Report on *"Checkout area design"*

July 29, 2010

74th European Study Group with Industry, 26th–30th April 2010

Department of Mathematics, University of Aveiro, Portugal

Problem presented by: *SONAE*

Study group contributors: *Joana Mota, Isabel Pereira, Diogo Pinheiro, Eugénio Rocha Manuel Scotto, Nélia Silva, Jorge Sá Esteves and João Luís Soares*

Report prepared by:

Cláudia Neves (e-mail: *claudia.neves@ua.pt*)

Fátima Ferreira (e-mail: *mmferrei@utad.pt*)

Isabel Pereira (e-mail: *isabel.pereira@ua.pt*)

Eugénio Rocha (e-mail: *eugenio@ua.pt*)

Manuel Scotto (e-mail: *mscotto@ua.pt*)

Executive Summary: In order to have a successful shopping experience in a supermarket, customers need to be able to efficiently find the articles they are looking for and to quickly check-out of the store. In order to serve their customers (checkout and pay) in a cost-effective way, retailers need to configure and size their check-outs solutions by defining: the number of checkout posts; type of checkout (traditional or and self-service); how many check-outs of each type. The main goal is to analyze the ideal check-out configuration for the desired service level. Motivate by the evidence that customers have preference for certain checkout counters, the focus is on the particular case of Vasco da Gama store.

1 Introduction

It is an almost consensual fact that supermarkets are of prominent importance in everyone's day life. Supermarkets have evolved in such a way that, nowday's, they constitute the last link in the chain of events starting at the production of an item and ending with the item being acquired by its final consumer. For most of us consumers they are the preferred place to buy groceries (food, beverages) and household items, daily or weekly, whether at our hometown or elsewhere.

Competition in the retail industry has never been greater. The supermarket industry is faced with the challenge of maintaining market share and profits while attempting new concepts and store formats in an effort to differentiate themselves from other types of retailers. In today's world, supermarkets are essentially retail stores trading in a highly competitive marketplace, where nearby competitors are often located within a walking distance of one another. New concepts and store formats, either emanating from technological developments or from a better understanding of the consumer needs or trends, arise continuously in an effort to differentiate stores among competitors.

Retailers are full aware that consistent failure to provide what a particular customer wants or what a particular customer may find elsewhere will drive this customer to a different supermarket next door, eventually creating new habits that will be hard to change (see e.g. [3]). The facts consistently indicate that today's customers are "super"-demanding and the experience of many retailers is that *customers want it all and they want it now!*.

On the other hand, there is the recognition that positive customer experiences also have a positive effect upon consumers habits, motives and attitudes, which might carry over to subsequent shopping decisions and behavior (like repeated purchase behavior and positive word-of-mouth communication). Hence, supermarket retail companies thrive continuously in order to impress their customers. A trip to a supermarket is looked upon as an outing that involves much more than just buying groceries.

The perception of an efficient service checkout has always played a key role to establishing competitive edges in attracting customers. Fast check-out is supposed to add memorable and satisfactory value to a purchasing experience and to boost customer satisfaction. On the opposite end, long waiting time for checkout may cause customers discomfort and bewilderment. A consumer dissatisfied with long waiting time may not give up its purchase at that time, but eventually will not return because of the perceived negative image.

Leading retail companies, like SONAE, acknowledge that deliverance of an excellent service is a winning strategy. Excellence in service is a profit strategy since it entails more customers purchasing, new customers arriving, and a smaller number of lost customers. Even upon entering a store, customers can perceive how much time they will spend in the store front by a single glance at the checkout traffic. This perception of long queues can perhaps be mitigated by adding sets of self-checkout counters with one communal queue. But there is a trade-off here. On the other hand, there are customers that dislike new technologies, the do-it-yourself discipline, and identify better with the traditional personal contact at checkout service. Therefore, the proper ratio of conventional parallel checkout servers and self-checkout counters concours to the key formula for maintaining the customer's perception at an excellent service level.

The quest to improve service in supermarkets, in terms of reducing the average waiting time in queue or the average number of customers in queue, is far from being a recent problem.

Since the seminal work of Erlang in 1909, it has been recognized that queueing models can be useful in making design, and operating, decisions in service systems. The enormous body of queueing research produced since Erlang has been employed with some success by professional systems analysts. Applications of queueing models of service systems related to the problems described here include: designing toll booths in bridges [4]; designing trunk lines for tele-marketing [1]; designing lanes for aircraft landing [9]; setting the number of doctors needed at an hospital emergency department [5], setting the number of representatives of an inbound call center [6].

However, to the best of our knowledge, there is not much literature on the application of the theory of queueing systems to the checkout area design of supermarkets. The reason would be that the problem had been uninteresting so far and it has gained a renovated scientific interest due to the increased complexity on the different types of checkout counters.

Planning a checkout area poses indeed a difficult problem, mainly because we cannot reduce queues by increasing the number of checkout positions in a straightforward manner. If by the customers viewpoint, a long waiting time to checkout connects with degrade of the quality of service, from the standpoint of retail management it is expensive to keep an underutilized server because every portion of space in the store front end is a real asset.

The main goal of this work is to assist SONAE in discerning a course of action that may actually lead to improvement in their customers' experiences at the checkout area. This is essentially an exploratory study which comprises an attempt to identify and to measure stylized features of Vasco da Gama customers when checking out.

When the company executives presented the case to us they made the following claims based on empirical evidence:

- *A considerable number of customers checkout from the store with few items.* On average 50% of clients check out from the store with 5 items or less.

- *Customers carrying few items tend to prefer self-service counters.* Vasco da Gama's group of four self-service counters, which represent 10% of whole set of counters, serve 19% of the customers carrying 5 items or less. Colombo's two-group of four self-service counters, which represent 13% of whole set of counters, serve 32% of the customers carrying 5 items or less.

