Operations Research Perspectives 5 (2018) 145–149

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

The expected discrimination frequency for two-server queues

Berenice Anne Neumann[a], Hendrik Baumann*[,b]

[a] Research Group Statistics and Stochastic Processes, University of Hamburg, 20146 Hamburg, Germany
[b] Institute of Applied Stochastics and Operations Research, Clausthal University of Technology, 38678 Clausthal-Zellerfeld, Germany

ARTICLE INFO

Keywords:
Queueing
Fairness measures
Multi-server queue
Combined queue vs. separate queues

2010 MSC:
60K25
68M20
60J28

ABSTRACT

Fairness measures for queues were introduced for measuring the individual satisfaction of human customers with respect to the waiting experience. The measure which performs best in some sense is the expected discrimination frequency (DF). In contrast to competing fairness measures, up to now, the DF has not been thoroughly analysed for multi-server systems. In particular, there are no results concerning the question whether or not in terms of the DF, combined queues are fairer than separate queues. In this note, we prove that under Markovian assumptions, combined queues are fairer and, furthermore, that this statement does not remain true for general queueing systems.

1. Introduction

Traditionally, the system performance of queueing systems is measured by characteristics such as waiting times, throughput, ... In recent years, fairness measures have been paid attention to. Considering fairness in queues has various reasons, and therefore, various kinds of fairness measures have been introduced.

In computer applications, it is a quite natural approach to consider the proportion of the response time of a job of size x to its size x. This quotient is referred to as the *slowdown*. For a queue with stochastic arrival process and stochastic service times, by considering stationary behaviour and taking the expectation, the (un)fairness of scheduling disciplines can be classified [1,2]. It turns out that the disciplines PS (processor sharing) and preemptive LCFS (last come, first served) can be regarded as some kind of fair with respect to the expected slowdown.

In many applications of queueing theory, human customers are involved (for example, supermarkets, waiting rooms at doctor's offices, check-in areas at airports, ...). Whereas slowdown-based considerations intend to find an abstract classification of fairness, for systems with human customers, psychological aspects become important: Human customers will judge the system by means of the 'perceived fairness'. Based on their satisfaction with their waiting experience, they will decide whether or not to revisit the facility providing the waiting system in the future (if they have a choice). Usually, human customers will not judge preemptive LCFS as a fair scheduling discipline, and hence, the slowdown-based classification of (un)fairness cannot be applied in this context.

Pychological studies [3] revealed that human customers perceive 'unfairness' if they are overtaken by other customers or if customers with a larger job size are allowed to leave the system earlier. Based on these findings, principles for measuring perceived fairness have been established [4]: For single-server queues, fairness measures should fulfill a *seniority preference principle* and a *service-requirement preference principle*. In their strong version, tests for these principles require that

- if two jobs have the same service requirement, the job which arrived earlier should be completed first,
- if two jobs arrive at the same time, the job with smaller service requirement should be completed first.

In both cases, interchanging the order of service of the two jobs under consideration should lead to a lower fairness/ higher unfairness. In order to analyse perceived fairness, order fairness [5], a slowdown-based measure [6], the measure RAQFM (resource allocation queueing fairness measure) [7] and the discrimination frequency (DF) [8] have been introduced, further analysis can be found in [9–13]. In some way, the DF performs best with respect to the principles established in [4], since it is the only measure introduced so far which satisfies the strong tests both for the seniority principle and the service requirement principle.

For multi-server systems, there is psychological evidence that human customers generally judge single-queue systems fairer than multi-queue systems, see [3]. A measure being appropriate for evaluating the fairness of multi-server and multi-queue systems should

* Corresponding author.
E-mail addresses: berenice.neumann@uni-hamburg.de (B.A. Neumann), hendrik.baumann@tu-clausthal.de (H. Baumann).

https://doi.org/10.1016/j.orp.2018.06.001
Received 8 January 2018; Received in revised form 27 April 2018; Accepted 12 June 2018
Available online 19 June 2018

reflect this judgement. For the RAQFM, an analysis has been performed in [10], yielding that for $G/D/k$ and $M/M/2$ models, the single-queue is fairer than the multi-queue. However, it is shown that this result does not hold for general $G/G/k$ queues. The main goal of this paper is to provide a similar analysis for the DF in the case of simple Markovian systems. We will focus on the FCFS discipline, but nevertheless, our results can be interpreted as a starting point for a future investigation of the impact of the scheduling discipline on the discrimination frequency in multi-server systems.

