TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC

ỨNG DỤNG LÝ THUYẾT HÀNG ĐỢI

TRONG VIỆC TỐI ƯU HOÁ THIẾT KẾ HỆ THỐNG

Giảng viên hướng dẫn	:	TS. Nguyễn Thị Ngọc Anh
Lớp	:	CTTN Toán Tin - K62

Sinh viên thực hiện	**MSSV**
Nguyễn Xuân Trường	20172881
Nguyễn Quang Huy	20173179

Hà Nội - 07/2020

Mục lục

Tóm tắt

Hầu hết các ngân hàng tại Việt Nam luôn xuất hiện hàng dài chờ đợi để được phục vụ. Nguyên nhân chính dẫn đến việc này là do việc thiết kế hệ thống hàng chờ của ngân hàng còn chưa đạt được tác dụng như mong muốn. Trong bài báo cáo này, chúng tôi sẽ phân tích rõ hơn về lý thuyết hàng đợi và sẽ đưa ra những cách để xây dựng một hệ thống hàng đợi tối ưu nhất cho các ngân hàng. Sau khi sử dụng các biện pháp tối ưu để thiết kế, sự hài lòng của khách hàng về chất lượng phục vụ tại ngân hàng đã tăng lên rõ rệt. Điều đó chứng tỏ rằng mô hình của bài toán hoàn toàn khả thi và các ngân hàng hoàn toàn có thể áp dụng những phương pháp này để giải quyết những vấn đề như hiện nay.

1 Mở đầu

Ngày nay, việc tham gia vào các điểm cung cấp dịch vụ như quầy sách, cửa hàng ăn, ngân hàng,... là rất phổ biến. Đặc điểm chung của khách hàng khi tham gia vào những điểm cung cấp dịch vụ này là họ phải tham gia vào hàng đợi. Tuy nhiên, theo một khảo sát mới đây [2], nhiều khách hàng tỏ ra không hài lòng về việc xếp hàng chờ đợi, đặc biệt là ở ngân hàng. Trong thực tế, phần lớn những lời phàn nàn đến từ việc các khách hàng chờ đợi trong một thời gian dài để xếp hàng hoặc mặc dù đã tới trước những khách hàng khác nhưng lại được phục vụ sau. Nguyên nhân cốt lõi của những vấn đề này là do việc thiết kế hệ thống hàng đợi của các ngân hàng chưa thực sự hiệu quả. Việc ứng dụng lý thuyết hàng đợi trong việc thiết kế hệ thống hàng đợi nhằm mục đích giảm thời gian chờ đợi của khách hàng, tối ưu hóa số trạm phục vụ và chi phí cơ hội bị mất đi của các ngân hàng khi các khách hàng tới nhưng không được phục vụ.

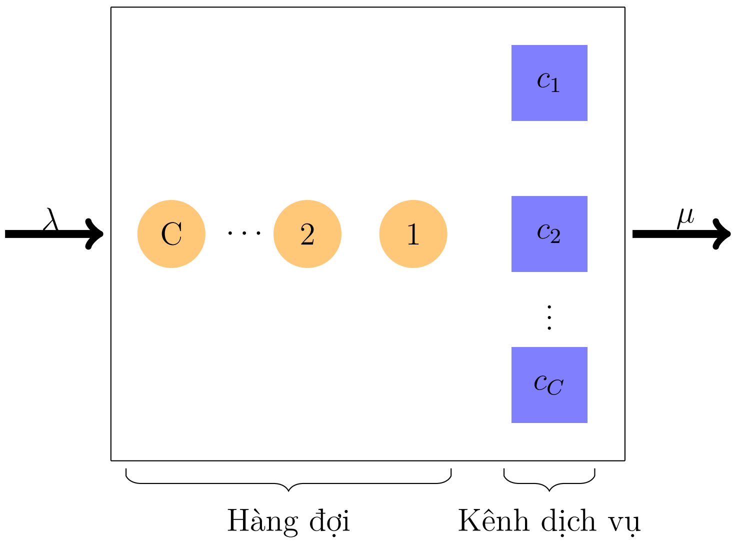
2 Phương pháp nghiên cứu

2.1 Lý thuyết hàng đợi

Lý thuyết hàng đợi là một nhánh của xác suất thống kê, được ứng dụng trong nhiều lĩnh vực khác nhau. Lý thuyết hàng đợi tập trung trả lời các câu hỏi như trung bình thời gian trong hàng đợi, trung bình thời gian phản hồi của hệ thống(thời gian trong hàng đợi và thời gian phục vụ), số lượng khách hàng trong hàng đợi,...