- *Customers tend to spend more time checking out in self-service stations than in traditional* *checkout counters*. More concretely, a customer carrying 5 items takes 50% more time in a self-service counter; a customer carrying 10 items takes 300% more time in a self-service counter; a customer carrying 13 items takes 400% more time in a self-service counter.

SONAE keeps data records of all scanning times for each individual item, at every checkout counter. The items (or articles) belong to certain categories of interest which are displayed in connection with the designated article, both in traditional and self-service checkouts. Comprehensive data-sets regarding scanning-service times were disclosed by SONAE for the purpose of this study. But these do not include records about payment-service time nor the payment method used by the customer (for instance, cash or card). Moreover, in the absence of information about customers behavior when in queue, we find this huge batch of data rather sparse eventually, at the enrollment of modeling Vasco da Gama store front which is typically a queueing system in the fair probabilistic sense.

We shall closely follow the recent work by Horst (2009) [8] which aim at discussing both theoretical and practical consequences of the introduction of self-checkout counters. Therein, comparative performance of self-checkout supermarkets against traditional supermarkets has been undertaken in terms of the expected sojourn time in the system and also by estimating the probability that more than n customers are in the queue. In Section 2 we introduce the mathematical contours of this study. We also present a slight modification to the work by Horst (2009) [8], that allows unequal preferences by the customers, tailored for SONAE's problem. For a comprehensive exposition on the probabilistic theory underpinning our work, the queueing theory, we refer the reader to [7]. Section 3 encloses analysis under SONAE's framework and results in terms of the most widely-used performance measures. Finally, in Section 4 we draw some concluding remarks and recommendations for subsequent work.

2 Model description

Queuing systems are easily recognized in any organized structure with customers arriving, customers waiting their turn for service, customers being served, and customers departing. Queuing systems are thus the natural probabilistic tool to model the customer flow behavior at supermarket checkout areas. Six basic components suffice to characterize completely a queueing system:

1. arrival pattern (or input pattern) of customers, characterized in view of the customer interarrival-time distribution, i.e., the duration of time intervals times between consecutive arrivals. These are usually assumed independent and identically distributed (positive) random variables;

2. service pattern, usually referred by the distribution of the service time taken by a customer, i.e., the distribution underlying the amount of time each costumer requires to be served. Likewise, service times are assumed independent and identically distributed (positive) random variables. Moreover, any random sequence of inter-arrival times is supposedly independent of the associated sequence of service times. Any interaction between customer and server only takes place during the service itself;

3. number of servers (or service stages), assumed to be identical and being able to serve only one customer at a time and being idle if and only if there are no customers waiting for service;

4. system capacity or the amount of buffer space in the queue, measured in terms of the number of customers allowed to enter the system;

5. queue discipline, referring to the manner in which the customers are taken from the queue to be served when a queue has formed. Departures from the waiting line to the server are usually assumed to be First In, First Out (FIFO) or First Come, First Served (FCFS), i.e., the service of a customer is initiated by order of arrival to the queue, rather than last in first out, or in random order;

6. number of service channels. Single-station queueing systems have only one queue line for all customers and one or several servers, so that, whenever a server is free the customer in the front of the queue goes directly to that server and start being served. Such systems are called (single channel or single line) multi-server queuing systems. Nevertheless, in traditional supermarket stores, generally each server has his (or her) own private queue in

front of the server. In this case we say that we have a multi-queue system, i.e., a collection of independent parallel single channel queues, either of single or multi-server type.

Each one of these basic characteristics are specifically addressed below, together with simplifying assumptions pertaining to SONAE's specific problem.

Another important characteristic to describe a queuing system is the customer acceptance. Customers acceptance discipline in the queue waiting line configuration may vary from patient, to balk (view the line, then leave), renege (join the line, then leave), jockey (join the line, then move to another line when you think it is moving faster), or collude (give your groceries to another customer). For computational reasons, we follow the usual assumption and further consider that SONAE's customers are patient, i.e., on arriving at a queueing system a customer stays in the system until being served, no matter how much longer the customer has to wait for service.

In the sequel, we shall use Kendall's notation $(A/B/c/d)$ to denote a queueing system characterized by quantities 1. up to 4. Hence, the letters A, B, c and d encapsulate the distribution of interarrival-times, distribution of the service time, the number of servers and system capacity, respectively. Since we assuming there are no constraints on the number of customers in the system, i.e., $d = \infty$, on the matter of simplicity we shall drop the d-component from the notation. Henceforth, we use the notation $(A/B/c)$ for a SONAE queueing system. For the service discipline, stated in point 5. above, we follow the usual rule at supermarket checkouts counters and consider departures from the queue are governed by the First Come First Served (FCFS) service discipline. We shall call on these assumptions, which we regard as mild yet reasonable restrictions, in order to develop our results. For a comprehensive exposition on the queuing theory underpinning our work we refer the reader to [7].

In the current framework, the arrival process, i.e., the distribution function underlying the physical process that is generating the input pattern is unknown. Moreover, usual measures of central tendency such as the average amount of arrivals to the queue per unit of time (average arrival rate) or the average time between successive arrivals (average inter-arrival time) cannot be adequately estimated, since SONAE does not keep record of time instants at which a customers enters either the store or a (self-)checkout queues. To circumvent this problem, we shall assume the arrivals originate from a large population of independent potential customers and that customers arrive individually, meaning that the arrival of one customer is independent of, or does not impact, the arrival of another costumer.

Now, let T be the random time between successive arrivals of customers. The probability that two or more customers arrive at the same instant is assumed negligible. Completely random arrivals are often addressed by means of a Poisson-exponential arrival process. Our framework does not escape the Poisson process' grasp and the random variable T thus follows an exponential distribution with mean $1/\lambda > 0$. Probability theory ascertains that the average time between arrivals is $1/\lambda$, if the average rate of customers entering the queueing system is λ (cf. [7], p.16). Intuitively one would expect a similar claim.

