

باسمہ تعالیٰ



پروژه‌ی درس آمار و احتمال مهندسی

آشنایی با سیستم‌های پردازش لبه‌ای

استاد درس
دکتر محمد‌مهدی مجاهدیان

دانشکده‌ی مهندسی برق
دانشگاه صنعتی شریف

تیم طراحی:
امیرحسن نقدی، علی صادقیان، محمد محمدیان، ماهان سرافرازی

۱۴۰۳ زمستان

آخرین مهلت تحويل:
۱۴۰۳ بهمن

فهرست مطالب

۱	مقدمه
۲	
۲	سرورهای شبکه
۲	۱.۱ سرور لبه
۲	۲.۱ سرور جمع آوری کننده
۲	۲.۲ سرور اصلی
۳	مدل‌سازی سیستم
۳	۱.۳ مدل‌سازی سرور لبه
۴	۲.۳ مدل‌سازی سرور جمع آوری کننده و اصلی
۴	۲.۳ یافتن نرخ خروجی سیستم
۴	۴ بهینه‌سازی نرخ پردازش سرور لبه
۴	۱.۴ تأثیرات تغییر M
۵	۲.۴ تحلیل هزینه‌ها
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 پردازش شده به سرورهای جمع‌آوری کننده 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 خطاهای و بازگشت‌ها است.

هدف این بخش، مدل‌سازی سیستم به صورت مرحله‌به‌مرحله، تحلیل عملکرد آن، و یافتن بهترین راه حل برای بهینه‌سازی هزینه‌هاست.

۲ سرورهای شبکه

۲.۱ سرور لبه^۲

هر سرور لبه، داده‌های ارسال شده از سنسورها را با نرخ λ دریافت می‌کند. هر سرور لبه دارای یک صفحه 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 جمع‌آوری کننده ارسال می‌کند. علاوه بر این، هر سرور لبه دارای حافظه‌ای به اندازه k است که وظیفه ذخیره آخرین k داده پردازش شده را بر عهده دارد. این حافظه برای ارسال مجدد داده‌ها در صورت لزوم استفاده می‌شود.

۲.۲ سرور جمع‌آوری کننده^۳

در سرور جمع‌آوری کننده، صفحی 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 پردازش شده‌های سرور لبه موجود باشد، به ابتدای صفحه آن سرور اضافه می‌شود. در صورتی که صفحه سرور لبه ظرفیت k خود را پر کرده باشد، آخرین کار موجود در صفحه حذف شده و کار بازگشتی به جای آن قرار می‌گیرد.

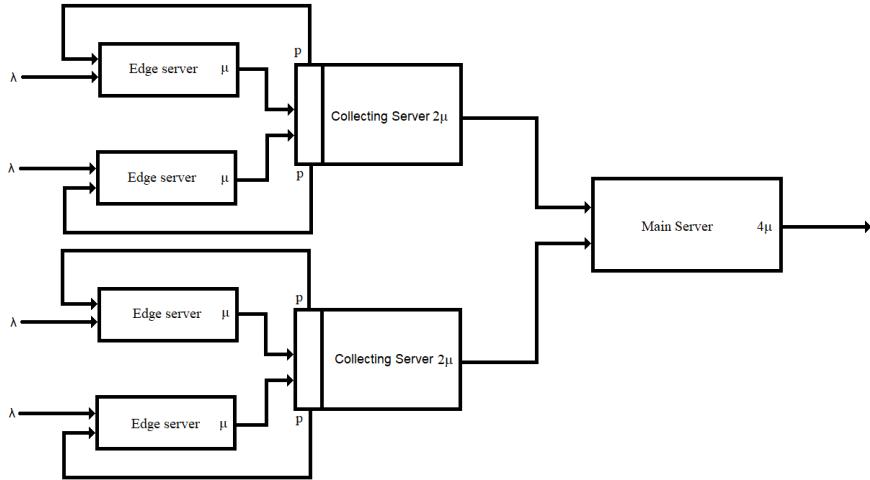
هر سرور جمع‌آوری کننده توان پردازشی برابر با $\mu_i = \sum_{j=1}^{L_i}$ دارد، که در آن L_i تعداد سرورهای لبه‌ای است که توسط سرور جمع‌آوری کننده پشتیبانی می‌شود.