Một hệ thống hàng đợi gồm các thành phần cơ bản sau (Hình 1):

- Quá trình vào, quá trình ra của hệ thống;
- Phân phối thời gian phục vụ;
- Số lượng kênh phục vụ;
- Khả năng của hệ thống;
- Nguyên tắc phục vụ.



Hình 1: Mô hình hàng đợi của ngân hàng.

2.2 Mô hình bài toán

Giả sử chúng ta cần thiết kế một hệ thống cho khách hàng trong đó có C trạm phục vụ khách hàng và số lượng khách hàng trong hệ thống là một số hữu hạn K. Số lượng khách hàng đến với hệ thống được xem như là một biến ngẫu nhiên tuân theo phân phối Poisson có cường độ trung bình là λ. Thời gian phục vụ tại mỗi trạm phục vụ khách hàng thì độc lập với nhau và tuân theo phân phối mũ có cường độ trung bình là μ.

Yêu cầu của bài toán thiết kế là làm thế nào để thiết kế được một trung tâm phục vụ khách hàng với các yêu cầu sau:

- Tối ưu chi phí hoạt động của trung tâm dịch vụ phục vụ khách hàng:

    · Chi phí cơ hội C_0 là thấp nhất.

    · Số lượng trạm phục vụ khách hàng C là nhỏ nhất.

- Nguyên tắc phục vụ: Ai đến trước phục vụ trước (FIFO).

2.3 Các kí hiệu cơ bản trong hệ thống hàng đợi

- Trạng thái hệ thống: Số lượng khách hàng trong hệ thống.

- Độ dài hàng đợi: Số khách hàng trong hàng chờ đang phải chờ để được phục vụ.
- P_n: Xác suất có đúng n khách hàng trong hệ thống khi ở trạng thái cân bằng.
- L: Số lượng khách hàng trung bình trong hệ thống.
- L_q: Số lượng khách hàng trung bình trong hàng đợi.
- W: Thời gian trung bình của khách hàng trong hệ thống.
- W_q: Thời gian trung bình của khách hàng trong hàng đợi.
- λ: Cường độ đến trung bình của khách hàng trên một đơn vị thời gian.
- μ: Cường độ phục vụ trung bình của ngân hàng cho một khách hàng trên một đơn vị thời gian.

2.4 Mô hình M/M/C/K

Theo chuẩn Kendall [1], mô hình bài toán tương đương với mô hình của lý thuyết hàng đợi kiểu M/M/C/K.

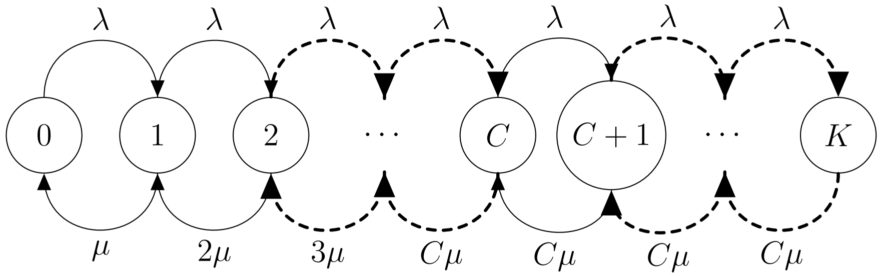
Tại thời điểm có $n-1$ khách hàng trong hàng đợi để thực hiện dịch vụ, nếu trong một khoảng λ đơn vị thời gian, thì sẽ có n khách hàng trong hàng đợi. Như vậy, trong hàng đợi sẽ chuyển từ trạng thái có $n-1$ khách hàng sang trạng thái có n khách hàng. Với thời gian phục vụ trung bình của mỗi trạm là μ khách hàng trong một đơn vị thời gian, thì sau một khoảng $n.\mu$ trạm phục vụ sẽ trở về trạng thái đang phục vụ $n-1$ khách hàng. Trạng thái hàng đợi sẽ chuyển từ trạng thái có n khách hàng sang trạng thái có $n-1$ khách hàng.