A good advantage in considering a Poisson process for the arrival process stems from the fact that the random split of a Poisson originates independent and yet again Poisson processes. In fact, what really matters is not the process of customer arrivals to the supermarket but the actual customer arrival process at each individual checkout counter. The latter justifies the focus on what happens at the store front, namely at the checkout counters, instead of looking back at the arrival process to the supermarket. Accordingly, we let k denote the number of checkout counters in the checkout area and consider that each arbitrary arrival customer chooses counter i with probability p_i^* defined as

$$p_i^\star := P\{\text{costumer chooses counter } i\}, \tag{1}$$

with $p_i^\star \leq 0$ for all $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} p_i^\star = 1$. A positive value $p_i^\star$ is assigned to the checkout position i if this system is operative, otherwise it is set at $p_i^\star = 0$. Therefore, the original Poisson arrival stream, with arrival rate $\lambda > 0$, is randomly split between k independent Poisson processes with intensities $\lambda_i := p_i^\star \lambda$, corresponding to the arrival streams to the queue in the k-th parallel checkout systems. This results in k single-channel Poisson arrival queues with arrival rates $\lambda_1, \lambda_2, \ldots, \lambda_k$, each one of those being formed in front of each checkout counter.

In the particular case of Vasco da Gama store, k is equal to 37 because there are 36 traditional checkouts (i.e., 36 single servers) in this supermarket and one self-checkout service that encloses a multi-server system with 4 servers. Figure 1 summarizes a decomposition of the arrival process in Vasco da Gama store front.

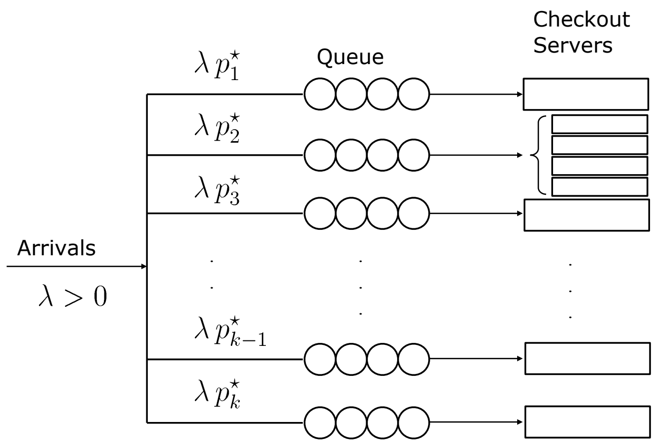


Figure 1: Scheme of the store front: decomposition of the arrival process.

Denoting by μ the average rate of serving customers, a measure of traffic congestion for a c-server system is the so-called traffic intensity given by $\rho = \lambda/(c\,\mu)$. When $\rho > 1$ the average number of arrivals into the system exceeds the maximum average service rate of the system, and we would expect that after some time the queue builds in, unless, at any particular instant, customers are not allowed to join. In case $\rho < 0.8$, we expect that the system does not present congestion. In such cases the system has a very small number of customers in queue and small number of customer waiting time in queue. These random quantities start to increase quite rapidly when the traffic intensity ρ exceeds 0.9, approaching an undesirable magnitude when the queue gets critically loaded (with ρ near 1) and bursting for $\rho > 1$, due to the fact that, on average, the number of arrivals into the system exceeds the aggregate service potential of the system. It is also known that, for $\rho < 1$, the queue length is typically larger when there is more variability in the interarrivals and/or in the service time distributions.

A value of ρ greater than one would prevent the system from settling down, hence a steady state regime is never attainable in this case. In order to attain steady-state results, ρ must be less than one. Steady-state results are important because they account for probabilistic statements or characteristics such as mean and variance that are valid at any particular instant we decide to look at the system. Figure 2 serves a mere illustrative purpose in this respect. It concerns a preliminary analysis of how the traffic intensity ρ changes, when only traditional checkouts are operative ($c = 1$). In order to have a rough idea about the total number of traditional checkouts (parallel servers) required to guarantee a steady-state solution, Figure 2 displays how the traffic intensity ρ varies as a function of the number k of traditional operative checkouts, given the arrival rates $\lambda = 100, 200, 250, 500, 750$ customers per hour. The mean service time per customer considered here is $1/\mu = 110.6$ seconds. This value was estimated from pooling all average service times from the traditional checkout counters operative between 10:00 a.m. and 12:00 a.m. on a busy Sunday at Vasco da Gama store. Following ideas from Horst (2009) [8], in this first approach in which we have consider that no checkout counter is preferred over another (i.e., $p_i^\star = 1/k$, all $i = 1, 2, \ldots, k$), we can observe that, with $k = 20$ traditional checkouts at disposal, the traffic intensity is less than 1 even if customers keep pouring in at the rate of 500 per hour. Although the 20 operative counters keep the system away from saturation, a value of ρ around 0.75 most likely entails long queues which, of course, is not at all desirable.

We should stress that, although from SONAE's records we could not estimate how the customer arrival process split into the several checkout queues, we conclude that customers clearly have preferences. This can be inferred from Figure 3 which displays the percentage of transactions at each traditional checkout counter. In fact, although the depicted percentages of transactions cannot be directly translated as the percentage of customers going to a certain checkout, the most striking feature is that the highest percentages rise on those checkout positions that are closer to the Oriente railway station, which in turn, marks the direction of the nearest car parking lots. Checkout positions 7 up to 10 refer to self-service counters and therefore were not represented here.

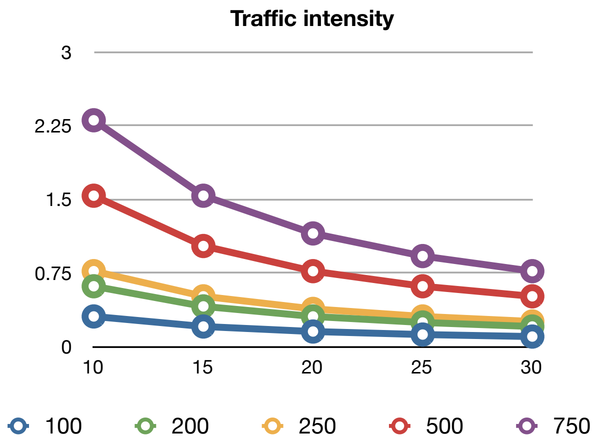


Figure 2: Traffic intensity as a function of the number of traditional checkout counters. Each line series corresponds to the designated average number of clients entering in the system per hour.