The structure will be as follows: In Section 2, we will describe the considered single-queue and multi-queue system, and restate the precise definition of the discrimination frequency. In Section 3, we will derive the expected DF for the single-queue system, and in Section 4, we will determine a lower bound for the expected discrimination frequency in the multi-queue system and prove that indeed, in terms of the DF, the single-queue system is fairer than the multi-queue system. In Section 5, we will present an example that for general (non-Markovian) systems this statement does not remain true. In Section 6, we will summarize our results, and we will outline possible directions of further research.

2. Basic terms and models under consideration

In this paper, we aim for comparing the expected discrimination frequency for an $M/M/2$-model and two $M/M/1$ models with separate queues. We briefly present both models and the precise definition of the discrimination frequency.

2.1. The M/M/2 model

Customers arrive according to a Poisson process with intensity λ. There are two identical servers, and the service times are independently and identically Exp(μ)-distributed. Furthermore, there is no restriction of the number of waiting customers, and the scheduling discipline is FCFS (first come, first served). Due to these modelling assumptions, the process $(N_t)_{t\geq 0}$ of the number N_t of customers in the system (waiting in the queue or being served) is a continuous-time Markov chain (CTMC). In case $\rho = \frac{\lambda}{2\mu} < 1$, the system is stable, and in the long-run, it will behave stationarily, that is, for any $k = 0, 1, 2, \ldots$, we have $\lim_{t\to\infty} \mathbb{P}(N_t = k) = \pi_k$, where $\pi = (\pi_k)_{k=0}^{\infty}$ is the stationary distribution. It is well-known [14, Section 3.5] that

$$\pi_0 = \frac{1-\rho}{1+\rho} \qquad \text{and} \qquad \pi_n = 2\rho^n\pi_0, \ \ n \geq 1.$$

Due to the PASTA property of the arrival process [15, Theorem VII.6.7], in the long-run, arriving customers will 'see' the stationary distribution, that is, with probability π_k, an arriving customer will find k other customers in the system. Note that for the stationary number N of customers in the system, we have $\mathbb{E}[N] = \sum_{n=0}^{\infty} n\pi_n = \frac{2\rho}{(1-\rho)(1+\rho)}$.

2.2. The multi-queue model

In order to model two separate queues, we consider two parallel $M/M/1$ models. Customers still arrive according to a Poisson process with parameter λ. Each arriving customer will join the first system with probability $\frac{1}{2}$, and the second one with probability $\frac{1}{2}$. Hence, the arrival process for each of both systems is a Poisson process with intensity $\frac{\lambda}{2}$. Both systems have one server, and the service times are independently and identically Exp(μ) distributed. Still, we assume infinite waiting capacity and FCFS as scheduling discipline. Let $(N_t^{(1)}, N_t^{(2)})_{t\geq 0}$ be the process of the number of customers in the first and in the second system, respectively. Due to the modelling assumptions, this process is again a CTMC, and furthermore, $(N_t^{(1)})_{t\geq 0}$ and $(N_t^{(2)})_{t\geq 0}$ are independent, and both are CTMCs. For $\rho = \frac{\lambda}{2\mu}$, we have stability, and

$$\lim_{t\to\infty} \mathbb{P}(N_t^{(1)} = k) = \lim_{t\to\infty} \mathbb{P}(N_t^{(2)} = k) = \pi_k, \qquad k = 0, 1, 2, \ldots,$$

where $\pi = (\pi_k)_{k=0}^{\infty}$ is the stationary distribution. Again, the exact shape of π is well-known [14, Section 3.2], we have $\pi_k = (1-\rho)\rho^k$ for all $k = 0, 1, 2, \ldots$. As for the $M/M/2$ model, we have the PASTA property, that is, in the long-run, with probability $\pi_k \cdot \pi_\ell$ an arriving customer sees k other customer in the first system, and ℓ other customers in the second system.