Như vậy, từ một bài toán hàng đợi được chuyển thành bài toán chuỗi trạng thái Markov rời rạc theo từng trạng thái của hàng đợi. Gọi n là khách hàng thứ n đến với dịch vụ, như vậy dựa vào trạng thái chuỗi Markov rời rạc. Mô hình hàng đợi M/M/C/K là một quá trình sinh chết với các giá trị sinh và giá trị chết thỏa mãn:

$$\lambda_n = \lambda \text{ nếu } n = 0, 1, ...$$

$$\mu_n = \begin{cases} n\mu \text{ nếu } n = 1, ..., C-1 \\ C\mu \text{ nếu } n = C, C+1, ..., K \end{cases}$$

Lý do mà $\mu_n = n\mu$ với $n = 1, ..., C-1$ là khi có ít hơn C khách hàng ở trong hệ thống , thì mỗi khách hàng sẽ được phục vụ và tốc độ phục vụ sẽ bằng tổng số khách hàng, Nhưng nếu có nhiều hơn C khách, thì chỉ có đúng C kênh phục vụ bận và do đó tốc độ phục vụ sẽ là $C\mu$. Ta có thể tính được xác suất khách hàng thứ n đến với dịch vụ là $p(n)$ như sau:



2.4.1 Xác suất khách hàng đến với dịch vụ

Theo [1] ta có một vài công thức sau đây:

Gọi n là khách hàng thứ n đến với dịch vụ, như vậy dựa vào trạng thái chuỗi Markov rời rạc, ta có thể tính được xác suất khách hàng thứ n đến với dịch vụ là $p(n)$ như sau:

- Nếu $n < C$:

$$n\mu p(n) = \lambda p(n-1) \tag{1}$$

$$p(n) = \frac{\lambda}{n.\mu} p(n-1) \tag{2}$$

$$p(n) = \frac{\rho^n}{n!} p(0) \tag{3}$$

Trong đó $\rho = \lambda/\mu$.

- Nếu $n \geq C$:

$$C\mu p(n) = \lambda p(n-1) \tag{4}$$

$$p(n) = \frac{\lambda}{C\mu} p(n-1) \tag{5}$$

$$p(n) = \frac{\rho^{n-c}}{C^{n-C}} p(C) \tag{6}$$

Trong đó $\rho = \lambda/\mu$.

Từ phương trình (5), để xác định giá trị $p(0)$ thì ta có:

$$p(C) = \frac{\rho}{C} p(C-1) \tag{7}$$

$$p(C) = \frac{\rho^C}{C!} p(0) \tag{8}$$

Do dó:

$$p(n) = \frac{\rho^n}{C^{n-C}C!} p(0) \tag{9}$$

Ta có:

$$\sum_{n=0}^{K} p(n) = 1 \tag{10}$$

$$\sum_{n=0}^{C-1} p(n) + \sum_{n=C}^{K} p(n) = 1 \tag{11}$$

$$\sum_{n=0}^{C-1} \frac{\rho^n}{n!} p(0) + \sum_{n=C}^{K} \frac{\rho^n}{C^{n-C}C!} p(0) = 1. \tag{12}$$

Từ đó suy ra:

$$p(0) = \left[\sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \sum_{n=C}^{K} \frac{\rho^n}{C^{n-C}C!} \right]^{-1}.$$

Như vậy, xác suất để khách hàng thứ n đến với dịch vụ $p(n)$ sẽ được tính như sau:

$$p(n) = \begin{cases} \dfrac{\rho^n}{n!} . p(0) \text{ nếu } 0 \leq n < C \\ \dfrac{\rho^n}{C^{n-C}C!} p(0) \quad \text{nếu} \quad n \geq C. \end{cases} \tag{13}$$

2.4.2 Xác suất khách hàng đến cần phải chờ

Trong trường hợp $n \geq C$, tức là số lượng khách hàng trong hệ thống nhiều hơn số lượng kênh phục vụ, khi đó khách hàng tiếp theo sẽ phải đợi. Xác suất của khách hàng cần phải chờ trong hệ thống được đưa cho bởi công thức sau:

$$C(C,\rho) = \sum_{n=C}^{K} p(n) = \sum_{n=C}^{K} \frac{\rho^n}{C^{n-C}C!} p(0). \tag{14}$$

