Queueing Systems

Polling systems with two alternating weary servers

--Manuscript Draft--

Manuscript Number:	
Full Title:	Polling systems with two alternating weary servers
Article Type:	Manuscript
Keywords:	polling systems; cyclic; alternating weary servers; exhaustive; gated; globally-gated; state-dependent arrival rates; stability; mean value analysis.
Corresponding Author:	Uri Yechiali, PhD Tel-Aviv University Tel-Aviv, ISRAEL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Tel-Aviv University
Corresponding Author's Secondary Institution:	
First Author:	Omri Avissar
First Author Secondary Information:	
Order of Authors:	Omri Avissar
	Uri Yechiali, PhD
Order of Authors Secondary Information:	

Polling systems with two alternating weary servers

Omri Avissar
omriavis@post.tau.ac.il

Uri Yechiali
uriy@post.tau.ac.il

Department of Statistics and Operations Research
School of Mathematical Sciences
Tel Aviv University, Israel

Abstract We introduce and study cyclic polling systems in which service times of customers increase after the completion of each cycle due to increased tiredness of the server. To prevent the system from exploding, the server must be deactivated to regain (some or all of) its efficiency while another server takes its place. Performing a "change of guard" takes some additional random time. This requires the determination of a "swapping policy" between the two servers. We model such systems under the gated, exhaustive, and globally-gated service regimes. In the case of swapping policies which call for a swap at the end of every fixed number of cycles, we show that, contrary to classical polling systems, the stability condition for the exhaustive regime differs from its counterpart for the gated regime. A single queue case with identical servers is further studied and analyzed. Assuming stability we show that, in the latter case, the maximal number of consecutive cycles a server can serve without resting under the gated regime is approximately double than that under the exhaustive regime. In addition, we construct an algorithm to obtain an optimal swapping policy for the case where two identical servers alternate every fixed number of cycles in a system operating under the exhaustive service regime.

Keywords polling systems; cyclic; alternating weary servers; exhaustive; gated; globally-gated; state-dependent arrival rates; stability; mean value analysis.

1 Introduction

We consider a polling system with N queues. In contrast to most classical polling models where a single server constantly cycles between the queues, we introduce a new model where there are two servers that alternate with each other according to some swapping policy. This "change of guard" is called for by the fact that service times increase from cycle to cycle due to increasing tiredness of the active server. For the system to remain stable it is necessary to replace a weary active server with a rested server while the former rests and gradually regains its efficiency. However, each swap between the servers takes time. Our goal is to analyze such systems and study stability-preserving symmetric swapping policies.

There exists a vast literature on the subject of polling systems and its applications in areas such as manufacturing, transportation, data reading, computer networks, telephone communications, etc. Many variations and extensions of the classic model (e.g. batch-service, fluid models, multiple servers, and different "smart customers" behaviors) have also been studied. We refer the reader to the recent comprehensive survey by Boon, van der Mei and Winands [1] and the 185 references there. The question of steady state stability is thoroughly studied by Flicker and Jaibi [9]. As the analysis of polling systems is complex, various analysis technics have been developed and utilized. Some of those technics are useful only if the service regimes satisfy a certain "branching property" (See Resing [12] for an elaborate treatment of polling systems and branching processes). Some earlier important works are Takagi [10] and Yechiali [14], where overviews of the commonly used analytic methods are presented. Pseudo conservation laws were introduced and developed in Boxma and Groenendijk [4]. See also Winands, Adan and van Houtum [13] and its introduction for a variety of computationally oriented approaches to calculate customers mean waiting times. The bulk of the above manuscripts deal with static server-visit policies. Dynamic server visit-order schemes, which

are analytically intricate, were originally studied in Browne and Yechiali [6] and [7].

The current work is motivated by the notion of shift scheduling. In section 2 we describe the model and introduce the notation used in the analysis. In section 3 we state the relevant stability condition as was introduced in Boxma, Ivanovs, Kosinski and Mandjes [5]. In sections 4 and 5 we define the laws of motion and construct formulas for calculating the means of the queue sizes at various time instants. We also indicate similarities between our model and a standard polling model. Section 6 is devoted to the derivation of the queue joint Probability Generating Functions (PGFs) at arrival and at departure epochs. In section 7 we remark on a method used in Boon, Van Wijk, Adan and Boxma [2] which allows calculation of the marginal queue PGFs in steady state. However, this method is not practical for our needs. The authors in [2] implemented a more practical "Mean Value Analysis" (MVA) approach. This MVA approach is mentioned in section 8, where we also present our optimality criterion. Section 9 consists of an elaboration on the relevant stability conditions, stated in section 3, for a single-queue model. In section 10 we utilize some MVA equations, derived in section 8, in order to analyze a single-queue model with identical servers (the case of not necessarily identical servers is also referred to). In addition, we present an algorithm to obtain a restricted optimal swapping policy between the two (identical) servers, for the case of an exhaustively-served single-queue system with exponential service times. We conclude in section 11 with some illustrative numerical results and possible extensions for our model.

2 The model

2.1 The basic model

Consider a polling system comprised of N queues $Q_1, Q_2, \ldots, Q_N$ served by a single server. For each queue, say Q_i, type-i customers arrive according to an independent Poisson process with rate $\lambda_i > 0$. There are two alternating servers in the system, dubbed "Server 1" and "Server 2". During each cycle exactly one of the servers is active, while the other remains inactive. An active server visits the queues in a cyclic manner, starting from Q_1, and incurring switch-over times when moving from Q_i to Q_{i+1}. The switching time from Q_i to Q_{i+1} is a random variable H_i having Laplace-Stieltjes transform (LST) $\widetilde{H}_i(\cdot)$. Switching times are independent of the servers' identity. At the end of a cycle (i.e. while switching from Q_N back to Q_1), the active server may be replaced by the inactive one. Performing such a "swap" requires additional H_0 units of time so that the actual switch-over time becomes $H'_N = H_N + H_0$. While visiting Q_i, the active server serves according to a pre-determined regime which can be gated, exhaustive, or globally-gated, while the inner order of service is FCFS. The **basic** service duration of a type-i customer is a random variable depending on the identity of the active server. If server 1 is active, the basic service duration of a type-i customer is G_i, where G_i is a positive random variable, drawn from a continuous Probability Distribution Function (PDF) having LST $\widetilde{G}_i(\cdot)$. If server 2 is active, the basic service duration of a type-i customer is K_i, where K_i is a positive random variable, drawn from a continuous PDF having LST $\widetilde{K}_i(\cdot)$.

The servers themselves get weary while active and must rest in order to continue operating in an efficient level. Both servers start at 0 "tiredness level" (**TL**). After each cycle in which a server is active, his **TL** increases by 1. After each cycle in which a server is not active, his **TL** decreases by 1 (to a minimum of 0). Let **ATL** be the **TL** of a cycle's active server. Let $\alpha > 1$ be a "fatigue parameter"- a constant factor corresponding to the deterioration of a server's efficiency. We assume that during each cycle, the **effective** service duration of type-i customers is the basic one multiplied by the factor $\alpha^{\mathbf{ATL}}$. We refer to the described model as the **"basic" model**. Combining the basic model with a specified swapping policy, results in a well-defined operating system.

For example, assume that we start the system with server 1 as the active server. During the first cycle **ATL** is 0, which means that type-i customers are being served according to the basic service duration G_i. By the cycle's end, the **TL** of server 1 increases from 0 to 1. Suppose server 1 continues to be active in the second and third cycles; the time to serve a type-i customer will become αG_i in the second cycle and $\alpha^2 G_i$ in the third. By the end of the third cycle the **TL** of server 1 becomes 3. Now, suppose a swapping occurs and server 2 becomes active during the **next two cycles**. Performing a swap requires additional H_0 units

of time just before Q_1 is revisited for the fourth time (now by server 2). The result will be **ATL**s of 0 and 1 so that type-i customers service durations in cycles number four and five are K_i and αK_i, respectively. By the end of the fifth cycle the **TL** of server 2 becomes 2. Swapping back to server 1 (who's **TL** has been reduced from 3 to 1 since the end of the third cycle) will result in an effective service duration of αG_i for the sixth cycle, after incurring an additional H_0 units of time at the end of the fifth cycle. When the sixth cycle in completed, the **TL** of server 2 has been reduced to 1 and the **TL** of server 1 (the last active server) has been increased to 2.

Before moving on, let us indicate a few points: Firstly, the switch-over times H_i are unaffected by **ATL**; secondly, the choice of the first server to be active does not incur the additional H_0 units of time; and lastly, choosing to never swap the servers will clearly result in the system's explosion.

2.2 Swapping policies and re-modeling of the system

We are interested in comparing various swapping policies in steady state. To this end, for a given swapping policy, we combine several sequential cycles into one big meta-cycle (the formers are referred to as sub-cycles, while the latter is called a cycle). We alter the arrival process in such a way that the resulting model will be equivalent to the original basic model (mentioned in section 2.1), under the given swapping policy. This will enable us to technically analyze the model by using recently developed techniques. We refer to the aforementioned resulting model as the **"new" model**.

To clarify the issue, let us consider a simple swapping policy in which we start the system by activating server 1 and perform a swap at the end of every **two** sub-cycles. This can be alternatively modeled by a polling system composed of 4 sub-cycles, for a total of $4N$ queues $Q_1, Q_2, \ldots, Q_{4N}$. Those queues are visited by a single server with the regular switching times during each sub-cycle and the added H_0 swapping time at the end of every even sub-cycle. The switch-over times corresponding to the $4N$ queues are the elements of a new vector of switch-over times, $\overline{H}_{new}$ (recall that $H'_N = H_N + H_0$):

$$\{H_1, H_2, \ldots, H_{N-1}, \mathbf{H_N}, H_1, H_2, \ldots, H_{N-1}, \mathbf{H'_N}, H_1, H_2, \ldots, H_{N-1}, \mathbf{H_N}, H_1, H_2, \ldots, H_{N-1}, \mathbf{H'_N}\}.$$

In the first and second sub-cycles server 1 is active while the **ATL**s are 0 and 1, respectively. In the third and fourth sub-cycles server 2 is active with **ATL**s 0 and 1. The service times corresponding to individual customers in the $4N$ queues are the elements of a new vector of service times, $\overline{GK}_{new}$:

$$\{G_1, G_2, \ldots, G_{N-1}, G_N, \alpha G_1, \alpha G_2, \ldots, \alpha G_{N-1}, \alpha G_N, K_1, K_2, \ldots, K_{N-1}, K_N, \alpha K_1, \alpha K_2, \ldots, \alpha K_{N-1}, \alpha K_N\}.$$

Note that at the end of the forth sub-cycle (the end of the first meta-cycle) the **TL** of server 1 is reduced back to 0. The same holds true for server 2 at the end of the second sub-cycle in the middle of the second (meta-)cycle. This allows us to replace the two tiring servers by a single non-tiring server.

We will only deal with (servers-)activation orders which are fully repetitive (i.e. can be modeled by using identical cycles), and possess the following two properties:

- Quasi-Fairness: Each cycle is composed of identical number of sub-cycles in which each of the servers is active.
- No Over-Rest: After his initial activation, every time a server's **TL** reaches zero, he is activated immediately.

In each fully repetitive activation order, during the first cycle, both servers reach their first activation sub-cycle with $\mathbf{TL} = 0$. We emphasize that the two mentioned properties are satisfied i.f.f. during each sequential cycle both servers continue to reach their first activation sub-cycle with $\mathbf{TL} = 0$.

In view of the above, we will concentrate on a family of symmetric swapping policies which call for a swap at the end of every T sub-cycles ($T = 1, 2, 3, \ldots$). For example, the "always swap" ($T = 1$) swapping policy means an activation order of 1,2 (without loss of generality, we always set server 1 to be the first active

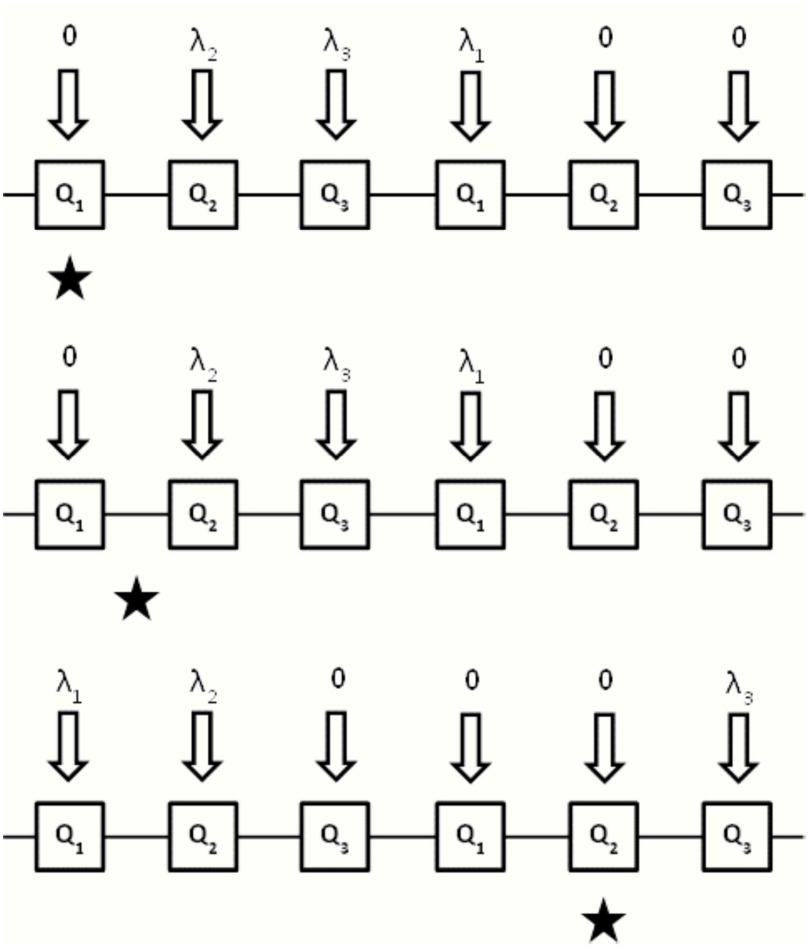
server), while the "swap at the end of every $T = 2$ sub-cycles" swapping policy means an activation order of 1,1,2,2. Note that the discussed family is a subset of the superset constructed of all activation orders which are fully repetitive and possess the "Quasi-Fairness" and "No Over-Rest" properties. Moreover, the later superset includes elements which do not belong to the former subset (e.g. the activation order 1,1,1,1,2,2,1,2,2,2).

Although we do not treat other types of activation orders, some of them can be modeled similarly (e.g. the activation order 1,2,2,1 is identical, in the long run, to the activation order 1,1,2,2 or 2,2,1,1; the activation order which starts with 1,1,1,2,1,2,2 and then continues with 1,2 indefinitely, is identical, in the long run, to the activation order 1,2 with basic service durations of αG_i for server 1 and αK_i for server 2), while other activation orders may result in the system's explosion (e.g. the activation orders 2,2,1 and 1,1,2,1).

The new model isn't yet fully equivalent to the basic model under a given swapping policy because customers' arrival rates to the new model have been quadrupled. In order to make the new model equivalent to the basic one, we incorporate the concept of **state-dependent** arrival rates, where a "state" corresponds to the position (a specific queue being visited or a specific switch-over) of the active server.

The service regime used in the basic model does not affect the service regime used in the new model, which we will consider to be **exhaustive** (we will elaborate on this point in remark 2.2). The basic model's service regime only affects the structure of the state-dependent arrival rates in the new model.
Let C be the total number of sub-cycles which are used to construct a cycle. Thus, each cycle is composed of CN queues. Consider now, the three different service regimes.

Gated: When the server **reaches** Q_i $(i = 1, \ldots, N)$ in sub-cycle γ $(\gamma = 1, \ldots, C)$ all arriving processes of type-i customers in the cycle are set to zero, save that of Q_i in sub-cycle $\gamma + 1$. In this way all and only type-i customers arriving after closing queue-i's gate in sub-cycle γ will be served in the next sub-cycle.
For example, assume $N = 3$, under the gated regime and the "always swap" (swapping) policy $(C = 2)$, some of the state-dependent arrival rates are illustrated in the following scheme (each star represents the position of the active server in the polling system directly above it):



Exhaustive: When the server **departs** from Q_i in sub-cycle γ all arriving processes of type-i customers in the cycle are set to zero, save that of Q_i in sub-cycle $\gamma+1$. In this way all and only type-i customers arriving after moving away from queue-i in sub-cycle γ will be served in the next sub-cycle.

Globally-Gated: When the server **reaches** Q_1 in sub-cycle γ all arriving processes of all different customer types in the cycle are set to zero, save those of sub-cycle $\gamma+1$. In this way all and only customers arriving after closing all the queues' gates in sub-cycle γ will be served in the next sub-cycle.

Now, let's define precisely the server's position. If the server is visiting Q_i we state that his position is V_i. If the server is switching (i.e. moving) from Q_i to Q_{i+1} we state that his position is M_i. Let P denote the vector of server's poison in the new model. Then,

$$P \equiv \{V_1, M_1, V_2, M_2, \ldots, V_N, M_N, \ldots, V_{CN-1}, M_{CN-1}, V_{CN}, M_{CN}\}.$$

Define $\lambda^{(p)}_{i+(\gamma-1)N}$ as the arrival rate to Q_i in sub-cycle γ while the server's position is $p \in P$.
If the basic model operates under the **gated regime**, the arrival rates to all queues in the new model satisfy:

$$\lambda^{(p)}_{i+(\gamma-1)N} = \begin{cases} \lambda_i & p \in \left[V_{i+(\gamma-2)N},\ M_{i-1+(\gamma-1)N}\right], \\ 0 & else, \end{cases} \tag{2.1}$$

where, if $p_1, p_2 \in P$, then $[p_1, p_2]$ is the closed interval of sequential positions starting at p_1 and ending at p_2. That is, the arrival rate to Q_i is λ_i during the time interval from the moment the server reaches Q_i in sub cycle $\gamma-1$ until it completes the switching period from Q_{i-1} to Q_i in sub-cycle γ.
If the basic model operates under the **exhaustive regime**, the arrival rates to all queues in the new model satisfy:

$$\lambda^{(p)}_{i+(\gamma-1)N} = \begin{cases} \lambda_i & p \in \left[M_{i+(\gamma-2)N},\ V_{i+(\gamma-1)N}\right], \\ 0 & else. \end{cases} \tag{2.2}$$

If the basic model operates under the **globally-gated regime**, the arrival rates to all queues in the new model satisfy:

$$\lambda^{(p)}_{i+(\gamma-1)N} = \begin{cases} \lambda_i & p \in \left[V_{1+(\gamma-2)N},\ M_{N+(\gamma-2)N}\right], \\ 0 & else. \end{cases} \tag{2.3}$$

It is easy to see that in the new model, each of the CN queues receive new customers during exactly one sub-cycle, which starts at various positions of the server. Actually, each original arrival process is active only while the server occupies $1/C$ of the available positions in a cycle.

