# On an Adaptive-Quasi-Deterministic Transmission Policy Queueing Model

Jacob Bergquist
*Dept. Industrial Eng. & Operations Res.*
& Center for Applied Probability
Columbia University
New York, NY 10027, USA
ORCID:0009-0001-7813-6499

Erol Gelenbe
*Inst. Theoretical & App. Informatics*
Polish Ac. Sci. 44-100 Gliwice, PL
CNRS I3S Lab., Univ. Côte d'Azur, FR
& Eng. Dept. Kings College London
ORCID:0000-0001-9688-2201

Karl Sigman
*Dept. Industrial Eng. & Operations Res.*
& Center for Applied Probability
Columbia University
New York, NY 10027, USA
ORCID:0000-0002-0126-3895

***Abstract*—We analyze, further and deeper, a recently proposed technique for addressing the Massive Access Problem (MAP), an issue in telecommunications which arises when too many devices transmit packets to a gateway in quick succession. This technique, the Adaptive-Quasi-Deterministic Transmission Policy (AQDTP) is a special case of "traffic shaping" which involves delaying some packets at the points of origin to alleviate congestion at the routers. One nice feature of AQDTP is that it loses no packets and allows an infinite buffer. In this work, to clarify the approach in a general queueing theory framework, and to move beyond the original telecommunications application, we frame these potential delays as time spent at a café by customers before proceeding to a service facility. We first present some sample-path results that significantly refine and expand upon what was shown in previous work, and then present further results under a general stationary ergodic stochastic framework. In the sample-path realm, we give conditions that ensure AQDTP will not change the total delay and sojourn time of any customer as compared to what that customer would have experienced if there was no café; but we also prove that AQDTP can never reduce the total delay. The difference is that, under AQDTP, some of that delay is spent at the café instead of in the queue/line at the service facility. In a stochastic framework, our focus is on stability and constructing proper stationary versions of the model. Under i.i.d. assumptions we dig deeper by proving Harris recurrence of an underlying two-dimensional Markov process, and explicitly find positive recurrent regeneration points.**

## I. Introduction

The widespread proliferation of the Internet of Things (IoT) has brought about new challenges in the field of telecommunications, particularly in the area of network access. One of the major challenges is the Massive Access Problem (MAP), which occurs when too many IoT devices simultaneously transmit data to a base station or IoT gateway, causing congestion and potentially untenable delays and loss of packets. A variety of techniques to address the MAP have been proposed over the years including adaptive routing ([7], [8]), Access Class Barring ([11], [16], [17]), Cognitive Machine-to-Machine communication ([1], [2]), and device clustering ([20], [15]). More proactive techniques include Joint Forecasting-Scheduling (JFS) and Priority based on Average Load (PAL) ([19], [21]. However, these techniques can involve time-consuming machine learning methods and require a significant amount of communication over the network to implement.

A simpler technique to address the MAP is that of *traffic shaping* of which there is a wide-ranging literature (see Section 7.2.7 in [14], and [13],[5],[4],[12]) where packets are purposely delayed at the origin to alleviate congestion. (This is not to be confused with "Traffic Policing", which involves preventative packet dropping, [6].)

A recent simple method of performing traffic shaping, the Quasi-Deterministic Transmission Policy (QDTP), was introduced and empirically investigated in the context of IoT devices in [10]. Following that, QDTP and the more general Adaptive-QDTP (AQDTP) were modeled and analyzed more formally and directly using some queueing theory in [9]. **An important feature of this method is that it applies to an infinite buffer, no packets are ever lost, and no apriori bounding constraints are placed on the inputs (arrivals, servicing).**

Here in the present paper, keeping much more to a direct queueing theory approach (to allow for a larger range of applications), we present deeper new results on AQDTP. We refer to the packets as *customers* and the initial delay facility as a *café*; one would prefer to spend most of the delay in a café rather than in a line somewhere; the notion of utility comes to mind in some applications; customers might even be willing to pay more for this: Business/First Class passengers at airports who can wait in lounges, for example.

We first present some general sample-path results that significantly refine and expand upon what was shown in [9]; in particular, the concluded inequality in Theorem 1 of [9] now becomes an equality in the present paper as part (a) in Proposition 3.2. We give conditions that ensure AQDTP will not change the total sojourn time of any packet as compared to what that packet would have experienced in the original FIFO model without delays; and we also prove that AQDTP can never reduce the total sojourn time. The difference is that, under AQDTP, some of that sojourn is spent at the café instead of in the buffer (line) at the service facility.

Next we move on to a stochastic framework under general stationary and ergodic input. Our focus is on stability and constructing proper stationary versions of the model. Under i.i.d. assumptions we dig deeper by proving Harris recurrence of an underlying two-dimensional Markov process, and explicitly find positive recurrent regeneration points (even when the system does not empty).

The layout of the paper is as follows: The AQDTP model is presented in Section II. Section III covers the sample-path results. Section IV introduces the stochastic setting. Section IV-B gives stability results, with Section IV-C specializing to i.i.d. input, Harris recurrence, and regeneration. Section IV-D contains some final remarks.