2.4.3 Xác suất khách hàng tới dịch vụ khi hàng đợi đầy

Vì hàng đợi với số lượng hữu hạn, khi hàng đợi đầy thì dịch vụ sẽ bị từ chối. Do vậy, xác suất P_r để một khách hàng bị từ chối dịch vụ được tính như sau:

$$P_r = \sum_{n=K}^{\infty} p(n). \tag{15}$$

Kết hợp với (9), ta thu được:

$$P_r = \sum_{n=K}^{\infty} \frac{\rho^n}{C^{n-C}.C!}.p(0) \tag{16}$$

2.5 Chi phí cơ hội

Chi phí cơ hội là chi phí thất thoát mà các ngân hàng phải gánh chịu khi mà khách hàng rời đi khi hàng đợi đã đầy. Theo (16), ta có xác suất khách hàng rời đi khi gặp hàng đợi đầy là:

$$P_r = \sum_{n=K}^{\infty} \frac{\rho^n}{C^{n-C}.C!}.p(0).$$

Gọi C_0 là trung bình chi phí cơ hội khi từ chối phục vụ khách hàng (hàng đợi đầy). C_0 được cho bởi công thức dưới đây:

$$C_0 = P_r \lambda c \tag{17}$$

$$= \frac{\rho^K}{C^{K-C}C!} p(0) \sum_{n=o}^{\infty} \left(\frac{\rho}{C}\right)^n \lambda c. \tag{18}$$

trong đó c là phí mà khách hàng phải chi trả khi tham gia vào hệ thống.

2.5.1 Số lượng khách hàng trung bình trong hàng đợi và trong hệ thống

Giả sử gọi L_q là số khách hàng trung bình trong hàng đợi. Số khách hàng trung bình L_q được tính như sau:

$$L_q = \sum_{n=C}^{K} (n-C) p(n) \tag{19}$$

Từ (9), ta có:

$$L_q = \sum_{n=C}^{K}(n-C)\frac{\rho^n}{C^{n-C}.C!}p(0) \tag{20}$$

$$L_q = p(0)\frac{\rho^C}{C!}\sum_{n=0}^{K-C-1} n\left(\frac{\rho}{C}\right)^n \tag{21}$$

Từ đó ta cũng tính được trung bình số khách hàng ở trong hệ thống:

$$L = L_q + \left(C - \sum_{n=0}^{C-1}(C-n)p(n)\right) = L_q + C - \sum_{n=0}^{C-1}(C-n)\frac{\rho^n}{n!}p(0). \tag{22}$$

2.5.2 Thời gian đợi trung bình của khách hàng

Giả sử gọi W_q là thời gian đợi trung bình của mỗi khách hàng trong hàng đợi khi đến với dịch vụ. Giá trị W_q được tính như sau:

$$W_q = \frac{L_q}{\lambda}. \tag{23}$$

Như vậy:

$$W_q = p(0).\frac{\rho^C}{\lambda.C!}\sum_{n=0}^{K-C-1} n\left(\frac{\rho}{C}\right)^n \tag{24}$$

Thời gian trung bình của khách hàng khi ở trong hệ thống là:

$$W = W_q + \frac{1}{\mu}. \tag{25}$$

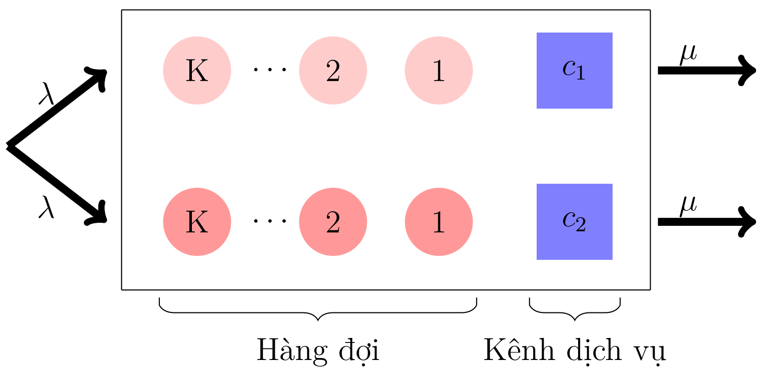
3 Áp dụng lý thuyết hàng đợi vào thiết kế hệ thống hàng đợi

