A SINGLE CHANNEL ON/OFF MODEL WITH TCP-LIKE CONTROL

MILAN BORKOVEC, AMITES DASGUPTA, SIDNEY RESNICK, AND GENNADY SAMORODNITSKY

ABSTRACT. We model behavior of a TCP-like source transmitting over a single channel to a server that processes work at constant rate r. Transmission by the source follows an *on/off* mechanism. When the overall load in the system is below a critical constant γ, transmission rates increase linearly but when the load exceeds γ, then transmission rates decrease geometrically fast. We study the system by means of an embedded Markov chain which gives the buffer content at the start of transmissions. Attention is paid to the time necessary to transmit a file of size L and both the tail behavior and expectation of the distribution of file transmission time are considered.

1. INTRODUCTION

Recent experimental work has shown that traffic carried in high-speed networks, for example, *Local Area Networks* (LAN) [34], *Wide Area Networks* (WAN) [22] and *Variable-Bit-Rate Video* (VAR Video) [4], has features like long-range dependence and self-similarity which are strikingly different from those of traditional voice traffic. Such observations have helped stimulate interest in the modeling and statistical analysis of traffic in modern data networks.

Modern data networks are robust and scalable and force congestion to stay with the end user. One of the networking features responsible for these good properties is the protocol TCP ([16]). TCP is the abbreviation for *Transport Control Protocol* and is responsible for controlling the rate of packet transmission by a sender and verifying the correct delivery of data packets to a destination. TCP adds support to detect error or lost packets and to trigger retransmission until the data is correctly and completely received. TCP controls transmission rates and roughly speaking can be thought to allow transmission rates to linearly increase until packet loss is detected by a lack of an acknowledgment at which time the rate of successive transmissions is halved until an acknowledgment comes back.

Our goal is to find a modest setting in which the effect of a TCP-like control can be mathematically studied. In particular, since TCP is a decentralized mechanism for controlling transmission rates, we are interested in seeing the effect of the control on a user who transmits a file. One of the interesting conclusions of our study is that the specific manner in which geometric decrease occurs does not affect the asymptotic form of the distribution of the time to transmit a file.

Natural models for high-speed network traffic include fluid queues fed by *on/off* sources with heavy-tailed *on-* and/or *off-*periods. Such models offer some mathematical tractability and an explanation of observed long range dependence in the packet count per unit time data. There is a large body of recent literature which uses such models for modeling network traffic. See, for example, [1, 6, 14, 28, 9, 17, 29, 34, 35, 15, 36] and the references therein.

The basic fluid model, which we call the *classical on/off* model, consists of a single idealized source feeding a server. The single channel of this model alternates between an *on* state, in

which the source transmits data at a constant rate, and an *off* state in which the source does not transmit. Durations of *on* and *off* periods are independent; *on* times are identically distributed, and so are *off* times. Experimental evidence ([8] or [35]) suggests that *on* and *off* periods are well modeled by heavy tailed distributions, i.e. distributions of the form

$$P(X > x) = x^{-\kappa} \ell(x) \quad x > 0,$$

where $1 < \kappa < 2$ and $\ell(x)$ is a slowly varying function; that is,

$$\lim_{t \to \infty} \frac{\ell(tx)}{\ell(t)} = 1, \quad \forall\, x > 0.$$

It is now well known that heavy tails for the *on* periods in the *on/off* model induce long range dependence in the transmission rate process.

The single channel model cannot of course model all the interactions occurring in an extensive network but allows mathematical tractability and gives an indication of local behavior in a network. Natural performance measures which have been considered to date include the buffer content process and the activity periods when the buffer content is in the stable state. We refer, for instance, to [6], [7], [24] and references therein.

Many theoretical models assume infinite buffer capacity for simplicity. The infinite capacity assumption is tolerated because one can choose a high threshold (the *overflow level*) and then make the assumption that when buffer content reaches this level, excess arriving work gets lost. The probability that this happens is the so-called *loss probability* or *loss fraction*. Some results on time to buffer overflow in such models with heavy tailed *on* periods are given in [12], [14], [23] and [36]. In [36] the influence of heavy-tailed input on loss fraction and mean buffer content for particular fluid queueing models is investigated.

In this paper, we present a single channel *on/off* model with TCP-like control mechanism which is designed to make buffer overflow extremely unlikely. The proposed model is in the spirit of the congestion avoidance achieved by TCP; see e.g. [19],[20] and [33]. The control mechanism presented in this article ensures that the buffer content process is controlled above a high threshold. In contrast to the result in the classical *on/off* model that heavy-tailed *on* periods leads to a heavy-tailed stationary distribution of the content process, here, the stationary content process distribution always has an exponential tail, no matter how heavy is the input tail.

Here is an informal description of our *on/off* model with control. Let $\{L_n\}$ be an iid sequence of non-negative random variables representing successive job sizes or file sizes needing to be transmitted by the source. Following transmissions there are *off* periods where no work is transmitted and the lengths of these *off* periods is represented by iid random variables $\{Y_n\}$. During an *on* period, if the buffer content is less than a threshold γ, the transmission rate linearly increases like $1 + \alpha t$ ($t \geq 0$, $\alpha \geq 0$). If the buffer content exceeds γ during an *on* period, then the transmission rate decreases exponentially fast like the function e^{-dt}, ($t \geq 0$ and $d > 0$). There is a constant service rate r, so that whenever the system is nonempty, work leaves the system at rate r. We assume $r \in (0, 1)$ in order that the empty buffer content increases immediately when load starts to be transmitted. The condition $r \in (0, 1)$ ensures in particular that the state 0 of the content process is not absorbing in the case $\alpha = 0$. The classical *on/off* model is included in this model by choosing $\gamma = \infty$ and $\alpha = 0$. Therefore, the *on/off* model proposed in this paper can be considered as an extension of the classical one and we refer to it as the generalized *on/off* model.

In studying the generalized *on/off* model we use techniques from the theory of Markov chains in discrete time with state space $[0, \infty)$ ([18] or [30]). Renewal considerations come from a particular Markov chain $\{X_n\}$ representing buffer content when a new transmission begins.

The paper is organized as follows. In Section 2 we construct the generalized *on/off* model with TCP-like control and show some basic properties. The model is built by first defining the cumulative input function and the transmission times for a file starting from an arbitrary

buffer content state. We also define the discrete-time Markov chain $\{X_n\}$ representing buffer content at the start of a new transmission. Stability questions phrased in terms of the stationary distribution of this chain are discussed in Section 3. Under appropriate conditions on L and Y, we show stationarity of $\{X_n\}$ and hence of the buffer content. Moreover, we prove that the stationary distribution of the buffer content is light-tailed. Unlike the classical *on/off* model without control, the transmission time of a file of length L is likely to be longer than L and Section 4 provides some quantification of the increase in transmission time. Several numerical results conclude this article.

There is a mushrooming literature on the influence of TCP in the data network literature and as a sample, we cite the following in addition to references already mentioned. Gilbert et al [11] study global synchronization of multiplexed traffic from TCP sources in a simplified model without queuing. The purpose of the paper is to gain understanding of how the TCP induced tendency towards oscillation and synchronization affect bottleneck links. Altman et al [2] study throughput (that is, the time average of the input rate process) of a TCP source transmitting an infinite file assuming packet losses occur at times constituting a stationary ergodic sequence. Padhye et al [21] also seek an analytic expression for throughput of a TCP source sending an infinite amount of data. Their answer is a function of the round trip time (the time until an acknowledgment from the destination is received for a transmitted packet) and loss rates. They assume that the events [packets lost in "round" i], $i \geq 1$, are independent events and use renewal reward techniques for calculating throughput. A survey of methods of controlling communication networks is given by Srikant [27].