## II. AQDTP QUEUEING MODEL

We start with a FIFO single-server queueing model with *unlimited* waiting space (queue/line/buffer), which for our purposes will be called the *nominal* queueing model. The $n^{th}$ arriving customer, $n \geq 0$, is denoted by $C_n$. The primitives of the model are the arrival times of customers $\{a_n : n \geq 0\}$ on the time axis $[0, \infty)$, with interarrival times

$$A_n = a_{n+1} - a_n,$$

and non-negative service times $\{S_n : n \geq 0\}$. **The only assumption on the arrival point process is that $a_n \leq a_{n+1},\ n \geq 0$, and $a_n \to \infty$ as $n \to \infty$.** We place no further conditions. The nominal model is what the customers would experience *if there was no traffic shaping*; $C_n$ would arrive at time $a_n$, wait in the line (if the server is busy) then get served and depart. Letting $L_n$ denote the delay (in queue/line) of $C_n$, it satisfies the classic Lindley recursion,

$$L_{n+1} = (L_n + S_n - A_n)^+,\ n \geq 0. \tag{1}$$

For the AQDTP model, traffic shaping is implemented in which the $a_n$ are delayed yielding new arrival times to the service facility, $t_n \geq a_n$, with interarrival times

$$T_n = t_{n+1} - t_n.$$

With $t_0 = a_0 = 0$, $\{t_n\}$ is defined recursively by

$$t_{n+1} = \max\{t_n + D_n, a_{n+1}\},\ n \geq 0, \tag{2}$$

where the strictly positive $\{D_n\}$ are an additional primitive (chosen by the designer) and can all be, for example, a deterministic constant $D$ (yielding the QDTP model) or more generally depend on $n$ and even be random variables. Since $t_{n+1} \geq t_n + D_n$, it follows that

$$T_n \geq D_n,\ n \geq 0, \tag{3}$$

meaning that consecutive arrivals are now separated in time by at least the amounts $D_n$, in effect keeping them spread out to avoid clustering. In fact, since the $D_n$ are assumed strictly positive, it follows that $t_{n+1} > t_n,\ n \geq 0$; $\{t_n\}$ forms a *simple* point process even if $\{a_n\}$ has batches, i.e., if $a_n = a_{n+1}$ for some values of $n$. Also note that from (2), it is possible that $t_n = a_n$ for some values of $n$. The service facility remains a FIFO single-server queue, but the delays in queue become $\{V_n\}$ defined by (recalling (1)),

$$V_{n+1} = (V_n + S_n - T_n)^+,\ n \geq 0. \tag{4}$$

We imagine that the *initial delays*

$$W_n = t_n - a_n \geq 0, \tag{5}$$

take place at a café, where $C_n$ first arrives at time $a_n$, then attends the queue at the service facility at time $t_n = a_n + W_n$. Thus for the AQDTP model, $C_n$ has a *total delay* (before starting service) of

$$Z_n = W_n + V_n, \tag{6}$$

and a *sojourn time* of

$$R_n = Z_n + S_n = W_n + V_n + S_n. \tag{7}$$

In essence, the AQDTP model is a two-stage in tandem model (but with only 1 server): $C_n$ arrives at the AQDTP model's café (stage 1) at time $a_n$, then moves to the service facility (stage 2) at time $t_n$, waits in its queue (if the server is busy), enters service at time

$$a_n^e = a_n + Z_n, \tag{8}$$

then finally departs the entire system at time

$$a_n^d = a_n + R_n. \tag{9}$$

Note that since $t_n = a_n + W_n$, we can write

$$T_n = t_{n+1} - t_n = A_n + W_{n+1} - W_n,\ n \geq 0. \tag{10}$$

The purpose of AQDTP is to reduce congestion (delay in line at the service facility) as compared to that in the nominal queue, by passing on some of the line delay to time spent in the café; **and no customers are to be lost, all are to be served.** The reduction, and how to achieve it, is illustrated in the next section.

## III. SAMPLE-PATH PROPERTIES OF AQDTP

Here we present some sample-path results, by giving conditions on the $D_n$ ensuring that delay at the service facility is reduced, $V_n < L_n$, but we also show that total delay (hence sojourn time too) can never be reduced, but is *shared* with the café (recall (6)). The idea is that one can control how much of total delay is allocated to each of the two, $W_n, V_n$, by choice of the $\{D_n\}$ used in the construction of $\{t_n\}$ from (2).

In Proposition 3.1 below, we include a proof of the recursion (11) below (given in [9] as Lemma 2) for completeness since it is a fundamental result. The rate result, Corollary 3.1, is new. In Proposition 3.2 below, we greatly refine and expand upon what was given in Theorem 1 in [9] in which only our current part (a) condition was considered and it concluded with $Z_n \leq L_n$ instead our new refined conclusion $Z_n = L_n$. All of parts (b)-(d) are entirely new.

**Proposition 3.1:** The initial delay sequence $W_n = t_n - a_n,\ n \geq 0$, satisfies a Lindley recursion,

$$W_{n+1} = (W_n + D_n - A_n)^+,\ n \geq 0; \tag{11}$$

the delay in queue/line recursion of a FIFO single-server queue with "service times" $D_n$ and interarrival times $A_n$.[1]

**Proof :** By induction on $n$: $W_0 = 0$ since $t_0 = a_0 = 0$, and $A_0 = a_1 - a_0 = a_1$. Thus $(W_0 + D_0 - A_0)^+ = (D_0 - a_1)^+$ and from (2) we have $t_1 = D_0$ if $D_0 > a_1$ in which case $(D_0 - a_1)^+ = D_0 - a_1 = t_1 - a_1 = W_1$, and $t_1 = a_1$ if $D_0 \leq a_1$ in which case $t_1 - a_1 = 0 = (D_0 - a_1)^+ = W_1$. Thus the recursion holds for $W_1$.

