

باسمه تعالی



پروژه‌ی درس آمار و احتمال مهندسی

آشنایی با سیستم‌های پردازش لب‌ای

استاد درس

دکتر محمد مهدی مجاهدیان

دانشکده‌ی مهندسی برق
دانشگاه صنعتی شریف

تیم طراحی:

امیرحسین نقدی، علی صادقیان، محمد محمدیان، ماهان سرافرازی

زمستان ۱۴۰۳

آخرین مهلت تحویل:

۱۴ بهمن ۱۴۰۳

فهرست مطالب

۲	۱ مقدمه
۲	۲ سرورهای شبکه
۲	۱.۲ سرور لبه
۲	۲.۲ سرور جمع آوری کننده
۲	۳.۲ سرور اصلی
۳	۳ مدل سازی سیستم
۳	۱.۳ مدل سازی سرور لبه
۴	۲.۳ مدل سازی سرور جمع آوری کننده اصلی
۴	۳.۳ یافتن نرخ خروجی سیستم
۴	۴ بهینه سازی نرخ پردازش سرور لبه
۴	۱.۴ تاثیرات تغییر μ
۵	۲.۴ تحلیل هزینه ها
۵	۵ نکات مهم!

۱ مقدمه

تصور کنید یک شبکه IoT^۱ پیچیده داریم که شامل تعداد زیادی سنسور IoT است. این سنسورها داده‌هایی را به سرورهای لبه ارسال می‌کنند تا پردازش اولیه روی آن‌ها انجام شود. سپس، کارهای پردازش‌شده به سرورهای جمع‌آوری‌کننده منتقل می‌شوند تا بررسی شوند و در صورت نیاز به سرورهای لبه بازگردند. در نهایت، کارهای تأییدشده به سرور اصلی ارسال می‌شوند، جایی که تمامی داده‌ها ذخیره و مدیریت می‌شوند. این معماری شبکه برای مدیریت مقادیر زیاد داده، پردازش سریع، و اطمینان از صحت داده‌ها طراحی شده است. با این حال، در این ساختار، چالش‌هایی وجود دارد که باید از طریق مدل‌سازی دقیق و شبیه‌سازی حل شوند. این چالش‌ها شامل مدیریت نرخ پردازش (μ) در سرورهای لبه، احتمال بازگشت کارها (p)، و هزینه‌های ناشی از خطاها و بازگشت‌ها است. هدف این بخش، مدل‌سازی سیستم به صورت مرحله‌به‌مرحله، تحلیل عملکرد آن، و یافتن بهترین راه‌حل برای بهینه‌سازی هزینه‌هاست.

۲ سرورهای شبکه

۱.۲ سرور لبه^۲

هر سرور لبه، داده‌های ارسال‌شده از سنسورها را با نرخ λ دریافت می‌کند. هر سرور لبه دارای یک صف با حداکثر ظرفیت k است که داده‌ها در آن ذخیره می‌شوند تا برای پردازش آماده شوند. سرور لبه، داده‌ها را با نرخ پردازشی μ پردازش کرده و پس از اتمام، آن‌ها را به سرور جمع‌آوری‌کننده ارسال می‌کند. علاوه بر این، هر سرور لبه دارای حافظه‌ای به اندازه k است که وظیفه ذخیره آخرین k داده پردازش‌شده را بر عهده دارد. این حافظه برای ارسال مجدد داده‌ها در صورت لزوم استفاده می‌شود.

۲.۲ سرور جمع‌آوری‌کننده^۳

در سرور جمع‌آوری‌کننده، صفی با ظرفیت نامحدود وجود دارد که تمامی کارها به ترتیب ورود در آن قرار می‌گیرند. هنگامی که یک کار به این سرور می‌رسد، با احتمال p ممکن است به سرور لبه بازگردانده شود. (این بازگشت کار به سرور لبه می‌تواند دلایل مختلفی باشد، از جمله مشکل در لینک ارتباطی بین سرور جمع‌کننده و سرور لبه، مشکل در پردازش داده در سرور لبه بخاطر انتخاب نادرست توان محاسباتی و ...). اگر کار بازگشتی در حافظه پردازش‌شده‌های سرور لبه موجود باشد، به ابتدای صف آن سرور اضافه می‌شود. در صورتی که صف سرور لبه ظرفیت k خود را پر کرده باشد، آخرین کار موجود در صف حذف شده و کار بازگشتی به جای آن قرار می‌گیرد. هر سرور جمع‌آوری‌کننده توان پردازشی برابر با $\sum_{i=1}^L \mu_i$ دارد، که در آن L تعداد سرورهای لبه‌ای است که توسط سرور جمع‌آوری‌کننده پشتیبانی می‌شود.