2. The model

In what follows we formally construct our single channel *on/off* model which we already introduced loosely in the introduction. The construction is basically realized in four steps.

Cumulative input function $I_x(t)$. We first model cumulative input by a sender transmitting an infinite file and commencing transmission when buffer content is x. This cumulative input t-time units from the beginning of the transmission is denoted by $I_x(t)$. We let $I_x(\cdot)$ depend crucially on x and the control level $\gamma > 0$. In particular, $dI_x(t)/dt$ should decrease exponentially of the form e^{-dt}, $d > 0$, for small t and $x > \gamma$ and increase linearly with $1 + \alpha t$, $\alpha > 0$, for $t \geq 0$ and $x < \gamma$.

Assume first $x \geq \gamma$. Begin by defining $t_0(x)$ to be the amount of time necessary for buffer content to go from x to γ or more formally

$$t_0(x) := \inf\{u > 0 : x + \int_0^u e^{-dw}dw - ru = \gamma\}$$
$$(2.1) \qquad = \inf\{u > 0 : x + d^{-1}(1 - e^{-du}) - ru = \gamma\}, \quad d > 0.$$

If the file size is infinite, then at time $t_0(x)$ the transmission rate is reset back to 1 again and transmission continues at an exponentially decreasing rate. This procedure is repeated at times $t_0(x) + n\,t_0(\gamma)$, $n = 1, 2, \ldots$.

For $x < \gamma$, let $t(x)$ be the amount of time required for buffer content to move from x to γ, i.e.

$$t(x) = \inf\{u > 0 : x + \int_0^u (1 + \alpha s)ds - ru = \gamma\}$$
$$(2.2) \qquad = \inf\{u > 0 : x + u + \frac{1}{2}\alpha u^2 - ru = \gamma\}$$
$$= (1 - r)\alpha^{-1}\left(-1 + \sqrt{1 + 2\alpha(\gamma - x)(1 - r)^{-2}}\right), \quad \alpha \geq 0.$$

If $t < t(x)$, then the cumulative input rate is

$$I_x(t) = \int_0^t (1 + \alpha u) du = t + \frac{1}{2} \alpha t^2.$$

At time $t(x)$, the input rate $1 + \alpha t(x)$ starts to decrease exponentially and is of the form

$$(1 + \alpha t(x)) e^{-d(\cdot)}$$

until $t(x) + t_\alpha(x)$, where $t_\alpha(x)$ is the time needed for buffer content, starting from γ, to again hit γ when initial input rate is $1 + \alpha t(x)$. More precisely,

$$
\begin{aligned}
t_\alpha(x) &= \inf\{u > 0 : \gamma + (1 + \alpha t(x)) \int_0^u e^{-dw} dw - ru = \gamma\} \\
&= \inf\{u > 0 : (1 + \alpha t(x))(1 - e^{-du})/d - ru = 0\}.
\end{aligned}
$$

At time $t(x) + t_\alpha(x)$, we assume the input rate is reset back to 1 again and we are back in the framework of $x \geq \gamma$ discussed at the beginning.

Having this construction in mind, we derive next the precise formula for the cumulative input function and some properties of associated auxiliary functions. Suppose again first that $x \geq \gamma$. From the definition of $t_0(x)$ in (2.1) we have

$$(2.3) \qquad d^{-1}(1 - e^{-dt_0(x)}) - rt_0(x) = \gamma - x \leq 0.$$

In $[0, t_0(x)]$, the sender transmits therefore

$$(2.4) \qquad d^{-1}(1 - e^{-dt_0(x)}) = \gamma - x + rt_0(x)$$

units of work. Since at time $t_0(x)$ the transmission rate is put back to 1 and transmission goes on at an exponentially decreasing rate, we thus have for

$$(2.5) \qquad t \in [t_0(x) + nt_0(\gamma), t_0(x) + (n+1)t_0(x)) =: J_n(x), \quad n = 0, 1, 2, \ldots$$

$I_x(t)$ is the work inputted in $[0, t_0(x)]$ plus the work inputted in $(t_0(x), t]$ or

$$I_x(t) = (\gamma - x + rt_0(x)) + nrt_0(\gamma) + d^{-1}\left(1 - e^{-d(t - (t_0(x) + nt_0(\gamma)))}\right).$$

To summarize, the cumulative input function for $x \geq \gamma$ is defined as

$$(2.6) \quad I_x(t) = \begin{cases} d^{-1}(1 - e^{-dt}), & 0 \leq t < t_0(x), \\ \gamma - x + rt_0(x) + nrt_0(\gamma) + d^{-1}(1 - e^{-d(t - nt_0(\gamma) - t_0(x))}), & t \in J_n(x), \end{cases}$$

where $J_n(x)$ is defined in (2.5) and with similar considerations for $x < \gamma$,

$$(2.7)$$

$$I_x(t) = \begin{cases} t + \alpha t^2/2, & 0 \leq t < t(x), \\ t(x) + \alpha t(x)^2/2 + (1 + \alpha t(x)) d^{-1}(1 - e^{-d(t - t(x))}), & t(x) \leq t < t(x) + t_\alpha(x), \\ t(x) + \alpha t(x)^2/2 + rt_\alpha(x) + nrt_0(\gamma) & t(x) + t_\alpha(x) + nt_0(\gamma) \leq t \\ \quad + d^{-1}(1 - e^{-d(t - t(x) - t_\alpha(x) - nt_0(\gamma))}), & \quad < t(x) + t_\alpha(x) + (n+1)t_0(\gamma), \end{cases}$$

where $n = 0, 1, 2, \ldots$.

Note that $t(x)$ is strictly decreasing in $x < \gamma$ and that

$$\lim_{x \uparrow \gamma} t(x) = 0.$$

Extending the definitions of $t(x)$ and $t_\alpha(x)$ to γ by continuity produces $t_\alpha(\gamma) = t_0(\gamma)$. Furthermore, if a transmission starts at γ, the transmission rate at time $t_0(\gamma)$ is $\exp\{-dt_0(\gamma)\}$, whereas if a transmission starts at $x < \gamma$, then when the buffer content moves to γ and then returns to γ, the input rate is $(1 + \alpha t(x)) \exp\{-dt_\alpha(x)\}$.

The next lemma gives some insight into the discontinuity of the input rate at time $t(x) + t_\alpha(x)$, $x \in [0, \gamma)$ and $t_0(x)$, $x \geq \gamma$, respectively, i.e. at the time when the content process hits the threshold γ for the first time from above and the input rate becomes 1 again. It is clear that the input rate always has a positive jump at the end of each cycle. The lemma shows that the size of the jump depends crucially on the initial state x.