Now suppose that for some $n \geq 1$, (11) holds for $W_n$, that is, $W_n = t_n - a_n = (W_{n-1} + D_{n-1} - A_{n-1})^+$. We will show that $W_{n+1} = t_{n+1} - a_{n+1} = (W_n + D_n - A_n)^+$ as well. By our hypothesis,

$$\begin{aligned}(W_n + D_n - A_n)^+ &= (t_n - a_n + D_n - (a_{n+1} - a_n))^+ \\ &= (t_n + D_n - a_{n+1})^+.\end{aligned}$$

From (2), if $t_n + D_n > a_{n+1}$, then $t_n + D_n = t_{n+1}$ and so $(t_n + D_n - a_{n+1})^+ = t_{n+1} - a_{n+1} = W_{n+1}$; whereas if $t_n + D_n \leq a_{n+1}$, then $(t_n + D_n - a_{n+1})^+ = 0 = a_{n+1} - a_{n+1} = t_{n+1} - a_{n+1} = W_{n+1}$; the recursion holds for $n+1$ as well. ■

**Corollary 3.1:** If a rate $0 < \lambda < \infty$ exists for $\{a_n\}$, that is, if $a_n/n \to \frac{1}{\lambda}$ as $n \to \infty$, and an average $d$ exists for the $D_n$, that is, $\frac{1}{n}\sum_{i=1}^{n} D_i \to d,$ as $n \to \infty$, where $d < \frac{1}{\lambda}$, then the rate of $\{t_n\}$ exists and is also equal to $\lambda$; $t_n/n \to \frac{1}{\lambda}$ as $n \to \infty$.

**Proof :** Because of the Lindley recursion representation (11), it is well known that under the conditions assumed, $W_n/n \to 0$ as $n \to \infty$ (see Lemma 6.1 on Page 134 in [23]). Hence, since $a_n = t_n + W_n$, the limit of $t_n/n$ is the same as $a_n/n$. ■

We now proceed to the main result in this Section.

**Proposition 3.2:** If $Z_0 = L_0$, then

(a) If $D_n \leq S_n,\ n \geq 0$, then $Z_n = L_n,\ n \geq 0$: total delay in AQDTP is identical to that in the nominal FIFO $G/G/1$ model.

(b) If $D_n = S_n,\ n \geq 0$, and if $V_0 = 0$, then $Z_n = L_n = W_n,\ n \geq 0$ : Every customer enters service immediately when arriving at the service facility; they spend no time delayed in the queue; all delay is spent at the café.

(c) If $D_n < S_n,\ n \geq 0$, then $Z_n = L_n,\ n \geq 0$, but for any $n \geq 1$, if $W_n > 0$ then $V_n > 0$ (equivalently if $V_n = 0$ then $W_n = 0$, i.e., $t_n = a_n$). Any customer who spends time at the café also spends time delayed in the queue; delay is shared.

(d) If $D_n > S_n,\ n \geq 0$, (and $V_0 = 0$), then: $V_n = 0,\ n \geq 0$, and thus $Z_n = W_n,\ n \geq 0$. All of the delay is spent at the café but $Z_n \geq L_n,\ n \geq 1$ with $Z_n > L_n$ if $L_n > 0$ : Total delay, hence sojourn time, is increased for each customer as compared to the nominal model. (But even in this case, in some queueing applications customers might prefer spending all their delay at the cafe even if at the expense of increasing total delay.)

**Proof :** For (a) it suffices (since by assumption $Z_0 = L_0$) to prove that if $Z_n = L_n$ for a given $n \geq 0$, then $Z_{n+1} = L_{n+1}$. To this end, assume that $Z_n = L_n$ for some $n$. Recalling (6) and (10) we have:

$$\begin{aligned}Z_{n+1} &= W_{n+1} + V_{n+1} \\ &= [W_n + D_n - A_n]^+ + [V_n + S_n - T_n]^+ \\ &= [W_n + D_n - A_n]^+ + [Z_n + S_n - A_n - W_{n+1}]^+ \\ &= [W_n + D_n - A_n]^+ + [L_n + S_n - A_n - W_{n+1}]^+.\end{aligned} \tag{12}$$

We consider two cases, (A) and (B):

(A) $W_{n+1} = (W_n + D_n - A_n)^+ = W_n + D_n - A_n > 0$. Then starting with the last line of (12), using our assumption that $L_n = Z_n = W_n + V_n$, and noting that $[V_n + S_n - D_n]^+ = V_n + S_n - D_n$ if $D_n \leq S_n$ yields

$$\begin{aligned}Z_{n+1} &= W_n + D_n - A_n + [L_n + S_n - A_n - W_{n+1}]^+ \\ &= W_n + D_n - A_n + [V_n + S_n - D_n]^+ \\ &= W_n + D_n - A_n + V_n + S_n - D_n \\ &= W_n + V_n + S_n - A_n \\ &= L_n + S_n - A_n \\ &= L_{n+1}.\end{aligned}$$

(B) $W_{n+1} = [W_n + D_n - A_n]^+ = 0$. Then starting with the last line of (12) immediately yields $Z_{n+1} = [L_n + S_n - A_n]^+ = L_{n+1}$.

Thus in both cases $Z_{n+1} = L_{n+1}$, and the proof of the first assertion is complete.