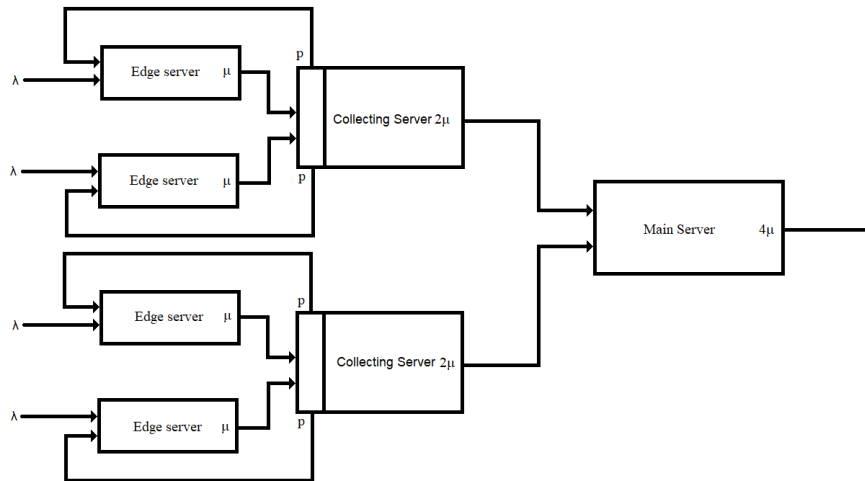
۳.۲ سرور اصلی^۴

سرور اصلی به تمامی سرورهای جمع‌آوری‌کننده متصل است و تمامی کارهای پردازش‌شده توسط این سرورها به سرور اصلی ارسال می‌شود. این سرور دارای یک صف با ظرفیت نامحدود است که تمامی کارها ورودی را مدیریت می‌کند.

توان پردازشی سرور اصلی برابر با $\sum_{j=1}^M \sum_{i=1}^{L_j} \mu_i^{L_j}$ است، که در آن:

- M : تعداد سرورهای جمع‌آوری‌کننده متصل به سرور اصلی
- L_j : تعداد سرورهای لبه متصل به سرور جمع‌آوری‌کننده j -م
- $\mu_i^{L_j}$: نرخ پردازش کارها در سرور لبه i -م متصل به سرور جمع‌آوری‌کننده j -م.

^۱ Internet of Things
^۲ Edge Server
^۳ Collecting Server
^۴ Main Server



شکل ۱: یک نمونه از شبکه با $M = 2, L = 2$

۳ مدل سازی سیستم

برای ساده تر کردن فرآیند مدل سازی، سیستم را به دو بخش تقسیم می کنیم:

۱.۳ مدل سازی سرور لبه

سرورهای لبه کارهای ارسال شده از سنسورهای IoT را با نرخ ورود λ دریافت می کنند. هر سرور دارای یک صف با حداکثر ظرفیت k است که داده ها در آن ذخیره می شوند تا برای پردازش آماده شوند. نرخ پردازش این سرورها برابر با μ است. علاوه بر صف، هر سرور لبه، حافظه ای با ظرفیت k دارد که آخرین کارهای پردازش شده را برای ارسال مجدد ذخیره می کند. احتمال بازگشت کارها از سرور جمع کننده به سرور لبه برابر با p است. هنگامی که کاری توسط سرور جمع کننده به سرور لبه ارجاع داده می شود، اگر کار در حافظه سرور لبه موجود باشد، وارد صف سرور لبه می شود. اگر کاری در صف سرور لبه وجود داشته باشد، کنار گذاشته شده و کار بازگشتی مورد بررسی قرار می گیرد. در صورتی که صف سرور لبه پر باشد، آخرین کار موجود در صف از صف حذف می شود.

پرسش شبیه سازی ۱. رفتار یک سرور لبه را مدل سازی کنید. می دانیم کارها به دلیل ماهیت گسسته شان از توزیع تصادفی پواسون با نرخ λ تولید می شوند و مدت زمانی که سرور لبه، کار را به سرانجام می رساند دارای توزیع تصادفی نمایی با میانگین μ است.

۱. با استفاده از فرض های ساده کننده، سرور لبه را مدل سازی کنید. سپس میانگین کارهای موجود در صف را بیابید. همچنین میانگین کارهایی که بدلیل نبود ظرفیت در صف سرور لبه دور ریخته می شوند را برحسب λ بیابید.