Lemma 2.1. *(a) Let $x \in [0, \gamma)$ and $i_x(t(x) + t_\alpha(x)) = dI_x(t)/dt \big|_{t=t(x)+t_\alpha(x)}$ be the input rate at $t(x) + t_\alpha(x)$. Then, $i_x(t(x) + t_\alpha(x)) = (1 + \alpha t(x)) e^{-dt_\alpha(x)}$ and $i_x(t(x) + t_\alpha(x))$ is strictly monotone increasing in x, i.e. for every $0 \leq x_1 < x_2 < \gamma$*

$$i_{x_1}(t(x_1) + t_\alpha(x_1)) < i_{x_2}(t(x_2) + t_\alpha(x_2)).$$

In particular, for every $x \in [0, \gamma)$

$$i_x(t(x) + t_\alpha(x)) < \lim_{y \uparrow \gamma} i_y(t(y) + t_\alpha(y)) = e^{-dt_0(\gamma)} < 1.$$

(b) Let $x \geq \gamma$ and $i_x(t_0(x)) = dI_x(t)/dt \big|_{t=t_0(x)}$ be the input rate at $t_0(x)$. Then, $i_x(t_0(x)) = e^{-dt_0(x)}$ and $i_x(t_0(x))$ is strictly monotone decreasing in x; that is, for every $\gamma \leq x_1 \leq x_2$

$$i_{x_1}(t_0(x_1)) > i_{x_2}(t_0(x_2)).$$

In particular, for every $x \geq \gamma$

$$i_x(t_0(x)) < \lim_{y \downarrow \gamma} i_y(t_0(y)) = e^{-dt_0(\gamma)} < 1.$$

Proof. Here, we prove only statement (a). (b) follows immediately from the fact that $t_0(x)$ is strictly monotone increasing in x for $x \geq \gamma$.

Let $0 \leq x_1 < x_2 < \gamma$ be arbitrary. Clearly, $t_\alpha(x_2) < t_\alpha(x_1)$. This follows from the definition of $t_\alpha(x)$ and the fact that $t(\cdot)$ is strictly decreasing in $[0, \gamma)$. The content process started at any $x < \gamma$ has at time $t(x) + t_\alpha(x)$ the rate

$$(2.8) \qquad i_x(t(x) + t_\alpha(x)) = (1 + \alpha t(x)) e^{-dt_\alpha(x)},$$

Consider the content process started at x_1. At time $t(x_1) + t_\alpha(x_2) < t(x_1) + t_\alpha(x_1)$, the content process is in state

$$\gamma + (1 + \alpha t(x_1)) \frac{1 - e^{-dt_\alpha(x_2)}}{d} - rt_\alpha(x_2) = \gamma + (1 + \alpha t(x_1)) \frac{r}{1 + \alpha t(x_2)} t_\alpha(x_2) - rt_\alpha(x_2)$$

$$(2.9) \qquad\qquad\qquad\qquad = \gamma + rt_\alpha(x_2) \left(\frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} - 1 \right) > \gamma.$$

The input rate at $t(x_1) + t_\alpha(x_2)$ is $(1 + \alpha t(x_1)) e^{-dt_\alpha(x_2)}$ and the extra time t^* needed for this rate to decrease to $(1 + \alpha t(x_2)) e^{-dt_\alpha(x_2)}$ satisfies

$$(1 + \alpha t(x_1)) e^{-d(t_\alpha(x_2) + t^*)} = (1 + \alpha t(x_2)) e^{-dt_\alpha(x_2)}$$

so that

$$(2.10) \qquad t^* := \frac{1}{d} \log \frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)}.$$

In what follows we show that at the time $t(x_1) + t_\alpha(x_2) + t^*$ the content process started at x_1 is still above γ, i.e. $t(x_1) + t_\alpha(x_2) + t^* < t(x_1) + t_\alpha(x_1)$ and thus $i_{x_1}(t(x_1) + t_\alpha(x_1)) < i_{x_1}(t(x_1) + t_\alpha(x_2) + t^*) = i_{x_2}(t(x_2) + t_\alpha(x_2))$.

Note that because of (2.8) and (2.10), the extra input from $t(x_1) + t_\alpha(x_2)$ to $t(x_1) + t_\alpha(x_2) + t^*$ is

$$
\begin{aligned}
(1 + \alpha t(x_1)) e^{-dt_\alpha(x_2)} \int_0^{t^*} e^{-du} du &= (1 + \alpha t(x_1)) e^{-dt_\alpha(x_2)} d^{-1} \left(1 - e^{-\log 1 + \alpha t(x_1)/1 + \alpha t(x_2)} \right) \\
&= (1 + \alpha t(x_1)) e^{-dt_\alpha(x_2)} d^{-1} \left(1 - \frac{1 + \alpha t(x_2)}{1 + \alpha t(x_1)} \right) \\
&= e^{-dt_\alpha(x_2)} \frac{\alpha (t(x_1) - t(x_2))}{d} .
\end{aligned}
$$
(2.11)

But,

$$
(1 + \alpha t(x_2)) \frac{1 - e^{-dt_\alpha(x_2)}}{d} = rt_\alpha(x_2)
$$

or equivalently

$$
e^{-dt_\alpha(x_2)} = 1 - \frac{rdt_\alpha(x_2)}{(1 + \alpha t(x_2))} .
$$
(2.12)

Therefore, by (2.11) and (2.12), the extra input is

$$
\frac{\alpha (t(x_1) - t(x_2))}{d} \left(1 - \frac{rdt_\alpha(x_2)}{(1 + \alpha t(x_2))} \right) = \frac{\alpha (t(x_1) - t(x_2))}{d} - \left(\frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} - 1 \right) rt_\alpha(x_2) .
$$
(2.13)

From (2.9) and (2.13), we conclude that the buffer state at time $t(x_1) + t_\alpha(x_2) + t^*$ is

$$
\begin{aligned}
\gamma + rt_\alpha(x_2) \left(\frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} - 1 \right) &+ \frac{\alpha (t(x_1) - t(x_2))}{d} - \left(\frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} - 1 \right) rt_\alpha(x_2) - r\, t^* \\
&= \gamma + \frac{\alpha (t(x_1) - t(x_2))}{d} - r\, t^* .
\end{aligned}
$$
(2.14)

Note now that

$$
\begin{aligned}
\frac{\alpha (t(x_1) - t(x_2))}{d} &= (1 + \alpha t(x_2)) d^{-1} \left(\frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} - 1 \right) \\
&\geq d^{-1} \left(\frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} - 1 \right) > \frac{1}{d} \log \frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} > \frac{r}{d} \log \frac{1 + \alpha t(x_1)}{1 + \alpha t(x_2)} = r\, t^* .
\end{aligned}
$$
(2.15)

Therefore, by the time $t(x_1) + t_\alpha(x_2) + t^*$ the state of the system is still above γ and so the time $t(x_1) + t_\alpha(x_1)$ has not been reached. $\qquad \square$

Transmission duration $\tau(x, l)$. The definition of the cumulative input function $I_x(\cdot)$ given above assumed transmission of a file of infinite size. When transmission of a file of size l commences with initial buffer content equal to x, the time necessary for transmitting the whole file is

$$
\tau(x, l) = \inf\{ t > 0 : I_x(t) = l \} =: I_x^{\leftarrow}(l)
$$
(2.16)

where $I_x^{\leftarrow}(\cdot)$ is the inverse of the continuous strictly increasing function $I_x(\cdot)$. Due to (2.6) and (2.7) $\tau(x, l)$ can be expressed more explicitly. For $x \geq \gamma$ we have

$$
\tau(x, l) = \begin{cases} -d^{-1} \log(1 - dl), & l \leq \gamma - x + rt_0(x), \\ t_0(x) + I_\gamma^{\leftarrow}(l - \gamma + x - rt_0(x)) \}, & \text{otherwise.} \end{cases}
$$
(2.17)

and for $x < \gamma$

(2.18)
$$
\tau(x, l) = \begin{cases}
\alpha^{-1}(-1 + \sqrt{1 + 2\alpha l}), & l \leq t(x) + \alpha t^2(x)/2, \\
t(x) - d^{-1} \log\left(1 - (1 + \alpha t(x))^{-1} d(l - t(x) - \alpha t(x)^2)\right), & t(x) + \alpha t(x)^2/2 < l \\
& \leq t(x) + \alpha t(x)^2/2 + rt_\alpha(x), \\
t(x) + t_\alpha(x) + I_\gamma^\leftarrow\left(l - t(x) - \alpha t(x)^2/2 - rt_\alpha(x)\right), & \text{otherwise.}
\end{cases}
$$

Straightforward but tedious analysis yields the following lemma.