For (b): We already are assuming that $Z_0 = L_0$. Thus if also $V_0 = 0$, then $0 = V_0 = Z_0 = W_0$ from (a) and so the recursions for $\{L_n\}$ and $\{W_n\}$ both start at 0 and hence yield identical processes $L_n = W_n,\ n \geq 0$. Thus from (a) it follows that $V_n = 0,\ n \geq 0$.

For (c): Suppose that $0 < W_n = W_{n-1} + D_{n-1} - A_{n-1}$. Then

$$\begin{aligned}V_n &= (V_{n-1} + S_{n-1} - A_{n-1} - W_n + W_{n-1})^+ \\ &= (V_{n-1} + S_{n-1} - D_{n-1})^+ \\ &= V_{n-1} + S_{n-1} - D_{n-1} > 0,\end{aligned}$$

because $S_{n-1} - D_{n-1} > 0,\ n \geq 1$, by assumption.

For (d): Since from (3), $T_n \geq D_n$, an upper bound

$$V_n \leq \overline{V}_n,\ n \geq 0,$$

is established by using the recursion

$$\overline{V}_{n+1} = (\overline{V}_n + S_n - D_n)^+,\ n \geq 0.$$

Thus if $D_n > S_n,\ n \geq 0$, then $S_n - D_n < 0,\ n \geq 0$, and the result $V_n = 0,\ n \geq 0$ follows. Thus $Z_n = W_n,\ n \geq 0$. But again using the assumption that $D_n > S_n,\ n \geq 0$, we obtain (by substituting each $S_n$ for $D_n$ in the recursion for $W_n$) that

[1]But its meaning is subtly different: In the AQDTP model, $W_n$ is the amount of time that the $n^{th}$ customer spends in the café; there is no queue nor service times at the café. This implies that as a whole, AQDTP can't be modeled as a classic tandem queue in which the first stage has service times $\{D_n\}$ and the second stage has service times $\{S_n\}$; for in that interpretation, $C_n$ would depart the café at time $a_n + W_n + D_n$ which is incorrect; they depart at time $t_n = a_n + W_n$.

$W_n \geq L_n,\ n \geq 0$, and the strict inequality, $D_n > S_n$, implies that $W_n > L_n$ whenever $L_n > 0$. ■

**Remark 3.1:** The arrival rate result in Corollary 3.1 is a significant one; one can always spread out the arrivals of a point process $\{a_n\}$ that has a rate $\lambda$ by simply adding a fixed amount (say a constant $c > 0$) to each interarrival time yielding new interarrival times $A_n + c$. But then the rate will decrease to

$$\frac{\lambda}{1+\lambda c}.$$

Using AQDTP, the arrival rate is preserved, and hence so is the departure rate which is the measure of how much demand has been processed per unit time in the long run. That departure rate often can not be allowed lowered in practice, or otherwise QoS will suffer. The way that AQDTP accomplishes this feat is to sometimes increase the interarrival times, but sometimes decrease ones that are large; so on average nothing has been altered.

Moreover, the result reveals a possible practical approach to defining $\{D_n\}$ so that its average $d$ does exist: Fix an appropriate parameter $b > 0$, and define $D_n = bS_n$, for then if the average of the $S_n$ exists, a reasonable assumption in practice, denoted by $1/\mu$, then so does the average $d$ of the $D_n$, and $d = b/\mu$. A stochastic version of Corollary 3.1 is given in Corollary 4.1 in the next section.

## IV. A STOCHASTIC FRAMEWORK

Here we assume that $\{(A_n, S_n, D_n) : n \geq 0\}$ forms a (general) stationary ergodic sequence of random variables, equivalently that $\{(a_n, (S_n, D_n))\} : n \geq 0\}$, forms a point-stationary ergodic marked point process. Since the random variables are stationary, we let $A = A_0$, $S = S_0$ and $D = D_0$ denote generic such ones. The **arrival rate**, $\lambda = \frac{1}{E(A)}$, is assumed positive and finite.

Our objective is to prove **stability** of the AQDTP model, by which we mean the existence of a unique limiting distribution, and an associated (proper) stationary ergodic version.

The following two conditions are referred to as the **stability conditions** for the AQDTP model:

$$0 < E(D) < E(A) < \infty. \tag{13}$$

$$0 < E(S) < E(A) < \infty. \tag{14}$$

As we will see over the next several sections, the first condition yields stability of $\{W_n\}$. Then adding in the second condition along with the first, will yield joint stability of $\{(W_n, V_n)\}$.

### A. Stability of $\{W_n\}$

A proof of the following is based directly on the classic Loynes' Lemma [18]; that and further applications of it can be found on Pages 131-137, Lemma 6.1, and Theorem 6.1 in [23].

**Proposition 4.1:** If stability condition (13) holds, then there exists a (2-sided; $n \in \mathbb{Z}$ instead of only $n \geq 0$) jointly stationary ergodic version of $\{(W_n, A_n, D_n)\}$ denoted by $\{(W_n^0, A_n^0, D_n^0)\} = \{(W_n^0, A_n^0, D_n^0) : n \in \mathbb{Z}\}$, such that

$$W_{n+1}^0 = (W_n^0 + D_n^0 - A_n^0)^+,\ n \in \mathrm{Z}. \tag{15}$$

$W_n$ converges in total variation, as $n \to \infty$ to the distribution of $W_0^0$, regardless of initial conditions, $W_0 = x \geq 0$. If $E(D) > E(A)$ then $\{W_n\}$ is unstable; $P(W_n \to \infty) = 1$.