Although we will compare the fairness (measured by the discrimination frequency), we briefly recapitulate that traditional performance measure favor the combined queue over the separate queue: Let N be the total stationary number of customers in the system. Then $N = N^{(1)} + N^{(2)}$ and $\mathbb{E}[N] = \frac{2\rho}{1-\rho}$, and this number is larger (by factor $1+\rho$) than the corresponding expected number of customers in the $M/M/2$ queue. Since Little's formula guarantees that the expected response (or sojourn) time of any 'black box' can be determined by $\frac{\mathbb{E}[N]}{\lambda}$, this result carries over to response times.

Note that there are different ways to 'choose' the queue an arriving customer joins. Here, we consider the 'coin toss'. A natural alternative is joining the shorter queue (if there is one). In this case, the stationary numbers of customers in the systems depend on each other. We leave this topic open for future research.

2.3. The discrimination frequency

The discrimination frequency was introduced in [8]. The intuitive concept behind it is to count the discriminating events a customer suffers from. These are *large jobs*, that are jobs which have a larger remaining service requirement at our job's time of arrival, but leave the system earlier, and *overtakes*, that are jobs which arrive after and leave before our marked job. Formally, in [8], the DF was defined as follows:

Definition 2.1. Let a_i, d_i, s_i be the arrival time, the departure time, and the service time of job J_i. Furthermore, let $s_j'(t)$ be the residual service time of J_j at time t (if J_j did not enter the system at time t, we have $s_j'(t) = s_j$). Then the amount $OV(i)$ of overtakes job J_i suffers from is

$$OV(i) := |\{j: (a_j \geq a_i \wedge d_j \leq d_i)\}|.$$

The amount $LJ(i)$ of large jobs that a job J_i suffers from is

$$LJ(i) := |\{j: (d_i \geq d_j > a_i \wedge s_j'(a_i) \geq s_i\}|.$$

The discrimination frequency of job J_i is

$$DF(i) = OV(i) + LJ(i).$$

The discrimination frequency of a system in steady state is the discrimination frequency of a stationary customer.

For stationary systems, the distribution of $OV(i)$, $LJ(i)$, and $DF(i)$ is identical for all customers i. We will refer to the number of overtakes, the number of large jobs, and the discrimination frequency of a randomly chosen customer as OV, LJ, and DF respectively. Hence, we will consider a 'tagged customer' who sees the stationary distribution of number of customers in the instant of his arrival, and we will pursue his way through the system, and count the number of overtakes and large jobs he suffers from.

3. The expected discrimination frequency for the combined queue

In order to compute $\mathbb{E}[DF]$, we determine $\mathbb{E}[LJ]$ and $\mathbb{E}[OV]$. Note that under FCFS, large jobs are only caused by customers which have entered the system before our tagged customer, and overtakes are only caused by customers which will enter the system after our tagged customer. Precisely, we will prove the following result in the next subsections.

Theorem 3.1. *For an $M/M/2$ model with a combined FCFS, the expected number of large jobs is $\mathbb{E}[LJ] = \frac{\rho^2}{(1-\rho)(1+\rho)}$, the expected number of overtakes in the $M/M/2$-system with a combined FCFS queue is $\mathbb{E}[OV] = \frac{\rho}{1+\rho}$, and the expected discrimination frequency is*

$$\mathbb{E}[DF] = \mathbb{E}[LJ] + \mathbb{E}[OV] = \frac{\rho}{(1-\rho)(1+\rho)}.$$

3.1. Comparing exponentially distributed random variables

The proofs of the formulas for $\mathbb{E}[LJ]$ and $\mathbb{E}[OV]$ rely on the comparison of exponentially distributed random variables. We use three well-known results:

1. Let $Z, Z_1, ..., Z_n$ be independent and exponentially distributed with parameter μ. Then $L = |\{k \in \{1,...,n\}: Z < Z_k\}|$ is uniformly distributed on $\{0, ..., n\}$ with expectation $\frac{n}{2}$. (Note that this statement holds true for any random variables independent and identically distributed random variables $Z, Z_1, ..., Z_n$ with continuous cumulative distribution function.)
2. Let Z_1, Z_2 be independent and exponentially distributed with parameters μ_1, μ_2, respectively. Then $\mathbb{P}(Z_1 < Z_2) = \mathbb{P}(Z_1 \le Z_2) = \frac{\mu_1}{\mu_1+\mu_2}$. In particular, for $\mu_1 = \mu_2$, this probability simplifies to $\frac{1}{2}$.
3. Let $Z, Z_1, Z_2, ...$ be independent where Z is exponentially distributed with some parameter α and $Z_1, Z_2, ...$ are exponentially distributed with parameter μ. According to the second result, the event $Z > Z_1$ occurs with probability $\frac{\mu}{\alpha+\mu}$. Furthermore, due to the memoryless property, we have $\mathbb{P}(Z > Z_1+...+Z_k | Z > Z_1+...+Z_{k-1}) = \frac{\mu}{\alpha+\mu}$ for all $k \ge 2$, and defining $Y = \sup\{k: Z > Z_1 + ...+Z_k\}$ (with $\sup \varnothing = 0$), we conclude that $\mathbb{P}(Y = k) = \left(\frac{\mu}{\alpha+\mu}\right)^k \cdot \left(1 - \frac{\mu}{\alpha+\mu}\right)$ for $k = 0, 1, 2, ...$, that is, Y is geometrically distributed with parameter $\frac{\mu}{\alpha+\mu}$.

3.2. The expected number of large jobs

At the instant of arrival, our tagged customer sees n other customers with probability π_n, where $\pi_0 = \frac{1-\rho}{1+\rho}$ and $\pi_n = 2\rho^n\pi_0$ for $n \ge 1$. If $n \le 1$, the service of the tagged customer starts immediately, and either no other customer leaves the system before our customer does, or a customer with smaller (residual) service time leaves before our customer. In both cases, there is no large job our tagged customer suffers from.

In case $n \ge 2$, according to the considerations concerning the comparison of exponentially distributed random variables, the number of customers causing large jobs is uniformly distributed on $\{0, ..., n\}$ with expectation $\frac{n}{2}$. At the instant in which the service of our tagged customer begins, the nth of these customers is still in service. With probability $\frac{1}{2}$, his residual service time is smaller than the service time of our tagged customer, and all of the n considered customers leave the system before our tagged customer does. With probability $\frac{1}{2}$, the residual service time of the nth customer is larger than the service time of our tagged customer, and only $n-1$ of the other customers leave the system before our customer does. In this case, the nth customer would definitely have been a large job, and therefore, we have to subtract 1 from the number of large jobs. Summarizing these considerations, we obtain

$$\mathbb{E}[LJ|N = n] = \frac{n}{2} - \frac{1}{2}\cdot 1 = \frac{n-1}{2}.$$

By applying total probability, we find

$$\begin{aligned}\mathbb{E}[LJ] &= \sum_{n=2}^{\infty} \pi_n \mathbb{E}[LJ|N=n] = \sum_{n=2}^{\infty} \frac{1-\rho}{1+\rho}\cdot 2\rho^n \cdot \frac{n-1}{2} \\ &= \frac{(1-\rho)\rho^2}{1+\rho} \sum_{n=2}^{\infty} (n-1)\rho^{n-2} = \frac{(1-\rho)\rho^2}{1+\rho} \sum_{n=0}^{\infty} (n+1)\rho^n \\ &= \frac{(1-\rho)\rho^2}{1+\rho} \cdot \frac{1}{(1-\rho)^2} = \frac{\rho^2}{(1-\rho)(1+\rho)}.\end{aligned}$$