Lemma 2.2. *The transmission time $\tau(x, l)$ defined in (2.17) and (2.18) is for fixed l continuous in x.*

Proof. The result follows from the continuity of $I_\gamma(\cdot), t(\cdot), t_\alpha(\cdot)$ and $t_0(\cdot)$. $\qquad\square$
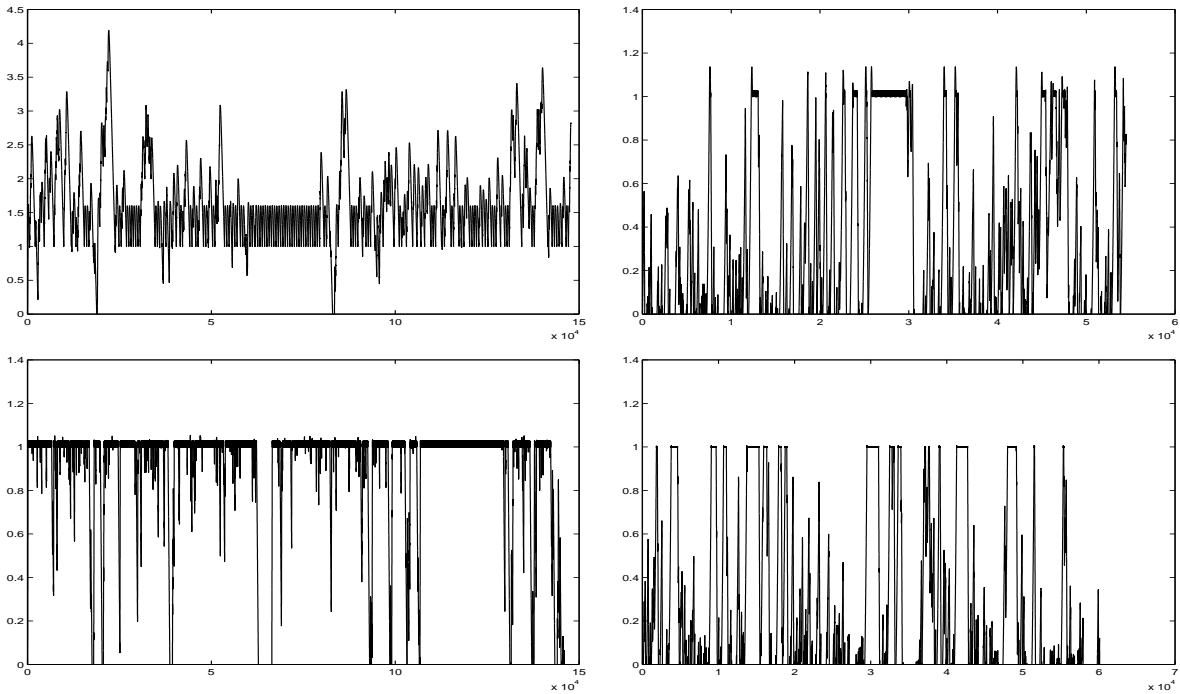

FIGURE 1. Simulated buffer contents with 300 *on/off* periods at each case. The chosen parameters are $\alpha = 0.1$, $\gamma = 1$, $d = 0.8$ (top) and 20 (bottom) as well as $r = 0.2$ (left) and 0.8 (right). Transmission loads are Pareto(2)-distributed and *off*-periods have Pareto distributions with index $\kappa = 2$. The maximal increasing stepsizes are $\delta^{(0.1)} = 0.72$ and $\delta = 0.60$ (left,top), $\delta^{(0.1)} = 0.06$ and $\delta = 0.03$ (right,top), $\delta^{(0.1)} = 0.04$ and $\delta = 0.02$ (left,bottom) respectively $\delta^{(0.1)} = 0.005$ and $\delta = 0.001$ (right,bottom).

The embedded Markov chain $\{X_n\}$. Let $\{L_n\}$ and $\{Y_n\}$ be the sequences of *on* and *off* periods as defined in the introduction. Given $\{L_n, Y_n\}$ and the transmission duration function $\tau(x, l)$, we may define a Markov chain $\{X_n, n \geq 0\}$ by the stochastic recurrence equation

(2.19)
$$
X_n = (X_{n-1} + L_n - r\tau(X_{n-1}, L_n) - rY_n)_+ , \quad n = 1, 2, \ldots,
$$

where $X_0 = x \geq 0$ is arbitrary. We will write $P_x(\cdot), E_x(\cdot)$ to denote probabilities and expectations computed under the assumption that $X_0 = x$. We will also denote the Markov kernel by

$$
P_x[X_1 \in dy] =: p(x, dy).
$$

Moreover, $P^n[x, \cdot]$ denotes the n-step transition probability of $\{X_n\}$, i.e. $P^n[x, \cdot] = P[X_n \in \cdot \,|\, X_0 = x]$ for every $x \geq 0$ and $n \in \mathbb{N}$. The stochastic recurrence equation defines the new state as a function of the old state and independent quantities and hence defines a Markov chain $\{X_n\}$ that represents buffer content when new transmissions commence. By Lemma 2.2 and the recurrence relation (2.19) it is readily seen that $\{X_n\}$ is weak Feller (i.e. the operator $p(x, \cdot)$ maps bounded continuous functions into bounded continuous functions). This property will be helpful in section 3 for showing the existence of a stationary distribution. In particular, we will see that under appropriate conditions on L and Y the process $\{X_n\}$ is (geometrically) ergodic. This implies that the transmission times $(\tau(X_{n-1}, L_n))$ converge uniformly to a well-specified, non-degenerate random variable $\tau(X, L)$.

A time scale for the Markov chain. Given the Markov chain $\{X_n\}$, we define a continuous time buffer content process $\{X(t), t \geq 0\}$. We will do this by defining a natural time scale. Set $S_0 = 0$ and recursively define

$$(2.20) \qquad S_n = S_{n-1} + \tau(X_{n-1}, L_n) + Y_n, \quad n = 1, 2, \ldots.$$

S_n is the time when the transmission labeled "n" begins. The cycle $[S_n, S_{n+1})$ consists of an *on* period of length $\tau(X_{n-1}, L_n)$ and an *off* period of duration Y_n.