As such, we need to assign different probabilities to split costumers between the different checkout counters, not making much sense to consider the uniform split $p_i^\star = 1/k$, all $i = 1, 2, \ldots, k$ as above. Since we are dealing with a mixed scenario (the store front encompasses both traditional and self-checkout servers), it would be rather unrealistic to say that every counter has the same probability of being chosen by an arriving customer. Hence the novelty in this work: the probability p_i of a customer approaching counter i is not the same for every counter, i.e., $p_i^\star \neq 1/k$ for at least one $i = 1, 2, \ldots, k$. Now, how to find these distinct probabilities p_i^* is another matter, surely of practical importance. The probabilities p_i^* could not be estimated on the basis of the available but that we shall defer this matter to the next section.

With respect to generalizing the service pattern, we note that the amount of time a customer spends in the server or checkout counter strongly depends on the method of payment, namely cash or card (or other that we shall no take into consideration). Only by itself, the payment factor can have huge impact upon the probabilistic statements underlying our analysis, the

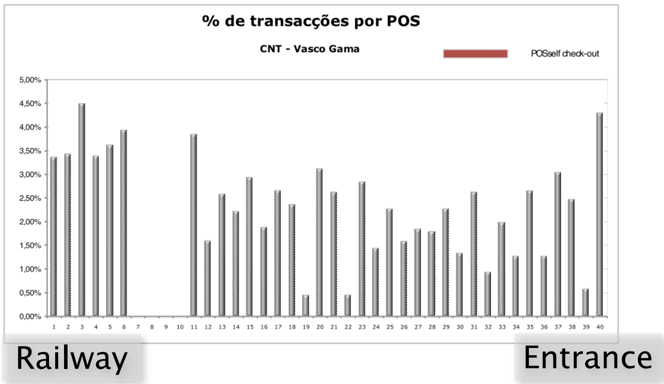


Figure 3: Percentage of transactions at each checkout counter of Vasco da Gama store.

obvious one being that we cannot assume service times to be exponentially distributed any more. Nevertheless, on account of simplicity we shall assume the time a customer spends paying cash and the time taken in a card payment are both exponentially distributed with expected value β_{cs} and β_{cr}, respectively. Whence, if the probability of a customer using cash payment is q, then the service time S is a random variable with hyper-exponential distribution (H_2). Corresponding power-moments of order $r = 1, 2$ are given by

$$E[S^r] \quad = \quad q\, r\beta_{cs}^r + (1-q)\, r\beta_{cr}^r = q\, \frac{r}{\mu_{cs}^r} + (1-q)\, \frac{r}{\mu_{cr}^r}. \tag{2}$$

This simple fact carries mathematical problems of computational nature which we recover in the next Section.

Due to their configuration, traditional checkout systems and self-service checkout counter systems are modeled by appropriate $(M/G/1)$ and $(M/G/4)$ queues, respectively. The notation M stands for the Poisson arrival process whereas the letter G stands for General (independent) service times, following an hyper-exponential distribution in this particular case.

3 Performance measures

We begin by recalling the assumption that the queuing system arises a steady-state scenario, i.e., that the traffic intensity ρ is less than one for every checkout system. This way we can proceed with the use of general queuing theory in order to study more thoroughly the behavior of the two types of checkout systems: traditional and self-service checkouts. Just by looking at the appropriate models, it should be possible to mimic the real behavior of the system and ascertain a certain desirable quality of service for the system. With the proper models, finding an appropriate balance between the cost of service and the amount of waiting is also tangible, under the umbrella of queueing theory.

In the previous section we have showed that each checkout counter is adequately described by a $(M/G/c)$ queueing system. In this section, we shall concentrate on the assessment of performance measures pertaining to such queueing systems. Performance measures are, in general, evaluated trough the computation of important system characteristics such as the number of customers in the system at an arbitrary time, L; the number of customers waiting for service at an arbitrary time (i.e., the queue length), L_q; the customer waiting time in queue (time elapsed since the arrival of a customer until it enters service), W_q; the customer waiting time in system (time elapsed since the arrival of a customer until it lives the system), W.

Performance measure involving waiting time distributions are particularly related with the view point of customers (in assessing customer satisfaction), whereas those regarding distributions of occupancy are more related with the standpoint of resource management.

Further on we shall take into account that, by virtue of the Law of Large Numbers (cf. [2]), performance measures are well approximated by sample means or averages taken upon the data available. Hence, even if the underlying theoretical distributions are unknown or do not possess closed-functional forms, expected-value measures offer a valid and fruitful way of assessing performance of the queueing system. Therefore, it is often satisfactory to provide a mean value performance analysis, in which case we are usually interested in obtaining:

- the expected number of customers in the system, $E[L]$;
- the expected queue size, $E[L_q]$;

- the expected waiting time in queue, $E[W_q]$;
- the expected waiting time system, $E[W]$.