3.3. The expected number of overtakes

The formula for $\mathbb{E}[OV]$ is due to results found in the work of Gordon [16]. Note that overtakes were defined in the sense of overtakes a customer suffers from. In the terminology of [16], this number corresponds to the number of *slips*. On the other hand, *skips* refer to the number of customers which are overtaken by a fixed customer. For a stationary system, the expected number of slips and skips coincide. For an $M/M/2$ system, a randomly chosen customer will skip at most one customer. This is the case if and only if the randomly chosen customer does not find an empty system (probability $1 - \pi_0$), and if, in the instant of starting to be served, his service requirement is smaller than the remaining service time of the customer at the other server. Since both (residual) service requirements are Exp(μ) distributed, this probability is $\frac{1}{2}$. In total, the expected number of skips, and thus of slips is

$$\mathbb{E}[OV] = \frac{1}{2}(1-\pi_0) = \frac{1}{2}\frac{2\rho}{1+\rho} = \frac{\rho}{1+\rho}.$$

This derivation can be interpreted as a special case of [16, p. 160 and p.165].

4. Lower bounds for the expected discrimination frequency for two servers with separate queues

As for the $M/M/2$ system, we consider $\mathbb{E}[LJ]$ and $\mathbb{E}[OV]$ separately in the subsections below. Note that we will only give a simple lower bound for $\mathbb{E}[LJ]$, but this lower bound enables us to prove that in terms of the discrimination frequency, the system with a combined queue is fairer than the system with separate queues.

Theorem 4.1. *For two parallel $M/M/1$-systems, we have* $\mathbb{E}[LJ] \ge \frac{\rho}{2(1-\rho)}$, $\mathbb{E}[OV] = \frac{\rho}{2(1-\rho)}$ *and*

$$\mathbb{E}[DF] \ge \frac{\rho}{(1-\rho)} \ge \mathbb{E}[DF_c],$$

where DF_c is the discrimination frequency in an $M/M/2$-system, that is, the corresponding system with a combined queue.

4.1. The expected number of large jobs

The tagged customer enters one of both systems, which are independent $M/M/1$ queues with arrival rate $\frac{\lambda}{2}$ and service rate μ. He can suffer from large jobs which occur in his own queue, and from large jobs which result from customers served in the other queue.

The expected number of large jobs which occur in the system the tagged customer joins coincides with the number of large jobs which arise in an $M/M/1$ queue with utilization $\rho = \frac{\lambda}{2\mu}$. Following [13], the expectation of this quantity can be derivated as follows: The expected number of customers found in the system by the tagged customer is $\frac{\rho}{1-\rho}$, and each of these customers will be a large job with probability $\frac{1}{2}$. Hence, the expected number of large jobs which occur in the queue our tagged customer has entered computes as $\frac{\rho}{2(1-\rho)}$. Since there may be large jobs in the other queue as well, we find

$$\mathbb{E}[LJ] \ge \frac{\rho}{2(1-\rho)}.$$

In a last step, we would have to determine the expected number of large jobs that occur in the other queue. Since the above result already enables us to prove Theorem 4.1, we omit any further considerations.

4.2. The expected number of overtakes

As pointed out above, in [16], the number of *skips* is the number of overtakes a customer performs, and the number of *slips* is the number of overtakes a customer suffers from. Hence, we are interested in the

expected number of slips of a randomly chosen customer, and according to more general results [16], this expected number is given by $\frac{\rho}{2(1-\rho)}$. Under our conditions, the derivation of this result simplifies, and in the next lines, we give a concise proof:

Our customer can overtake at most the N customers that are in the other queue at the instant of his arrival. Either at least all of these N customers are served during our customers sojourn time in which case no slips occur or only $k \leq N$ customers are served implying that the number of slips is $N - k$. We can now describe the number of slips by $N - \min\{N, Y\}$ with Y being the number of customers served during the sojourn time of the tagged customer provided that the system never runs empty.

