برسد. به بیان دیگر در صورتی که سیستم را فقط شامل این $i+j$ پیام در نظر بگیریم، ارسال $m_{i,j}$ در شرایطی آغاز می‌شود که این سیستم فرضی خالی شود.

برای خارج شدن این $i+j$ پیام از سیستم $\binom{i+j}{i}$ ترتیب مختلف وجود دارد. بدیهی است که متوسط زمان خارج شدن پیام‌ها، در ترتیب‌های مختلف متفاوت است. از آنجا که ما به دنبال روشی برای استخراج حدبالای متوسط زمان تأخیر پیام‌ها هستیم، هدف ما تعیین مسیری است که در آن مسیر مقدار متوسط زمان خارج شدن پیام‌ها از سیستم بیشینه است.

می‌دانیم که نرخ انقضای پیام‌های مرتبه یک همواره بیش از نرخ انقضای پیام‌های مرتبه صفر است، یعنی $g_1 > g_0$. در این شرایط، ادعا می‌کنیم که از بین ترتیب‌های مختلف خارج شدن این $i+j$ پیام از سیستم، ترتیبی که در آن j پیام مرتبه یک همگی قبل از i پیام مرتبه صفر از سیستم خارج شوند، دارای بیشترین مقدار متوسط زمان خارج شدن پیام‌ها از سیستم است. با توجه به اینکه پیام‌های مرتبه صفر نرخ انقضای کمتری دارند، در نتیجه احتمال اینکه زمان انقضای آنها، قبل از پیام‌ها پس از پیام‌های مرتبه یک، باعث می‌شویم که امید ریاضی زمان خارج شدن پیام‌ها از سرور بیشینه شود. مسیر مورد نظر در شکل ۹ مشخص داده شده است.



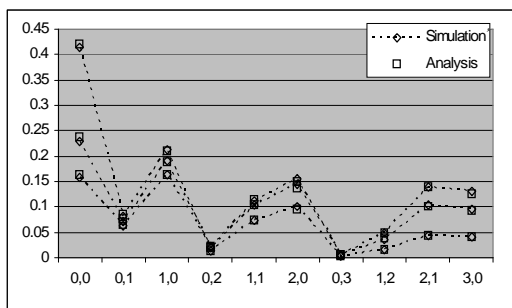
شکل (۹): مسیر خروج پیام‌ها با زمان تأخیر بیشینه

مسیر مورد نظر را می‌توانیم یک زنجیره‌ی مارکوف یک‌بعدی برای سیستم در نظر بگیریم. در این مسیر نرخ سرویس‌دهی سیستم در هر حالت مشخص است. در این شرایط می‌توان با استفاده از نتایج به دست آمده در [5,6,9]، مقدار متوسط زمان تأخیر پیام‌ها و احتمال از دست رفتن پیام‌ها در این مسیر را محاسبه کرد. با توجه به این که این مسیر بیشترین زمان تأخیر را بین تمامی مسیرها دارد، مقادیر محاسبه شده، حدبالایی برای متوسط این مقادیر بر روی تمامی مسیرها است.

با توجه به این که $g_0 > g_1 > \dots > g_{N-1} > g_N$ ، m_{i_0, i_1, \dots, i_N} زمانی بیشترین تأخیر متوسط را تجربه خواهد کرد، که ترتیب خارج شدن پیام‌هایی که جلوی m_{i_0, i_1, \dots, i_N} حضور دارند، به این ترتیب باشد که i_N پیام مرتبه N در ابتدا، پس از آن i_{N-1} پیام مرتبه $N-1$ و به همین ترتیب در نهایت i_0 پیام مرتبه 0 از سیستم خارج شوند. در نتیجه در این مسیر می‌توانیم سیستم را به صورت یک زنجیره‌ی مارکوف یک-بعدی در نظر بگیریم و از نتایج به دست آمده در [5,6,9]، برای به دست آوردن حد بالای متوسط زمان تأخیر پیام‌ها و حد بالای احتمال از دست رفتن آنها استفاده کنیم.

۵- شبیه‌سازی

برای تحقیق نتایج به دست آمده، شبیه‌سازی سیستم به کمک ابزار GPSS World [13] انجام گرفته است. مقادیر متوسط احتمالات حضور سیستم در حالت‌های مختلف در چندین اجرا، با مقدار به دست آمده از تحلیل‌های انجام شده مقایسه شده است. طول زمان شبیه‌سازی نیز به اندازه کافی طولانی در نظر گرفته شده است. سرور با قابلیت ارسال 10Kbps و طول متوسط پیام‌ها 16Kbytes در نظر گرفته شده‌اند و دیگر پارامترهای شبیه‌سازی براساس m تعیین شده‌اند. نمودار شکل ۱۱ مقادیر احتمال حضور در حالت‌های مختلف سیستم را نشان می‌دهد. در شبیه‌سازی انجام شده، مقادیر پارامترهای شبیه‌سازی $I_0 = 0.4I_0$ ، $g_1 = 2g_0$ ، $g_0 = 0.1m$ و $B = 3$ در نظر گرفته شده است و شبیه‌سازی در سه حالت $I_0 = m$ ، $I_0 = 0.8m$ و $I_0 = 0.5m$ انجام گرفته است. همان‌گونه که مشاهده می‌شود، نتایج حاصل از تحلیل به خوبی با نتایج شبیه‌سازی سازگار است.