Proposition 4.1 allows us to construct a stationary ergodic version of the point process $\{t_n\}$, and it has the same rate $\lambda$ as $\{a_n\}$:

**Corollary 4.1:** If stability condition (13) holds, then

$$t_n^0 = a_n^0 + W_n^0 \tag{16}$$

defines a point-stationary ergodic version of $\{t_n\}$, that is, $T_n^0 = t_{n+1}^0 - t_n^0$ defines a stationary ergodic sequence of interarrival times. Moroever,

$$E(T^0) = \frac{1}{\lambda};$$

$\{t_n\}$ has rate $\lambda$, the same as $\{a_n\}$.

**Proof :** Defining $t_n^0 = W_n^0 + a_n^0$, so that $T_n^0 = t_{n+1}^0 - t_n^0 = A_n^0 + W_{n+1}^0 - W_n^0$ indeed yields a stationary ergodic sequence of interarrival times since it is a function of $\{W_n^0\}$ an already proven to be stationary ergodic sequence; $\{t_n^0\}$ is thus a point-stationary ergodic version of $\{t_n\}$. That it has rate $\lambda$ follows immediately:
$E(T_n^0) = E(A_n^0) + E(W_{n+1}^0 - W_n^0) = \frac{1}{\lambda} + 0 = \frac{1}{\lambda}$. ■

### B. Stability of AQDTP

From Proposition 4.1 and Corollary 4.1 we can replace $\{(W_n, A_n, T_n, S_n, D_n)\}$ by a two-sided stationary ergodic joint version, $\{(W_n^0, A_n^0, T_n^0, S_n^0, D_n^0)\}$, in the following total delay recursion so that it jointly uses stationary ergodic versions of input:

$$Z_{n+1} = (W_n^0 + D_n^0 - A_n^0)^+ + (V_n + S_n^0 - T_n^0)^+,\ n \geq 0. \tag{17}$$

The first piece on the right, derived from (15), already forms a stationary ergodic sequence. We now deal with the second piece. Recalling from Corollary 4.1 that $E(T_n^0) = \frac{1}{\lambda}$, and our stability condition (14), $\lambda < \mu$, we can analogously obtain, using Proposition 4.1 methods, on the second piece, a jointly stationary ergodic pair $\{(W_n^0, V_n^0) : n \in \mathrm{Z}\}$, yielding a stationary ergodic version $\{Z_n^0\}$ of $\{Z_n\}$ satisfying

$$Z_{n+1}^0 = (W_n^0 + D_n^0 - A_n^0)^+ + (V_n^0 + S_n^0 - T_n^0)^+,\ n \in \mathrm{Z}. \tag{18}$$

We can also jointly throw in $\{S_n^0\}$ to obtain a stationary ergodic sojourn time sequence via $R_n^0 = Z_n^0 + S_n^0$. Analogous to Proposition 4.1, we thus obtain:

**Theorem 4.1:** For the AQDTP model with stationary ergodic input satisfying the stability conditions (13) and (14), there always exists a unique stationary ergodic version of total delay and sojourn time. $(W_n, V_n)$ converges in total variation to the joint distribution of $(W^0, V^0)$ regardless of initial conditions, and $Z_n$ converges in total variation to the distribution of $W_0 + V_0$, regardless of initial conditions.

*C. I.I.D. Input Case: Harris recurrence and regeneration of AQDTP*

Here we focus on the special case when each of the following two input sequences, $\{A_n\}$ and $\{(S_n, D_n)\}$, are i.i.d. and independent. Note that we are allowing the two random variables $S_n$ and $D_n$ to be dependent for each $n$, because in applications $D_n$ might even be a function of $S_n$. **In our framework here, a priori, place no restrictions on the kind of dependency/correlation allowed between $S_n$ and $D_n$.**

The above i.i.d assumptions, which we will refer to as *the i.i.d. input case*, in particular imply that the nominal FIFO queueing model forms a FIFO $GI/GI/1$ queue. Moreover, the café delay recursion, $W_{n+1} = (W_n + D_n - A_n)^+$, now endowed with i.i.d. input, implies that $\{W_n : n \geq 0\}$ forms a Markov chain.

Since $T_n = t_{n+1} - t_n = A_n + W_{n+1} - W_n$ we can re-write the recursion for $\{V_n\}$ by using the Markov chain $\{W_n\}$ to drive it:

$$V_{n+1} = (V_n + S_n - T_n)^+ \tag{19}$$

$$= \Big((V_n + S_n - A_n - (W_{n+1} - W_n)\Big)^+ \tag{20}$$

$$= \Big(V_n + S_n - A_n - ((W_n + D_n - A_n)^+ - W_n)\Big)^+. \tag{21}$$

Focusing on (21), and recalling the i.i.d. assumptions, it is immediate that for $M_n \stackrel{\text{def}}{=} (W_n, V_n)$,

$$\{M_n : n \geq 0\}, \text{ forms a Markov chain on } \mathbb{R}_+^2. \tag{22}$$

We next dive deeper. The Markov chain $\{M_n\}$ turns out to be Harris ergodic; Proposition 4.2. (For basics on Harris recurrence/ergodicity, see [3] Chapter VII, Section 3, including Proposition 3.13, Page 205.)

A key feature of Harris recurrent Markov chains is that **they always form regenerative processes**. In Proposition 4.3 we explicitly find two different kinds of regeneration points, **Type I** and **Type II**, which are exhaustive; they cover all the ground. The first type are visits to the empty state, the second type are more elaborate.