۲. نمودار تعداد کارهای موجود در صف و نرخ کارهایی که دور ریخته شده اند را به ازای حالت های مختلف $0.5\mu < \lambda < 2\mu$ بر حسب زمان رسم کنید (می توانید نرخ μ را به دلخواه انتخاب کنید). برای رسیدن به حالت ایستای زنجیره مارکف، نتایج شبیه سازی را برای هر حالت به ازای $N = 100000$ رسم کنید (N برابر تعداد داده های ارسالی از سنسورها است). برای پیدا کردن میانگین تعداد کارها در صف، نیاز است که بخشی از ابتدای داده ها را نادیده بگیرد. میانگین حاصل از شبیه سازی را با آنچه که بصورت مدل سازی بدست آورده اید مقایسه کنید. آیا مدل سازی با شبیه سازی تطابق دارد؟ در صورتی که بین شبیه سازی و مدل سازی اختلاف وجود داشته باشد، آن را توجیه کنید. دلیل حذف بخش ابتدایی داده ها این است که در ابتدای شبیه سازی، سیستم ممکن است هنوز به حالت پایدار نرسیده باشد و نتایج این دوره ممکن است نمایانگر رفتار واقعی سیستم در طولانی مدت نباشند.

۲.۳ مدل سازی سرور جمع آوری کننده اصلی

سرورهای جمع آوری کننده کارهای پردازش شده را از سرورهای لبه دریافت می کنند. این سرورها دارای صفی با ظرفیت نامحدود هستند و احتمال بازگشت کارها از آن ها به سرور لبه برابر با p است. کارهای تأیید شده از سرورهای جمع آوری کننده به سرور اصلی ارسال می شوند، جایی که تمامی کارها مدیریت و پردازش می گردند. توان پردازشی سرور اصلی برابر با مجموع توان پردازشی سرورهای جمع کننده است.

پرسش شبیه سازی ۲. رفتار سیستم چند سرور جمع آوری کننده و یک سرور اصلی را مدلسازی کنید (سرورهای جمع کننده را در حالتی که تحت بارها گوناگون در نظر بگیرید $(\mu < \lambda, \lambda > \mu)$). می توانید از فرض های ساده کننده استفاده کنید ولی نباید نتایج با شبیه سازی تفاوت زیادی داشته باشد. برای شبیه سازی به یاد داشته باشید که بخشی از ابتدای داده ها را باید دور بریزید.

۳.۳ یافتن نرخ خروجی سیستم

پس از مدل سازی هر بخش، نرخ خروجی^۵ کل سیستم را به صورت تابعی از پارامترهای λ ، μ و p محاسبه کنید. همچنین، مقدار کارهای بازگشتی و نرخ از دست دادن کارها به دلیل پر شدن صف ها را بیابید.

پرسش شبیه سازی ۳. حال مدل سازی سیستم خود را کامل کنید و برای حالت های زیر امتحان کنید.

۱. فرض کنید سیستم دارای ۳ سرور جمع آوری کننده است که هر کدام به ۵ سرور لبه که مستقل از یکدیگر مشغول کار هستند متصل هستند. هر سرور لبه دارای توان پردازشی μ است و کارها با نرخ λ وارد سیستم می شوند همچنین ظرفیت سیستم های لبه k می باشد. همچنین احتمال بازگشت کارها به سیستم لبه از سیستم جمع آوری کننده را p در نظر بگیرید. حال نرخ خروج کارها را به ازای مقادیر زیر محاسبه کنید. دقت کنید برای صحیح بودن نرخ خروجی، شبیه سازی را برای $N = 100000$ بار انجام دهید و سپس میانگین نرخ خروجی ها را اعلام کنید.

$$\mu = 10, \lambda = 3, k = 10, p = 0.1$$

$$\mu = 10, \lambda = 6, k = 10, p = 0.1$$

$$\mu = 10, \lambda = 9, k = 10, p = 0.1$$

۲. نتایج شبیه سازی را با مدلسازی نهایی خود مقایسه کنید. در صورتی که تفاوتی وجود دارد، آن را توجیه کنید.

۴ بهینه سازی نرخ پردازش سرور لبه

در این بخش، هدف تحلیل و یافتن بهترین مقدار برای نرخ پردازش (μ) سرور لبه است، به گونه ای که هزینه کل سیستم به حداقل برسد.