To conclude, with the new adjusted arrival rates (in accordance with the chosen regime), new switching times and new service times for the single server, the new (exhaustively-served) model is equivalent to the basic model with the aforementioned swapping policy.

To complete the presentation of the new model, we introduce additional notation. For $i = 1, 2, \ldots, N$, let $l = i + (\gamma-1)N \in \{1, 2, \ldots, CN\}$ denote the index corresponding to the new CN queues. Let S_l and it's LST, $\widetilde{S}_l(\cdot)$, correspond to the l-th element in the new vector of CN switch-over times $\overline{H}_{new}$, and let B_l and it's LST, $\widetilde{B}_l(\cdot)$, correspond to the l-th element in the new vector of CN service times, $\overline{GK}_{new}$. Let $j \in \{1, 2, \ldots, CN\}$, and note that the sub-cycle number corresponding to the system states V_j and M_j is $\gamma = \left\lceil \frac{j}{N} \right\rceil$ (e.g. $\lambda^{(p)}_{j+(\gamma-1)N} = \lambda^{(p)}_{j+(\lceil \frac{j}{N} \rceil - 1)N}$).

Remark 2.1. When calculating expressions related to the globally-gated regime, we sometimes use the notation "$(l \ mod \ N)$" or "$(j \ mod \ N)$". This notation should always be considered modulo N. For example, for $l = 2N$, $(l \ mod \ N) = N$.

Remark 2.2. In the new model, under the gated regime $\lambda^{(V_l)}_l = 0 \ \forall l = 1, \ldots, CN$, while under the globally-gated regime $\lambda^{(p)}_l = 0 \ \forall l = 1, \ldots, CN$ and $p \in [V_{l-(l \ mod \ N)+1},\ V_l]$. In other words, under the gated and

globally-gated regimes, customers arrival rate to Q_l equals zero, from the moment Q_l's gate closes until the end of V_l. Hence, the type-l customers present at the arrival epoch to Q_l are exactly the type-l customers who are supposed to be served during V_l. Since there are no new type-l customers arriving during V_l, we can assume that a basic model which operates under any of the three various service regimes, translates to a new model which operates under the **exhaustive** service regime with the relevant state-dependent arrival rates. Note that, in the sequel, when referring to the new model under the gated (exhaustive; globally-gated) regime, we mean the new model **which originated from a basic model** operating under the gated (exhaustive; globally-gated) regime.

Remark 2.3. When discussing a new model which originated from a basic model consisting of a single queue, we use the notation $\lambda \equiv \lambda_1$, $H \sim H_1$, $G \sim G_1$ and $K \sim K_1$.

2.3 The compact model

Under a "swap at the end of every $T \geq 2$ sub-cycles" policy, each cycle of the resulting new model consists of $2T$ sub-cycles. However, assuming that in the basic model **both servers are identical**, each of the first consecutive T sub-cycles share the same **effective** service durations with its counterpart in the last consecutive T sub-cycles. Since, within a cycle, swapping occurs only at the end of the T-th sub-cycle and the $2T$-th sub-cycle, the new model's cycle can practically be "cut" in half into two separated identically structured "half-cycles". By setting a "half-cycle" to be a full cycle in the new model, we create a "compact" form of the new model, which is easier to analyze, since it is essentially a new model with half the number of queues ($l, j = 1, \ldots, \frac{CN}{2}$; $\frac{CN}{2} \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of integer numbers). We refer to the compact form of a new model as the **"compact" model**.

We now state some points of interest regarding the compact model. Firstly, as long as we discuss a basic model with identical servers which operate under a "swap at the end of every $T \geq 2$ sub-cycles" policy, we can always consider a compact model from a new model. Generally speaking, this holds true since in the new model, under all service regimes, when the server changes his position, the arrival rates to queues whose next visit period will occur **after** the following sub-cycle do not change. This prevents overlapping between different arrival rates in the compact model.

Secondly, we can also consider a compact model in case we discuss a basic model with identical servers which operate under the "always swap" policy, but we have to be careful. In the new model, under all service regimes, each queue is empty at a server's departure epoch from it. However, only in a basic model operating under the exhaustive regime, does the equivalent queue in the next sub-cycle is also empty at the same epoch. The lack of this property in a basic model operating under the gated or globally-gated regime, contradicts the new model's exhaustive service regime assumption. Therefore, under the "always swap" policy, the service regime used in the compact model is **the same** as the service regime used in the basic model.

Thirdly, when discussing a basic model with identical servers which operate under an "always swap" policy, the compact model does not include state-dependent arrival rates. Moreover, if the basic model consists of a single queue and operates under the exhaustive regime, the resulting compact model will be an M/G/1 system with multiple server vacations (see Levy and Yechiali [11]) of length $H'_N = H_N + H_0$ (the resulting new model will, of course, also be equivalent).

Lastly, consider the **new** model which originated from a basic model operating under the "swap at the end of every T sub-cycles" policy and the **compact** model which originated from the same basic model under the "swap at the end of every $2T$ sub-cycles" policy. Clearly, both models have the same number ($2T$) of sub-cycles per cycle. Comparing the structure of both models emphasizes the differences between them (e.g. choosing to swap earlier thus "paying" the additional H_0 units of time in order to avoid increased effective service durations), and allows for a convenient arguments comparison (i.e. the expected cycle's times, the expected visit time in the l-th visited queue etc.). However, one does not gain nor loss any new information that cannot be extracted from comparing two new models, or better yet, an easier analyzed two compact models.

3 Stability

Based on [5], due to the state-dependent arrival rates, the condition for stability in the new model is that the Perron-Frobenius eigenvalue of the matrix $(\mathbf{R}-\mathbf{I_{CN}})$ should be strictly less than 0, where $\mathbf{I_{CN}}$ is the identity matrix of order CN and

$$\mathbf{R} \equiv \left(\lambda_l^{(V_j)} E(B_l)\right) = \begin{pmatrix} \lambda_1^{(V_1)} E(B_1) & \cdots & \lambda_1^{(V_{CN})} E(B_1) \\ \vdots & \ddots & \vdots \\ \lambda_{CN}^{(V_1)} E(B_{CN}) & \cdots & \lambda_{CN}^{(V_{CN})} E(B_{CN}) \end{pmatrix}. \tag{3.1}$$

In other words, let $\pi \in \mathbb{R}$ where $\mathbb{R}$ denotes the set of real numbers. The above stability condition means that the maximal π root of the equation

$$\det\left(\mathbf{R} - (\pi+1) * \mathbf{I_{CN}}\right) = 0, \tag{3.2}$$

should be negative. Note that the (l,j) element of the matrix $(\mathbf{R}-\mathbf{I_{CN}})$ corresponds to the expected change in the number of customers in Q_l, during an average service time of a type-l customer while the server visits Q_j. Another point of interest is that at list half out of the values of $\lambda_l^{(V_j)}$ are zeroes. More accurately, only $1/C$ out of the $(CN)^2$ elements of $\mathbf{R}$ are non-zeroes.

Remark 3.1. Since a compact model is equivalent to a certain new model, they also share the same stability condition. This can also be seen by directly calculating the stability condition for the compact model (where $, j = 1, \ldots, \frac{CN}{2}$). Note that the stability condition does not change if the model in question operates under the gated or globally-gated regime. Thus, even under those regimes and an "always swap" policy, directly calculating the stability condition for the compact model will result in the classical "total traffic-intensity in system must be less than 1" stability condition.

4 Laws of motion

Define θ_l to be the length of a busy period in a regular M/G/1 queueing system with type-l customers, constant $\lambda_l^{(V_l)}$ arrival rate, and service times B_l. Then, $E(\theta_l) = \frac{E(B_l)}{1-\rho_l}$, where
$\rho_l \equiv \lambda_l^{(V_l)} E(B_l)$ is the fraction of time the server is visiting Q_l.
The LST of θ_l is the root in $(0,1]$ of the equation (see e.g. Cohen [8]):

$$\widetilde{\theta}_l(w) = \widetilde{B}_l\left[w + \lambda_l^{(V_l)}\left(1 - \widetilde{\theta}_l(w)\right)\right].$$

Define

$$D_l \equiv \begin{cases} \theta_l & \text{exhaustive,} \\ B_l & \text{gated or globally-gated,} \end{cases} \tag{4.1}$$

with LST,

$$\widetilde{D}_l(w) \equiv \begin{cases} \widetilde{\theta}_l(w) & \text{exhaustive,} \\ \widetilde{B}_l(w) & \text{gated or globally-gated.} \end{cases} \tag{4.2}$$

Under the exhaustive regime $\lambda_l^{(V_l)} = \lambda_{l+(\gamma-1)N}$. But under both the gated and globally-gated regimes $\lambda_l^{(V_l)} = 0$, implying that $\theta_l = B_l$, and $\widetilde{\theta}_l(w) = \widetilde{B}_l(w)$.

Let $A_l^{(p)}(\Omega)$ denote the number of Poisson arrivals to Q_l during a (random) time interval of length Ω (with LST $\widetilde{\Omega}(w)$) while the server's position is p throughout that time interval. Define X_l^j as the number of customers in Q_j $(j = 1, \ldots, CN)$ at the moment the server polls Q_l, and define Y_l^j as the number of customers in Q_j at the moment the server departs from Q_l. Moreover, let $D_{l1}, D_{l2}, D_{l3}, \ldots$ be a sequence of i.i.d. random variables all distributed like D_l. Noting that under the gated and globally-gated regimes $\lambda_l^{(V_l)} = \lambda_l^{(M_l)} = 0$, and under the exhaustive regime $\lambda_l^{(M_l)} = 0$, the evolution of the process is as follows:

$$Y_l^j = \begin{cases} X_l^j + A_j^{(V_l)}(\sum_{k=1}^{X_l^l} D_{lk}) & l \neq j, \\ 0 & l = j, \end{cases} \tag{4.3}$$

$$X_{l+1}^{j} = \begin{cases} X_l^j + A_j^{(V_l)}(\sum_{k=1}^{X_l^l} D_{lk}) + A_j^{(M_l)}(S_l) & l \neq j, \\ 0 & l = j. \end{cases} \tag{4.4}$$

5 First moments

From the aforementioned laws of motion (4.3) and (4.4), it is easy to observe that:

$$E(Y_l^j) = \begin{cases} E(X_l^j) + \lambda_j^{(V_l)} E\left(X_l^l\right) E\left(D_l\right) & l \neq j, \\ 0 & l = j, \end{cases} \tag{5.1}$$

$$E(X_{l+1}^j) = \begin{cases} E(X_l^j) + \lambda_j^{(V_l)} E\left(X_l^l\right) E\left(D_l\right) + \lambda_j^{(M_l)} E\left(S_l\right) & l \neq j, \\ 0 & l = j. \end{cases} \tag{5.2}$$

Recursive substitution in equation (5.1) yields,

$$E\left(X_l^j\right) = \sum_{r=j+1}^{l-1} [\lambda_j^{(V_r)} E\left(X_r^r\right) E\left(D_r\right) + \lambda_j^{(M_r)} E\left(S_r\right)]. \tag{5.3}$$

Note that equation (5.3) holds for all three regimes. Now, let us consider the case where $l = j$ for the various regimes:
For the **gated regime**,

$$E\left(X_j^j\right) = \lambda_{j-(\gamma-1)N} \sum_{r=j-N}^{j-1} [E\left(X_r^r\right) E\left(B_r\right) + E\left(S_r\right)]. \tag{5.4}$$

Note that $\lambda_{j-(\gamma-1)N}$ is simply an alternative way to represent the original queue's constant λ_j. Thus, $E\left(X_j^j\right)$ is actually the number of customers arriving to the original queue, $Q_{j-(\gamma-1)N}$, counting from the beginning of the last time the server visited $Q_{j-(\gamma-1)N}$ until it's return. This is similar to classical polling systems operating under the gated regime in which $E\left(X_i^i\right) = \lambda_i E(\mathbf{C})$, were $\mathbf{C}$ stands for the cycle duration. Also note that $E\left(X_l^j\right)$ will always be zero unless $\{[j+1, l-1] \cap [j-N, j-1]\} \neq \emptyset$, where $\emptyset$ denotes the empty set. This can be interpreted as having less than N switch-over periods while cycling from Q_l to Q_j in the new model. From a combinatorial point of view, having a choice of CN places for "l", and N places for "j", means that, similarly to the case of the matrix $\mathbf{R}$ in section 3 (see equation (3.1) and its following explanation), only $1/C$ out of the $(CN)^2$ values of $E\left(X_l^j\right)$ are non-zeroes.
For the **exhaustive regime**,

$$E\left(X_j^j\right) = \lambda_{j-(\gamma-1)N} \left[E\left(S_{j-N}\right) + \sum_{r=j-N+1}^{j-1} [E\left(X_r^r\right) E\left(\theta_r\right) + E\left(S_r\right)] \right]. \tag{5.5}$$

$E\left(X_j^j\right)$ is actually the number of customers arriving to the original queue, $Q_{j-(\gamma-1)N}$, counting from the last time the server **departed** from $Q_{j-(\gamma-1)N}$ until it's return epoch to $Q_{j-(\gamma-1)N}$. Denote by $\mathbf{SC}$ the elapsed time period between the last time the server **polled** $Q_{j-(\gamma-1)N}$ until it's return epoch to $Q_{j-(\gamma-1)N}$. Then, $E\left(X_j^j\right)$ is actually the number of customers arriving to $Q_{j-(\gamma-1)N}$, during $\left[1 - \lambda_{j-(\gamma-1)N} E\left(B_{j-N}\right)\right] E(\mathbf{SC})$. Note that

$$\begin{aligned} E(X_j^j) &= \lambda_{j-(\gamma-1)N} \left[E(\mathbf{SC}) - E(V_{j-N})\right] \\ &= \lambda_{j-(\gamma-1)N} \left[E(\mathbf{SC}) - \lambda_{j-(\gamma-1)N} E(B_{j-N}) E(\mathbf{SC})\right] \\ &= \lambda_{j-(\gamma-1)N} [1 - \lambda_{j-(\gamma-1)N} E(B_{j-N})] E(\mathbf{SC}). \end{aligned}$$

This is similar to classical polling systems operating under the exhaustive regime in which $E\left(X_i^i\right) = \lambda_i [1 - \lambda_i E(B_i)] E(\mathbf{C})$. For similar reasons to those stated regarding the gated regime, only $1/C$ out of the $(CN)^2$ values of $E\left(X_l^j\right)$ are non-zeroes.

Finally, for the **globally-gated regime**,

$$E\left(X_j^j\right) = \lambda_{j-(\gamma-1)N} \sum_{r=j-(j \bmod N)+1-N}^{j-(j \bmod N)} \left[E\left(X_r^r\right) E\left(B_r\right) + E\left(S_r\right)\right]. \tag{5.6}$$

$E\left(X_j^j\right)$ is actually the number of customers arriving to the original queue, $Q_{j-(\gamma-1)N}$, during the sub-cycle which is followed by sub-cycle $\gamma = \left\lceil \frac{j}{N} \right\rceil$. This is similar to classical polling systems operating under the globally-gated regime in which $E\left(X_i^i\right) = \lambda_i E(\mathbf{C})$. Again, for similar reasons to those stated regarding the gated regime, only $1/C$ out of the $(CN)^2$ values of $E\left(X_l^j\right)$ are non-zeroes.

Remark 5.1. One can utilize equation (5.3) in order to derive expressions for $E(X_l^j)$s for $l \neq j$, some of which can than be used to show additional similarities between our model and classical polling systems. However, explicitly expressing the results requires arguments which do not have equivalent representations in classical polling systems (as is the case of the $E(X_j^j)$s which are expressed by the expected length of certain sub-cycles which start at various epochs, e.g. $\mathbf{SC}$). Generally speaking, this holds true since the underling system operates under varying conditions (i.e. the tiredness effects and swapping policy dictates different parameters for each "basic model cycle"). We will not elaborate further on this subject.