۳.۲ سرور اصلی^۴

سرور اصلی به تمامی سرورهای جمع‌آوری کننده متصل است و تمامی کارهای پردازش شده توسط این سرورها به سرور اصلی ارسال می‌شود. این سرور دارای یک صفحه با ظرفیت نامحدود است که تمامی کارها ورودی را مدیریت می‌کند.

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

- M : تعداد سرورهای جمع‌آوری کننده متصل به سرور اصلی
- L_j : تعداد سرورهای لبه متصل به سرور جمع‌آوری کننده j -م
- μ_i : نرخ پردازش کارها در سرور لبه i -م متصل به سرور جمع‌آوری کننده j -م.



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

۲ مدل‌سازی سیستم

برای ساده‌تر کردن فرآیند مدل‌سازی، سیستم را به دو بخش تقسیم می‌کنیم:

۱.۳ مدل‌سازی سرور لبه

سرورهای لبه کارهای ارسال شده از سنسورهای IoT را با نرخ ورود λ دریافت می‌کنند. هر سرور دارای یک صفحه با حداکثر ظرفیت k است که داده‌ها در آن ذخیره می‌شوند تا برای پردازش آماده شوند. نرخ پردازش این سرورها برابر با μ است.

علاوه بر صفحه، هر سرور لبه، حافظه‌ای با ظرفیت k دارد که آخرین کارهای پردازش شده را برای ارسال مجدد ذخیره می‌کند.

احتمال بازگشت کارها از سرور جمع‌کننده به سرور لبه برابر با p است. هنگامی که کاری توسط سرور جمع‌کننده به سرور لبه ارجاع داده می‌شود، اگر کار در حافظه سرور لبه موجود باشد، وارد صفحه سرور لبه می‌شود. اگر کاری در صفحه سرور لبه وجود داشته باشد، کنار گذاشته شده و کار بازگشته شده و کار بازگشته مورد بررسی قرار می‌گیرد. در صورتی که صفحه سرور لبه پر باشد، آخرین کار موجود در صفحه از صفحه حذف می‌شود.

پرسش شبیه‌سازی ۱. رفتار یک سرور لبه را مدل‌سازی کنید. می‌دانیم کارها به دلیل ماهیت گسسته‌شان از توزیع تصادفی پوآسون با نرخ λ تولید می‌شوند و مدت زمانی که سرور لبه، کار را به سرانجام می‌رساند دارای توزیع تصادفی نمایی با میانگین μ است.

۱. با استفاده از فرض‌های ساده کننده، سرور لبه را مدل‌سازی کنید. سپس میانگین کارهای موجود در صفحه را بیابید. همچنین میانگین کارهایی که بدلیل نبود ظرفیت در صفحه سرور لبه دور ریخته می‌شوند را برحسب λ بیابید.

۲. نمودار تعداد کارهای موجود در صفحه و نرخ کارهایی که دور ریخته شده‌اند را به ازای حالت‌های مختلف ابتدای داده‌ها را نماید. میانگین حاصل از شبیه‌سازی را با آنچه که بصورت مدل‌سازی بدست آورده‌اید مقایسه کنید. آیا مدل‌سازی با شبیه‌سازی تطابق دارد؟ در صورتی که بین شبیه‌سازی و مدل‌سازی اختلاف وجود داشته باشد، آن را توجیه کنید. دلیل حذف بخش ابتدایی داده‌ها این است که در ابتدای شبیه‌سازی، 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 مدل‌سازی سرور جمع آوری‌کننده و اصلی

سرورهای جمع آوری‌کننده کارهای پردازش شده را از سرورهای لبه دریافت می‌کنند. این سرورها دارای صفحه با ظرفیت نامحدود هستند و احتمال بازگشت کارها از آن‌ها به سرور لبه برابر با p است. کارهای تأییدشده از سرورهای جمع آوری‌کننده به سرور اصلی ارسال می‌شوند، جایی که تمامی کارها مدیریت و پردازش می‌گردند. توان پردازشی سرور اصلی برابر با مجموع توان پردازشی سرورهای جمع‌کننده است.