The continuous time buffer content process $\{X(t), t \geq 0\}$. The buffer content process of the generalized *on/off* model with control can be now defined by interpolating between the points $\{S_n\}$. If $S_{n-1} \leq t \leq S_{n-1} + \tau(X_{n-1}, L_n)$ for some $n = 1, 2, \ldots,$

$$(2.21) \qquad X(t) = X_{n-1} + I_{X_{n-1}}(t - S_{n-1}) - r(t - S_{n-1})$$
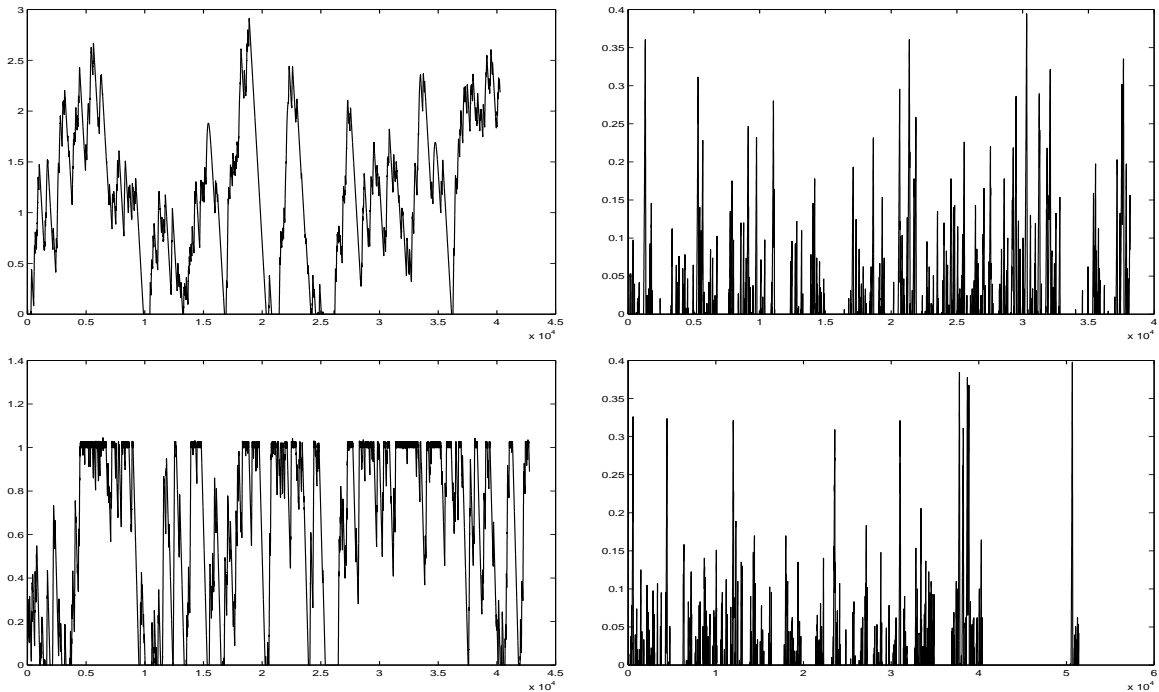

FIGURE 2. Simulated buffer contents with 300 *on/off* periods at each case. The chosen parameters are $\alpha = 0.1$, $\gamma = 1$, $d = 0.8$ (top) and 20 (bottom) as well as $r = 0.2$ (left) and 0.8 (right). Transmission loads are Expo(4)-distributed and *off*-periods have Pareto distributions with index $\kappa = 2$. The maximal increasing step sizes are the same as in Figure 1.

and if $S_{n-1} + \tau(X_{n-1}, L_n) < t < S_n$ then

$$(2.22) \qquad X(t) = (X(S_{n-1} + \tau(X_{n-1}, L_n)) - r(t - S_{n-1} - \tau(X_{n-1}, L_n)))_+ \ .$$

Figure 1 and 2 present several simulations of the buffer content process for different parameters.

The present model is a generalization of the standard *on/off* model. The standard model is recovered by setting $\gamma = \infty$ and $\alpha = 0$. Note that if we retain $\gamma < \infty$ but set $\alpha = 0$ then we get the model which has transmissions proceeding with constant rate inputs when initial buffer content is $x < \gamma$. In this case, for $x \geq \gamma$

$$(2.23) \qquad I_x(t) = \begin{cases} d^{-1}(1 - e^{-dt}), & 0 \leq t < t_0(x), \\ \gamma - x + rt_0(x) + nrt_0(\gamma) & t_0(x) + nt_0(\gamma) \leq t \\ \quad + d^{-1}(1 - e^{-d(t - nt_0(\gamma) - t_0(x))}), & < t_0(x) + (n+1)t_0(\gamma), \end{cases}$$

and

$$(2.24) \qquad \tau(x, l) = \begin{cases} -\log(1 - dL)/d, & l \leq \gamma - x + rt_0(x), \\ t_0(x) + I_\gamma^\leftarrow(l - \gamma + x - rt_0(x))\}, & \text{otherwise,} \end{cases}$$

while for $x < \gamma$

$$(2.25)\, I_x(t) = \begin{cases} t, & 0 \leq t < (\gamma - x)/(1 - r), \\ (\gamma - x)/(1 - r) + nrt_0(\gamma) & (\gamma - x)/(1 - r) + nt_0(\gamma) \leq t \\ \quad + d^{-1}(1 - e^{-d(t - nt_0(\gamma) - (\gamma - x)/(1 - r))}), & < (\gamma - x)/(1 - r) + (n+1)t_0(\gamma) \end{cases}$$

and

$$(2.26) \quad \tau(x, l) = \begin{cases} l, & l \leq (\gamma - x)/(1 - r), \\ (\gamma - x)/(1 - r) + I_\gamma^\leftarrow(L - (\gamma - x)/(1 - r)), & \text{otherwise.} \end{cases}$$

One of our objectives is to study the transmission durations $\{\tau_n = \tau(X_{n-1}, L_n), n \geq 1\}$. Since the underlying philosophy behind TCP is to force congestion to the end users, it is of interest to see to what extent the control responsible for keeping buffer content stable also prolongs transmission times for the sender. In what follows we show that there exists a limit distribution for τ_n as n tends to infinity and we discuss how this distribution depends on L. Such discussion is framed in terms of tail behavior and expectation. We study thus the Markov chain $\{X_n\}$ in more detail.

Note that while the Markov chain $\{X_n\}$ may increase with positive probability, it does so only in a very controlled way. Because of the construction of $\{X_n\}$, if $X_n \geq \gamma$, then

$$X_{n+1} - X_n \leq \sup_{t>0} \left(\int_0^t e^{-dw} dw - rt \right) = \sup_{t>0} \left(\frac{1 - e^{-dt}}{d} - rt \right) =: \delta = \frac{1 - r + r \log r}{d}.$$

Moreover, whenever $X_{n-1} < \gamma$ then $X_n \leq \gamma + \sup_{t>0} \left((1 + \alpha t(0))d^{-1}(1 - e^{-dt}) - rt \right) =: \gamma + \delta^{(\alpha)}$. Straightforward calculations show that

$$(2.27) \qquad \delta^{(\alpha)} := \frac{(1 + \alpha t(0))(1 - r(1 + \alpha t(0))^{-1}) + r \log r(1 + \alpha t(0))^{-1}}{d} \geq \delta^{(0)} = \delta.$$

These simple results will be very useful and may already give an idea why the process $\{X_n\}$ remains stable.

3. Existence of a stationary distribution for $\{X_n\}$

In this section we consider the Markov structure of $\{X_n\}$ in more detail. Subjects of interest are the existence (and uniqueness) of a stationary distribution, support and tail behavior of the stationary distribution as well as ergodicity. For an introduction to Markov chain theory we refer to [18] or [30].

A little reflection shows that if L is always rather small and the *off* period is zero, then the content process will always increase and there is no hope the Markov chain will be stable. In what follows we assume that L must be sufficiently large with positive probability.

Assumption 3.1. *The distribution of L satisfies*

$$(3.1) \qquad P[L > 1/d] > 0.$$

Assumption 3.1 guarantees that the process $\{X_n\}$ dips below $\gamma + \delta$ in finite time no matter how large was the initial state.