6 PGFs of Joint queue length at state change epochs

For **each service regime**, define the joint PGF of $\left\{X_{l+1}^j\right\}_{j=1}^{CN}$ as,

$$\begin{aligned}
\widehat{X}_{l+1}\left(z_1, z_2, \ldots, z_{CN}\right) &\equiv E\left(\prod_{j=1}^{CN} z_j^{X_{l+1}^j}\right) \\
&= E_{X_l} E\left(\prod_{j=1}^{CN} z_j^{X_{l+1}^j} | X_l\right) = E_{X_l} E\left(\prod_{\substack{j=1 \\ j\neq l}}^{CN} z_j^{X_l^j + A_j^{(V_l)}\left(\sum_{k=1}^{X_l^l} D_{lk}\right) + A_j^{(M_l)}(S_l)} | X_l\right) \\
&= E_{X_l}\left(\prod_{\substack{j=1 \\ j\neq l}}^{CN} z_j^{X_l^j} * \left[E\left(\prod_{\substack{j=1 \\ j\neq l}}^{CN} z_j^{A_j^{(V_l)}(D_l)}\right)\right]^{X_l^l}\right) * E\left(\prod_{\substack{j=1 \\ j\neq l}}^{CN} z_j^{A_j^{(M_l)}(S_l)}\right).
\end{aligned} \tag{6.1}$$

Note that,

$$E\left(\prod_{j=1}^{CN} z_j^{A_l^{(P)}(\Omega)}\right) = \widetilde{\Omega}\left[\sum_{j=1}^{CN} \lambda_j^{(P)}(1-z_j)\right].$$

Then, using equation (6.1),

$$\begin{aligned}
&\widehat{X}_{l+1}\left(z_1, z_2, \ldots, z_{CN}\right) \\
&= E_{X_l}\left(\prod_{\substack{j=1 \\ j\neq l}}^{CN} z_j^{X_l^j} * \left(\widetilde{D}_l\left[\sum_{\substack{j=1 \\ j\neq l}}^{CN} \lambda_j^{(V_l)}\left(1-z_j\right)\right]\right)^{X_l^l}\right) * \widetilde{S}_l\left[\sum_{\substack{j=1 \\ j\neq l}}^{CN} \lambda_j^{(M_l)}\left(1-z_j\right)\right].
\end{aligned} \tag{6.2}$$

We conclude, from equation (6.2) that,

$$\widehat{X}_{l+1}\left(z_1,z_2,\ldots,z_{CN}\right)$$
$$=\widehat{X}_l\left(z_1,z_2,\ldots,z_{l-1},\widetilde{D}_l\left[\sum_{\substack{j=1\\ j\neq l}}^{CN}\lambda_j^{(V_l)}\left(1-z_j\right)\right],z_{l+1},\ldots,z_{CN}\right)*\widetilde{S}_l\left[\sum_{\substack{j=1\\ j\neq l}}^{CN}\lambda_j^{(M_l)}\left(1-z_j\right)\right]. \tag{6.3}$$

In a similar manner we obtain the PGF $\widehat{Y}_l\left(z_1,z_2,\ldots,z_{CN}\right)$:

$$\widehat{Y}_l\left(z_1,z_2,\ldots,z_{CN}\right)\equiv E\left(\prod_{j=1}^{CN}z_j^{Y_l^j}\right)=\widehat{X}_l\left(z_1,z_2,\ldots,z_{l-1},\widetilde{D}_l\left[\sum_{\substack{j=1\\ j\neq l}}^{CN}\lambda_j^{(V_l)}\left(1-z_j\right)\right],z_{l+1},\ldots,z_{CN}\right). \tag{6.4}$$

Thus, equation (6.3) can be rewritten as

$$\widehat{X}_{l+1}\left(z_1,z_2,\ldots,z_{CN}\right)=\widehat{Y}_l\left(z_1,z_2,\ldots,z_{CN}\right)*\widetilde{S}_l\left[\sum_{\substack{j=1\\ j\neq l}}^{CN}\lambda_j^{(M_l)}\left(1-z_j\right)\right]. \tag{6.5}$$

Namely, the number of customers at the various queues at an instant of server's visit to Q_{l+1} is the sum of the number of customers at server's departure from Q_l plus the number of arrivals to Q_{l+1} during the switch-over time S_l. We now distinguish between the three different regimes.

The gated regime

According to equation (4.2), for the gated regime, $\widetilde{D}_l\left(\cdot\right)=\widetilde{B}_l\left(\cdot\right)$. Let us consider arguments from equation (6.3), starting with the expression $\widetilde{B}_l\left[\sum_{\substack{j=1\\ j\neq l}}^{CN}\lambda_j^{(V_l)}\left(1-z_j\right)\right]$. In accordance with the calculations of $E\left(X_j^j\right)$ in section 5, the only queues whose arrival rates are positive for a given $p=V_l$ are the N queues which will be visited right after the current Q_l. Moreover, those arrival rates are precisely the constant ones of the original queues. Hence, the mentioned expression can be rewritten as $\widetilde{B}_l\left[\sum_{j=l+1}^{l+N}\lambda_{j-(\gamma-1)N}\left(1-z_j\right)\right]$. The same holds true regarding $\widetilde{S}_l\left[\sum_{\substack{j=1\\ j\neq l}}^{CN}\lambda_j^{(M_l)}(1-z_j)\right]$, since $\lambda_j^{(V_l)}=\lambda_j^{(M_l)}\ \forall l,j$. Therefore, the last expression can be rewritten as $\widetilde{S}_l\left[\sum_{j=l+1}^{l+N}\lambda_{j-(\gamma-1)N}(1-z_j)\right]$.

Equation (6.4) thus becomes

$$\widehat{Y}_l\left(z_1,z_2,\ldots,z_{CN}\right)=\widehat{X}_l\left(z_1,z_2,\ldots,z_{l-1},\widetilde{B}_l\left[\sum_{j=l+1}^{l+N}\lambda_{j-(\gamma-1)N}\left(1-z_j\right)\right],z_{l+1},\ldots,z_{CN}\right), \tag{6.6}$$

and equation (6.5) becomes

$$\widehat{X}_{l+1}\left(z_1, z_2, \ldots, z_{CN}\right) = \widehat{Y}_l\left(z_1, z_2, \ldots, z_{CN}\right) * \widetilde{S}_l\left[\sum_{j=l+1}^{l+N} \lambda_{j-(\gamma-1)N}\left(1 - z_j\right)\right]. \tag{6.7}$$

The exhaustive regime

According to equation (4.2), for the exhaustive regime, $\widetilde{D}_l\left(\cdot\right) = \widetilde{\theta}_l\left(\cdot\right)$. Let us consider arguments from equation (6.3), starting with the expression $\widetilde{\theta}_l\left[\sum_{\substack{j=1 \\ j \neq l}}^{CN} \lambda_j^{(V_l)}\left(1 - z_j\right)\right]$. In accordance with the calculations of $E\left(X_j^j\right)$ in section 5, the only queues whose arrival rates are positive for a given $p = V_l$ are the current Q_l and the following $N-1$ queues. As before, those arrival rates are precisely the constant ones of the original queues. Hence, the mentioned expression can be rewritten as $\widetilde{\theta}_l\left[\sum_{j=l+1}^{l+N-1} \lambda_{j-(\gamma-1)N}\left(1 - z_j\right)\right]$. Regarding $\widetilde{S}_l\left[\sum_{\substack{j=1 \\ j \neq l}}^{CN} \lambda_j^{(M_l)}(1 - z_j)\right]$, the only queues whose arrival rates are positive for a given $p = M_l$ are the next N queues to be visited (again with the original queues' arrival rates). Therefore, the last expression can be rewritten as $\widetilde{S}_l\left[\sum_{j=l+1}^{l+N} \lambda_{j-(\gamma-1)N}(1 - z_j)\right]$.
Equation (6.4) thus becomes

$$\widehat{Y}_l\left(z_1, z_2, \ldots, z_{CN}\right) = \widehat{X}_l\left(z_1, z_2, \ldots, z_{l-1}, \widetilde{\theta}_l\left[\sum_{j=l+1}^{l+N-1} \lambda_{j-(\gamma-1)N}\left(1 - z_j\right)\right], z_{l+1}, \ldots, z_{CN}\right), \tag{6.8}$$

and equation (6.5) becomes

$$\widehat{X}_{l+1}\left(z_1, z_2, \ldots, z_{CN}\right) = \widehat{Y}_l\left(z_1, z_2, \ldots, z_{CN}\right) * \widetilde{S}_l\left[\sum_{j=l+1}^{l+N} \lambda_{j-(\gamma-1)N}\left(1 - z_j\right)\right]. \tag{6.9}$$

The globally-gated regime

For the globally-gated regime, $\widetilde{D}_l\left(\cdot\right) = \widetilde{B}_l\left(\cdot\right)$. Consider the expression $\widetilde{B}_l\left[\sum_{\substack{j=1 \\ j \neq l}}^{CN} \lambda_j^{(V_l)}\left(1 - z_j\right)\right]$ from equation (6.3). In accordance with the calculations of $E\left(X_j^j\right)$ in section 5, the only queues whose arrival rates are positive for a given $p = V_l$ are the N queues composing the previous sub-cycle (which is followed by sub-cycle $\left\lceil \frac{l}{N} \right\rceil$). Moreover, those arrival rates are the precisely constant ones of the original queues. Hence, the mentioned expression can be rewritten as $\widetilde{B}_l\left[\sum_{j=l-(l \bmod N)+1-N}^{l-(l \bmod N)} \lambda_{j-(\gamma-1)N}\left(1 - z_j\right)\right]$. The same holds true regarding $\widetilde{S}_l\left[\sum_{\substack{j=1 \\ j \neq l}}^{CN} \lambda_j^{(M_l)}(1 - z_j)\right]$, since $\lambda_j^{(V_l)} = \lambda_j^{(M_l)}\ \forall l, j$. Therefore, the latter expression can be rewritten as $\widetilde{S}_l\left[\sum_{j=l-(l \bmod N)+1-N}^{l-(l \bmod N)} \lambda_{j-(\gamma-1)N}(1 - z_j)\right]$.
Equation (6.4) thus becomes

$$\widehat{Y}_l\left(z_1, z_2, \ldots, z_{CN}\right) = \widehat{X}_l\left(z_1, z_2, \ldots, z_{l-1}, \widetilde{B}_l\left[\sum_{j=l-(l \bmod N)+1-N}^{l-(l \bmod N)} \lambda_{j-(\gamma-1)N}\left(1 - z_j\right)\right], z_{l+1}, \ldots, z_{CN}\right), \tag{6.10}$$

and equation (6.5) becomes

$$\widehat{X}_{l+1}(z_1, z_2, \ldots, z_{CN}) = \widehat{Y}_l(z_1, z_2, \ldots, z_{CN}) * \widetilde{S}_l\left[\sum_{j=l-(l \bmod N)+1-N}^{l-(l \bmod N)} \lambda_{j-(\gamma-1)N}(1-z_j)\right]. \tag{6.11}$$

7 Marginal queue PGF in steady state

As stated in [2], although the steady state marginal queue length distributions at customer's arrival and departure epochs are the same, they differ from the distribution of the steady state marginal queue length at an arbitrary moment. In other words, the PASTA property does not hold. The authors in [2] circumvented this problem by relaying on the arrival rate's "fixation" during a given position for the server.
Specifically, define $\widehat{L}_j(z)$ as the PGF of Q_j length at an arbitrary moment, and $\widehat{L}_{j|p}(z)$ as the PGF of Q_j length during an arbitrary moment under the condition that the server resides in positions p. Weighting over the relative expected time the server occupies the different positions during a steady state cycle, yields the following relation:

$$\widehat{L}_j(z) \equiv E\left(z^{L_j}\right) = \sum_{p=V_1}^{M_{CN}} \frac{E(p)}{E(\mathbf{C})}\widehat{L}_{j|p}(z) = \sum_{l=1}^{CN}\left[\frac{E(V_l)}{E(\mathbf{C})}\widehat{L}_{j|p=V_l}(z) + \frac{E(M_l)}{E(\mathbf{C})}\widehat{L}_{j|p=M_l}(z)\right]. \tag{7.1}$$

Note that p (i.e. V_l and M_l) serves, according to the context, either as an indicator of the system state, or as a random variable measuring the time length of the system state.

In order to compute equation (7.1), one needs to find $\widehat{L}_{j|p}(z)$, $E(V_l)$ and $E(\mathbf{C})$ (obviously, $E(M_l) = E(S_l)\ \forall l$). $E(V_l)$ and $E(\mathbf{C})$ can be calculated using equations (8.1) – (8.4) which will be presented in section 8. For the sake of brevity, we refer the reader to [2] for an elaboration on the calculation of $\widehat{L}_{j|p}(z)$, which can be implemented in our model using equations (2.1) – (2.3). However, we note that the calculation of $\widehat{L}_{j|p}(z)$ in [2] requires the use of explicit expressions for the PGFs (6.6) – (6.11). Using recursive substitutions, the latter can be expressed as a function of $\widehat{X}_1(\cdot)$ (or any other $\widehat{X}_l(\cdot)$) and the known LSTs $\widetilde{B}_l(\cdot)$ and $\widetilde{S}_l(\cdot)$. However, there is no known simple explicit expression for $\widehat{X}_1(\cdot)$. Based on [12], Boxma [3] expressed $\widehat{X}_1(\cdot)$ as an infinite product of arguments from the framework of "Multiple Branching Processes with Immigration". Nevertheless, this presentation still results in technically intractable mathematical model. We remark that this problem does not necessarily prevents one from using the resulting intractable expressions for some general proofs (e.g. for convergence).

Remark 7.1. In [2], the authors also addressed the issue of finding the LSTs of the waiting time distributions. The state-dependent arrival rates imply that the distributional form of Little's Law does not hold. As a result, the authors in [2] developed a generalization of the distributional form which can be applied. As in the case of the marginal queue PGFs in steady state, this method leads to expressions which include an infinite product. Moreover, the use of this method is accompanied by a considerable increase in complexity, due to the need to add additional queues in models with zero arrival rates (as is in our case).

8 Mean Value Analysis

As noted in section 7, finding the marginal queues' lengths PGF in steady state is intractable. As explained in [2], one can use a "Mean Value Analysis" (MVA) approach in order to calculate the expected waiting time of type-l customers in steady state (and the corresponding expected queues lengths). The use of the MVA approach relays, among other thing, on deriving explicit expressions for $E(V_l)$ and $E(\mathbf{C})$. We now show how to compute them.

Each of the type-l customers present at the arrival epoch to Q_l initiates a (possibly degenerate) regular M/G/1 busy period. The number of the mentioned type-l customers is determined by the state-dependent

arrival rate, accumulated from the last epoch at which the server departs (the empty) Q_l. This yields:

$$E\left(V_l\right)=\frac{E\left(B_l\right)}{1-\lambda_l^{(V_l)}*E\left(B_l\right)}*\left[\lambda_l^{(M_l)}E\left(S_l\right)+\sum_{r=l+1}^{l+CN-1}\left(\lambda_l^{(V_r)}E\left(V_r\right)+\lambda_l^{(M_r)}E\left(S_r\right)\right)\right].$$

Under the **gated regime** this means

$$E\left(V_l\right)=\lambda_{l-(\gamma-1)N}*E\left(B_l\right)*\sum_{r=l-N}^{l-1}\left(E\left(V_r\right)+E\left(S_r\right)\right). \tag{8.1}$$

Under the **exhaustive regime**,

$$E\left(V_l\right)=\frac{\lambda_{l-(\gamma-1)N}*E\left(B_l\right)}{1-\lambda_{l-(\gamma-1)N}*E\left(B_l\right)}*\left[E\left(S_{l-N}\right)+\sum_{r=l-N+1}^{l-1}\left(E\left(V_r\right)+E\left(S_r\right)\right)\right]. \tag{8.2}$$

Under the **globally-gated regime**,

$$E\left(V_l\right)=\lambda_{l-(\gamma-1)N}*E\left(B_l\right)*\sum_{r=l-(l\ mod\ N)+1-N}^{l-(l\ mod\ N)}\left(E\left(V_r\right)+E\left(S_r\right)\right). \tag{8.3}$$

Obviously, under each service regimes,

$$E\left(\mathbf{C}\right)=\sum_{l=1}^{CN}\left(E\left(V_l\right)+\ E\left(S_l\right)\right). \tag{8.4}$$

The idea behind the MVA approach is to express $E\left(Lq_l\right)$ (the expected number of type-l customers in the system, excluding a potential type-l customer in service), and $E(Wq_l)$ (the expected time a type-l customer waits from his arrival epoch until his service starts), using a set of equations (linear in the method's arguments) and, by implementing Little's Law, derive explicit expressions for them. In section 10 we will use the MVA approach in order to analyze some basic cases derived from our model. In the current section we mostly present the equations which will be used in that analysis. For a complete presentation of the MVA approach, for exhaustively served polling systems which include state-dependent arrival rates, we refer the reader to [2]. We note that the MVA approach, in conjunction with equations $(2.1)-(2.3)$, can be used to (numerically) study intricate cases derived from our model.

8.1 Some MVA equations

This would require some explanations and notation, which we will present simultaneously.
Let $p,p_1,p_2\in P$ be system states. The fraction of time the system spends in a given state p within a cycle is (recall that p serves both as an indicator of the server's position, or as the duration that the server stays in that position)

$$\rho^{(p)}\equiv\frac{E\left(p\right)}{E\left(\mathbf{C}\right)}.$$

The mean arrival rate of type-l customers to the system is

$$\overline{\lambda}_l\equiv\sum_{p=V_1}^{M_{CN}}\rho^{(p)}*\lambda_l^{(p)}=\frac{1}{E\left(\mathbf{C}\right)}\sum_{j=1}^{CN}\left[E\left(V_j\right)*\lambda_l^{(V_j)}+E\left(S_j\right)*\lambda_l^{(M_j)}\right].$$

Under the **gated regime** this means

$$\overline{\lambda}_l=\frac{\lambda_{l-(\gamma-1)N}}{E\left(\mathbf{C}\right)}*\sum_{j=l-N}^{l-1}\left[E\left(V_j\right)+E\left(S_j\right)\right]. \tag{8.5}$$

Under the **exhaustive regime**,

$$\overline{\lambda}_l = \frac{\lambda_{l-(\gamma-1)N}}{E(\mathbf{C})} * \left[E(S_{l-N}) + \sum_{j=l-N+1}^{l-1} [E(V_j) + E(S_j)] + E(V_l) \right]. \tag{8.6}$$

Under the **globally-gated regime**,

$$\overline{\lambda}_l = \frac{\lambda_{l-(\gamma-1)N}}{E(\mathbf{C})} * \sum_{j=l-(l \bmod N)+1-N}^{l-(l \bmod N)} [E(V_j) + E(S_j)]. \tag{8.7}$$

Note that under **all** regimes (either in a new or compact model form), adding together the $\overline{\lambda}_l$s representing all "duplications" of an original basic model queue, will sum up to the latter's original λ. That is,

$$\sum_{r=1}^{C} \overline{\lambda}_{l+(r-1)N} = \frac{\lambda_{l-(\gamma-1)N}}{E(\mathbf{C})} * E(\mathbf{C}) = \lambda_{l-(\gamma-1)N}.$$

By little's law,

$$E(Lq_l) = \overline{\lambda}_l E(Wq_l). \tag{8.8}$$

Define $E(R_p)$ as the expected residual duration of time the system spends in state p, assuming it is currently at a random epoch within p. The expected residual time until the end of the service of the currently served type-l customer is,

$$E(R_{B_l}) = \frac{E(B_l{}^2)}{2E(B_l)}.$$

In addition,

$$E(R_{S_l}) \equiv E(R_{M_l}) = \frac{E(S_l{}^2)}{2E(S_l)}.$$

The rest of the MVA equations are presented under the assumption that the basic model consists of a **single queue**. $E\left(Lq_l^{(p)}\right)$ is the expected number of type-l customers in the system excluding a potential type-l customer in service, assuming the system is currently at a random epoch within p. For $p \neq V_l$ it can be calculated by building up the number of type-l customers, counting from the last departure epoch from Q_l until the current random point in state p.
Assuming **(globally) gated regime**,

$$E\left(Lq_l^{(M_{l-1})}\right) = \lambda\left(E(V_{l-1}) + E\left(R_{S_{l-1}}\right)\right), \tag{8.9}$$

and

$$E\left(Lq_l^{(V_{l-1})}\right) = \lambda E\left(R_{V_{l-1}}\right), \tag{8.10}$$

while $E\left(Lq_l^{(p)}\right) = 0$ for $p \neq V_{l-1}, M_{l-1}, V_l$.
Assuming **exhaustive regime**,

$$E\left(Lq_l^{(M_{l-1})}\right) = \lambda E\left(R_{S_{l-1}}\right), \tag{8.11}$$

while $E\left(Lq_l^{(p)}\right) = 0$ for $p \neq M_{l-1}, V_l$.