These quantities relate each other in such a way that we know them all once we compute the expected waiting time in queue, $E[W_q]$. More concretely, the following equality holds:

$$E[L] = E[L_q] + E[\text{number of busy servers}] = E[L_q] + c\rho.$$

In addition, the celebrated Little's law ensures that

$$E[L_q] = \lambda E[W_q] \quad \text{and} \quad E[L] = \lambda E[W].$$

Before moving on to mean value issues, we note that, if we let $L(t)$ denote the total number of customers in the system at time $t \geq 0$, then the probability

$$p_n(t) := P\{L(t) = n\}, \quad n = 0, 1, 2, \ldots.$$

remains constant for any arbitrary time point t after a steady-state is reached. Whence, after a steady-state is reached, the latter rephrases as

$$p_n := P(L = n), \quad n = 0, 1, 2, \ldots.$$

For the particular case of a $(M/M/c)$ system, i.e., a queueing systems with exponential service times, p_n possesses a well-known closed-form expressions (see, e.g. [7], Chapter 2). However, we are now dealing with a general service time G which renders more complex calculations than the exponential service times and no close-form solutions are actually known. The main problem in working with $(M/G/c)$ systems arises from the fact that the number of customers in the system $L(t)$ does not constitute a Markov process (usually inherited by the lack of memory of the exponential distribution; see [7], Section 1.8 for details). The probability per time unit for a departure of a customer now depends also on the time the customer in service has already spent in the service and this information is not contained in the random variable $L(t)$. Again, assuming the k queueing systems, the two major (steady-state) expected-value measures,

$$E[L^{(i)}] = \sum_{n=0}^{\infty} n\, p_n,$$
$$E[L_q^{(i)}] = \sum_{n=c^{(i)}}^{\infty} (n - c^{(i)})\, p_n,$$

with $c^{(i)} = 1$ if $i \leq k_1$ and $c^{(i)} = 4$ if $k_1 < i \leq k_1 + k_2 =: k$, require a different strategy of calculation (since they depend on p_n). Section 5.1.1 of [7] expounds two possible derivations of these measures pertaining to $(M/G/1)$ systems.

The first derivation hinges on the arrival times and the PASTA property (Poisson Arrivals See Time Averages). This property essentially states that *the steady-state distribution of the number of customers in the system seen by arriving customers, say* $(\pi_n)_{n\in\mathrm{N}}$*, is just the steady-state distribution of the number of customers in the system,here denoted by* $(p_n)_{n\in\mathrm{N}}$. As a consequence, *the average number of customers in queue as seen by an arriving customer is the same as the time-average number of customers in queue.*

Moreover, for an arriving customer it only matters the remaining service time of the customer being served and not his/her total service time. Whence,

$$E[W_q] = E[L_q]\, E[S] + P\{\text{server busy}\} E[\text{residual service time} \,|\, \text{server busy}].$$

Due their to good properties, there exist exact results to compute the steady-state average waiting time in queue of $(M/M/c^{(i)})$ and $(M/G/1)$ queueing systems with traffic intensity $\rho_i = \lambda_i/\mu$, where $\lambda_i = \lambda\, p_i^\star$ and $p_i^\star$ as defined in (1) for $i = 1, 2, \ldots, k$. In fact,

$$E[W_q^{(M/M/c^{(i)})}] = \frac{\left(c^{(i)}\rho_i\right)^{c^{(i)}}}{c^{(i)}!\, c^{(i)}\mu\,(1-\rho_i)^2} \left[\sum_{j=0}^{c^{(i)}-1} \frac{\left(c^{(i)}\rho_i\right)^j}{j!} + \frac{\left(c^{(i)}\rho_i\right)^{c^{(i)}}}{c^{(i)}!(1-\rho_i)}\right]^{-1}, \quad i = 1, 2, \ldots, k,$$

There are no closed form expressions to compute $E(W_q^{(M/G/c^{(i)})})$ like the above. But we may resort to approximations proposed in the literature such as, e.g., Allen-Cunneen approximation (cf. [10], p. 123):

$$E[W_q^{(M/G/c^{(i)})}] = \left(\frac{1+\tau_s^2}{2}\right) E[W_q^{(M/M/c^{(i)})}],$$

with $\tau_s = \sigma_S/E[S]$ denoting the coefficient of variation the service time distribution. In the particular case of $c^{(i)} = 1$, all i, the above gives rise to the equality:

$$E[W_q^{(M/G/1)}] = \frac{\lambda_i\, E[S^2]}{2(1-\rho_i)}, \qquad i = 1, 2, \ldots, k,$$

hence the exact result for the steady-state average waiting time in queue of $(M/G/1)$.

As a consequence, for the $(M/G/1)$ system with arrival rate $\lambda\, p_i^\star$ and general hyper-exponential distribution considered, for which $E[S^r]$ is as defined in (2), for $r = 1, 2$, it then follows that the expected queue size and the expected sojourn time becomes

$$\begin{aligned}
E[W_q^{(i)}] &= \frac{\lambda\, p_i^\star\, E[S^2]}{2\big(1-\lambda\, p_i^\star E[S]\big)},\\
E[L_q^{(i)}] &= \frac{(\lambda\, p_i^\star)^2 E[S^2]}{2\big(1-\lambda\, p_i^\star E[S]\big)}\\
E[W^{(i)}] &= \frac{\lambda\, p_i^\star E[S^2]}{2\big(1-\lambda\, p_i^\star E[S]\big)} + E[S],\\
E[L^{(i)}] &= \frac{(\lambda\, p_i^\star)^2\, E[S^2]}{2\big(1-\lambda\, p_i^\star E[S]\big)} + \lambda\, p_i^\star E[S],
\end{aligned}$$

for $i = 1, 2, \ldots, k_1$, known as the "Pollaczeck-Khintchine" formulae.

The second derivation we have announced stems from analyzing the queue at departure points, which gives rise to a discrete-time Markov chain. The latter enables not only to derive identical formula for the expected system size $E[L]$ to the one presented before, but also steady-state system-size probabilities $(\pi_n)_{n\in N}$ which, due to PASTA property, turn out to be analogues to the steady-state probabilities $(p_n)_{n\in N}$ above.

This steady-state probabilities, along with the steady-state distribution of the waiting time in queue, are relevant to compute the the probability of encountering the system in certain states, such as empty, $p_0 = P(L = 0)$, having all server busy, $P(L \geq c)$, having more than a certain number of customers, say n, $P(L > n) = 1 - \sum_{i=0}^{n-1} \pi_i$, or of a customer having to wait more than a certain time x in queue, $P(W_q > x)$.