**Proposition 4.2:** For the i.i.d. input case satisfying the stability conditions (13) and (14), the Markov chain $M_n = (W_n, V_n)$ is Harris ergodic.

**Proof :** From Theorem 4.1, $\{M_n\}$ is ergodic and converges in total variation to a limiting stationary probability distribution $\pi$, regardless of initial conditions on $M_0$. Thus for $A \subset \mathbb{R}_+^2$, if $\pi(A) > 0$, then regardless of initial conditions, by ergodicity,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{M_i \in A\} = \pi(A) > 0, \text{ wp1};$$

$A$ is visited infinitely often. Thus $\pi$ serves as a recurrence measure; $\{M_n\}$ is positive Harris recurrent by definition. ■

To proceed further, we need an important Lemma:

**Lemma 4.1:** Suppose the stability conditions (13) and (14) hold. Then either

$$P(A > \max\{S, D\}) > 0. \textbf{ Type 1} \tag{23}$$

or

$$P(D > S) > 0 \textbf{ Type 2} \tag{24}$$

must hold. (A natural sufficient condition for obtaining (23) would be that the interarrival time distribution has unbounded support, $P(A > x) > 0,\ x \geq 0$.))

**Proof :** If (23) does not hold, then (24) must hold, for if it did not, then $P(D \leq S) = 1$ implying that $S = \max\{S, D\}$, and thus (23) is equivalent to $P(A > S) > 0$ which indeed holds from the stability condition (14); we get a contradiction. ■

**Proposition 4.3:**

Assume the stability conditions, (13) and (14).

1) **Type I Regeneration**: If (23) also holds, then the successive times when $M_n = (0, 0)$ can be chosen as positive recurrent regeneration points. In particular, total delay, $Z_n = W_n + V_n$, forms a positive recurrent regenerative process, with visits to state 0.
2) **Type II Regeneration**: If (23) does not hold, then (24) does hold (by Lemma 4.1) and in this case positive recurrent regeneration points can be found for $\{M_n\}$ of the form (in distribution upon regeneration) $(X, 0)$ where the construction of the random variable $X$ is given explicitly below in Algorithm 4.1.

**Proof :**

[**Type I Regeneration**]. Since the recursion for $\{W_n\}$ describes a stable $GI/GI/1$ queue, $P_\pi(W_0 = 0) > 0$. Thus there exists a $B > 0$ such that $P_\pi(W_0 = 0,\ V_0 \leq B) > 0$. By Harris recurrence, the event $\{W_n = 0, V_n \leq B\}$ thus occurs infinitely often and does so a positive proportion of time. For a fixed sufficiently small $\delta > 0$, the assumed (23) implies $p = P(A_n > \max\{S_n, D_n\} + \delta) > 0$. If we define $k = [B/\delta]$ (the smallest integer $\geq B/\delta$), and define the event

$$F_n^k = \{A_{n+i} > \max\{S_{n+i}, D_{n+i}\} + \delta,\ 0 \leq i \leq k-1\},$$

then whenever the event $\{W_n = 0, V_n \leq B\}$ occurs, the event $F_n^k$ is independent of it and will occur with probability $p^k = P(F_n^k) > 0$.

Using (20), suppose that for some $n$, both events $\{W_n = 0, V_n \leq B\}$, and $F_n^k$ occur. Then since $W_{n+1} = (W_n + D_n - A_n)^+$, we conclude that $W_{n+i} = 0,\ 0 \leq i \leq k$, implying that

$$\begin{aligned} V_{n+1} &= \Big(\ V_n + S_n - A_n - (W_{n+1} - W_n)\ \Big)^+ \\ &= (V_n + S_n - A_n)^+ \\ &\leq \Big(B - \delta\Big)^+, \end{aligned}$$

and we can continue forward in time step-by-step to obtain $V_{n+2} \leq (B - 2\delta)^+, \cdots, V_{n+k} \leq (B - k\delta)^+ = 0$. Thus we have $W_{n+k} = V_{n+k} = 0$. By the Borel-Cantelli Lemma, the event $\{W_n = 0,\ F_n^k\}$ will occur infinitely often with a positive proportion of times $\geq p^k P_\pi(W_n = 0, V_n \leq B) > 0$.

(The regenerative cycle length distribution is aperiodic: given that $M_n = 0$, there is a positive probability $P(A_n > \max\{S_n, D_n\})$, that $M_{n+1} = 0$ as well.) The proof of Type I regeneration is complete.

**Proof Type II Regeneration**. First, note that since in general (from (3)), $T_n \geq D_n,\ n \geq 0$, we have

$$V_{n+1} = (V_n + S_n - T_n)^+ \leq (V_n + S_n - D_n)^+,\ n \geq 0.$$

We thus can define a new upper bound process $\{\hat{V}_n\}$ by using the recursion

$$\hat{V}_{n+1} = (\hat{V}_n + S_n - D_n)^+,\ n \geq 0, \tag{25}$$

for which it follows that

$$V_n \leq \hat{V}_n,\ n \geq 0,\ \text{if } V_0 = \hat{V}_0. \tag{26}$$

Now choose $B > 0$ sufficiently large so that $P_\pi(W_0 = 0,\ V_0 \leq B) > 0$ which implies the event $\{W_n = 0,\ V_n \leq B\}$ will happen infinitely often. Choose a $\delta > 0$ such that $P(D > S + \delta) > 0$. Define $k = \lceil B/\delta \rceil$, and $F_n^k = \{\{D_{n+i} > S_{n+i} + \delta\},\ 0 \leq i \leq k-1\}$. Now suppose that for some $n$, both the events $\{W_n = 0,\ V_n \leq B\}$ and $F_n^k$ occur. Then similar to the proof of Proposition 4.2 (we use (25) and (26) and set $\hat{V}_n = V_n$), we have $\hat{V}_{n+k} = 0$ and hence $V_{n+k} = 0$.