Now, assuming **(globally) gated regime**, $E(R_{V_l})$ consists of the expected residual serving time of the type-l customer currently being served and the sum of all expected service times of the other type-l customers in Q_l. We thus write

$$E(R_{V_l}) = E(R_{B_l}) + E\left(Lq_l^{(V_l)}\right) E(B_l). \tag{8.12}$$

For **all** service regimes, we have

$$E(Lq_l) = \sum_{p=V_1}^{M_C} \rho^{(p)} E\left(Lq_l^{(p)}\right).$$

For the **(globally) gated regime** this means (recall that $\left(Lq_l^{(p)}\right) = 0$ for $p \neq V_{l-1}, M_{l-1}, V_l$),

$$E\left(Lq_l\right) = \rho^{(V_{l-1})} E\left(Lq_l^{(V_{l-1})}\right) + \rho^{(M_{l-1})} E\left(Lq_l^{(M_{l-1})}\right) + \rho^{(V_l)} E\left(Lq_l^{(V_l)}\right). \tag{8.13}$$

For the **exhaustive regime** this means (recall that $E\left(Lq_l^{(p)}\right) = 0$ for $p \neq M_{l-1}, V_l$),

$$E\left(Lq_l\right) = \rho^{(M_{l-1})} E\left(Lq_l^{(M_{l-1})}\right) + \rho^{(V_l)} E\left(Lq_l^{(V_l)}\right). \tag{8.14}$$

For **all** service regimes, the fraction of type-l customers arriving during p is $\frac{\lambda_l^{(p)} E(p)}{\sum_{p=V_1}^{M_C} \rho^{(p)} * \lambda_l^{(p)}} = \frac{\rho^{(p)} \lambda_l^{(p)}}{\overline{\lambda}_l}$. Conditioning on the system state in which a type-l customer arrives to the a system yields,

$$E\left(Wq_l\right) = \sum_{p=V_1}^{M_C} Prob\left(\begin{array}{c} arrival\ occurs \\ during\ p \end{array}\right) * E\left(Wq_l \,\middle|\, \begin{array}{c} arrival\ occurs \\ during\ p \end{array}\right).$$

Assuming **(globally) gated regime** we get,

$$\begin{aligned} E\left(Wq_l\right) = & \frac{\rho^{(V_{l-1})} * \lambda}{\overline{\lambda}_l} * \left[E\left(R_{V_{l-1}}\right) + E\left(S_{l-1}\right) + E\left(Lq_l^{(V_{l-1})}\right) * E\left(B_l\right)\right] \\ & + \frac{\rho^{(M_{l-1})} * \lambda}{\overline{\lambda}_l} * \left[E\left(R_{S_{l-1}}\right) + E\left(Lq_l^{(M_{l-1})}\right) * E\left(B_l\right)\right]. \end{aligned} \tag{8.15}$$

Assuming **exhaustive regime** results in,

$$E\left(Wq_l\right) = \frac{\rho^{(V_l)} \lambda}{\overline{\lambda}_l} * \left[E\left(R_{B_l}\right) + E\left(Lq_l^{(V_l)}\right) E\left(B_l\right)\right] + \frac{\rho^{(M_{l-1})} * \lambda}{\overline{\lambda}_l} * \left[E\left(R_{S_{l-1}}\right) + E\left(Lq_l^{(M_{l-1})}\right) * E\left(B_l\right)\right]. \tag{8.16}$$

For example, under the exhaustive regime, a customer arriving during V_l will enter service after waiting the residual time until the currently served type-l customer departs the system, plus all the service times of the type-l customers present at his arrival epoch. Hence, under the exhaustive regime,

$$E\left(Wq_l \,\middle|\, \begin{array}{c} arrival\ occurs \\ during\ V_l \end{array}\right) = E\left(R_{B_l}\right) + E\left(Lq_l^{(V_l)}\right) E\left(B_l\right).$$

8.2 Optimality criterion

Generally speaking (and under all three regimes), type-i customers from the basic model can only arrive to queues $Q_i, Q_{i+N}, \ldots, Q_{i+(C-1)N}$ in the new model. The fraction of type-i customers present in $Q_l \in \left\{Q_i, Q_{i+N}, \ldots, Q_{i+(C-1)N}\right\}$ is $\frac{E(L_l)}{\sum_{r=0}^{C-1} E(L_{i+rN})}$, where $E\left(L_l\right)$ is the expected total number of type-l customers in the system. The expected sojourn time of an arbitrary type-i customer in the basic system is,

$$E\left(W_i\right) = \frac{\sum_{r=0}^{C-1} \left[E\left(L_{i+rN}\right) * \left(E\left(Wq_{i+rN}\right) + E(B_{i+rN})\right)\right]}{\sum_{r=0}^{C-1} E\left(L_{i+rN}\right)} \ \forall i = 1, \ldots, N.$$

By little's law $E\left(L_l\right) = \overline{\lambda}_l * \left[E\left(Wq_l\right) + E(B_l)\right]$, so we can write

$$E\left(W_i\right) = \frac{\sum_{r=0}^{C-1} \left[\overline{\lambda}_{i+rN} * \left(E\left(Wq_{i+rN}\right) + E(B_{i+rN})\right)^2\right]}{\sum_{r=0}^{C-1} \left[\overline{\lambda}_{i+rN} * \left(E\left(Wq_{i+rN}\right) + E(B_{i+rN})\right)\right]} \ \forall i = 1, \ldots, N.$$

Define $E(W_l^{new}) \equiv E\left(Wq_l\right) + E(B_l)$ to be the expected sojourn time of an arbitrary type-l customer in the new model. We conclude that

$$E\left(W_i\right) = \frac{\sum_{r=0}^{C-1} \left[\overline{\lambda}_{i+rN} * E^2\left(W_{i+rN}^{new}\right)\right]}{\sum_{r=0}^{C-1} \left[\overline{\lambda}_{i+rN} * E\left(W_{i+rN}^{new}\right)\right]} \ \forall i = 1, \ldots, N. \tag{8.17}$$

In section 10 we will discuss basic models which consist of a single queue. In those cases, $E\left(W_1\right)$ will measure the system performances.

9 Stability for a single queue case

In section 3 we stated the following stability condition: The maximal π root of equation (3.2) should be negative. We now examine the general structure of this stability condition for a new model which originated from a basic model consisting of a single **exhaustively-served** queue under "swap at the end of every T sub-cycles" ($T = 1, 2, 3, \dots$) policy. Recall that $E(G)$ ($E(K)$) is the expected basic service duration of server 1 (server 2). We have:

$$\det\left(\mathbf{R} - (\pi{+}1) * \mathbf{I_{CN}}\right) =$$

$$\begin{vmatrix} \lambda\alpha^0\ E(G) - (\pi{+}1) & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & & & & \vdots \\ 0 & & \lambda\ \alpha^{T-1}E(G) - (\pi{+}1) & & & 0 \\ 0 & & & \lambda\ \alpha^0 E(K) - (\pi{+}1) & & 0 \\ \vdots & & & & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \lambda\ \alpha^{T-1}E(K) - (\pi{+}1) \end{vmatrix}.$$

So,

$$\det\left(\mathbf{R} - (\pi{+}1) * \mathbf{I_{CN}}\right) = 0 \Longrightarrow \prod_{r=0}^{T-1} \left[(\lambda\alpha^r\ E(G) - (\pi{+}1)) * (\lambda\alpha^r\ E(K) - (\pi{+}1))\right] = 0.$$

Hence, the collection of all π solutions is given by

$$\bigcup_{r=0}^{T-1} \{\{\pi{=}\lambda\alpha^r\ E(G) - 1\} \cup \{\pi{=}\lambda\alpha^r\ E(K) - 1\}\}.$$

Set $E^{max}(B) \equiv \max\left[E(G), E(K)\right]$.
Then,

$$\max_{\mathbb{R}} \left[\pi | \pi \in \bigcup_{r=0}^{T-1} \{\{\pi{=}\lambda\alpha^r\ E(G) - 1\} \cup \{\pi{=}\lambda\alpha^r\ E(K) - 1\}\}\right] < 0,$$

yields the stability condition for the exhaustive regime:

$$\alpha^{T-1}\lambda E^{max}(B) < 1. \tag{9.1}$$

Failure to meet this condition means that the server would eventually get "stuck" in some Q_l, $l = 1, 2, \dots, C$.

We define **"zero TL stability"** as the stability condition of a new model under the "always swap" policy, namely, $T = 1 \Longrightarrow \lambda E^{max}(B) < 1$. Given "zero **TL** stability", in a stable system $\alpha \in (1, UB)$, where

$$UB = \begin{cases} \infty & T = 1, \\ \frac{1}{[\lambda E^{max}(B)]^{\frac{1}{T-1}}} & T \in [2, \infty), \end{cases}$$

and T^{max}, the maximal T for which the corresponding swapping policy still produces a stable system, is

$$T^{max} = \max_{T \in \mathbb{Z}} \left[T | \, T < \frac{\ln\left(\frac{\alpha}{\lambda E^{max}(B)}\right)}{\ln(\alpha)}\right] = \left\lceil \frac{\ln\left(\frac{1}{\lambda E^{max}(B)}\right)}{\ln(\alpha)} \right\rceil.$$

We now examine the general structure of the discussed stability condition for a new model, which originated from a basic model consisting of a single **gatedly-served** queue under "swap at the end of every T sub-cycles" ($T = 1, 2, 3, \dots$) policy.

Remark 9.1. For a basic model consisting of a single queue, the gated regime and the globally-gated regime converge. In such cases we will refer to the service regime as "(globally) gated".

We have:

$$\det\left(\mathbf{R}-(\pi+1)*\mathbf{I_{CN}}\right)=$$
$$\begin{vmatrix} -(\pi+1) & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \lambda\,\alpha^{0}E(G) \\ \lambda\,\alpha^{1}E(G) & -(\pi+1) & & & & & & & 0 \\ 0 & \lambda\,\alpha^{2}E(G) & \ddots & & & & & & 0 \\ \vdots & & \ddots & -(\pi+1) & & & & & \vdots \\ 0 & & & \lambda\,\alpha^{T-1}E(G) & -(\pi+1) & & & & 0 \\ \vdots & & & & \lambda\alpha^{0}\,E\left(K\right) & \ddots & & & \vdots \\ 0 & & & & & \ddots & -(\pi+1) & & 0 \\ 0 & & & & & & \lambda\,\alpha^{T-2}E(K) & -(\pi+1) & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \lambda\,\alpha^{T-1}E(K) & -(\pi+1) \end{vmatrix}.$$

This means,

$$\begin{aligned}\det\left(\mathbf{R}-(\pi+1)*\mathbf{I_{CN}}\right)&=(\pi+1)^{2T}-\prod_{r=1}^{T}\left[\lambda\,\alpha^{r-1}E\left(G\right)*\lambda\alpha^{r-1}E\left(K\right)\right]\\&=(\pi+1)^{2T}-\left[\lambda\sqrt{E\left(G\right)E\left(K\right)}\right]^{2T}*\alpha^{T(T-1)}.\end{aligned}$$

So,

$$\begin{aligned}\det\left(\mathbf{R}-(\pi+1)*\mathbf{I_{CN}}\right)=0&\Longrightarrow(\pi+1)^{T}=\alpha^{\frac{T(T-1)}{2}}\left[\lambda\sqrt{E\left(G\right)E\left(K\right)}\right]^{T}\\&\Longrightarrow\pi=\alpha^{\frac{(T-1)}{2}}\lambda\sqrt{E\left(G\right)E\left(K\right)}-1.\end{aligned}$$

Then,

$$\max_{\mathbb{R}}\left[\pi|\,\pi=\alpha^{\frac{(T-1)}{2}}\lambda\sqrt{E\left(G\right)E\left(K\right)}-1<0\right],$$

yields the stability condition for the (globally) gated regime:

$$\alpha^{\frac{(T-1)}{2}}\lambda\sqrt{E\left(G\right)E\left(K\right)}<1. \tag{9.2}$$

Given "zero **TL** stability" (i.e. $T=1\Longrightarrow\lambda\sqrt{E\left(G\right)E\left(K\right)}<1$), in a stable system $\alpha\in(1,UB)$, where

$$UB=\begin{cases}\infty & T=1,\\ \frac{1}{\left[\lambda^{2}E(G)E(K)\right]^{\frac{1}{T-1}}} & T\in[2,\infty),\end{cases}$$

and

$$T^{max}=\max_{T\in\mathbb{Z}}\left[T|\,T<\frac{\ln\left(\frac{\alpha}{\lambda^{2}E(G)E(K)}\right)}{\ln\left(\alpha\right)}\right]=\left\lceil\frac{\ln\left(\frac{1}{\lambda^{2}E(G)E(K)}\right)}{\ln\left(\alpha\right)}\right\rceil.$$

The stability condition for the exhaustive regime (equation (9.1)), states that the highest traffic-intensity produced by a queue would not exceed 1. In the case of the (globally) gated regime (equation (9.2)), we need only demand that the (unweighed) geometric mean of all traffic-intensities produced by the queues would not exceed 1. For a given α, this generally translates to a **higher T^{max} under the gated regime than under the exhaustive regime** (note that the stability condition of the exhaustive regime is a sufficient condition for the stability of the gated regime). One way to look at it is to observe that, under the (globally) gated regime, no matter how slow the tired server is, he never gets "stuck" in a queue. After a cycle in which a server operates at his highest incurred **TL**, he is always replaced by a "fresh" server, with $\mathbf{TL}=0$.

We now compare the two stability conditions for the case of identical servers ($B \sim G \sim K$). For the **exhaustive** regime, $E^{max}(B) = E(B)$ leads to

$$\alpha^{T-1}\lambda E^{max}(B) < 1 \Longrightarrow \alpha^{T-1}\lambda E(B) < 1.$$

For the **(globally) gated** regime, $E(G)E(K) = E^2(B)$ implies

$$\alpha^{\frac{(T-1)}{2}}\lambda\sqrt{E(G)E(K)} < 1 \Longrightarrow \alpha^{\frac{(T-1)}{2}}\lambda E(B) < 1.$$

Both regimes' "zero **TL** stability" conditions converge to

$$\lambda E(B) < 1.$$

We conclude that, given "zero **TL** stability", in a stable system $\alpha \in (1, UB)$, where

$$UB = \begin{cases} \infty & T = 1, \\ \frac{1}{[\lambda E(B)]^{\frac{1}{T-1}}} & T \in [2,\infty) \\ & exhaustive, \\ \left(\frac{1}{[\lambda E(B)]^{\frac{1}{T-1}}}\right)^2 & T \in [2,\infty) \\ & (globally)\ gated, \end{cases}$$

and

$$T^{max} = \begin{cases} T_E^{max} \equiv \max_{T\in\mathbb{Z}}\left[T|\, T < \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)}\right] = \left\lceil \frac{\ln\left(\frac{1}{\lambda E(B)}\right)}{\ln(\alpha)} \right\rceil & exhaustive, \\ T_G^{max} \equiv \max_{T\in\mathbb{Z}}\left[T|\, T < \frac{\ln\left(\frac{\alpha}{[\lambda E(B)]^2}\right)}{\ln(\alpha)}\right] = \left\lceil 2 * \frac{\ln\left(\frac{1}{\lambda E(B)}\right)}{\ln(\alpha)} \right\rceil & (globally)\ gated. \end{cases} \tag{9.3}$$

We can write,

$$T_G^{max} = \begin{cases} 2T_E^{max} & 0 \le T_E^{max} - \frac{\ln\left(\frac{1}{\lambda E(B)}\right)}{\ln(\alpha)} < 0.5, \\ 2T_E^{max} - 1 & 0.5 \le T_E^{max} - \frac{\ln\left(\frac{1}{\lambda E(B)}\right)}{\ln(\alpha)} < 1. \end{cases} \tag{9.4}$$

We now interpret the results for the case of identical servers. In classical polling systems consisting of a single queue (with service durations B), the stability condition is identical under the exhaustive and (globally) gated regimes. Namely, it is the same as the "zero **TL** stability" condition, $\lambda E(B) < 1$. This holds true since $\lambda E(B)$ is the expected number of customers arriving to the queue during an expected service time of a single customer (which takes place during a certain visit period V'). Under the exhaustive regime, those arriving customers are served during the same visit period in which they arrived (V'). This requires an average of $E(B)$ units of time per customer, so $\lambda E(B) \ge 1$ means that the server will eventually get "stuck" in the queue. Under the (globally) gated regime, those arriving customers are served during the next visit period to their arrival ($V'+1$). This requires an average of $E(B)$ units of time per customer during which the number of new arrivals will be $(\lambda E(B))^2$, and so on. Thus, $\lambda E(B) \ge 1$ means that the number of customers served each visit period will tends to infinity in the long run (in a kind of "snow ball affect"). In our model the ergodic behavior of the system, under the exhaustive regime, follows the same logic. If during any visit period $\lambda E(B_l) \ge 1$ (which occurs i.f.f $\alpha^{T-1}\lambda E(B) \ge 1$), the system will explode in the long run due to the fact that each arriving customer is served during an average of $E(B_l)$ units of time. However, for the (globally) gated regime, the logic differs. This results from the fact that, during the visit period $V'+1$, the effective service duration isn't the same as in the service period V'. Each arriving customer in V' is served during an average of $E(B_{l+1})$ units of time. Since the server never gets "stuck" in

a queue, after a cycle in which he operates at his highest incurred **TL**, he is always "refreshed" (**TL** $= 0$). Thus, $T_G^{max} \geq T_E^{max}$.