Since our customer joins an ordinary $M/M/1$ queue, his sojourn time is exponentially distributed with parameter $\mu(1-\rho)$, and due to all service times being exponentially distributed with parameter μ, we can use the third property from Section 3.1, and find that Y is geometrically distributed with parameter $\frac{\mu}{\mu+\mu(1-\rho)} = \frac{1}{2-\rho}$. Since N is geometrically distributed with parameter ρ, $\min\{Y, N\}$ is geometrically distributed with parameter $\rho \cdot \frac{1}{2-\rho} = \frac{\rho}{2-\rho}$. In total, the expected number of slips and skips is given by

$$\mathbb{E}[OV] = \mathbb{E}[N] - \mathbb{E}[\min\{N, Y\}] = \frac{\rho}{1-\rho} - \frac{\frac{\rho}{2-\rho}}{1-\frac{\rho}{2-\rho}}$$
$$= \frac{\rho}{1-\rho} - \frac{\rho}{2-2\rho} = \frac{\rho}{2(1-\rho)}.$$

5. Counterexample: Deterministic arrival process and deterministic service times

Whereas for Markovian systems, the single-queue system is fairer (in terms of the DF) than the multi-queue system, unfortunately, this statement does not hold for arbitrary systems.

Consider the following example with two servers: Let the service time be 1 for all customers, and let there be two arrivals at time 0, two arrivals at time 2, two arrivals at time 4, and so on, that is, the inter-arrival times alternate deterministically between 0 and 2. First assume that we have a single queue for both servers. Then each arrival suffers from one large job and one overtake (caused by the customer which arrived at the same time). Hence, the expected discrimination frequency for a randomly chosen customer is 2.

Now consider the multi-queue system, where each job joins each queue with probability $\frac{1}{2}$. Still, each customer can be discriminated at most twice. But with probability $\frac{1}{4}$ both customers arriving at the same time are assigned to the same queue. In this case, the first of these customers is not discriminated at all. Hence, for a randomly chosen customer, there is a positive probability that his DF is 0, and it follows that the expected DF is < 2. This behaviour of the discrimination frequency is not desired as it contrasts the psychological findings presented in [3], where it has been reported that human cusomters judge the single-queue to be fairer than the multi-queue. There are two possibilities to deal with this problem:

- We could restrict the class of models under consideration. For example, we could only allow independent interarrival times with continuous distribution. It seems reasonable to hope that under these restrictions no counterexamples can be constructed since then the events that two customers enter the system at the same time, or leave the system at the same time occur with probability 0.
- We could think about changing the definition of the discrimination frequency since the only reason for the above counterexample to work is that for overtakes, we not only count the jobs that leave before the tagged customer, but also the jobs that leave at the same time. So, an idea would be not to count the customers leaving at the same time. In case of continuous distributions (as for the models discussed in Sections 3 and 4), the results are not effected at all. On the other hand, the discrimination frequency was originally developed for evaluating the fairness of scheduling disciplines for single-server queues, and therefore, the most important feature should be that criteria for fairness measures are met for single-server queues.

Both ways require further research. In the first case, we have to analyse a quite large class of queueing models, and in the second case, we have to recheck whether or not slight adjustments of the definition of the discrimination frequency change the fact that it meets the basic criteria for fairness measures such as the strong service-requirement preference test and the strong seniority preference test (see [4], [8]).

6. Conclusion and further research

For queueing systems with two servers, psychological studies suggest that customers will judge the system with a combined queue as 'fairer' than that with separate queues. Hence, if a function shall measure the fairness of multi-server queueing systems, it should reflect this judgement. For the fairness measure RAQFM, this analysis has been performed in [3], and in this paper, we have provided some considerations for the fairness measure DF (discrimination frequency). Indeed, the main results of this paper show that under Markovian assumptions (Poisson input and exponential service times) the combined queue is fairer than the separate queues in terms of the DF. Unfortunately, this statement does not remain true for arbitrary queueing systems. However, for single-server queueing systems, the measure DF performs better than RAQFM (see [8]) with respect to properties for fairness measures which were established based on psychological foundings.

Hence, future research could intend to adjust the measure DF in such a way that it does not lose the desired properties for single-server systems, see Section 5. Alternatively, more models could be analysed in order to investigate whether or not the measure DF meets the fairness judgements for some 'large' class of multi-server queues.