شکل (۱۱): نمودار احتمال حضور در حالت‌های مختلف سیستم به

ترتیب برای $I_0 = 0.5m$ و $I_0 = 0.8m$ ، $I_0 = m$

در شبیه‌سازی دوم، اثر تغییرات نرخ انقضای پیام‌ها بر روی احتمالات حضور در حالت‌های مختلف سیستم بررسی شد. نمودار شکل ۱۲ مقادیر احتمال حضور در حالت‌های مختلف سیستم را نشان می‌دهد. در شبیه‌سازی انجام شده، مقادیر پارامترهای شبیه‌سازی $I_0 = m$ ، $I_0 = 0.4I_0$ ، $g_1 = 2g_0$ و $B = 3$ در نظر گرفته شده است. شبیه‌سازی در سه حالت $g_0 = 0.2m$ ، $g_0 = 0.1m$ و $g_0 = 0.05m$ انجام

حالت $(i_0, \dots, i_d, \dots, i_N)$ به حالت $(i_0, \dots, i_d - 1, \dots, i_N)$ است، نیاز داریم. در نتیجه به کمک رابطه‌ی ۱۶ و با محاسبه‌ی مقادیر احتمال ترتیب‌های مختلف سرویس گرفتن پیام‌ها، وقتی که سیستم در حالت (i_0, i_1, \dots, i_N) حضور دارد، می‌توانیم مطابق رابطه ۱۷، مقدار $m_{i_0, i_1, \dots, i_d, \dots, i_N}$ یعنی نرخ کل سرویس‌دهی سیستم در حالت (i_0, i_1, \dots, i_N) را محاسبه نماییم.

$$m_{i_0, i_1, \dots, i_d, \dots, i_N} = \sum_{d=0}^N P_{i_0, \dots, i_d, \dots, i_N} [m + i_0 g_0 + \dots + (i_d - 1) g_d + \dots + i_N g_N] \quad (17)$$

مشابه آنچه که برای حالت $N = 1$ مطرح شد، بر اساس معادلات تعادلی محلی، رابطه ۱۸، نسبت نرخ‌های مختلف حرکت از حالت (i_0, i_1, \dots, i_N) به حالت‌های $(i_0 - 1, i_1, \dots, i_N)$ تا $(i_0, i_1, \dots, i_N - 1)$ را در اختیار ما می‌گذارد.

$$\frac{X_{d' i_0, i_1, \dots, i_N}}{X_{d i_0, i_1, \dots, i_N}} = \frac{X_{d' i_0, \dots, i_{d'} - 1, \dots, i_N}}{X_{d i_0, \dots, i_{d'} - 1, \dots, i_N}} \quad (18)$$

که در آن $0 \leq d', d'' \leq N$.

با توجه به اینکه $m_{i_0, i_1, \dots, i_d, \dots, i_N}$ نرخ کل سرویس‌دهی سیستم در حالت (i_0, i_1, \dots, i_N) ، مطابق رابطه‌ی ۱۷ برای ما مشخص است، به کمک رابطه‌ی ۱۸ می‌توانیم مقادیر مختلف $X_{d i_0, i_1, \dots, i_N}$ را به دست آوریم. رابطه‌ی ۱۹، این امکان را در اختیار ما قرار می‌دهد، که به کمک یک روش بازگشتی به محاسبه مقدار $X_{d i_0, i_1, \dots, i_N}$ بپردازیم.

$$X_{d i_0, i_1, \dots, i_N} = \frac{\prod_{d'=0}^N X_{d' i_0, \dots, i_{d'} - 1, \dots, i_N}}{\sum_{d''=0}^N \left[\frac{X_{d'' i_0, \dots, i_{d''} - 1, \dots, i_N}}{X_{d i_0, \dots, i_{d''} - 1, \dots, i_N}} \prod_{d'=0}^N X_{d' i_0, \dots, i_{d'} - 1, \dots, i_N} \right]} m_{i_0, \dots, i_d, \dots, i_N} \quad (19)$$