Note that under an "always swap" policy, the average amount of service time per customer is identical in both queues comprising the new model ($E(B_1) = E(B_2) = E(B)$). Indeed, for $T = 1$, the stability conditions for the exhaustive regime and (globally) gated regime are identical (and equal to the "zero **TL** stability" condition $\lambda E(B) < 1$).

We remind that, as stated in remark 3.1, the stability condition for a compact model is the same as in the equivalent new model.

10 Analysis of a single queue case

10.1 The exhaustive regime, N=1

Consider a system composed of a single exhaustively-served queue and two identical servers which operates under the "swap every T sub-cycles" ($T = 1, 2, \dots$) policy. Furthermore, assume "zero **TL** stability". Consider the **compact models**, under a given $T \leq T_E^{max} = \left\lceil \frac{\ln\left(\frac{1}{\lambda E(B)}\right)}{\ln(\alpha)} \right\rceil$.

Our initial goal is to find an explicit expression for the optimality criterion, $E(W_1)$, which depends only on the initial parameters $H,\ H_0,\ B,\ \alpha,\ \lambda$ and T.
From equation (8.17), using (8.6),

$$E(W_1) = \frac{\sum_{r=1}^{T} \overline{\lambda}_r E^2(W_r^{new})}{\sum_{r=1}^{T} \overline{\lambda}_r E(W_r^{new})} = \frac{\sum_{r=1}^{T} (E(S_{r-1}) + E(V_r)) E^2(W_r^{new})}{\sum_{r=1}^{T} (E(S_{r-1}) + E(V_r)) E(W_r^{new})}. \tag{10.1}$$

From equation (8.2),

$$E(V_l) = \frac{\lambda E(B_l) E(S_{l-1})}{1 - \lambda E(B_l)}. \tag{10.2}$$

Note that since $B_l = \alpha^{l-1} B$ and $S_l = \begin{cases} H & l < T \\ H + H_0 & l = T \end{cases}$, equation (10.2) means $E(V_2) < E(V_3) < \dots < E(V_T)\ \forall T$.
Substitution of equation (10.2) into equation (10.1) yield,

$$E(W_1) = \frac{\sum_{r=1}^{T} \frac{E(S_{r-1})}{1-\lambda E(B_r)} [E(Wq_r) + E(B_r)]^2}{\sum_{r=1}^{T} \frac{E(S_{r-1})}{1-\lambda E(B_r)} [E(Wq_r) + E(B_r)]}. \tag{10.3}$$

Combining equations (8.8), (8.11), (8.14) and (8.16) yields,

$$E(Wq_l) = \frac{E\left({S_{l-1}}^2\right)}{2E(S_{l-1})} + \frac{\lambda E\left({B_l}^2\right)}{2(1 - \lambda E(B_l))} = E\left(R_{S_{l-1}}\right) + E\left(Wq_{M(\lambda)/G(B_l)/1}\right). \tag{10.4}$$

Equation (10.4) reflects a decomposition property which exists due to the absence of correlation between the different visit periods and the properties of the Poisson arrival rates. Equation (10.4) and the above explanation also holds for the case of not necessarily identical servers. Note that we can view the current model (with identical servers) as an M/G/1 system with multiple vacations of duration H, where the server's **TL** increases after each vacation. From this viewing point, each T-th vacation is a special extended vacation (lasting an additional H_0 units of time) from which the server returns at full strength (i.e. **TL** $= 0$).
Substitution of $B_l = \alpha^{l-1} B$ and $S_l = \begin{cases} H & l < T \\ H + H_0 & l = T \end{cases}$ into equation (10.4) yields,

$$E(Wq_l) = \begin{cases} \frac{E(H^2) + E({H_0}^2) + 2E(H)E(H_0)}{2(E(H) + E(H_0))} + \frac{\lambda E(B^2)}{2(1-\lambda E(B))} & l = 1, \\ \frac{E(H^2)}{2E(H)} + \frac{\lambda \alpha^{2(l-1)} E(B^2)}{2(1-\lambda \alpha^{l-1} E(B))} & l > 1. \end{cases} \tag{10.5}$$

Note that $E(Wq_1)$ is unaffected by α. Also note that for $T=1$,

$$E(W_1)=E(W_1^{new})=E(Wq_1)+E(B)=\frac{E(H^2)+E(H_0{}^2)+2E(H)E(H_0)}{2(E(H)+E(H_0))}+\frac{\lambda E(B^2)}{2(1-\lambda E(B))}+E(B).$$

In accordance with section 2.3, the last equation is equivalent to the mean sojourn time in an M/G/1 system with multiple server vacations of duration $H+H_0$.
Substitution of equations (10.4) and (10.5) into equation (10.3) yields,

$$E(W_1)=\frac{\left\{\begin{array}{l}\frac{E(H)+E(H_0)}{1-\lambda E(B)}*\left[\frac{E(H^2)+E(H_0{}^2)+2E(H)E(H_0)}{2(E(H)+E(H_0))}+\frac{\lambda E(B^2)}{2(1-\lambda E(B))}+E(B)\right]^2\\ +\sum_{r=2}^{T}\ \frac{E(H)}{1-\lambda\alpha^{r-1}E(B)}*\left[\frac{E(H^2)}{2E(H)}+\frac{\lambda\alpha^{2(r-1)}E(B^2)}{2(1-\lambda\alpha^{r-1}E(B))}+\alpha^{r-1}E(B)\right]^2\end{array}\right\}}{\left\{\begin{array}{l}\frac{E(H)+E(H_0)}{1-\lambda E(B)}*\left[\frac{E(H^2)+E(H_0{}^2)+2E(H)E(H_0)}{2(E(H)+E(H_0))}+\frac{\lambda E(B^2)}{2(1-\lambda E(B))}+E(B)\right]\\ +\sum_{r=2}^{T}\ \frac{E(H)}{1-\lambda\alpha^{r-1}E(B)}*\left[\frac{E(H^2)}{2E(H)}+\frac{\lambda\alpha^{2(r-1)}E(B^2)}{2(1-\lambda\alpha^{r-1}E(B))}+\alpha^{r-1}E(B)\right]\end{array}\right\}}. \tag{10.6}$$

For the sake of completeness, we note that equation (8.4) leads to

$$E(\mathbf{C})=\sum_{k=1}^{T}\frac{E(H)}{1-\lambda\alpha^{k-1}E(B)}+\frac{E(H_0)}{1-\lambda E(B)}, \tag{10.7}$$

and equation (8.6) leads to

$$\overline{\lambda}_l=\left\{\begin{array}{ll}\frac{\lambda}{E(\mathbf{C})}\left(\frac{E(H)+E(H_0)}{1-\lambda E(B)}\right) & l=1,\\ \\ \frac{\lambda}{E(\mathbf{C})}\left(\frac{E(H)}{1-\lambda\alpha^{l-1}E(B)}\right) & 1<l\leq T.\end{array}\right. \tag{10.8}$$

Remark 10.1. The above results can be extended for the more general case of not necessarily identical servers in a strait forward way (assuming $\lambda E^{max}(B)<1$, and $T\leq T^{max}=\left\lceil\frac{\ln\left(\frac{1}{\lambda E^{max}(B)}\right)}{\ln(\alpha)}\right\rceil$). This is done by simply recalling that the aforementioned case is composed of two "compact model cycles", whose only difference lies in $B_l=\left\{\begin{array}{ll}G_{i=l} & 1\leq l\leq T\\ K_{i=l-T} & 1+T\leq l\leq 2T\end{array}\right.$.
The results are:

$$E(W_1)=\frac{\sum_{r=1}^{T}\left[\frac{E(S_{r-1})}{1-\lambda E(B_r)}\left[E(Wq_r)+E(B_r)\right]^2+\frac{E(S_{r+T-1})}{1-\lambda E(B_{r+T})}\left[E\left(Wq_{r+T}\right)+E(B_{r+T})\right]^2\right]}{\sum_{r=1}^{T}\left[\frac{E(S_{r-1})}{1-\lambda E(B_r)}\left[E(Wq_r)+E(B_r)\right]+\frac{E(S_{r+T-1})}{1-\lambda E(B_{r+T})}\left[E\left(Wq_{r+T}\right)+E(B_{r+T})\right]\right]}$$

$$=\frac{\sum_{r=1}^{T}E(S_{r-1})\left[\begin{array}{c}\frac{1}{1-\lambda\alpha^{r-1}E(G)}\left[E(Wq_r)+\alpha^{r-1}E(G)\right]^2\\ +\frac{1}{1-\lambda\alpha^{r-1}E(K)}\left[E\left(Wq_{r+T}\right)+\alpha^{r-1}E(K)\right]^2\end{array}\right]}{\sum_{r=1}^{T}E(S_{r-1})\left[\begin{array}{c}\frac{1}{1-\lambda\alpha^{r-1}E(G)}\left[E(Wq_r)+\alpha^{r-1}E(G)\right]\\ +\frac{1}{1-\lambda\alpha^{r-1}E(K)}\left[E\left(Wq_{r+T}\right)+\alpha^{r-1}E(K)\right]\end{array}\right]},$$

where

$$E(S_l)=E(S_{l+T})=\left\{\begin{array}{ll}E(H)+E(H_0) & l=T,\\ E(H) & 1\leq l<T,\end{array}\right.$$

and

$$E\left(Wq_l\right)=\begin{cases}\frac{E\left(H^2\right)+E\left({H_0}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))}+\frac{\lambda E\left(G^2\right)}{2(1-\lambda E(G))} & l=1,\\ \frac{E\left(H^2\right)}{2E(H)}+\frac{\lambda\alpha^{2(l-1)}E\left(G^2\right)}{2\left(1-\lambda\alpha^{l-1}E(G)\right)} & 1<l\leq T,\\ \frac{E\left(H^2\right)+E\left({H_0}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))}+\frac{\lambda E\left(K^2\right)}{2(1-\lambda E(K))} & l=T+1,\\ \frac{E\left(H^2\right)}{2E(H)}+\frac{\lambda\alpha^{2(l-T-1)}E\left(K^2\right)}{2\left(1-\lambda\alpha^{l-T-1}E(K)\right)} & T+1<l\leq 2T.\end{cases}$$

In addition,

$$E\left(\mathbf{C}\right)=E\left(H\right)*\sum_{r=1}^{T}\left(\frac{1}{1-\lambda\alpha^{r-1}E\left(G\right)}+\frac{1}{1-\lambda\alpha^{r-1}E\left(K\right)}\right)+E\left(H_0\right)*\left(\frac{1}{1-\lambda E\left(G\right)}+\frac{1}{1-\lambda E\left(K\right)}\right),$$

and

$$\overline{\lambda}_l=\begin{cases}\frac{\lambda}{E(\mathbf{C})}\left(\frac{E(H)+E(H_0)}{1-\lambda E(G)}\right) & l=1,\\ \frac{\lambda}{E(\mathbf{C})}\left(\frac{E(H)}{1-\lambda\alpha^{l-1}E(G)}\right) & 1<l\leq T,\\ \frac{\lambda}{E(\mathbf{C})}\left(\frac{E(H)+E(H_0)}{1-\lambda E(K)}\right) & l=T+1,\\ \frac{\lambda}{E(\mathbf{C})}\left(\frac{E(H)}{1-\lambda\alpha^{l-1}E(K)}\right) & T+1<l\leq 2T.\end{cases}$$

We now concentrate our efforts on finding an efficient algorithm to obtain an optimal swapping policy for the case of identical servers. Define,

$$\xi_r\equiv\frac{E\left(S_{r-1}\right)}{1-\lambda E\left(B_r\right)}=\begin{cases}\frac{E(H)+E(H_0)}{1-\lambda E(B)} & r=1,\\ \frac{E(H)}{1-\lambda\alpha^{r-1}E(B)} & r>1.\end{cases}$$

So, we can rephrase equation (10.3) as

$$E\left(W_1\right)=\frac{\sum_{r=1}^{T}\xi_r*E^2(W_r^{new})}{\sum_{r=1}^{T}\xi_r*E(W_r^{new})}.$$

The following observation is crucial: Since $E\left(W_r^{new}\right)=E\left(Wq_r\right)+\alpha^{r-1}E\left(B\right)$, equation (10.5) implies that $E(W_r^{new})$ is positive and unaffected by $T\ \forall r=1,2,\ldots,T$. Clearly, the same holds for $\xi_r\ \forall r=1,2,\ldots,T$. Let $E_T\left(W_1\right)$ be $E\left(W_1\right)$ under a given $T=1,2,\ldots,T_E^{max}$, and let

$$T^{opt}\equiv\left\{T\middle|\,E_T\left(W_1\right)\leq E_l\left(W_1\right)\ \forall l=1,2,\ldots,T_E^{max}\right\}.$$

In the following, we will make use of the integer numbers T^1 and T^2, which we assume satisfy $1\leq T^1<T^2\leq T_E^{max}$.
If $E_{T^2}\left(W_1\right)\geq E_{T^1}\left(W_1\right)$ we can write

$$\frac{\sum_{r=1}^{T^2}\xi_r*E^2\left(W_r^{new}\right)}{\sum_{r=1}^{T^2}\xi_r*E\left(W_r^{new}\right)}\geq\frac{\sum_{r=1}^{T^1}\xi_r*E^2\left(W_r^{new}\right)}{\sum_{r=1}^{T^1}\xi_r*E\left(W_r^{new}\right)}$$

$$\Longrightarrow\sum_{k=T^1+1}^{T^2}\left[\begin{array}{c}\xi_k*E\left(W_k^{new}\right)\\ *\left(\sum_{r=1}^{T^1}\xi_r*E\left(W_r^{new}\right)E\left(W_k^{new}\right)\right)\end{array}\right]\geq\sum_{k=T^1+1}^{T^2}\left[\begin{array}{c}\xi_k*E\left(W_k^{new}\right)\\ *\left(\sum_{r=1}^{T^1}\xi_r*E\left(W_r^{new}\right)E\left(W_r^{new}\right)\right)\end{array}\right].$$

Note that the only difference between the two sides of the last inequality lies in $E\left(W_k^{new}\right)$ versus $E\left(W_r^{new}\right)$. Now, $\alpha > 1$ means $E\left(W_l^{new}\right) < E\left(W_{l+1}^{new}\right)\ \forall l = 2, 3, \ldots,\ T_E^{max} - 1$. That is,

$$\frac{E\left(H^2\right)}{2E\left(H\right)} + \frac{\lambda\alpha^{2(l-1)}E\left(B^2\right)}{2\left(1-\lambda\alpha^{l-1}E\left(B\right)\right)} + \alpha^{l-1}E\left(B\right) < \frac{E\left(H^2\right)}{2E\left(H\right)} + \frac{\lambda\alpha^{2l}E\left(B^2\right)}{2\left(1-\lambda\alpha^{l}E\left(B\right)\right)} + \alpha^{l}E\left(B\right).$$

In other words, $E\left(W_2^{new}\right) < E\left(W_3^{new}\right) < \cdots < E\left(W_{T_E^{max}}^{new}\right)$. So $E\left(W_1^{new}\right) \leq E\left(W_{T^1+1}^{new}\right)$ means $E\left(W_1^{new}\right) < E\left(W_l^{new}\right)\ \forall l = T^1+2, T^1+3, \ldots, T_E^{max}$. Moreover:

- $E\left(W_1^{new}\right) < E\left(W_2^{new}\right) \Longrightarrow E_1\left(W_1\right) < E_l\left(W_1\right)\ \forall l = 2, 3, \ldots., T_E^{max} \Longrightarrow T^{opt} = \{1\}$.
- $E\left(W_1^{new}\right) = E\left(W_2^{new}\right) \Longrightarrow E_1\left(W_1\right) = E_2\left(W_1\right) < E_l\left(W_1\right)\ \forall l = 3, 4, \ldots., T_E^{max} \Longrightarrow T^{opt} = \{1, 2\}$.
- And $\begin{cases} E\left(W_2^{new}\right) < E\left(W_1^{new}\right) \Longrightarrow E_1\left(W_1\right) > E_2\left(W_1\right) \Longrightarrow \{1\} \notin T^{opt}, \\ \left(W_{T'-1}^{new}\right) < E\left(W_1^{new}\right) \leq E\left(W_{T'}^{new}\right) \text{ for some } T' = 3, 2, \ldots., T_E^{max} \\ \qquad \Longrightarrow E_{T^l}\left(W_1\right) < E_{T^{l+1}}\left(W_1\right)\ \forall l = T'-1, T', \ldots, T_E^{max} - 1 \\ \qquad \Longrightarrow \{T', T'+1, \ldots, T_E^{max}\} \notin T^{opt}. \end{cases}$

  This means

$$\begin{cases} E\left(W_2^{new}\right) < E\left(W_1^{new}\right) \text{ for } T_E^{max} = 2 \Longrightarrow T^{opt} = \{2\}, \\ E\left(W_2^{new}\right) < E\left(W_1^{new}\right) \leq E\left(W_3^{new}\right) \text{ for } T_E^{max} \geq 3 \Longrightarrow T^{opt} = \{2\}, \\ E\left(W_{T'-1}^{new}\right) < E\left(W_1^{new}\right) \leq E\left(W_{T'}^{new}\right) \text{ for some } T' = 4, 5, \ldots., T_E^{max} \\ \qquad \Longrightarrow \{\{1\} \cup \{T', T'+1, \ldots, T_E^{max}\}\} \notin T^{opt}, \\ E\left(W_1^{new}\right) > E\left(W_{T_E^{max}}^{new}\right) \Longrightarrow \{1\} \notin T^{opt}. \end{cases}$$

Remark 10.2. $\left(W_{T''}^{new}\right) < E\left(W_1^{new}\right)$ means that $E_l\left(W_1\right) < E_1\left(W_1\right)\ \forall\ l = 2, 3, \ldots, T''$. So $\left(W_{T'-1}^{new}\right) < E\left(W_1^{new}\right) \leq E\left(W_{T'}^{new}\right)$ for some $T' = 3, 4, \ldots., T_E^{max}$ means $E_l\left(W_1\right) < E_1\left(W_1\right)\ \forall\ l = 2, 3, \ldots, T'-1$.