۱.۴ تاثیرات تغییر μ

- افزایش μ : کارها سریع تر پردازش می شوند و احتمال از دست دادن کارها به دلیل پر شدن صف کاهش می یابد. اما احتمال بازگشت کارها (p) افزایش می یابد و هزینه بازگشت ها بیشتر می شود.
- کاهش μ : احتمال بازگشت کارها (p) کاهش می یابد، اما نرخ از دست دادن کارها به دلیل پر شدن صف بیشتر می شود.

^۵Throughput

۲.۴ تحلیل هزینه‌ها

برای هر داده از دست‌رفته، هزینه A دلار و برای هر بازگشت، هزینه B دلار فرض کنید. هدف یافتن مقدار بهینه μ است که هزینه کل زیر را کمینه کند:

$$= (B \times \text{تعداد بازگشت کارها}) + (A \times \text{تعداد کارها از دست‌رفته}) = \text{هزینه کل}$$

پرسش شبیه‌سازی ۴. حال مدل‌سازی سیستم خود را کامل کنید. مقدار بهینه μ را به ازای $A > B$ ، $B > A$ و $A = B$ بیابید.

۱. فرض کنید سیستم دارای ۱ سرور اصلی، ۳ سرور جمع‌کننده و هر سرور جمع‌کننده دارای ۵ سرور لبه است. توان پردازشی هر سرور جمع‌کننده برابر مجموع توان پردازشی سرورها لبه متصل به آن سرور جمع‌کننده است. همچنین توان پردازشی سرور اصلی برابر مجموع توان پردازشی سرورها جمع‌کننده است. تمام سرورهای لبه برای سادگی یکسان هستند. رابطه بین p و μ را بصورت زیر در نظر بگیرید.

$$p(\mu) = 2 \left(\frac{1}{1 + e^{-0.25\mu}} - 0.5 \right)$$

همچنین $(A = 0.5, B = 1)$ ، $(A = 1, B = 1)$ و $(A = 1, B = 0.5)$ شبیه‌سازی را انجام دهید و سعی کنید p بهینه را بیابید. نتایج خود را با مدل‌سازی تطابق دهید و در صورت وجود مغایرت آن را توجیه کنید.

۵ نکات مهم!

لطفاً به نکات زیر دقت کنید:

- این پروژه بخشی از نمره‌ی شما در این درس را تشکیل خواهد داد.
- می‌توانید پروژه را در قالب گروه‌های ۲ نفره انجام دهید. فرمی برای ثبت گروه‌ها در اختیار شما قرار خواهد گرفت. دقت داشته باشید که در هنگام تحویل پروژه باید تمامی اعضای گروه به تمامی بخش‌ها مسلط باشند و در نهایت همه‌ی اعضای یک گروه نمره‌ی واحدی را دریافت خواهند کرد.
- تمامی شبیه‌سازی‌ها باید با کمک زبان Python انجام شود. شما تنها مجاز به استفاده از کتابخانه‌های `plotly`، `random`، `scipy`، `numpy` و `matplotlib` هستید. اگر روی عنوان هر کتابخانه کلیک کنید، به راهنمای آن کتابخانه هدایت می‌شوید.
- تحویل پروژه به صورت گزارش و کدهای نوشته‌شده است. گزارش باید شامل پاسخ پرسش‌ها، تصاویر و نمودارها و نتیجه‌گیری‌های لازم باشد. توجه کنید که قسمت عمده بارم شبیه‌سازی را گزارش شما و نتیجه‌ای که از خروجی کد می‌گیرید دارد. همچنین تمیزی گزارش بسیار مهم است. کدها و گزارش را در یک فایل فشرده‌شده در سامانه‌ی درس‌افزار آپلود کنید.
- اگر برای پاسخ به پرسش‌ها، از منبعی (کتاب، مقاله، سایت و...) کمک گرفته‌اید، حتماً به آن ارجاع دهید.
- نوشتن گزارش کار با \LaTeX نمره‌ی امتیازی دارد.
- پرسش‌های شبیه‌سازی با رنگ سبز مشخص شده‌اند.
- بخش‌های تئوری گزارش که در قالب پرسش‌ها طرح شده‌اند را می‌توانید روی کاغذ بنویسید و تصویر آن‌ها را در گزارش خود بیاورید، ولی توصیه‌ی برادرانه می‌کنم که این کار را نکنید!
- در صورت مشاهده‌ی تقلب، نمره‌ی هردو فرد صفر منظور خواهد شد.

موفق باشید!