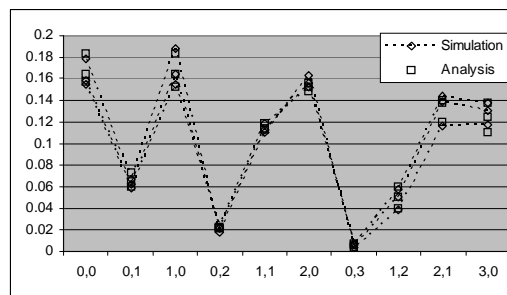
در نهایت برای استفاده از رابطه‌ی بازگشتی ۱۹ باید توجه کنیم که

$$X_{d 0, 0, \dots, i_d, 0, 0} = m + (i_d - 1) g_d$$

برای به دست آوردن حد بالای متوسط زمان انتظار پیام‌ها در صف و همچنین حد بالای احتمال از دست رفتن پیام‌ها، در شرایطی که پیام‌ها امکان ارسال مجدد تا N مرتبه را داشته باشند، پیام m_{i_0, i_1, \dots, i_N} که در حالت (i_0, i_1, \dots, i_N) وارد سیستم شده است را در نظر می‌گیریم.

- [1] Bodic, G., *Multimedia Messaging Service: An Engineering Approach to MMS*, John Wiley & Sons Publishing, 2003.
- [2] Gross, D., Harris, C. M., *Fundamentals of Queuing Theory*, 3rd ed. New York, USA John Wiley & Sons Publishing, 1998.
- [3] Zerfos, P., Meng, X., Wong, Starsky H. Y., *A Study Of The Short Message Service Of A Nationwide Cellular Network*, IMC06 October 25–27, 2006, Rio de Janeiro, Brazil.
- [4] Haung, Y.R., Lin, Y.B., *A bandwidth-on-demand strategy for GPRS*, IEEE Transactions on Wireless Communications, 4(4): 1394-1399, 2005.
- [5] Ghaderi, M., Keshav, S., *Multimedia Messaging Service: System Description And Performance Analysis*, In Proceedings of First International Conference on wireless Internet. IEEE, July 2005.
- [6] Haung, Y.R., Ho, J. M., *Overload Control For Short Message Transfer In GPRS/UMTS Networks*, 2005.
- [7] Markou, M., Panayiotou, C.G., *Dynamic control and optimization of buffer size in wireless networks*, 2005.
- [8] Markou, M., Panayiotou, C.G., *Dynamic control and optimization of buffer size for short message transfer in GPRS/UMTS networks*, IEEE International Conference on Information and Communication Technologies: From Theory to Applications, vol. 1, pp. 209–210, Apr 2004.
- [9] Huang, Y.R., *Determining the optimal buffer size for Short Message Transfer in a heterogeneous GPRS/UMTS network*, IEEE Trans. Vehicular Technology, 52(1), January 2003.
- [10] Lau V.K.N., Maric S.V., *Mobility of queued call requests of call-queueing technique for cellular systems*, IEEE Trans. Vehicular Technology, 47, pp. 480-488, Mar 1998.
- [11] McMillan D., *Delay analysis of a cellular mobile priority queueing system*, IEEE/ACM Trans. Networking, vol. 3, pp. 310-319, June 1995.
- [12] OMA, *Multimedia messaging service; architecture overview v1.2*, Dec. 2003.
- [13] <http://www.minutemansoftware.com>

گرفته است. همان‌گونه که مشاهده می‌شود، در این نمودار نیز نتایج حاصل از تحلیل به خوبی با نتایج شبیه‌سازی سازگار است.



شکل (۱۲): نمودار احتمال حضور در حالت‌های مختلف سیستم به

ترتیب برای $g_0 = 0.05m$ و $g_0 = 0.1m$ و $g_0 = 0.2m$

۶- نتیجه گیری

در این مقاله، مدلی از یک مرکز سرویس پیام چندرسانه‌ای با در نظر گرفتن ترافیک پیام‌های ارسال مجدد شده ارائه شد. با استفاده از نتایج تحلیل مدل پیشنهادی، روشی برای استخراج حدبالای متوسط زمان انتظار پیام‌ها در صف و حدبالای احتمال از دست رفتن پیام‌ها توضیح داده شد. همچنین مدل ارائه شده و نتایج به دست آمده، برای در برگرفتن ترافیک پیام‌های ارسال مجدد شده به تعداد مراتب دلخواه، گسترش داده شد.

در ادامه‌ی این کار به دنبال این هستیم که با استفاده از نتایج به دست آمده در این مقاله، با به کار بردن مکانیزم‌هایی همچون اولویت‌دهی به پیام‌ها و تعیین مقدار بهینه‌ی تأخیر پیام‌ها قبل از ارسال مجدد آنها، کارایی مرکز سرویس پیام چندرسانه‌ای را بهبود بخشیم.