Proposition 3.2. *If (3.1) holds, then for all $x \geq 0$*

$$(3.2) \qquad P_x\Big(\bigcap_{j=1}^{\infty}[X_j > \gamma + \delta]\Big) = 0.$$

Proof. Condition (3.1) implies that for all $x \geq \gamma$

$$0 < \eta := P[L_1 > 1/d] \leq P[L_1 > (1 - e^{-dt_0(x)})/d]$$
$$\leq P_x[X_1 \leq \gamma + \delta] = p(x, [0, \gamma + \delta]).$$

The reason for the inequalities is that if $L_1 > (1 - e^{-dt_0(x)})/d$, then a transmission starting when content level is $x \geq \gamma$ lasts long enough for the control to push the content level to γ and then the most that the content level can be by the start of the next transmission is $\gamma + \delta$. This implies that eventually the Markov chain is below $\gamma + \delta$ since for every $x \geq 0$

$$P_x\Big(\bigcap_{j=1}^{\infty}[X_j > \gamma + \delta]\Big) = \lim_{N \to \infty} P_x\Big(\bigcap_{j=1}^{N}[X_j > \gamma + \delta]\Big)$$
$$= \lim_{N \to \infty} \int \cdots \int_{\substack{y_j > \gamma + \delta \\ j=1,\dots,N}} p(x, dy_1)p(y_1, dy_2)\dots p(y_{N-1}, dy_N)$$
$$(3.3) \qquad \leq \lim_{N \to \infty} (1 - \eta)^{N-1} P_x[X_1 > \gamma + \delta] = 0.$$

$\square$

Theorem 3.3. *Let $\{X_n\}$ be the process defined in (2.19) and suppose L satisfies Assumption 3.1. Then, the process $\{X_n\}$ has a stationary distribution π. Moreover, for every $x \geq \gamma + \delta^{(\alpha)}$*

$$(3.4) \qquad \pi((x, \infty)) \leq \begin{cases} 0, & P[L \leq r\, t_0(\gamma)] = 0, \\ P[L \leq 1/d]^{-\gamma/\delta^{(\alpha)} - 2} e^{-\log P[L \leq 1/d]^{-1/2\delta^{(\alpha)}} x}, & else. \end{cases}$$

Remark 3.4. In the standard *on/off* model where $\gamma = \infty$ and $\alpha = 0$, a common condition for stability is

$$(3.5) \qquad E\{(1 - r)L - rY\} < 0$$

which is called the *negative drift condition*. Compare this to Theorem 3.3 and we see that because of the effect of the control, a drift condition is not required for stability in the presence of the *on/off* control. The control mechanism in our model ensures that $\{X_n\}$ has negative drift above a sufficiently high level. Moreover, the tail of the stationary distribution π of $\{X_n\}$ is always

light-tailed (even in the case when L is heavy-tailed). This is a completely different result from the standard case where the stationary distribution becomes heavy-tailed too (see [12] and [13]).

We postpone the proof of Theorem 3.3 until after the next two propositions. The consequences of Theorem 3.3 are not needed in their proofs.

We discuss first the support of any stationary of $\{X_n\}$ under different conditions on L and Y.

Proposition 3.5. *Suppose the assumptions of Theorem 3.3 hold.*
(a) If

$$(3.6) \qquad P[L < rt_0(\gamma)] = 0,$$

then the support of any stationary distribution lies in $[0, \gamma + \delta^{(\alpha)}]$.

(b) If $Y = 0$ a.s. then the support of any stationary distribution is a subset of $[\gamma, \infty)$. If additionally $P[L < rt_0(\gamma)] = 0$, then the support of any stationary distribution is a subset of $[\gamma, \gamma + \delta^{(\alpha)}]$.

(c) If for some $\eta > 0$

$$P[Y \geq \eta/r] = 0 \quad and \quad P[L \leq t(0) + \alpha t(0)^2/2] = 0,$$

where $t(x)$ is defined in (2.2), then the support of any stationary distribution is a subset of $[(\gamma - \eta) \vee 0, \infty)$.

Remark 3.6. The simulations in Figure 1 and 2 suggest weaker conditions suffice for any stationary distribution to concentrate on either neighborhoods of 0 or infinity.

Proof. (a) From Assumption 3.1 and Proposition 3.2, we know that for all $x \geq 0$,

$$P_x\left(\bigcup_{j=0}^{\infty}[X_j \leq \gamma + \delta^{(\alpha)}]\right) = 1.$$

So if

$$(3.7) \qquad M := \inf\{n \geq 0 : X_n \leq \gamma + \delta^{(\alpha)}\},$$

it follows that

$$P_x[M < \infty] = 1, \quad \text{for all } x \geq 0.$$

We claim

$$(3.8) \qquad P_x\left(\bigcap_{j=M}^{\infty}[X_j \leq \gamma + \delta^{(\alpha)}]\right) = 1, \quad \text{for all } x \geq 0,$$

which will show that any stationary distribution is supported in $[0, \gamma + \delta^{(\alpha)}]$. It suffices to show for $j \geq 0$

$$P_x[X_{M+j} \leq \gamma + \delta^{(\alpha)}] = 1$$

which we do by induction on j. The result is clearly true for $j = 0$. Suppose it is true for j and we prove it for $j+1$ as follows. By the strong Markov property, we have for $j \geq 1$,

$$P_x[X_{M+j+1} \leq \gamma + \delta^{(\alpha)}] = P_x[X_{M+j} \leq \gamma, X_{M+j+1} \leq \gamma + \delta^{(\alpha)}]$$
$$+ P_x[X_{M+j} \in (\gamma, \gamma + \delta^{(\alpha)}], X_{M+j+1} \leq \gamma + \delta^{(\alpha)}]$$
$$= P_x[X_{M+j} \leq \gamma] + \int_{h \in (0, \delta^{(\alpha)}]} P_{\gamma+h}[X_1 \leq \gamma + \delta^{(\alpha)}]P_x[X_{M+j} \in \gamma + dh].$$

Recall

$$t_0(\gamma) = \inf\{u > 0 : \gamma + (1 - e^{-du})/d - ru = \gamma\}$$
$$= \inf\{u > 0 : \gamma + h + (1 - e^{-du})/d - ru = \gamma + h\}$$

for any $h \geq 0$. Thus, starting from $\gamma + h$, the control will force the contents process back to $\gamma + h$ provided there is a sufficiently large job size requiring the transmission time to exceed $t_0(\gamma)$. Thus, the job size needs to exceed

$$\frac{1 - e^{-dt_0(\gamma)}}{d} = rt_0(\gamma).$$

Since (3.6) holds, we have

$$P_{\gamma+h}[X_1 \leq \gamma + \delta^{(\alpha)}] = 1, \quad \text{for all } h \in [0, \delta^{(\alpha)}],$$

and thus

$$\begin{aligned} P_x[X_{M+j+1} \leq \gamma + \delta^{(\alpha)}] &= P_x[X_{M+j} \leq \gamma] + P_x[X_{M+j} \in (\gamma, \gamma + \delta^{(\alpha)}]] \\ &= P_x[X_{M+j} \leq \gamma + \delta^{(\alpha)}] = 1, \end{aligned}$$

as required.

(b) If $X(0) = x < \gamma$, then starting at time 0, $X(\cdot)$ is increasing because $r < 1$. For this case, since there are no off periods, the contents process must eventually hit γ, and then never go below γ. If $X(0) = x \geq \gamma$, the process is already above γ and therefore, the support of any stationary distribution of $\{X_n\}$ must be in $[\gamma, \infty)$.