If $E_{T^2}\left(W_1\right) \leq E_{T^1}\left(W_1\right)$ we can write

$$\sum_{k=T^1+1}^{T^2} \left[\begin{array}{c} \xi_k * E\left(W_k^{new}\right) \\ * \left(\sum_{r=1}^{T^1} \xi_r * E\left(W_r^{new}\right) E\left(W_k^{new}\right)\right) \end{array} \right] \leq \sum_{k=T^1+1}^{T^2} \left[\begin{array}{c} \xi_k * E\left(W_k^{new}\right) \\ * \left(\sum_{r=1}^{T^1} \xi_r * E\left(W_r^{new}\right) E\left(W_r^{new}\right)\right) \end{array} \right].$$

Remark 10.3. Observe that a necessary (but not sufficient) condition for $E_{T^2}\left(W_1\right) \leq E_{T^1}\left(W_1\right)$, is $E\left(W_{T^1+1}^{new}\right) < E\left(W_1^{new}\right)$. Note that, assuming $E\left(W_{T^2}^{new}\right) < E\left(W_1^{new}\right)$, any T^1 and T^2 satisfy this necessary condition.

Assume $E_{T'}\left(W_1\right) \leq E_{T'-1}\left(W_1\right)$, this implies that

$$\sum_{r=1}^{T'-1} \xi_r * E\left(W_r^{new}\right) E\left(W_{T'}^{new}\right) \leq \sum_{r=1}^{T'-1} \xi_r * E\left(W_r^{new}\right) E\left(W_r^{new}\right). \tag{10.9}$$

Consider $E_{T'-1}\left(W_1\right) < E_{T'-2}\left(W_1\right)$ where $T'-2 \geq 1$. This holds true i.f.f.,

$$\sum_{r=1}^{T'-2} \xi_r * E\left(W_r^{new}\right) E\left(W_{T'-1}^{new}\right) < \sum_{r=1}^{T'-2} \xi_r * E\left(W_r^{new}\right) E\left(W_r^{new}\right).$$

Since $E\left(W_{T'}^{new}\right) > E\left(W_{T'-1}^{new}\right)$, we can deduce from equation (10.9) that

$$\sum_{r=1}^{T'-2} \xi_r * E\left(W_r^{new}\right) E\left(W_{T'-1}^{new}\right) < \sum_{r=1}^{T'-2} \xi_r * E\left(W_r^{new}\right) E\left(W_{T'}^{new}\right) < \sum_{r=1}^{T'-2} \xi_r * E\left(W_r^{new}\right) E\left(W_r^{new}\right),$$

where the rightmost and leftmost elements imply that $E_{T'-1}(W_1) < E_{T'-2}(W_1)$. So $E_{T'}(W_1) \leq E_{T'-1}(W_1) \Longrightarrow E_{T'-1}(W_1) < E_{T'-2}(W_1)\ \forall T' = 3,4,\ldots,T_E^{max}$. This means that $E_{T'}(W_1) \leq E_{T'-1}(W_1)$ leads to $E_{T'-1}(W_1) < E_{T'-2}(W_1) < \cdots < E_2(W_1) < E_1(W_1)$.

To conclude, assuming "zero **TL** stability", the system has the following general form:

$$\begin{cases} E_1(W_1) > E_2(W_1) > \cdots > E_{T'-2}(W_1) > E_{T'-1}(W_1) \geq E_{T'}(W_1), \\ E_{T'}(W_1) < E_{T'+1}(W_1) < \cdots < E_{T_E^{max}}(W_1) < E_{T_E^{max}}(W_1). \end{cases}$$

Where $E\left(W_{l-1}^{new}\right) < E\left(W_1^{new}\right) \leq E\left(W_l^{new}\right)$ for some $l \geq T'+1$, and

$$T^{opt} = \begin{cases} \{T'-1, T'\} & E_{T'-1}(W_1) = E_{T'}(W_1), \\ \{T'\} & E_{T'-1}(W_1) > E_{T'}(W_1). \end{cases}$$

Remark 10.4. While moving from $l=1$ to $l = T_E^{max}$, $E_l(W_1)$ decreases down to $E_{T^{opt}}(W_1)$ and then increases. By using equation (10.6) to check whether an $E_l(W_1)$ is in the increasing part or in the decreasing part, one can use a binary search method in order to find T^{opt} (which is non-empty and contains at most two consecutive elements). Knowing where to locate $E\left(W_1^{new}\right)$ on the $E\left(W_l^{new}\right)$ axis would narrow down the search.

Now, for $1 < T \leq T_E^{max}$, the inequality $E\left(W_1^{new}\right) \leq E\left(W_T^{new}\right)$ means

$$\frac{E\left(H^2\right) + E\left({H_0}^2\right) + 2E(H)E(H_0)}{2\left(E(H)+E(H_0)\right)} + \frac{\lambda E\left(B^2\right)}{2\left(1-\lambda E(B)\right)} + E(B) \leq \frac{E\left(H^2\right)}{2E(H)} + \frac{\lambda\alpha^{2(T-1)}E\left(B^2\right)}{2\left(1-\lambda\alpha^{T-1}E(B)\right)} + \alpha^{T-1}E(B)$$

$$\Longrightarrow \frac{E\left({H_0}^2\right) - E(H_0)\left[\frac{Var(H)}{E(H)} - E(H)\right]}{E(H)+E(H_0)} \leq \frac{\left\{\begin{array}{c} 2E(B)\left(\alpha^{T-1}-1\right) + \lambda\left(Var(B) - E^2(B)\right) \\ * \left[\alpha^{2(T-1)}\left(1-\lambda E(B)\right) - \left(1-\lambda\alpha^{T-1}E(B)\right)\right] \end{array}\right\}}{\left(1-\lambda\alpha^{T-1}E(B)\right)\left(1-\lambda E(B)\right)}.$$

Assume $B{\sim}exp(\cdot)$. This means $Var(B) = E^2(B)$, so the last inequality simplifies to

$$\frac{E\left({H_0}^2\right) - E(H_0)\left[\frac{Var(H)}{E(H)} - E(H)\right]}{E(H)+E(H_0)} \leq \frac{2E(B)\left(\alpha^{T-1}-1\right)}{\left(1-\lambda\alpha^{T-1}E(B)\right)\left(1-\lambda E(B)\right)}. \tag{10.10}$$

For any given $B{\sim}exp(\cdot)$, λ and α, $E\left({H_0}^2\right) - E(H_0)\left[\frac{Var(H)}{E(H)} - E(H)\right] \leq 0 \Longrightarrow \frac{E\left({H_0}^2\right)}{E(H_0)} \leq \frac{Var(H)}{E(H)} - E(H)$ would mean that $E\left(W_1^{new}\right) < E\left(W_2^{new}\right)$. So this is a sufficient (but not necessary) condition for $T^{opt} = 1$ (note that this sufficient condition never holds if $H{\sim}exp(\cdot)$).

In case $\frac{E\left({H_0}^2\right)}{E(H_0)} > \frac{Var(H)}{E(H)} - E(H)$, isolating T from equation (10.10) leads to

$$T \geq \frac{\ln\left\{\alpha\left[\frac{2E(B)(E(H)+E(H_0))+(1-\lambda E(B))\left[E\left({H_0}^2\right)-E(H_0)\left(\frac{Var(H)}{E(H)}-E(H)\right)\right]}{2E(B)(E(H)+E(H_0))+\lambda E(B)(1-\lambda E(B))\left[E\left({H_0}^2\right)-E(H_0)\left(\frac{Var(H)}{E(H)}-E(H)\right)\right]}\right]\right\}}{\ln(\alpha)} \equiv TBOUND.$$

Note that since $\lambda E(B) < 1$ and $\alpha > 1$, $TBOUND > 1$. Also note that $TBOUND$ monotonically decreases in $\alpha > 1$.

Clearly, we are interested in $\lceil TBOUND \rceil$. **We claim:** $\lceil TBOUND \rceil \leq T_E^{max} + 1$.

Proof: We first show that $TBOUND < \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)}$, for otherwise

$$\frac{\ln\left\{\alpha\left[\frac{2E(B)(E(H)+E(H_0))+(1-\lambda E(B))\left[E\left({H_0}^2\right)-E(H_0)\left(\frac{Var(H)}{E(H)}-E(H)\right)\right]}{2E(B)(E(H)+E(H_0))+\lambda E(B)(1-\lambda E(B))\left[E\left({H_0}^2\right)-E(H_0)\left(\frac{Var(H)}{E(H)}-E(H)\right)\right]}\right]\right\}}{\ln(\alpha)} \geq \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)}$$

$$\Longrightarrow \begin{bmatrix} \lambda E(B)\, 2E(B)\,(E(H)+E(H_0)) \\ +\lambda E(B)\,(1-\lambda E(B)) \\ *\left(E\left({H_0}^2\right)-E(H_0)\left(\frac{Var(H)}{E(H)}-E(H)\right)\right) \end{bmatrix} \geq \begin{bmatrix} 2E(B)\,(E(H)+E(H_0)) \\ +\lambda E(B)\,(1-\lambda E(B)) \\ *\left(E\left({H_0}^2\right)-E(H_0)\left(\frac{Var(H)}{E(H)}-E(H)\right)\right) \end{bmatrix}$$

$$\Longrightarrow \lambda E(B) \geq 1.$$

This defies the "zero **TL** stability" condition.

Next, recall that $T_E^{max} = \max_{T\in\mathbb{Z}} \left[T \middle| T < \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)} \right]$ (see equation (9.3)). Now,

- If $\frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)} \in \mathbb{Z}$ then $\lceil TBOUND \rceil \leq \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)}$ and $T_E^{max} = \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)} - 1$

  so $\lceil TBOUND \rceil \leq T_E^{max} + 1$.

- If $\frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)} \notin \mathbb{Z}$ then $\lceil TBOUND \rceil \leq \left\lceil \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)} \right\rceil$ and $T_E^{max} = \left\lceil \frac{\ln\left(\frac{\alpha}{\lambda E(B)}\right)}{\ln(\alpha)} \right\rceil - 1$

  so, again, $\lceil TBOUND \rceil \leq T_E^{max} + 1$.

This completes the proof.

To summarize, assuming "zero **TL** stability", $B{\sim}exp(\cdot)$ and $\frac{E\left({H_0}^2\right)}{E(H_0)} > \frac{Var(H)}{E(H)} - E(H)$:

$$\begin{cases} 2 \leq \lceil TBOUND \rceil \leq T_E^{max} + 1, \\ \lceil TBOUND \rceil \leq T_E^{max} \Longleftrightarrow E\left(W^{new}_{\lceil TBOUND \rceil - 1}\right) < E\left(W_1^{new}\right) \leq E\left(W^{new}_{\lceil TBOUND \rceil}\right), \\ \lceil TBOUND \rceil = T_E^{max} + 1 \Longleftrightarrow E\left(W^{new}_{T_E^{max}}\right) < E\left(W_1^{new}\right). \end{cases}$$

To conclude, assuming $B{\sim}exp(\cdot)$, an $o(\log(T^{max}))$ algorithm for finding T^{opt} is:

Setp1: IF $\lambda E(B) \geq 1$ THEN $T^{opt} = \emptyset$ ELSE

Setp2: IF $T_E^{max} = 1$ THEN $T^{opt} = \{1\}$ ELSE

Setp3: IF $\frac{E\left({H_0}^2\right)}{E(H_0)} \leq \frac{Var(H)}{E(H)} - E(H)$ THEN $T^{opt} = \{1\}$ ELSE

Setp4: IF $TBOUND = 2$ THEN $T^{opt} = \{1, 2\}$ ELSE

Setp5: IF $\lceil TBOUND \rceil = 2$ THEN $T^{opt} = \{1\}$ ELSE

Setp6: IF $\lceil TBOUND \rceil = 3$ THEN $T^{opt} = \{2\}$ ELSE

Setp7: Use a binary search method in order to find all elements of
$T^{opt} = \text{argmin}\{ E_T(W_1) | T = 2, 3, \ldots, \lceil TBOUND \rceil - 1 \}$ (cf. remark 10.4).

Remark 10.5. Steps 4-6 follow the same logic as step 7. They deal with simpler cases which do not require farther calculations.

10.2 The (globally) gated regime with identical servers, N=1

Consider a system composed of a single queue, operating under the (globally) gated regime, with two identical servers switching according to the "swap at the end of every T sub-cycles" ($T = 1, 2, \ldots$) policy. Furthermore, assume "zero **TL** stability". Consider the **compact models**, under a given $2 \leq T \leq T_G^{max} = \left\lceil 2 * \frac{\ln\left(\frac{1}{\lambda E(B)}\right)}{\ln(\alpha)} \right\rceil$ (we will later relax the $T \neq 1$ assumption).

We begin by finding expressions for $E(V_l)$. Equation (8.1) states that

$$E(V_l) = \lambda E(B_l)\left[E(V_{l-1}) + E(S_{l-1})\right] \forall l = 1, \ldots, T. \tag{10.11}$$

Recursive substitution in equation (10.11) yields,

$$E(V_1) = \frac{\sum_{k=1}^{T} E(S_k)\,\lambda E(B_1)\prod_{r=k+1}^{T}\lambda E(B_r)}{1-\prod_{r=1}^{T}\lambda E(B_r)}. \tag{10.12}$$

Since $B_l = \alpha^{l-1}B$ and $E(S_l) = \begin{cases} E(H)+E(H_0) & l=T \\ E(H) & 1\le l<T \end{cases}$, equation (10.12) can be rewritten as

$$E(V_1) = \frac{E(H)\sum_{k=1}^{T}(\lambda E(B))^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}} + E(H_0)\,\lambda E(B)}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}}.$$

For the same reason, equation (10.11) also implies

$$E(V_l) = \lambda\alpha^{l-1}E(B)\left[E(V_{l-1}) + E(H)\right] \ \forall l = 2, 3, \ldots, T.$$

Recursive substitution in the last expression yields

$$E(V_l) = E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} + E(V_1)(\lambda E(B))^{l-1}\alpha^{\frac{(l-1)l}{2}} \ \forall l = 2, 3, \ldots, T. \tag{10.13}$$

Substitution of $E(V_1)$, from equation (10.12), into equation (10.13) yields

$$\begin{aligned} E(V_l) = \quad & E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} \\ & + \frac{E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} + E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}}}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}} \ \forall l = 2, 3, \ldots, T. \end{aligned} \tag{10.14}$$

Remark 10.6. To avoid confusion regarding the use of "modulo T" marks in the context of recursive substitution, we have found an expression for $E(V_1)$ (which is the only case where the modulo context differs from the non-modulo context) and then used it to express all other $E(V_l)$s. This created a de-facto distinction between cases which are not essentially different ($l = 1$ versus $l = 2, 3, \ldots, T$). This is made evident by setting $l = 1$ in equation (10.13), which yields the identity $E(V_1) \equiv E(V_1)$. This is also true regarding the five cases which are used to express $E(Wq_l)$ in the following paragraph.

We now find expressions for $E(Wq_l)$. Tediously combining equations (8.8), (8.9), (8.10), (8.12), (8.13) and (8.15) yields the following five cases:

1. For $l = 1$ and $T = 2$,

$$\begin{aligned} E(Wq_1) = \quad & \frac{E(H)+E(H_0)}{E(V_2)+E(H)+E(H_0)} * \left[\begin{array}{c} \left(\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))}\right) \\ *(1+\lambda E(B))+\lambda E(B)\,E(V_2) \end{array}\right] \\ & + \frac{E(V_2)}{E(V_2)+E(H)+E(H_0)} * \left[\frac{E\left(B^2\right)\alpha}{2E(B)}(1+\lambda E(B)) + E(H) + E(H_0)\right] \\ & + \frac{(\lambda E(B)\,\alpha)^2(1+\lambda E(B))}{(E(V_2)+E(H)+E(H_0))\left(1-(\lambda E(B))^4\alpha^2\right)} \\ & * \left\{\begin{array}{c} E(V_1)\left(E(H)+\frac{E\left(B^2\right)}{2E(B)}\right)+\frac{E\left(H^2\right)}{2} \\ +(\lambda E(B))^2\left[\begin{array}{c} E(V_2)\left(E(H)+E(H_0)+\frac{E\left(B^2\right)\alpha}{2E(B)}\right) \\ +\left(\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2}\right) \end{array}\right] \end{array}\right\}. \end{aligned}$$

2. For $l = 1$ and $T \geq 3$ we have

$$
\begin{aligned}
E\left(Wq_1\right) = \quad & \frac{E\left(H\right)+E\left(H_0\right)}{E\left(V_T\right)+E\left(H\right)+E\left(H_0\right)} * \left[\begin{array}{c} \left(\frac{E\left(H^2\right)+E\left({H_0}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))} \right) \\ *\left(1+\lambda E\left(B\right)\right)+\lambda E\left(B\right)E\left(V_T\right) \end{array} \right] \\
& + \frac{E\left(V_T\right)}{E\left(V_T\right)+E\left(H\right)+E\left(H_0\right)} * \left[\frac{E\left(B^2\right)\alpha^{T-1}}{2E\left(B\right)}\left(1+\lambda E\left(B\right)\right)+E\left(H\right)+E(H_0) \right] \\
& + \frac{\left\{ \begin{array}{c} \sum_{k=1}^{T-1} \left\{ * \left[E\left(V_k\right)\left(E\left(H\right)+\frac{E\left(B^2\right)\alpha^{k-1}}{2E(B)} \right)+\frac{E\left(H^2\right)}{2} \right] \right\} \left(\frac{(\lambda E(B))^{2(2T-k)}(1+\lambda E(B))\alpha^{2T(T-1)-k(k-1)}}{E(V_T)+E(H)+E(H_0)} \right) \\ +\frac{(\lambda E(B))^{2T}(1+\lambda E(B))\alpha^{T(T-1)}}{E(V_T)+E(H)+E(H_0)} * \left[\begin{array}{c} E\left(V_T\right)\left(E\left(H\right)+E\left(H_0\right)+\frac{E\left(B^2\right)\alpha^{T-1}}{2E(B)} \right) \\ +\left(\frac{E\left(H^2\right)+E\left({H_0}^2\right)+2E(H)E(H_0)}{2} \right) \end{array} \right] \end{array} \right\}}{1-\left(\lambda E\left(B\right)\right)^{2T}\alpha^{T(T-1)}} \\
& + \sum_{k=2}^{T} \left\{ * \left[E\left(V_{k-1}\right)\left(E\left(H\right)+\frac{E\left(B^2\right)\alpha^{k-2}}{2E(B)} \right)+\frac{E\left(H^2\right)}{2} \right] \left(\frac{(\lambda E(B))^{2(T-k+1)}(1+\lambda E(B))\alpha^{T(T-1)-k(k-3)-2}}{E(V_T)+E(H)+E(H_0)} \right) \right\}.
\end{aligned}
$$