Bây giờ chúng ta sẽ áp dụng các kiến thức của lý thuyết hàng đợi để thiết kế một hệ thống xếp hàng của một ngân hàng. Nhưng đầu tiên ta sẽ nói về một số vấn đề của các ngân hàng hiện tại. Sau khi đi khảo sát thực tế, chúng tôi đã nhận thấy rằng các ngân hàng thường không chỉ có một hàng đợi mà là có đến hai hoặc ba hàng đợi. Ở đây chúng tôi gọi mỗi hàng đợi là một kênh dịch vụ vì ở mỗi hàng sẽ có các nhân viên phục vụ riêng và hoàn toàn độc lập với nhau. Hiển nhiên thời gian phục vụ cho mỗi khách hàng của từng nhân viên là không giống nhau. Do đó nguyên tắc "Đến trước phục vụ trước" (FIFO) sẽ bị vi phạm. Điều này dẫn đến việc không hài lòng ở khách hàng gây nên nhiều vấn đề và thậm chí là thiệt hại về kinh tế. Khi đó câu hỏi đặt ra sẽ là:" Thiết kế với bao nhiêu cửa số dịch vụ là tối ưu nhất và chi phí cơ hội là tối thiểu?". Chúng tôi sẽ phân tich dưới đây và đưa ra câu trả lời. Lưu ý rằng các bài toán dưới đây sẽ có giả thiết rằng hàng chờ của các ngân hàng là hữu hạn (Mô hình $M/M/C/K$) vì điều này hoàn toàn đúng trong thực tế.

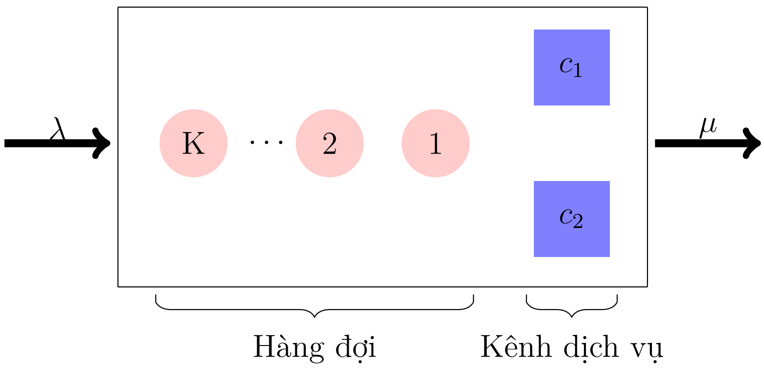
3.1 Mô hình có một hàng đợi và hai hàng đợi

Giả sử rằng ta có hai mô hình hàng đợi. Mô hình thứ nhất có một hàng đợi và hai kênh dịch vụ nghĩa là mô hình $M/M/2/K$ (hình 2). Mô hình thứ hai có hai hàng đợi và mỗi hàng đợi có một kênh dịch vụ (hình 3), các kênh dịch vụ này hoạt động hoàn toàn độc lập với nhau. Cường độ khách hàng đến là $\lambda = 16$ khách trên một đơn vị thời gian. Ta cũng xem rằng cường độ phục vụ của mỗi dịch vụ là $\mu = 20$. Khi đó xác suất khách hàng tham gia vào một hàng chờ là $1/2$ và ta sẽ có hai mô hình hàng đợi $M/M/1/K$ với $\lambda_1 = \lambda_2 = \lambda/2 = 8$. Bây giờ ta sẽ so sánh hai mô hình trên dựa vào một vài tiêu chí cụ thể.

Mô hình bài toán:



Hình 2: Mô hình hai hàng đợi $M/M/1/K$.



Hình 3: Mô hình hàng đợi $M/M/2/K$.