(c) If $X_0 = x < \gamma$, then since $P[L_1 > t(0) + \alpha t(0)^2/2] = 1$, and it takes

$$t(x) + \alpha t(x)^2/2 < t(0) + \alpha t(0)^2/2 < L_1$$

units of work for the contents process to climb past γ, the contents process must be within $[\gamma, \gamma + \delta^{(\alpha)}]$ at time $\tau(x, L_1)$. Thus, at time $\tau(x, L_1) + Y_1$, the contents is at least $\gamma - rY_1 \geq \gamma - \eta$. So for $x < \gamma$, we have $p(x, [\gamma - \eta, \infty)) = 1$. A similar argument holds also for $X_0 = x \geq \gamma$. Therefore in all cases

$$p(x, [\gamma - \eta, \infty)) = 1, \quad \text{for all } x \geq 0.$$

$\square$

The next proposition is crucial for proving (3.4). It shows that all marginals distributions of $\{X_m; m \geq n\}$ for n large enough have exponential decay.

Proposition 3.7. *Let $\{X_n\}$ be the process defined in (2.19) and suppose L satisfies Assumption 3.1. Then, for every $x \geq \gamma + \delta^{(\alpha)}$, $y \geq 0$ and $n > \lceil \frac{x - \gamma - \delta^{(\alpha)}}{2\delta^{(\alpha)}} \rceil$,*

$$(3.9) \qquad P_y[X_n > x] \leq \begin{cases} 0, & P[L \leq 1/d] = 0, \\ P[L \leq 1/d]^{-\gamma/\delta^{(\alpha)} - 2} e^{-\log P[L \leq 1/d]^{-1/2\delta^{(\alpha)}} x}, & else. \end{cases}$$

Proof. Let $x \geq \gamma + \delta^{(\alpha)}$ and $y \geq 0$ arbitrary. Define $K := \lceil \frac{x - \gamma - \delta^{(\alpha)}}{2\delta^{(\alpha)}} \rceil + 1 \geq 1$ and choose $n \geq K$ arbitrary. From previous considerations we know that

$$(3.10) \qquad \{L_j > 1/d\} \subseteq \{X_j \leq \gamma + \delta^{(\alpha)}\}, \quad j = 1, 2, 3, \dots.$$

As a consequence

$$(3.11) \qquad \{X_n > x\} \subseteq \left(\cup_{j=0}^{K-1} \{L_{n-j} > 1/d\} \right)^c = \cap_{j=0}^{K-1} \{L_{n-j} \leq 1/d\}$$

and thus because of the independence of the events $\{L_1 \leq 1/d\}, \{L_2 \leq 1/d\}, \dots$

$$P_y[X_n > x] \quad \leq \quad P[L \leq 1/d]^K.$$

Obviously, $P_y[X_n > x] = 0$ if $P[L \leq 1/d] = 0$. Furthermore, for $P[L \leq 1/d] > 0$ we have

$$P_y[X_n > x] \leq e^{\left(\frac{x/2 - \gamma}{\delta^{(\alpha)}} - 2 \right) \log P[L \leq 1/d]} = P[L \leq 1/d]^{-\gamma/\delta^{(\alpha)} - 2} e^{-\log P[L \leq 1/d]^{-1/2\delta^{(\alpha)}} x}.$$

It remains therefore to show (3.11). Assume there exists $j \in \{0, 1, ..., K-1\}$ such that $L_{n-j} > 1/d$. Clearly, by (3.10)

$$X_{n-j} \leq \gamma + \delta^{(\alpha)} .$$

Since the process $\{X_n\}$ above γ increases at most δ in one step we conclude that a crude upper bound for X_n is

$$
\begin{aligned}
\gamma + \delta^{(\alpha)} + j\,\delta &\leq \gamma + \delta^{(\alpha)} + (K-1)\,\delta^{(\alpha)} \\
&\leq \gamma + \delta^{(\alpha)} + \left(\frac{x - \gamma - \delta^{(\alpha)}}{2\delta^{(\alpha)}} \right) \delta^{(\alpha)} \\
&\leq x ,
\end{aligned}
$$

which implies (3.11) and the proof is complete. $\qquad\square$

Proof of Theorem 3.3. The proof is an application of Theorem 2 of [32] which require drift conditions be checked for $\{X_n\}$. Recall that $\{X_n\}$ is weak Feller since for every bounded continuous function g

$$E_x\left\{g(X_1)\right\} = \int_0^\infty \int_0^\infty g\left(\max\left(0, x + l - r\tau(x, l) - ry\right)\right) F_Y(dy) F_L(dl)$$

is continuous in x by Lemma 2.2.

Next fix $N \in \mathbb{N}$ and define $\varepsilon := P[L > 1/d]/2 > 0$ and the set $A := [0, a]$, where $a > \gamma$ is chosen sufficiently large so that for $x \in A^c$

$$(3.12) \qquad \frac{P[L > 1/d](1 + (\gamma + \delta^{(\alpha)})^N) + P[L \leq 1/d](1 + (x + \delta^{(\alpha)})^N)}{1 + x^N} \leq 1 - \epsilon,$$

which is possible since the left side of (3.12) converges to $P[L \leq 1/d]$ as $x \to \infty$. The set A is compact.

Set $g(x) = 1 + x^N$. Clearly for any $x \in A$

$$(3.13) \qquad E_x\left\{g(X_1)\right\} \leq 1 + (a + \delta^{(\alpha)})^N < \infty.$$

Moreover, because of (3.12), for any $x \in A^c$

$$
\begin{aligned}
E_x\left\{g(X_1)\right\} &= E_x\left\{g(X_1)1_{\{L_1 > 1/d\}}\right\} + E_x\left\{g(X_1)1_{\{L_1 \leq 1/d\}}\right\} \\
&\leq (1 + (\gamma + \delta^{(\alpha)})^N)P[L > 1/d] + (1 + (x + \delta^{(\alpha)})^N)P[L \leq 1/d] \\
&\leq (1 - \epsilon)(1 + x^N) \\
(3.14) \qquad &= (1 - \epsilon)g(x).
\end{aligned}
$$

Because of (3.13), (3.14) and the weak Feller property all assumptions in Theorem 2 of [32] are fulfilled and the process $\{X_n\}$ has a stationary distribution π.

The upper bounds for the tail of the stationary distribution follow immediately from Proposition 3.5(a), Proposition 3.7 and the fact that $\{L < rt_0(\gamma)\} = \left\{L < d^{-1}(1 - e^{-dt_0(\gamma)})\right\} \subseteq \{L \leq 1/d\}$. $\qquad\square$

Theorem 3.3 guarantees the existence of a stationary distribution for $\{X_n\}$. However, no information about the uniqueness of π and the convergence of the n-step transition probabilities is provided. The minimal Assumption 3.1 of the theorem does not suffice to get hold of this deeper level of complexity. In what follows we study the notion of ergodicity and hence also ϕ-irreducibility for some σ-finite measure ϕ. The main result is stated in the next theorem.

Theorem 3.8. *Let $\{X_n\}$ be the process defined in (2.19) and suppose L satisfies Assumption 3.1. Moreover, assume that the random variables L and Y satisfy one of the following two conditions:*

(a) $P[L - r\alpha^{-1}(-1 + \sqrt{1 + 2\alpha L}) - rY < 0] > 0$,

(b) $P[Y \geq (\gamma + \delta^{(\alpha)})/r] > 0$.