3. For $l = 2$ we have

$$
\begin{aligned}
E\left(Wq_2\right) = \quad & \frac{E(H)}{E\left(V_1\right)+E\left(H\right)} * \left[\frac{E\left(H^2\right)}{2E\left(H\right)}\left(1+\lambda E\left(B\right)\alpha\right)+\lambda E\left(B\right)\alpha E\left(V_1\right) \right] \\
& + \frac{E\left(V_1\right)}{E\left(V_1\right)+E\left(H\right)} * \left[\frac{E\left(B^2\right)}{2E\left(B\right)}\left(1+\lambda E\left(B\right)\alpha\right)+E\left(H\right) \right] \\
& + \frac{\left\{ \begin{array}{c} \sum_{k=1}^{T-1} \left\{ * \left[E\left(V_k\right)\left(E\left(H\right)+\frac{E\left(B^2\right)\alpha^{k-1}}{2E(B)} \right)+\frac{E\left(H^2\right)}{2} \right] \right\} \left(\frac{(\lambda E(B))^{2(T-k+1)}(1+\lambda E(B)\alpha)\alpha^{T(T-1)-k(k-1)}}{E(V_1)+E(H)} \right) \\ +\frac{(\lambda E(B))^{2}(1+\lambda E(B)\alpha)}{E(V_1)+E(H)} * \left[\begin{array}{c} E\left(V_T\right)\left(E\left(H\right)+E\left(H_0\right)+\frac{E\left(B^2\right)\alpha^{T-1}}{2E(B)} \right) \\ +\left(\frac{E\left(H^2\right)+E\left({H_0}^2\right)+2E(H)E(H_0)}{2} \right) \end{array} \right] \end{array} \right\}}{1-\left(\lambda E\left(B\right)\right)^{2T}\alpha^{T(T-1)}}.
\end{aligned}
$$

4. For $l = 3 \leq T$ we have

$$
\begin{aligned}
E\left(Wq_3\right) = \quad & \frac{E(H)}{E\left(V_2\right)+E\left(H\right)} * \left[\frac{E\left(H^2\right)}{2E\left(H\right)}\left(1+\lambda E\left(B\right)\alpha^2\right)+\lambda E\left(B\right)\alpha^2 E\left(V_2\right)\right] \\
& + \frac{E\left(V_2\right)}{E\left(V_2\right)+E\left(H\right)} * \left[\frac{E\left(B^2\right)\alpha}{2E\left(B\right)}\left(1+\lambda E\left(B\right)\alpha^2\right)+E\left(H\right)\right] \\
& + \frac{\left(\lambda E\left(B\right)\alpha\right)^2\left(1+\lambda E\left(B\right)\alpha^2\right)}{E\left(V_2\right)+E\left(H\right)} * \left[\ E\left(V_1\right)\left(E\left(H\right)+\frac{E\left(B^2\right)}{2E(B)}\right)+\frac{E\left(H^2\right)}{2}\ \right] \\
& + \frac{\left\{\begin{array}{l} \sum_{k=1}^{T-1}\left\{\begin{array}{l}\left(\frac{\left(\lambda E(B)\right)^{2(T-k+2)}\left(1+\lambda E(B)\alpha^2\right)\alpha^{T(T-1)-k(k-1)+2}}{E(V_2)+E(H)}\right) \\ * \left[E\left(V_k\right)\left(E\left(H\right)+\frac{E\left(B^2\right)\alpha^{k-1}}{2E(B)}\right)+\frac{E\left(H^2\right)}{2}\right]\end{array}\right\} \\ + \frac{\left(\lambda E(B)\right)^4\left(1+\lambda E(B)\alpha^2\right)\alpha^2}{E(V_2)+E(H)} * \left[\begin{array}{c} E\left(V_T\right)\left(E\left(H\right)+E\left(H_0\right)+\frac{E\left(B^2\right)\alpha^{T-1}}{2E(B)}\right) \\ +\left(\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2}\right)\end{array}\right]\end{array}\right\}}{1-\left(\lambda E\left(B\right)\right)^{2T}\alpha^{T(T-1)}}.
\end{aligned}
$$

5. For $l = 4, 5, \ldots, T$ we have

$$
\begin{aligned}
E\left(Wq_l\right) = \quad & \frac{E(H)}{E\left(V_{l-1}\right)+E\left(H\right)} * \left[\frac{E\left(H^2\right)}{2E\left(H\right)}\left(1+\lambda E\left(B\right)\alpha^{l-1}\right)+\lambda E\left(B\right)\alpha^{l-1} E\left(V_{l-1}\right)\right] \\
& + \frac{E\left(V_{l-1}\right)}{E\left(V_{l-1}\right)+E\left(H\right)} * \left[\frac{E\left(B^2\right)\alpha^{l-2}}{2E\left(B\right)}\left(1+\lambda E\left(B\right)\alpha^{l-1}\right)+E\left(H\right)\right] \\
& + \sum_{k=2}^{l-1}\left\{\begin{array}{l}\left(\frac{\left(\lambda E(B)\right)^{2(l-k)}\left(1+\lambda E(B)\alpha^{l-1}\right)\alpha^{l(l-3)-k(k-3)}}{E(V_{l-1})+E(H)}\right) \\ * \left[\ E\left(V_{k-1}\right)\left(E\left(H\right)+\frac{E\left(B^2\right)\alpha^{k-2}}{2E(B)}\right)+\frac{E\left(H^2\right)}{2}\ \right]\end{array}\right\} \\
& + \frac{\left\{\begin{array}{l} \sum_{k=1}^{T-1}\left\{\begin{array}{l}\left(\frac{\left(\lambda E(B)\right)^{2(T-k+l-1)}\left(1+\lambda E(B)\alpha^{l-1}\right)\alpha^{T(T-1)-k(k-1)+(l-1)(l-2)}}{E(V_{l-1})+E(H)}\right) \\ * \left[E\left(V_k\right)\left(E\left(H\right)+\frac{E\left(B^2\right)\alpha^{k-1}}{2E(B)}\right)+\frac{E\left(H^2\right)}{2}\right]\end{array}\right\} \\ + \frac{\left(\lambda E(B)\right)^{2(l-1)}\left(1+\lambda E(B)\alpha^{l-1}\right)\alpha^{(l-1)(l-2)}}{E(V_{l-1})+E(H)} * \left[\begin{array}{c} E\left(V_T\right)\left(E\left(H\right)+E\left(H_0\right)+\frac{E\left(B^2\right)\alpha^{T-1}}{2E(B)}\right) \\ +\left(\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2}\right)\end{array}\right]\end{array}\right\}}{1-\left(\lambda E\left(B\right)\right)^{2T}\alpha^{T(T-1)}}.
\end{aligned}
$$

Substitution of $E\left(V_1\right)$ and $E\left(V_l\right)$ $\forall l = 2, 3, \ldots, T$, from equations (10.12) and (10.14) respectively, into the above five cases, results in expressions for $E\left(Wq_l\right)$ $\forall l = 1, 2, \ldots, T$ which depend only on the initial parameters $H,\ H_0,\ B,\ \alpha,\ \lambda$ and T. For example, in case $T = 2$ we have

$$
E\left(V_1\right) = \frac{E\left(H\right)\lambda E\left(B\right)\left(1+\lambda E\left(B\right)\alpha\right)+\left(H_0\right)\lambda E\left(B\right)}{1-\left(\lambda E\left(B\right)\right)^2\alpha}.
$$

$$
E\left(V_2\right) = \frac{E\left(H\right)\lambda E\left(B\right)\left(1+\lambda E\left(B\right)\right)\alpha+E\left(H_0\right)\left(\lambda E\left(B\right)\right)^2\alpha}{1-\left(\lambda E\left(B\right)\right)^2\alpha}.
$$

$$E\left(Wq_1\right) = \frac{\left(E\left(H\right)+E\left(H_0\right)\right)\left(1-\left(\lambda E\left(B\right)\right)^2\alpha\right)}{E\left(H\right)\left(1+\lambda E\left(B\right)\alpha\right)+E\left(H_0\right)} * \left\{ \begin{array}{c} \left(\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))}\right)\left(1+\lambda E\left(B\right)\right) \\ +\left(\frac{E(H)(\lambda E(B))^2(1+\lambda E(B))\alpha+E(H_0)(\lambda E(B))^3\alpha}{1-(\lambda E(B))^2\alpha}\right) \end{array} \right\}$$
$$+ \frac{E\left(H\right)\lambda E\left(B\right)\left(1+\lambda E\left(B\right)\right)\alpha+E\left(H_0\right)\left(\lambda E\left(B\right)\right)^2\alpha}{E\left(H\right)\left(1+\lambda E\left(B\right)\alpha\right)+E\left(H_0\right)} * \left[\frac{E\left(B^2\right)\alpha}{2E\left(B\right)}\left(1+\lambda E\left(B\right)\right)+E\left(H\right)+E\left(H_0\right)\right]$$
$$+ \frac{\left(\lambda E\left(B\right)\alpha\right)^2\left(1+\lambda E\left(B\right)\right)}{\left(E\left(H\right)\left(1+\lambda E\left(B\right)\alpha\right)+E\left(H_0\right)\right)\left(1+\left(\lambda E\left(B\right)\right)^2\alpha\right)}$$
$$* \left\{ \begin{array}{c} \left(\frac{E(H)\lambda E(B)(1+\lambda E(B)\alpha)+(H_0)\lambda E(B)}{1-(\lambda E(B))^2\alpha}\right)\left(E\left(H\right)+\frac{E\left(B^2\right)}{2E(B)}\right)+\frac{E\left(H^2\right)}{2} \\ +\left(\lambda E\left(B\right)\right)^2\left[\begin{array}{c} \left(\frac{E(H)(\lambda E(B))(1+\lambda E(B))\alpha+E(H_0)(\lambda E(B))^2\alpha}{1-(\lambda E(B))^2\alpha}\right)\left(E\left(H\right)+E\left(H_0\right)+\frac{E\left(B^2\right)\alpha}{2E(B)}\right) \\ +\ \frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2} \end{array} \right] \end{array} \right\}.$$

Recall that we assumed $T \neq 1$ at the beginning of section 10.2. We now relax this assumption (note that, as mentioned in section 2.3, since the discussed model does not include state-dependent arrival rates, it can also be treated as a classical polling system). For the case of $T = 1$ (i.e. the "always swap" policy), we can simply use the expressions obtained for $T = 2$ after setting $\alpha = 1$ and replacing any H which didn't originated from $S_{T=2}$ with $H + H_0$. This would result in $E\left(V_1\right) = E\left(V_2\right)$ and $E\left(Wq_1\right) = E\left(Wq_2\right)$ which represent the respective $E\left(V_1\right)$ and $E\left(Wq_1\right)$ for the $T = 1$ case. The resulting expressions are

$$E\left(V_1\right) = \frac{\left(E\left(H\right)+E\left(H_0\right)\right)\lambda E\left(B\right)\left(1+\lambda E\left(B\right)\right)}{1-\left(\lambda E\left(B\right)\right)^2}. \tag{10.15}$$

$$E\left(Wq_1\right) = \frac{E\left(V_1\right)}{E\left(V_1\right)+E\left(H\right)+E\left(H_0\right)} * \left[\frac{E\left(B^2\right)}{2E\left(B\right)}\left(1+\lambda E\left(B\right)\right)+E\left(H\right)+E\left(H_0\right)\right]$$
$$+ \frac{E\left(H\right)+E\left(H_0\right)}{E\left(V_1\right)+E\left(H\right)+E\left(H_0\right)} * \left[\ \frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))}\left(1+\lambda E\left(B\right)\right)+\lambda E\left(B\right)E\left(V_1\right)\ \right]$$
$$+ \frac{\sum_{k=2}^{5}\left(\lambda E\left(B\right)\right)^k * \left[\begin{array}{c} E\left(V_1\right)\left(E\left(H\right)+E\left(H_0\right)+\frac{E\left(B^2\right)}{2E(B)}\right) \\ +\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2} \end{array} \right]}{\left(E\left(V_1\right)+E\left(H\right)+E\left(H_0\right)\right)\left(1-\left(\lambda E\left(B\right)\right)^4\right)}. \tag{10.16}$$

Equation (10.16) means that, for the case of $T = 1$,

$$E\left(W_1\right) = E\left(W_1^{new}\right) = E\left(Wq_1\right)+E\left(B\right)$$
$$= \left(\frac{1+\sum_{k=1}^{3}\left(\lambda E\left(B\right)\right)^k}{1-\left(\lambda E\left(B\right)\right)^4}\right) * \left\{ \begin{array}{c} \lambda E\left(B\right)\left[\frac{E\left(B^2\right)}{2E(B)}+E\left(H\right)+E\left(H_0\right)\right] \\ +\left(1-\lambda E\left(B\right)\right)\left[\frac{E\left(H^2\right)+E\left(H_0{}^2\right)+2E(H)E(H_0)}{2(E(H)+E(H_0))}\right] \end{array} \right\}+E\left(B\right). \tag{10.17}$$

We can thus calculate $E\left(W_1\right)$, for any $1 \leq T \leq T_G^{max}$, using equation (8.17). Namely, using

$$E\left(W_1\right) = \frac{\sum_{r=1}^{T}\overline{\lambda}_r E^2\left(W_r^{new}\right)}{\sum_{r=1}^{T}\overline{\lambda}_r E\left(W_r^{new}\right)} = \frac{\sum_{r=1}^{T}\left(E\left(V_{r-1}\right)+E\left(S_{r-1}\right)\right)\left(E\left(Wq_r\right)+E\left(B_r\right)\right)^2}{\sum_{r=1}^{T}\left(E\left(V_{r-1}\right)+E\left(S_{r-1}\right)\right)\left(E\left(Wq_1\right)+E\left(B_r\right)\right)}$$
$$= \frac{\sum_{r=1}^{T}\left[\left(E\left(V_{r-1}\right)+E\left(H\right)\right)\left(E\left(Wq_r\right)+\alpha^{r-1}E\left(B\right)\right)^2\right]+E\left(H_0\right)\left(E\left(Wq_1\right)+E\left(B\right)\right)^2}{\sum_{r=1}^{T}\left[\left(E\left(V_{r-1}\right)+E\left(H\right)\right)\left(E\left(Wq_r\right)+\alpha^{r-1}E\left(B\right)\right)\right]+E\left(H_0\right)\left(E\left(Wq_1\right)+E\left(B\right)\right)}.$$

For the sake of complete presentation, note that

$$E\left(\mathbf{C}\right)=\sum_{k=1}^{T}\left(E\left(S_{k-1}\right)+E(V_{k-1})\right)=\sum_{k=1}^{T}E\left(V_{k-1}\right)+TE\left(H\right)+E\left(H_0\right), \tag{10.18}$$

and

$$\overline{\lambda}_l=\begin{cases}\frac{\lambda(E(V_T)+E(H)+E(H_0))}{E(\mathbf{C})} & l=1,\\ \\ \frac{\lambda(E(V_{l-1})+E(H))}{E(\mathbf{C})} & l>1.\end{cases} \tag{10.19}$$

Remark 10.7. The above results can be extended for the more general case of not necessarily identical servers using the same approach which was described in remark 10.1 (assuming $\lambda\sqrt{E\left(G\right)E\left(K\right)}<1$ and $T\leq T^{max}=\left\lceil\frac{\ln\left(\frac{1}{\lambda^2E(G)E(K)}\right)}{\ln(\alpha)}\right\rceil$). The $T=1$ case does not require a special treatment.

The current (globally) gated regime case is much more complicated than the exhaustive regime case studied in section 10.1. This is mainly due to the fact that, unlike in section 10.1, here the lengths of the different visit periods are (positively) correlated with each other. Such correlations allays exist, except in the case of a single queue exhaustive regime.

For the rest of the current section, we assume that $E(H)>0$. This is done for the sake of a clear presentation. The general case of $E(H)\geq 0$ will be addressed in remarks 10.8 and 10.9.
We conclude this section with some observations regarding the expected visit periods. From equations (10.12) and (10.15):

$$E\left(V_1\right)=\begin{cases}\frac{(E(H)+E(H_0))\lambda E(B)(1+\lambda E(B))}{1-(\lambda E(B))^2} & T=1,\\ \\ \frac{E(H)\sum_{k=1}^{T}(\lambda E(B))^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}}+E(H_0)\lambda E(B)}{1-(\lambda E(B))^T\alpha^{\frac{T(T-1)}{2}}} & T=2,3,\ldots,T_G^{max}.\end{cases} \tag{10.20}$$

Let $E_T\left(V_l\right)$ be $E\left(V_l\right)$ under a given $T=1,2,\ldots,T_G^{max}$. Assuming $T_G^{max}\geq 2$ (i.e. $\alpha^{\frac{1}{2}}\lambda E\left(B\right)<1$), $E_1\left(V_1\right)<E_2\left(V_1\right)$ means

$$\begin{aligned}&\frac{\left(E\left(H\right)+E\left(H_0\right)\right)\lambda E\left(B\right)\left(1+\lambda E\left(B\right)\right)}{1-\left(\lambda E\left(B\right)\right)^2}<\frac{E\left(H\right)\sum_{k=1}^{2}\left(\lambda E\left(B\right)\right)^{2-k+1}\alpha^{\frac{2(2-1)-k(k-1)}{2}}+E\left(H_0\right)\lambda E\left(B\right)}{1-\left(\lambda E\left(B\right)\right)^2\alpha^{\frac{2(2-1)}{2}}}\\ &\Longrightarrow E\left(H_0\right)\left(1-\left(\lambda E\left(B\right)\right)^2\alpha-\lambda E\left(B\right)\left(\alpha-1\right)\right)<E\left(H\right)\left(\alpha-1\right)\left(1+\lambda E\left(B\right)\right).\end{aligned} \tag{10.21}$$

It follows that, if $\frac{1}{\alpha}\leq\lambda E\left(B\right)<\frac{1}{\sqrt{\alpha}}$, equation (10.21) is always true. So, for $T_G^{max}=2$ (which implies $\alpha\lambda E(B)\geq 1$), $E_1\left(V_1\right)<E_2\left(V_1\right)$.