Ta có kết quả thu được như sau:

Số hàng chờ	λ	μ	L_q	L	W_q	W
1	16	20	0.152	0.952	0.010	0.060
2	8	20	0.267	0.667	0.033	0.083

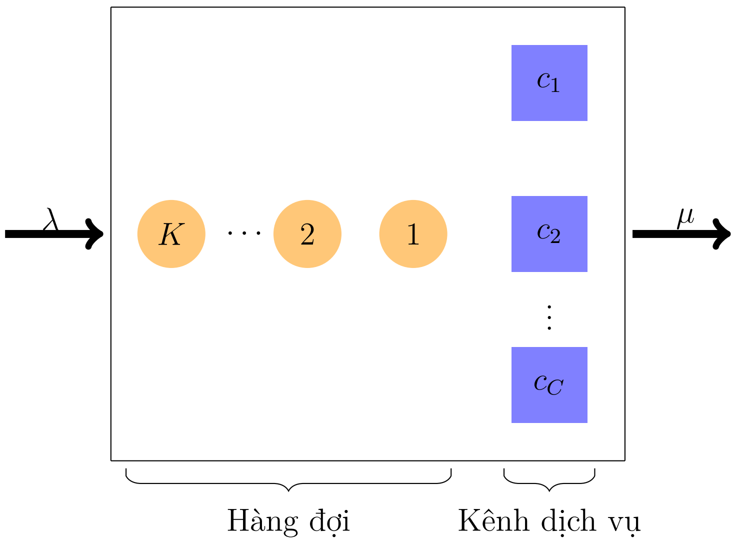
Làm tương tự với hai mô hình là mô hình có 3 hàng đợi $M/M/1/K$ và mô hình $M/M/3/K$, ta thu được kết quả:

Số hàng chờ	λ	μ	L_q	L	W_q	W
1	16	20	0.019	0.819	0.001	0.051
3	$\frac{16}{3}$	20	0.097	0.364	0.018	0.068

Từ bảng trên ta có thể thấy việc mô hình chia thành nhiều hàng chờ giúp cho việc giảm bớt độ dài hàng chờ trong cả hệ thống tuy nhiên điều đáng bàn là việc này cũng làm cho thời gian đợi trung bình W của một khách hàng trở nên lâu hơn so với mô hình một hàng chờ có số kênh dịch vụ bằng với nó. Do đó khi thiết kế hệ thống, ta sẽ ưu tiên việc xây dựng nhiều kênh dịch vụ hơn là chia thành nhiều hàng chờ.

3.2 Mô hình có nhiều dịch vụ

Ta sẽ tiếp tục với ví dụ trên nhưng thay bằng việc chia thành nhiều hàng đợi, ta sẽ giữ nguyên một hàng đợi và mở nhiều kênh dịch vụ. Mô hình bài toán lúc này như sau:



Hình 4: Mô hình hàng đợi có nhiều kênh dịch vụ.

Sử dụng công thức (14), chạy chương trình, ta thu được kết quả như sau:

Số dịch vụ	1	2	3	4	5
Xs chờ của khách hàng $C(c,\rho)$	0.800	0.229	0.052	0.009	0.001

Từ bảng trên ta có thể thấy rằng việc tăng thêm các dịch vụ sẽ giúp làm giảm xác suất chờ của khách hàng. Điều này hoàn toàn dễ hiểu bởi vì khi ta mở thêm nhiều kênh phục vụ, số khách hàng được phục vụ sẽ tăng lên nhiều hơn, qua đó làm giảm xác suất chờ của khách hàng.

3.3 Tối ưu chi phí

Qua các mục trên ta có thể thấy rằng mô hình đem lại lợi nhuận cao nhất cho ngân hàng chính là mô hình với một hàng đợi và nhiều cổng dịch vụ. Trong mục này ta sẽ quyết định xem thiết kế bao nhiêu cổng dịch vụ là tối ưu nhất. Để đơn giản hoá vấn đề, trong phần dưới đây, chúng tôi chỉ áp dụng cho mô hình có hàng đợi vô hạn, túc là mô hình $M/M/C$. Ta gọi c_q là chi phí để phục vụ cho một khách hàng có trong hệ thống trong một giờ, c_s là tiền lương phải chi trả cho nhân viên ở các cổng dịch vụ trong một giờ. Tổng chi phí phải chi trả trong một giờ là:

$$f = xc_s + Lc_q.$$

Trong đó x là số kênh dịch vụ. Ta cần tìm x sao chi chi phí f là nhỏ nhất. Theo (22), cho $K \to \infty$ ta có:

$$L = \frac{\lambda}{x\mu - \lambda.}$$

Suy ra:

$$f = x * c_s + \frac{\lambda\mu}{x\mu - \lambda}c_q$$

Để f nhỏ nhất thì:

$$\begin{aligned} &f'_x = 0 \\ \Leftrightarrow &c_s - \frac{\lambda\mu}{(x\mu - \lambda)^2}c_q = 0 \\ \Rightarrow &x = \frac{1}{\mu}(\lambda + \sqrt{\frac{c_q\lambda\mu}{c_s}}). \end{aligned}$$