Tail probabilities such as the probability of encountering more than n customers in queue,

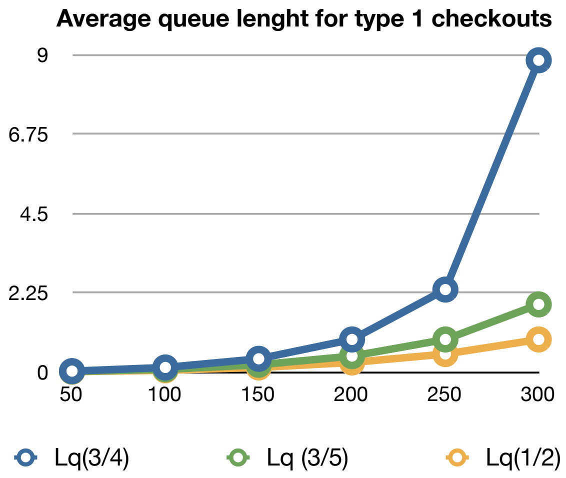
$$P(L > n) = 1 - \sum_{i=0}^{n-1} \pi_i, \tag{3}$$

can translate the probability of a supermarket being too crowded, provided we consider that the system is crowded with more than n customers waiting in queue at a certain checkout counter. The π_n's can be obtained by solving recursion equations with initial value π_0 corresponding to the fraction of time the system (checkout) is empty.

Due to the lack of useful information for estimating the correct checkout queues parameters estimated, to provide an possible/effective performance analysis of traditional checkout queues we shall make use of prior information in [8] about expected time of payment, given payment is either by card or cash. In his work, Horst (2009)[8], claims that paying cash takes on average 19 seconds while card payments linger to 26 seconds on average. This claim is supported on a report by the De Nederlandesche Bank. We shall borough strength from these estimates to obtain an estimate for the proportion of transactions using cash. Since SONAE claims that a payment takes on average 20 seconds, regardless of the payment method, we then set $q = 6/7$, i.e., approximately 86% of the transactions in Vasco da Gama store are cash transactions. Adopting the estimated value of 110.6 seconds for the average time needed to scan the articles in a traditional checkout, we obtain that $\mu_{cs} = 1/(110.6 + 19)$ and $\mu_{cr} = 1/(110.6 + 26)$, whence $E[S] = 130.6$ and $E[S^2] = 34124.72$. Note that these are all estimates despite we are using the same notation for their theoretical counterparts.

At this point, it is worthy to note that, with respect to traditional checkout, customers seem to prefer the checkout positions nearby the railway station plus the one next to the entrance of Vasco da Gama store (cf. Figure 3). Due to their high preference, these 9 traditional checkout counters are called *type 1 systems*. Traditional checkout systems belonging to the complementary checkout region are designated by *type 2 systems*. We also note that we do not know the percentage of time these positions are operative during a day's work but we are assuming the system on a steady-state regime anyway. Figure 4 displays the average number of customers in the queue, $E[L_q]$, and the expected sojourn time, $E[W]$, for several values of probability of preference assigned to type 1 (traditional) checkouts. In mean terms, if there are at most 250 customers arriving the checkout area per hour, we should expect to find less than three customers queueing in, even if 75% of customers choose counters of type 1. Given the same conditions, i.e., $\lambda = 250$ and $p^\star = 3/4$, the expected sojourn time is at most 6.75 minutes. With more than 300 customers per hour, with a demand of 75% for type 1 checkouts, both these quantities tend to disrupt, an indicator that the system is near saturation.

We now turn to the analysis of self-checkout systems, characterized by $(M/G/c)$, with $c = 4$. The fact that we have now multiple service channels is an impending problem of practical importance since it entails we are constrained to the departure time instants. The ulterior mathematical reason why it is difficult to derive the necessary results is connected to the impossibility of using imbedded Markov chains as in the previous $(M/G/1)$ queueing systems. The study of a $(M/G/c)$ system reduces to the study of the queue behavior. Hence, calculations are confined to the average queue length at departure points. Although Little's formula still applies, the average number of customers in the queue, at the departure of a predecessor customer from the service, entails knowledge about the probability of n customers in queue just after departure. Knowing such queueing probabilities is virtually impossible from SONAE's data records. As a matter of fact, we encounter a similar difficulty with a $(M/G/1)$ system (for traditional checkouts) if we want to carry on with the calculation of probability (3) for specific values of n because it depends on π_n's, which in turn, depend on the number of customers who arrive during a service time interval. To circumvent this difficulty, we shall reduce our service time to a two-point discrete random variable S such that $P\{S = 129.6\} = 6/7$ and $P\{S = 136.6\} = 1/7$, thus yielding overall average service time $E[S] = 130.6$ and associated second moment $E[S^2] = 17062.37$. The underlying assumption is that the servers are so sharp now that the service rate μ_{cs} (*resp.* μ_{cr}) is not an average rate but an exact (deterministic) number $\mu_{cs} = 1/129.6$ (*resp.* $\mu_{cr} = 1/136.6$). Such an assumption upon the service times (with only two possible times of 129.6 seconds for cash payments and 136.6 for card payments) has severe implications because it clearly imposes constraints upon the number of items customers wish to buy. Nevertheless, the dichotomy of service times fixes the problem of not knowing the number of customers arriving at departure



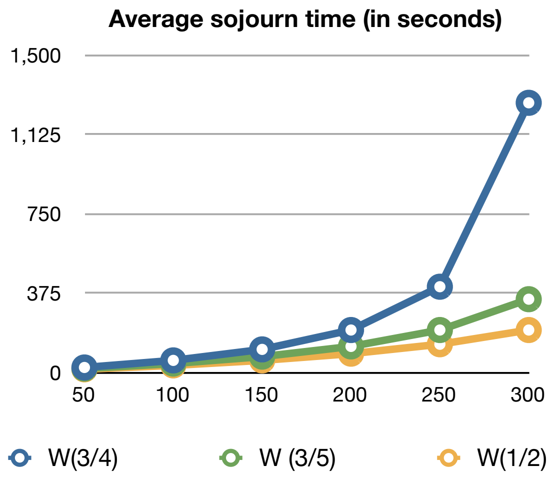


Figure 4: Performance measures for a counter of type 1, plotted against the arrival rate $\lambda = 50(50)300$ customers by the hour.