For $T=2,3,\ldots,T_G^{max}-1$, $E_T\left(V_1\right)<E_{T+1}\left(V_1\right)$ means

$$\begin{aligned}&\frac{\left[\begin{array}{c}E\left(H\right)\sum_{k=1}^{T}\left(\lambda E\left(B\right)\right)^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}}\\ +E\left(H_0\right)\lambda E\left(B\right)\end{array}\right]}{1-\left(\lambda E\left(B\right)\right)^T\alpha^{\frac{T(T-1)}{2}}}<\frac{\left[\begin{array}{c}E\left(H\right)\sum_{k=1}^{T}\left(\lambda E\left(B\right)\right)^{T-k+2}\alpha^{\frac{(T+1)T-k(k-1)}{2}}\\ +E\left(H\right)\left(\lambda E\left(B\right)\right)^{T-(T+1)+2}\alpha^{\frac{(T+1)T-(T+1)T}{2}}\\ +E\left(H_0\right)\lambda E\left(B\right)\end{array}\right]}{1-\left(\lambda E\left(B\right)\right)^{T+1}\alpha^{\frac{(T+1)T}{2}}}\\ &\Longrightarrow\frac{\left[\begin{array}{c}E\left(H\right)\sum_{k=1}^{T}\left(\lambda E\left(B\right)\right)^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}}\\ +E\left(H_0\right)\lambda E\left(B\right)\end{array}\right]}{1-\left(\lambda E\left(B\right)\right)^T\alpha^{\frac{T(T-1)}{2}}}<\frac{\left[\begin{array}{c}E\left(H\right)\sum_{k=1}^{T}\left(\lambda E\left(B\right)\right)^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}}*\alpha^T\lambda E\left(B\right)\\ +E\left(H\right)\lambda E\left(B\right)\\ +E\left(H_0\right)\lambda E\left(B\right)\end{array}\right]}{1-\left(\lambda E\left(B\right)\right)^T\alpha^{\frac{T(T-1)}{2}}*\alpha^T\lambda E\left(B\right)}.\end{aligned} \tag{10.22}$$

In this presentation it is readily observed that $\alpha^T\lambda E\left(B\right)\geq 1\Longrightarrow E_T\left(V_1\right)<E_{T+1}\left(V_1\right)$.

From equation (10.14),

$$E_T(V_l) = E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} + \frac{\left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} \\ +E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right]}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}} \quad \forall \begin{array}{l} T=2,3,\ldots,T_G^{max} \text{ and} \\ l=2,3,\ldots,T. \end{array} \tag{10.23}$$

For $T=2,3,\ldots,T_G^{max}-1$ and $l=2,3,\ldots,T$, $E_T(V_l) < E_{T+1}(V_l)$ means

$$\left\{\begin{array}{c} E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} \\ +\frac{\left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} \\ +E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right]}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}} \end{array}\right\}$$

$$< \left\{\begin{array}{c} E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} \\ +\frac{\left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k+1}\alpha^{\frac{(T+1)T+l(l-1)-k(k-1)}{2}} \\ +E(H)(\lambda E(B))^{T+l-(T+1)+1}\alpha^{\frac{(T+1)T+l(l-1)-(T+1)T}{2}} \\ +E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right]}{1-(\lambda E(B))^{T+1}\alpha^{\frac{(T+1)T}{2}}} \end{array}\right\}$$

$$\Longrightarrow \quad \frac{\left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} \\ +E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right]}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}}$$

$$< \frac{\left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} * \alpha^{T}\lambda E(B) \\ +E(H)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \\ +E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right]}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}} * \alpha^{T}\lambda E(B)}. \tag{10.24}$$

In this presentation it is readily observed that $\alpha^T\lambda E(B) \geq 1 \Longrightarrow E_T(V_l) < E_{T+1}(V_l)$.
From equations (10.21), (10.22) and (10.24) we **conclude** that $\alpha^T\lambda E(B) \geq 1 \Longrightarrow E_T(V_l) < E_{T+1}(V_l)\ \forall T = 1,2,\ldots,T_G^{max}-1$ and $l=1,2,\ldots,T$.

Define $T^{mid} \equiv \min_{T\in\mathbb{Z}}\left\{T|\,\alpha^T\lambda E(B)\geq 1\right\} = \left\lceil \frac{ln(\frac{1}{\lambda E(B)})}{ln(\alpha)} \right\rceil$. Note that $T^{mid} \geq 1$.Intriguingly, $T^{mid} = T_E^{max}$. So equation (9.4) also describes the relation between T_G^{max} and T^{mid} (hence the "mid" superscript). According to the last conclusion, we can write that, for $E(H) \neq 0$ (see remark 10.8),

$$E_{T^{mid}}(V_l) < E_{T^{mid}+1}(V_l) < \cdots < E_{T_G^{max}}(V_l)\ \ \forall l = 1,2,\ldots,T.$$

Remark 10.8. As stated above, the last conclusion was obtained under the assumption that $E(H) \neq 0$. For the case of $E(H)=0$ and $\alpha^T\lambda E(B) > 1$ the conclusion is still valid. However, for the case of $E(H)=0$ and $\alpha^T\lambda E(B) = 1$ we obtain

$$E_{T^{mid}}(V_l) = E_{T^{mid}+1}(V_l) < E_{T^{mid}+2}(V_l)\cdots < E_{T_G^{max}}(V_l)\ \ \forall l = 1,2,\ldots,T.$$

For $T = 2, 3, \ldots, T_G^{max}$, $E_T(V_1) < E_T(V_2)$ means

$$\frac{\left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}} \\ +E(H_0)\lambda E(B) \end{array}\right]}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}}$$

$$< \left\{ \begin{array}{c} E(H)(\lambda E(B))^{2-2+1}\alpha^{\frac{(2-2+1)(2+2-2)}{2}} \\ \\ +\left[\begin{array}{c} \left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+2-k}\alpha^{\frac{T(T-1)+2(2-1)-k(k-1)}{2}} \\ +E(H_0)(\lambda E(B))^{2}\alpha^{\frac{2(2-1)}{2}} \end{array}\right] \\ \hline 1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}} \end{array}\right] \end{array} \right\}$$

$$\Longrightarrow \quad \frac{\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}} \\ +E(H_0)\lambda E(B) \end{array}}{1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}}}$$

$$< \left\{ \begin{array}{c} E(H)\alpha\lambda E(B) \\ \\ +\left[\begin{array}{c} \left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T-k+1}\alpha^{\frac{T(T-1)-k(k-1)}{2}} \\ +E(H_0)\lambda E(B) \end{array}\right] \\ \hline 1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}} \end{array}\right] * \alpha\lambda E(B) \end{array} \right\}. \tag{10.25}$$

In this presentation it is readily observed that $\alpha\lambda E(B) \geq 1 \Longrightarrow E_T(V_1) < E_T(V_2)$. So, for $T_G^{max} = 2$ (which implies $T^{mid} = 1$), $E_T(V_1) < E_T(V_2)$.

For $T = 3, 4, \ldots, T_G^{max}$ and $l = 2, 3, \ldots, T-1$, $E_T(V_l) < E_T(V_{l+1})$ means

$$\left\{ \begin{array}{c} E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} \\ \\ +\left[\begin{array}{c} \left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} \\ +E(H_0)(\lambda E(B))^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right] \\ \hline 1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}} \end{array}\right] \end{array} \right\}$$

$$< \left\{ \begin{array}{c} E(H)\sum_{k=2}^{l}(\lambda E(B))^{l-k+2}\alpha^{\frac{(l-k+2)(l+k-1)}{2}} \\ +E(H)(\lambda E(B))^{l-(l+1)+2}\alpha^{\frac{(l-(l+1)+2)(l+l+1-1)}{2}} \\ \\ +\left[\begin{array}{c} \left[\begin{array}{c} E(H)\sum_{k=1}^{T}(\lambda E(B))^{T+l-k+1}\alpha^{\frac{T(T-1)+(l+1)l-k(k-1)}{2}} \\ +E(H_0)(\lambda E(B))^{l+1}\alpha^{\frac{(l+1)l}{2}} \end{array}\right] \\ \hline 1-(\lambda E(B))^{T}\alpha^{\frac{T(T-1)}{2}} \end{array}\right] \end{array} \right\}$$

$$\Longrightarrow \left\{ \begin{array}{c} E\left(H\right)\sum_{k=2}^{l}\left(\lambda E\left(B\right)\right)^{l-k+1}\alpha^{\frac{(l-k+1)(l+k-2)}{2}} \\ \\ +\left[\begin{array}{c}\left[\begin{array}{c} E\left(H\right)\sum_{k=1}^{T}\left(\lambda E\left(B\right)\right)^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} \\ +E\left(H_0\right)\left(\lambda E\left(B\right)\right)^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right] \\ \hline 1-\left(\lambda E(B)\right)^{T}\alpha^{\frac{T(T-1)}{2}} \end{array}\right] \end{array} \right\}$$

$$< \left\{ \begin{array}{c} E\left(H\right)\sum_{k=2}^{l}\left(\lambda E\left(B\right)\right)^{l-k+1}\alpha^{\frac{(l-k+2)(l+k-1)}{2}} * \alpha^{l}\lambda E\left(B\right) \\ \\ +E\left(H\right)\alpha^{l}\lambda E\left(B\right) \\ \\ +\left[\begin{array}{c}\left[\begin{array}{c} E\left(H\right)\sum_{k=1}^{T}\left(\lambda E\left(B\right)\right)^{T+l-k}\alpha^{\frac{T(T-1)+l(l-1)-k(k-1)}{2}} \\ +E\left(H_0\right)\left(\lambda E\left(B\right)\right)^{l}\alpha^{\frac{l(l-1)}{2}} \end{array}\right] \\ \hline 1-\left(\lambda E(B)\right)^{T}\alpha^{\frac{T(T-1)}{2}} \end{array}\right] * \alpha^{l}\lambda E\left(B\right) \end{array} \right\}. \tag{10.26}$$

In this presentation it is readily observed that $\alpha^{l}\lambda E\left(B\right) \geq 1 \Longrightarrow E_T\left(V_l\right) < E_T\left(V_{l+1}\right)$.
From equations (10.25) and (10.26) we **conclude** that for $T = T^{mid}+1, T^{mid}+2 \ldots, T_G^{max}$ and $l = T^{mid}, T^{mid}+1, \ldots, T-1$, $E_T\left(V_l\right) < E_T\left(V_{l+1}\right)$. According to the last conclusion we can write that, for $E(H) \neq 0$ (cf. remark 10.9),

$$E_T\left(V_{T^{mid}}\right) < E_T\left(V_{T^{mid}+1}\right) < \cdots < E_T\left(V_{T_G^{max}}\right) \ \forall T = T^{mid}+1, T^{mid}+2, \ldots, T_G^{max}.$$

Remark 10.9. According to equation (8.1), V_l's length is composed of (i) the service times of all type-l customers arriving during V_{l-1}, and (ii) the service times of all type-l customers arriving during M_{l-1}. This means that, for $l = 2, 3, \ldots, T_G^{max}$,

$$E(V_l) = \lambda[E(V_{l-1}) + E(H)]\alpha^{l-1}E(B).$$

Generally speaking, (ii) keeps increasing while moving from $l = 1$ to $l = T_G^{max}$. We can nullify this parameter-dependent affect by assuming $E(H) = 0$. The expected amount of time spent per served customer in $E(V_l)$ is $\alpha^{l-1}E(B)$. The arrival rate of this workload to the system during $E(V_{l-1})$ is $\lambda\alpha^{l-1}E(B)$. Recall the equality between T^{mid} and T_E^{max} which is embodied in the fact that $\lambda\alpha^{l-1}E(B) \geq 1$ for $l = T^{mid}+1, T^{mid}+2, \ldots, T_G^{max}$. Hence, without relevance to $E(H)$, the single server spends more (expected) time in V_l than in V_{l-1} for $l = T^{mid+2}, T^{mid+3}, \ldots, T_G^{max}$. Note that, similarly to remark 10.8,

$$E(V_{T^{mid}}) \leq E(V_{T^{mid}+1}),$$

where

$$E(V_{T^{mid}}) = E(V_{T^{mid}+1}) \iff \left\{\{H=0\} \cap \{\lambda\alpha^{T^{mid}}E(B) = 1\}\right\}.$$

Remark 10.10. For a general $E(H) \geq 0$, T^{opt} may be bigger than T^{mid}. To see this, recall that T^{mid} and T_G^{max} are not affected by H_0. In a system which is stable under "zero **TL** stability", continuously increasing $E(H_0)$ will eventually results in $T^{opt} = T_G^{max}$ since swapping the servers becomes more "costly" (both directly, by some prolonged switch-over periods, and indirectly, by prolonged visit periods).

11 Concluding remarks

In this paper we introduced a new polling system comprised of two alternating weary servers, which operates under either the gated, exhaustive, or globally-gated regime. The tradeoff between the tiredness effects on

the servers and the "swapping cost" (H_0 units of time) can be illustrated by observing some numerical results. Consider the following basic model consisting of a single queue and identical servers, where we take $\alpha > 1$ as a variable:

$$\underbrace{\lambda \;=\; 1}_{arrival\ rate}\ ;\ \underbrace{H \sim \exp(1)}_{swich-over\ time}\ ;\ \underbrace{H_0 \sim \exp(0.5)}_{swapping\ time}\ ;\ \underbrace{B \sim \exp(4)}_{service\ time}\ .$$

Using standard polling techniques (see e.g. [14]), we have accurately calculated $E(W_1)$ for this model under the "always swap" policy. This results in $E(W_1) = 3\frac{2}{3}$ for the (globally) gated regime and in $E(W_1) = 2\frac{2}{3}$ for the exhaustive regime. Next, for each of the "swap at the end of every T sub-cycles" policies where $T = 2, 3, 4$, we used the MVA approach to numerically calculate the value of α, **for which the resulting system's $E(W_1)$ obtains the same value as under the "always swap" policy**. The results are summarized in the following table:

T	Exhaustive: $E(W_1) = 2\frac{2}{3}$	(Globally) Gated: $E(W_1) = 3\frac{2}{3}$
2	$\alpha = 2.5$	$\alpha \cong 2.335$
3	$\alpha \cong 1.63$	$\alpha \cong 1.692$
4	$\alpha \cong 1.4$	$\alpha \cong 1.451$

As was expected, under both regimes, a smaller α allows swapping the servers less often and still gain the same (or better) expected sojourn time as in the "always swap" policy. Clearly, in each case, increasing (decreasing) α will result in a worse (better) expected sojourn time than in the "always swap" policy.

Although we have only uncovered the "tip of the iceberg", one can consider various ways in which to extend the presented new polling system. Aside from the inclusion of additional regimes (e.g. mixed), one can (i) change the way the fatigue parameter and tiredness levels affect the service time distributions; (ii) differ between the tiredness effects on each server; (iii) develop a broader scope of criterions for comparison between the different swapping policies, etc. Another extension can be to include tiredness effects on switch-over times. Note that doing so does not affect the stability condition under the three discussed regimes (nor under the mixed regime for that matter).

References

1. Boon, M.A.A., van der Mei, R.D., Winands, E.M.M., (2011). *Applications of polling systems.* Surveys in Operations Research and Management Science **16 (2)**, 67-82.

2. Boon, M.A.A., Van Wijk, A.C.C., Adan, I.J.B.F & Boxma, O.J. (2010). *A polling model with smart customers.* Queueing Systems **66**, 239-274.

3. Boxma, O.J. (1994). *Polling systems.* In: From universal morphisms to megabytes: A baayen space odyssey- Liber amicorum for P.C. baayen. CWI, Amsterdam, 215-230.

4. Boxma, O.J., Groenendijk, W.P. (1987). *Pseudo-conservation laws in cyclic serving systems.* J. of Applied Probability **24**, 949-964.

5. Boxma, O.J., Ivanovs, J., Kosinski, K., Mandjes, M.R.H. (2011). *Lévy-driven polling systems and continuous-state branching processes.* Stochastic Systems **1 (2)**, 411-436.

6. Browne, S., Yechiali, U., (1989). *Dynamic priority rules for cyclic-type queues.* Advances in Applied Probability **21**, 432-450.

7. Browne, S., Yechiali, U., (1989). *Dynamic routing in polling systems.* In: M. Bonatti (Ed.) Teletraffic Science for New Cost-Effective Systems, Networks and Services. North-Holland, 1455-1466.

8. Cohen, J.W., (1969). *The single server queue.* North-Holland, Amsterdam.

9. Flicker, C., Jaibi, M.R., (1993). *Monotonicity and stability of periodic polling models.* Queueing Systems **15**, 211-238.

10. Takagi, H., (1986). *Analysis of polling system.* The MIT Press.

11. Levy, Y., Yechiali, U., (1975). *Utilization of idle time in an M/G/1 queueing system.* Management Science **22 (2)**, 202-211.

12. Resing, J.A.C. (1993). *Polling systems and multiple branching processes.* Queueing Systems **13**, 409-426.

13. Winands, E.M.M., Adan, I.J.B.F., van Houtum, G.J. (2006). *Mean value analysis for polling systems.* Queueing Systems **54**, 35-44.

14. Yechiali, U., (1993). *Analysis and control of polling systems.* Performance Evaluation of Computer and Communication Systems. Springer-Verlag, 630-650.