Với ví dụ như ở trên, ta có $\lambda = 16$, $\mu = 20$, cho $c_q = 4$, $c_s = 20$ (đơn vị: ngàn đồng), ta tính được số cổng dịch vụ là $c = x \approx 1$. Vậy chỉ nên mở 1 cổng dịch vụ để phục vụ khách hàng.

4 Thảo luận

Với những phân tích nêu trên ta có thể thấy rằng có nhiều cách thiết kế mô hình dựa trên lý thuyết hàng đợi sao cho chi phí bỏ ra của các ngân hàng là thấp nhất. Và mô hình tối ưu nhất chính là mô hình một hàng đợi và có nhiều kênh dịch vụ. Tuy nhiên việc quyết định bao nhiêu kênh dịch vụ còn phụ thuộc vào lượng khách hàng cũng như chi phí chi trả của các ngân hàng.

Ví dụ trên được thu thập dữ liệu từ việc khảo sát thực tế tại một ngân hàng. Cách thiết kế này hoàn toàn có thể áp dụng cho các mô hình kinh doanh có hàng đợi khác như là cửa hàng ăn, xếp hàng vào kho, khám chữa bệnh,... Đối với mỗi mô hình kinh doanh ta sẽ lựa chọn các phương pháp thích hợp nhất.

5 Tổng kết

Bài báo cáo này của chúng tôi đã đưa ra những phân tích mô hình toán học để xác định các tham số quan trọng như độ dài hàng đợi trung bình, thời gian chờ trung bình, chi phí cơ hội,... Nhà thiết kế hoàn toàn có thể sử dụng những công cụ trên để tìm ra giải pháp tối ưu cho việc thiết kế hệ thống của họ.

Ngoài ra, ứng dụng lý thuyết hàng đợi vào thiết kế hệ thống tại các ngân hàng cũng đã được chúng tôi giới thiệu. Giải pháp đề xuất của bài toán cũng đã giúp giảm thời gian chờ đợi của khách hàng, cải thiện chất lượng phục vụ và đem lại lợi nhuận cao cho các ngân hàng. Do đó, việc áp dụng lý thuyết xếp hàng trong thiết kế dịch vụ sẽ giúp các nhà thiết kế lựa chọn giải pháp tốt nhất thỏa mãn nhu cầu của khách hàng bằng cách giảm thiểu thời gian chờ đợi của khách hàng và chi phí vận hành của dịch vụ thiết kế.

Lời cảm ơn

Cuối cùng, chúng tôi xin bày tỏ lòng biết ơn sâu sắc đến *TS. Nguyễn Thị Ngọc Anh*, người đã tận tình giảng dạy và hướng dẫn trong học phần *Các mô hình ngẫu nhiên và ứng dụng* để chúng em có thể hoàn thành bài báo cáo này.

Tuy đã cố gắng thực hiện chuẩn mực nhất có thể, nhưng do thời gian hoàn thành bài báo cáo và vấn đề hiểu biết của chúng em còn rất nhiều hạn chế, nên bài báo cáo không thể tránh khỏi việc tồn tại những thiếu sót và cần bổ sung. Chúng tôi kính mong sẽ nhận được sự góp ý từ cô và mọi người để có thể hoàn thiện chủ đề cũng như hoàn thiện kiến thức của mình hơn nữa.

Chúng tôi xin chân thành cảm ơn!

Tài liệu tham khảo

Tài liệu

[1] Richard M Feldman and Ciriaco Valdez-Flores. *Applied probability and stochastic processes*. Springer Science & Business Media, 2009.

[2] Huimin Xiao and Guozheng Zhang. The queuing theory application in bank service optimization. In *2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM)*, volume 2, pages 1097–1100. IEEE, 2010.