The first step in this direction would be a generalization of the analysis presented in Sections 3 and 4 to systems with more than two servers. Since most of our results easily extend to this case, we strongly conjecture that the main result remains true. Nevertheless, for the system with separate queues an exact term for the expected number of large jobs would be desirable, and more 'intelligent' queueing strategies (e.g., choose the shorter queue) could be considered. In further steps, Semi-Markovian models, that is, models with $M/G/\bullet$- or $G/M/\bullet$-assumptions, could be analysed.

Another interesting task is the determination of higher moments of the DF or its complete distribution. As pointed out in Section 3, the distributions of the numbers of skips and overtakes (slips) do not coincide, and therefore, approaches focusing on overtakes instead of skips will have to be developed.

Finally, we want to remark that the measure DF was originally defined in order to analyse the impact of the scheduling discipline on the fairness. For single-server queues, some comparisons are given in [13], and it is natural to investigate which of these results extend to multi-server systems.

Acknowledgements

We acknowledge support by Open Access Publishing Fund of Clausthal University of Technology.

Furthermore, we would like to thank the anonymous referees for their comments on our manuscript.

References

[1] Wierman A, Harchol-Balter M. Classifying scheduling policies with respect to unfairness in an m/GI/1. SIGMETRICS Perform Eval Rev 2003;31(1):238–49.

[2] Wierman A. Fairness and classifications. SIGMETRICS Perform Eval Rev 2007;34(4):4–12.
[3] Rafaeli A, Barron G, Haber K. The Effects of Queue Structure on Attitudes. J Service Res 2002;5(2):125–39.
[4] Avi-Itzhak B, Levy H, Raz D. Quantifying fairness in queueing systems - principles, approaches, and applicability. Probab Eng Inf Sci 2008;22(4):495–517.
[5] Avi-Itzhak B, Levy H. On measuring fairness in queues. Adv Appl Probab 2004;36(3):919–36.
[6] Avi-Itzhak B, Brosh E, Levy H. SQF: A slowdown queueing fairness measure. Perform Eval 2007;64(9):1121–36.
[7] Avi-Itzhak B, Levy H, Raz D. A resource allocation queueing fairness measure: properties and bounds. Queueing Syst 2007;56(2):65–71.
[8] Sandmann W. A discrimination frequency based queueing fairness measure with regard to job seniority and service requirement. Proceedings of the 1st EuroNGI Conference on Next Generation Internet. IEEE Computer Society Press; 2005. p. 106–13.
[9] Raz D, Avi-Itzhak B, Levy H. A Resource-Allocation Queueing Fairness Measure. Proceedings of Sigmetrics 2004/Performance 2004 Joint Conference on Measurement and Modelling of Computer Systems. ACM; 2004. p. 130–41.
[10] Raz D, Avi-Itzhak B, Levy H. Fairness Considerations of Scheduling in Multi-Server and Multi-Queue Systems. Proceedings of the 1st International Conference on Performance Evaluation Methodolgies and Tools. New York, NY, USA: ACM1-59593-504-5; 2006.
[11] Sandmann W. Analysis of a Queueing Fairness Measure. Proceedings of the 13th GI/ITG Conference on Measurement, Modelling and Evaluation of Computer and Communication Systems. VDE Verlag; 2006. p. 219–31.
[12] Sandmann W. Scheduling to improve queue justice. Proceedings of the 20th European conference on modelling and simulation ECMS. 2006. p. 372–7.
[13] Sandmann W. Quantitative fairness for assessing perceived service quality in queues. Oper Res 2013;13(2):153–86.
[14] Kleinrock L. Queueing Systems - Volume I: Theory. New York, Chicester, Brisbane, Toronto: John Wiley and Sons; 1975.
[15] Asmussen S. Applied Probability and Queues. 2nd Applications of Mathematics vol. 51. New York, Berlin, Heidelberg, Hong Kong, London, Milan, Paris, Tokyo: Springer; 2003.
[16] Gordon ES. New problems in queues – social injustice and server production managment. Ph.D. thesis. Massachusetts Institute of Technology, Dept. of Nuclear Engineering; 1987.