times and we find (cf. [7], Example 5.3):

$$\begin{aligned} P\{L > 3\} &= P\{\text{at least 3 customers in queue}\} \\ &= 1 - \pi_0 - \pi_1 - \pi_2 - \pi_3 \\ &= 1 - p_0 - p_1 - p_2 - p_3 \\ &= 1 - (1-\rho)\Big\{\frac{1}{c_0} - 1 + \Big(\frac{1-c_1}{c_0} - 1\Big)\frac{1}{c_0} + \frac{1}{c_0}\Big[\frac{1-c_1}{c_0}\Big(\frac{1-c_1}{c_0} - 1\Big) - \frac{c_2}{2c_0}\Big]\Big\}, \end{aligned}$$

where

$$c_j = \frac{6}{7}\exp\Big\{-\frac{\lambda_i}{\mu_{cs}}\Big\}\Big(\frac{\lambda_i}{\mu_{cs}}\Big)^j + \frac{1}{7}\exp\Big\{-\frac{\lambda_i}{\mu_{cr}}\Big\}\Big(\frac{\lambda_i}{\mu_{cr}}\Big)^j, \quad i = 1, 2, \ldots, k_1,\ j = 0, 1, 2.$$

In Figure 5 we depict the probability of a queue with more than two customers in a traditional checkout system of type 1. Two different values of preference probability are chosen, 3/4 and 1/2, the latter corresponding to the situation where a customer virtually throws a coin in order to decide which checkout region to use (between type 1 and type 2 regions). In this case, even if there is arriving customers at the rate of 300 per hour, 8 traditional checkout servers would suffice to keep below 10% the probability of encountering more than 2 customers in queue.

Let us now consider the case in which a customer decides for a type 1 region with probability 3/4. An arrival rate of 200 customers per hour requires at least 8 servers to ensure the probability of finding more than 2 customers queueing in a type 1 system stays below 10%. On the other hand, we should impose an upper limit for the number of type 1 serves. There are 36 traditional checkout positions in the whole Vasco da Gama store. If the type 1 checkout region encompasses more than 12, say, traditional servers, those servers not so close to the railway hand-side should become less appealing. Hence, other subtypes of checkouts would be necessary.

Altogether, with the 3/4 probability assigned to the preference of customers for a type 1 region enclosing 10 traditional checkouts, the probability of finding more than 2 customers in queue is below 0.3, even with an input traffic of 300 customers per hour.

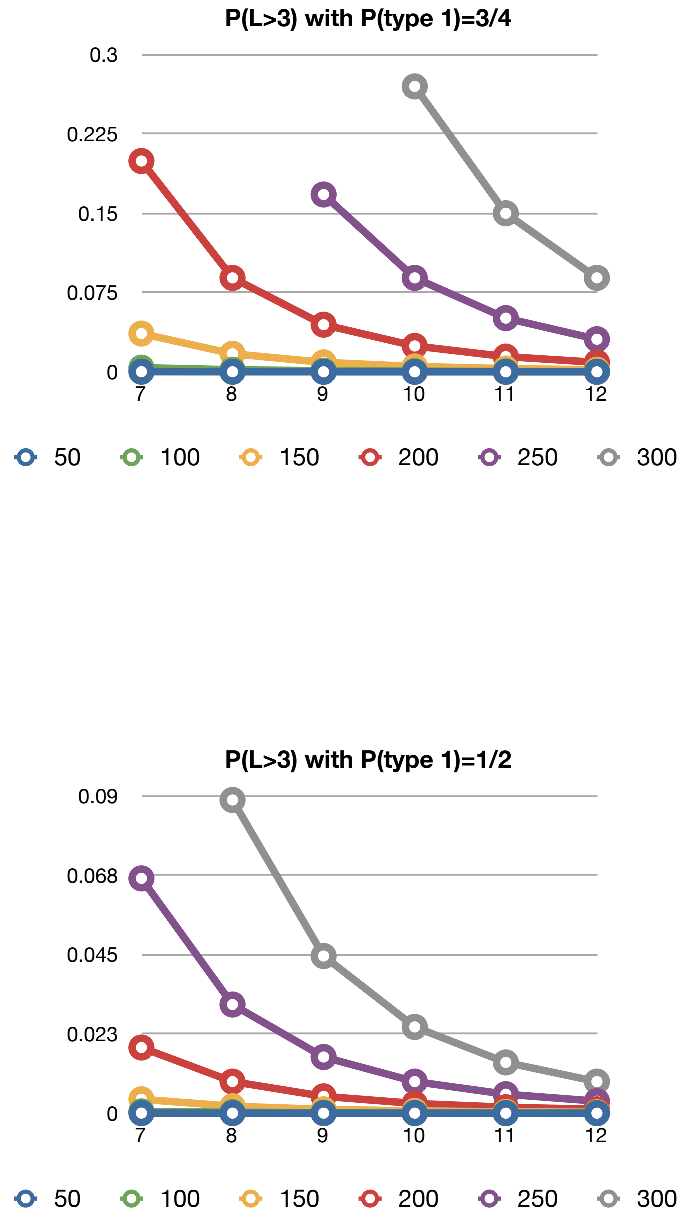


Figure 5: Probability of finding at least three customers queueing in a type I checkout counter for different probability of choosing checkout region of type 1, all plotted against the number of traditional checkout checkout servers, given the arrival rate $\lambda = 50(50)300$ customers by the hour.