پرسش شبیه‌سازی ۲. رفتار سیستم چند سرور جمع آوری‌کننده و یک سرور اصلی را مدل‌سازی کنید (سرورهای جمع‌کننده را در حالتی که تحت بارها گوناگون در نظر بگیرید ($\mu < \lambda > \lambda$)). می‌توانید از فرض‌های ساده کننده استفاده کنید ولی نباید نتایج با شبیه‌سازی تفاوت زیادی داشته باشد. برای شبیه‌سازی به یاد داشته باشید که بخشی از ابتدای داده‌ها را باید دور ببریزید.

۳.۳ یافتن نرخ خروجی سیستم

پس از مدل‌سازی هر بخش، نرخ خروجی^۵ کل سیستم را به صورت تابعی از پارامترهای λ ، μ و m محاسبه کنید. همچنین، مقدار کارهای بازگشتی و نرخ از دست دادن کارها به دلیل پرشدن صفحه را بیابید.

پرسش شبیه‌سازی ۳. حال مدل‌سازی سیستم خود را کامل کنید و برای حالت‌های زیر امتحان کنید.

۱. فرض کنید سیستم دارای ۳ سرور جمع آوری‌کننده 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 می‌شوند همچنین ظرفیت سیستم‌های لبه k می‌باشد. همچنین احتمال بازگشت کارها به سیستم لبه از سیستم جمع آوری‌کننده را p در نظر بگیرید. حال نرخ خروج کارها را به ازای مقادیر زیر محاسبه کنید. وقت کنید برای صحیح بودن نرخ خروجی، شبیه‌سازی را برای $N = 100000$ بار انجام دهید و سپس میانگین نرخ خروجی‌ها را اعلام کنید.

$$\mu = 10, \lambda = 2, k = 10, p = 0.1 \quad \bullet$$

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

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

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

۴ بهینه‌سازی نرخ پردازش سرور لبه

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

۱.۴ تاثیرات تغییر μ

• افزایش μ : کارها سریع‌تر پردازش می‌شوند و احتمال از دست دادن کارها به دلیل پرشدن صفحه کاهش می‌یابد. اما احتمال بازگشت کارها (p) افزایش می‌یابد و هزینه بازگشت‌ها بیشتر می‌شود.

• کاهش μ : احتمال بازگشت کارها (p) کاهش می‌یابد، اما نرخ از دست دادن کارها به دلیل پرشدن صفحه بیشتر می‌شود.

Throughput^۵

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:

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

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

۱. فرض کنید سیستم دارای ۱ سورور اصلی، ۳ سورور جمع‌کننده و هر سورور جمع‌کننده دارای ۵ سورور لبه است. توان پردازشی هر سورور جمع‌کننده برابر مجموع توان پردازشی سورورها لبه متصل به آن سورور جمع‌کننده است. همچنین توان پردازشی سورور اصلی برابر مجموع توان پردازشی سورورها جمع‌کننده است. تمام سورور‌های 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 $(A = 1, B = 0.5)$ و $(A = 1, B = 1)$ 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:

۱. این پروژه بخشی از نمره‌ی شما در این درس را تشکیل خواهد داد.
۲. 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 تحويل پروژه باید تمامی اعضای گروه به تمامی بخش‌ها مسلط باشند و در نهایت همه‌ی اعضای یک گروه نمره‌ی واحدی را دریافت خواهند کرد.
۳. تمامی شبیه‌سازی‌ها باید با کمک زبان Python انجام شود. شما تنها مجاز به استفاده از کتابخانه‌های `matplotlib`, `random`, `scipy`, `numpy` هستید. اگر روی عنوان هر کتابخانه کلیک کنید، به راهنمای آن کتابخانه هدایت می‌شوید.
۴. تحويل پروژه به صورت گزارش و کدهای نوشته شده 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 فشرده شده در سامانه‌ی درس افزار آپلود کنید.
۵. اگر برای پاسخ به پرسش‌ها، از منبعی (کتاب، مقاله، سایت و...) کمک گرفته‌اید، حتماً به آن ارجاع دهید.
۶. نوشن گزارش کار با \LaTeX نمره‌ی امتیازی دارد.
۷. پرسش‌های شبیه‌سازی با رنگ سبز مشخص شده‌اند.
۸. 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!
۹. در صورت مشاهده‌ی تقلب، نمره‌ی هردو فرد صفر منظور خواهد شد.

موفق باشید!