Meanwhile, the random variable $X = W_{n+k}$ was constructed from only i.i.d. $\{(D_{n+i}, A_{n+i}) : 0 \leq i \leq k-1\}$, *conditional on* $F_n^k$, and is independent of all else; that is how $M_n$ regenerates; next we give a more explicit algorithm for the construction of such as $X$.

**Algorithm 4.1:**

1) Let $\{(S_i, D_i) : 0 \leq i \leq k-1\}$ denote $k$ i.i.d. pairs conditional on each pair satisfying $F_0^k = \{D_i > S_i + \delta\},\ 0 \leq i \leq k-1$.
2) Independently, let $\{A_i : 0 \leq i \leq k-1\}$ be i.i.d..
3) Use as input $\{(A_i, D_i : 0 \leq i \leq k-1\}$ (starting with $W_0 = 0$) in the recursion $W_{n+1} = (W_n + D_n - A_n)^+,\ 0 \leq n \leq k-1$.
4) Set $X = W_k$. Then when a regeneration occurs for $\{M_n\}$ at a time $n+k$, it is distributed as $(X, 0)$.

■

### D. Some Final Remarks

In Proposition 4.3, although the stability conditions imply $P(A > S) > 0$ and $P(A > D) > 0$, they are not strong enough to imply $P(A > \max\{S, D\}) > 0$, when $S$ and $D$ are dependent. Individually, each of $\{V_n\}$ and $\{W_n\}$ will empty infinitely often, a positive proportion of times; but in general, they do not do so at the same time $n$; hence the need to derive more involved regeneration points in such a case. We illustrate here with a **Counterexample:**

Choose $P(A = 2.6) = 1$ and choose

$$(S, D) = \begin{cases} (2,3) & \text{w.p. } 0.5\ , \\ (3,2) & \text{w.p. } 0.5. \end{cases}$$

Then

$$P(A > S) = P(S = 2) = 0.5,$$

and

$$P(A > D) = P(D = 2) = 0.5.$$

But

$$P(A > \max\{S, D\}) = P(A > 3) = 0.$$

To see that $M_n \neq (0,0)$ for $n > 0$, we will show that $W_n$ and $V_n$ move/alternate in opposite directions. Suppose $W_{n+1} - W_n \leq 0$ for some $n$ which can happen only when $(S_n, D_n) = (3, 2)$. Then $T_n = 2.6 + W_{n+1} - W_n \leq 2.6$ and thus

$$V_{n+1} = (V_n + 3 - T_n)^+ \geq (V_n + .4)^+ = V_n + .4;$$

hence $V_{n+1} - V_n \geq 0.4$. Thus if $W_{n+1} - W_n \leq 0$, then $V_{n+1} - V_n > 0$, and if $V_{n+1} - V_n \leq 0$, then $W_{n+1} - W_n > 0$; $M_n \neq (0,0)$ for $n > 0$.

To explicitly characterize the regeneration points of Type II, we choose any $b > 0$ such that $P_\pi(W_0 = 0, V_0 \leq b) > 0$, then find a minimal such $b$. Suppose the event $\{W_n = 0, V_n \leq b\}$ occurs. We can then condition on alternating $\{(S_{n+i}, D_{n+i}) : 0 \leq i \leq m-1\} = \{(2,3),(3,2),(2,3),\ldots,(3,2)\}$, for any length $m$, which occurs with positive probability $(1/2)^m$. Thus $W_{n+1} = 0.4, W_{n+2} = 0, W_{n+3} = 0.4, \ldots$ and so on, alternating between $0.4$ and $0$. Thus $T_{n+i} = 3$ for even $i$ and $T_{n+i} = 2.2$ for odd $i$. Thus $V_{n+i}$ goes down by $1$ and up by $0.8$ until we have $V_{n+i} = 0$ for some $i$. If $V_{n+i} = 0$, we must have $W_{n+i} = 0.4$ (since $M_n \neq (0,0)$ for $n > 0$), and hence $P_\pi(W_0 = 0, V_0 = 0.4) > 0$ holds. Thus, as regeneration points we can take those consecutive times $n$ such that $M_n = (0.4, 0)$.

There are other queueing model examples of this sort of phenomena: the classic FIFO $GI/GI/c$ queue with $c \geq 2$ can be stable but such that the system will never be found empty by an arrival. (One needs $P(A > S) > 0$ for it to empty.) For $c = 2$, for example, just take $A_n = 1.5,\ n \geq 0,\ S_n = 2,\ n \geq 0$. Then $\rho = \lambda/\mu = 4/3 < 2$, so stability holds. But all arriving customers (after $n = 0$) will find one server free, but the other busy. Nonetheless, for any stable ($\rho < c$) FIFO $GI/GI/c$ queue, regeneration points can be found (see, for example, Chapter 7, Section 2, Page 344 in [3]). For other classic examples, see [22] and [24].