4 Conclusions and recommendations

Long waiting time for checkout may cause customers discomfort and bewilderment. Customers really want fast checkout at supermarkets (see [8], p. 6), while retailers face the need to provide successful shopping experiences. One way to fulfill this purpose consists on placing the so-called self-checkout systems, thus creating a sensible alternative for those customers carrying baskets and shopping only a few items, usually less than 5 items (as mentioned above). Self-checkout counters should in fact provide a convenient and appealing way out of the supermarket for light shopping. Therefore a trade-off that has to be made. If by the one hand, communal queues give every customer their fair turn at the self-checkouts so they do not wait a moment longer than is fair every time they shop (and with this, the stress of making the wrong choice and paying for it is eliminated), by the other hand, when cost of waiting and do-it-yourself implies the cost of losing a customer (because long waiting time drives the customer away), then measurement becomes more difficult, rendering store management a more complicated task.

An alternative way to manage the queueing system by measuring, analyzing and minimizing its combined total costs, the supermarkets may try to manage the service system by setting threshold parameters for system operating characteristics, and then use faster servers, more servers, automation of the service activity or some combination of strategies to achieve those parameters for store *crowdiness.*

For example, a consumer products store may decide to open another cash register checkout station when the mean number of customers in line at the first five registers goes, say, over three; or consider the supermarket as too crowded if the probability of more than 3 customers in a queue is greater than 0.05.

This work is an attempt to modeling the store front configuration for one particular store of SONAE's group: the Vasco da Gama store. There is a total of 40 checkout servers available in this store, 4 of which are self-checkout counters attached to a single queueing line. In principle, all self-checkout servers remain operative for the entire period of the day the store is open.

Our starting point is the modeling of a checkout area by Horst (2009) [8] regarding the random split of the arrival process to the checkout area into k Poisson subprocesses underlying the arrivals at each checkout system (or counter). Motivated by the most striking feature in

Figure 3 that customers are more inclined to proceed to those checkout systems closely located to the railway station, also near to the car parking sites, we have considered two specific regions, assigned with specific probability of preference. This is our main contribution to the application of queueing theory for analyzing the problem of congestion in supermarket checkouts. We are particularly concerned with modeling the average waiting time of a customer in queue, viewed as an important component of the user service. Due to sparsity of information, in the sense that we do not have any data available for the (single) queue attached to each server (traditional or self-checkout), we were forced to reduce our approach to the case of determinist service rates. Bearing our primary assumption in mind, that the number of customers in queue depends on the time a precedent customer has already spent in the server, the latter was found the most fruitful approach.

The amount of time a customer spends in the server or checkout counter strongly depends on the method of payment, namely cash or card. The latest balance of the De Nederlandesche Bank gives 19 seconds for the mean duration of cash payments, whereas card payments can rise up to 26 seconds on average. On account of simplicity, other methods of payment were disregarded in the current analysis. Furthermore, we have considered traditional checkouts are mutually independent and are allocated within two regions of preference, according to their spatial positions in the Vasco da Gama store front. The so-called type 1 region encloses traditional checkout systems on the nearest end to the railway station and the one adjacent to the entrance in the supermarket (see Figure 3). If there are 75% customers going for the type 1 region then sojourn time in the checkout area is approximately 6.75 minutes with an input stream of 250 customers per hour, and rising up to 21.24 minutes, approximately, when instilling 300 customers per hour (on average), given all $k_1 = 9$ type 1 traditional checkouts are operative. This is the sort of information displayed in Figure 4. Furthermore, Figure 5 shows that, at a rate of 250 customers arriving the checkout area per hour, the probability of finding at least 3 customers in queue already surpasses 15% with 9 type 1 checkouts being operative for some time.

It is worthwhile to mention at this point that these results are somewhat uninformative because they follow from a deterministic setup, carried by the replacement of averages by exact equalities. What we are actually saying is that a randomly chosen customer takes exactly 19 seconds paying cash, instead of saying a randomly chosen customer is expected to take 19 seconds if paying in cash. The use of such a straightforward approach is the main reason why we have not studied more thoroughly the somewhat similar case of self-service checkouts.

Taking all into consideration, subsequent enhancement of the present results depend on the monitoring of queue behavior. Moreover, it is crucial to discern and identify stable periods of the whole system (store front or checkout area), providing a more accurate estimation of service rates. This involves record of time values useful to pinpoint time intervals during which the checkout counters are actually operative. Having a grasp at the percentage of cash payments at Vasco da Gama, as well as the percentage of customers moving towards type 1 traditional checkouts and percentage of customers going for self-checkout servers, would improve the current results. Valuable information in the present framework is the number of arrivals during any inter-departure period. Having this sort of information available would avoid the ultimate reduction to $(M/D/c)$ systems that encapsulate deterministic service times.

References

[1] Andrews, B. H. and H. L. Parsons (1989). L. L. Bean chooses a telephone agent scheduling system, *Interfaces*, 19(6), 1–9.

[2] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

[3] Codrington, G. (2002). Welcome to the Future World, *Future Business: Marketing, Media, and Business in S.A., April*, 38–40.

[4] Edie, L. C. (1954). Traffic delays at toll booths, *Operations Research*, 107–138.

[5] Green, L., J. Giulio, R. Green and J. Soares (2006). Using queueing theory to increase the effectiveness of physician staffing in the emergency department, *Academic Emergency Medicine*, 13(1), 61-68.

[6] Green, L., P. Kolesar and J. Soares (2001). Improving the SIPP approach for staffing service systems that have cyclic demands, *Operations Research*, 49(4), 549-564.

[7] Gross, D., Shortle, J. F., Thompson, J. M. and Harris, C. (2008). *Fundamentals of Queueing Theory*, 4th ed., Wiley Series in Probability and Statistics. Wiley, New Jersey.

[8] Horst, N. (2009). *The Revolution of the Checkout Area*, Mater thesis in Business Mathematics and Informatics, Vrije Universiteit, Amsterdam.

[9] Koopman, B. O. (1972). Air-terminal queues under time-dependent conditions, *Operations Research*, 20(6), 1089–1114.

[10] Whitt, W. (1993). Approximations for the GI/G/m Queue. *Production and Operations Managment* , **2**, 114–161.