## Acknowledgement

Erol Gelenbe's research in this paper has been supported by the European Commission's H2020 DOSS Project under Grant Agreement No. 101120270.


## References


[1] Adnan Aijaz and A. Hamid Aghvami. "Cognitive Machine-to-Machine Communications for Internet-of-Things: A Protocol Stack Perspective". In: *IEEE Internet of Things Journal* 2.2 (2015), pp. 103–112. DOI: 10.1109/JIOT.2015.2390775.

[2] Adnan Aijaz et al. "CRB-MAC: A Receiver-Based MAC Protocol for Cognitive Radio Equipped Smart Grid Sensor Networks". In: *IEEE Sensors Journal* 14.12 (2014), pp. 4325–4333. DOI: 10.1109/JSEN.2014.2346430.

[3] Soren Asmussen. *Applied Probability and Queues (2nd Edition)*. Springer, 2003.

[4] D. O. Awduche. "MPLS and Traffic Engineering in IP Networks". In: *IEEE Comm. Mag.* 37.12 (1999), 42–47.

[5] H. Brandt. *ATM*. John Wiley & Sons, Chichester, England, 2001.

[6] CISCO. "Comparing Traffic Policing and Traffic Shaping for Bandwidth Limiting". In: *Cisco Tech Notes, Document ID: 19645* (2005).

[7] Erol Gelenbe and Edith Ngai. "Adaptive QoS routing for significant events in wireless sensor networks". In: *2008 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems* (2008), pp. 410–415. DOI: 10.1109/MAHSS.2008.4660035.

[8] Erol Gelenbe and Edith Ngai. "Adaptive Random Re-Routing for Differentiated QoS in Sensor Networks". In: *The Computer Journal* 53.7 (2010), pp. 1052–1061. DOI: 10.1093/comjnl/bxp082.

[9] Erol Gelenbe and Karl Sigman. "IoT Traffic Shaping and the Massive Access Problem". In: *IEEE International Conference on Communications* (2022), pp. 2732–2737. DOI: 10.1109/ICC45855.2022.9839054.

[10] Erol Gelenbe et al. "Diffusion Analysis Improves Scalability of IoT Networks to Mitigate the Massive Access Problem". In: *2021 29th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)* (2021), pp. 1–8. DOI: 10.1109/MASCOTS53633.2021.9614289.

[11] Fayezeh Ghavimi and Hsiao-Hwa Chen. "M2M Communications in 3GPP LTE/LTE-A Networks: Architectures, Service Requirements, Challenges, and Applications". In: *IEEE Communications Surveys & Tutorials* 17.2 (2015), pp. 525–549. DOI: 10.1109/COMST.2014.2361626.

[12] J. Helzer and L. Xu. "Congestion control for multmedia streaming with self-limiting sources". In: *13TH IEEE International Conference on Network Protocols (ICNP'05), Nov. 6-9*. 2005. URL: http://csr.bu.edu/icnp2005/posters/helzer.pdf.

[13] IETF. "An Architecture for Differentiated Services". In: *ETF RFC 2475* (1988).

[14] ITU. "Traffic control and congestion control in B-ISDN". In: *ITU-T Recommendation I.371* (2004).

[15] Liang Liang et al. "A Cluster-Based Congestion-Mitigating Access Scheme for Massive M2M Communications in Internet of Things". In: *IEEE Internet of Things Journal* 5.3 (2018), pp. 2200–2211. DOI: 10.1109/JIOT.2018.2828030.

[16] Shao-Yu Lien et al. "Cooperative Access Class Barring for Machine-to-Machine Communications". In: *IEEE Transactions on Wireless Communications* 11.1 (2012), pp. 27–32. DOI: 10.1109/TWC.2011.111611.110350.

[17] Tzu-Ming Lin et al. "PRADA: Prioritized Random Access With Dynamic Access Barring for MTC in 3GPP LTE-A Networks". In: *IEEE Transactions on Vehicular Technology* 63.5 (2014), pp. 2467–2472. DOI: 10.1109/TVT.2013.2290128.

[18] R.M. Loynes. "The stability of queues with nonindependent inter-arrival and service times". In: *Proceedings of Cambridge Philosophical Society* 58 (1962), pp. 497–520.

[19] Mert Nakip et al. "Joint Forecasting-Scheduling for the Internet of Things". In: *2019 IEEE Global Conference on Internet of Things (GCIoT)* (2019), pp. 1–7. DOI: 10.1109/GCIoT47977.2019.9058408.

[20] Ieryung Park, Dohyun Kim, and Dongsoo Har. "MAC Achieving Low Latency and Energy Efficiency in Hierarchical M2M Networks With Clustered Nodes". In: *IEEE Sensors Journal* 15.3 (2015), pp. 1657–1661. DOI: 10.1109/JSEN.2014.2364055.

[21] Volkan Rodoplu et al. "Multi-Channel Joint Forecasting-Scheduling for the Internet of Things". In: *IEEE Access* 8 (2020), pp. 217324–217354. DOI: 10.1109/ACCESS.2020.3038358.

[22] Karl Sigman. "Regeneration in Tandem Queues with Multi-server Stations". In: *Journal of Applied Probability* 25 (1988), pp. 391–403.

[23] Karl Sigman. *Stationary Marked Point Processes: An Intuitive Approach*. Chapman and Hall: New York, 1995.

[24] Karl Sigman. "The Stability of Open Queueing Networks". In: *Stochastic Processes And Their Applications* 35.1 (1988), pp. 11–25.