Then, the process $\{X_n\}$ *has a unique stationary distribution* π *and* 0 *is an atom.* $\{X_n\}$ *is geometric ergodic, i.e. there exists a* $\rho < 1$ *such that for every* $x \geq 0$

$$(3.15) \qquad\qquad \rho^{-n}\|P^n[x, \cdot] - \pi\| \to 0, \quad as\ n \to \infty,$$

where $\|\cdot\|$ *denotes total variation of signed measures on* $B([0, \infty))$. *Moreover, the process is strong mixing and, as shown in Theorem 3.3, (3.4) holds.*

Remark 3.9. (a) In the case $\alpha = 0$ condition (a) of Theorem 3.8 simplifies to $P[(1-r)L - rY < 0] > 0$. A sufficient (but not necessary) condition for this case is that there exists $\eta > 0$ such that

$$(3.16) \qquad\qquad P[Y \geq \eta/r] > 0 \quad and \quad P[L < \eta/(1 - r)] > 0$$

since $\{L < \eta/(1 - r), Y \geq \eta/r\} \subseteq \{(1 - r)L - rY < 0\}$.
(b) Conditions (a) and (b) imply that $Y \neq 0$ a.s.. If $Y = 0$ a.s. then $\{0\}$ can not be hit starting from $x > 0$. See also Proposition 3.5.

In order to prove Theorem 3.8 we show that under the assumptions (a) or (b) of Theorem 3.8 the process $\{X_n\}$ hits zero a.s. in finite time.

Proposition 3.10. *Under the assumptions of Theorem 3.8, the process* $\{X_n\}$ *has an atom at* $\{0\}$; *that is, for every* $x \in [0, \infty)$,

$$P_x[\inf\{j \geq 1 : X_j = 0\} < \infty] = 1.$$

Proof. Suppose first that Condition (a) holds. Suppose a transmission of a file of size l commences when the buffer content is x. In a buffer with no control (i.e. $\gamma = \infty$)

$$\tau = \tau(x, l) = \alpha^{-1}(-1 + \sqrt{1 + 2\alpha l})$$

because we solve the equation $l = \alpha\tau^2/2 + \tau$. Clearly, the transmission time increases when the control is present since the input rate decreases above γ. Therefore, we conclude that

$$(3.17) \qquad\qquad \tau = \tau(x, l) \geq \alpha^{-1}(-1 + \sqrt{1 + 2\alpha l}).$$

Assumption (a) implies that there exists $\varepsilon > 0$ such that

$$(3.18) \qquad\qquad P[L - r\alpha^{-1}(-1 + \sqrt{1 + 2\alpha L}) - rY \leq -\varepsilon] > 0.$$

Set $K := \lceil(\gamma + \delta^{(\alpha)})/\varepsilon\rceil + 1$ and define for every $n \geq 1$

$$U_n := 1_{[L_n > 1/d] \cap_{j=1}^K [L_{n+j} - r\alpha^{-1}(-1 + \sqrt{1 + 2\alpha L_{n+j}}) - rY_{n+j} \leq -\varepsilon]}.$$

Since $\{L_n\}$ and $\{Y_n\}$ are independent iid sequences, $\{U_n\}$ is strictly stationary and in fact K-dependent, and because of (3.18)

$$(3.19) \qquad E\{U_n\} = P[L > 1/d]P[L - r\alpha^{-1}(-1 + \sqrt{1 + 2\alpha L}) - rY \leq -\varepsilon]^K > 0.$$

Furthermore, because of the strict stationarity, for any initial x

$$(3.20) \qquad\qquad \frac{U_1 + \cdots + U_n}{n} \to E\{U_1\}, \quad P_x - a.s.\ as\ n \to \infty.$$

Combining (3.19) and (3.20) we conclude that for almost all ω, there exists $N = N(\omega) < \infty$ such that $U_N(\omega) = 1$. If $X_{N-1} \geq \gamma$, then as in Proposition 3.2, $L_N > 1/d$ implies $X_N \leq \gamma + \delta^{(\alpha)}$ while if $X_{N-1} < \gamma$, the same conclusion holds. Thus

$$X_N(\omega) \leq \gamma + \delta^{(\alpha)}$$

and using (3.17) and (3.18), we get

$$X_{N+1}(\omega) = (X_N(\omega) + L_{N+1}(\omega) - r\tau(X_N(\omega), L_{N+1}(\omega)) - rY_{N+1}(\omega))_+$$

$$\leq \left(\gamma + \delta^{(\alpha)} + L_{N+1}(\omega) - r(-1 + \sqrt{1 + 2\alpha L_{N+1}(\omega)})/\alpha - rY_{N+1}(\omega)\right)_+$$

$$\leq \max\{0, \gamma + \delta^{(\alpha)} - \varepsilon\} = \gamma + \delta^{(\alpha)} - \varepsilon$$

and continuing this procedure K times convinces us that $X_{N+K}(\omega) = 0$ which finishes the proof.

Next suppose condition (b) holds. In this case the statement follows immediately from the fact that for every $x \in [0, \infty)$ and $\omega \in \{L > 1/d, Y \geq (\gamma + \delta^{(\alpha)})/r\}$

$$x + L(\omega) - r(\tau(L(\omega), x) + Y(\omega)) \leq \gamma + \delta^{(\alpha)} - rY(\omega) \leq 0.$$

$\square$

We are now prepared to consider Harris-recurrence, irreducibility and regeneration.

Corollary 3.11. *Under the assumptions of Theorem 3.8, the process $\{X_n\}$ is Harris-recurrent with regeneration set $\{0\}$, regenerative and $\delta_{\{0\}}$-irreducible.*

Proof. Because of Proposition 3.10, $\{0\}$ is a Harris-recurrent one-point set and hence a regeneration set (see e.g. [3], p.151). Furthermore, for every $A \in \mathcal{B}([0, \infty))$ with $0 \in A$ and all $x \in [0, \infty)$ we have

$$\begin{aligned} 1 &= P_x[\inf\{j \geq 1 : X_j = 0\} < \infty] \leq P_x[\inf\{j \geq 1 : X_j \in A\} < \infty] \\ &= P_x(\bigcup_{n=1}^{\infty}\{X_n \in A\}) \leq \sum_{n=1}^{\infty} P_x[X_n \in A], \end{aligned}$$

and thus $\{X_n\}$ is $\delta_{\{0\}}$-irreducible. $\square$

Proof of Theorem 3.8. The inequalities (3.13) and (3.14) from the proof of Theorem 3.3 are clearly still valid. Because of this, Corollary 3.11 and the weak Feller property all assumptions in Theorem 4 of [31] are fulfilled and the statements follow. $\square$

4. Transmission time τ

In this section we study the transmission time τ when the process $\{X_n\}$ is stationary (i.e. $X_n \sim \pi$ for every $n \geq 0$). In what follows X is a generic random variable with distribution π and independent of L and Y. The lemma below states that stationarity of $\{X_n\}$ implies stationarity of the sequence of transmission times $\{\tau_n\} := \{\tau(X_{n-1}, L_n)\}$. Further, the geometric ergodic property of $\{X_n\}$ guarantees that the transmission times τ_n converges uniformly to the invariant distribution.

Lemma 4.1. *Suppose L satisfies Assumption 3.1. Then with respect to P_π, $\{\tau_n\} := \{\tau(X_{n-1}, L_n)\}$ is a stationary sequence. Moreover, if $\{X_n\}$ satisfies the conditions in Theorem 3.8 then for every $y \geq 0$*

$$(4.1) \qquad P_y[\tau_n \in A] \to P_\pi[\tau(X, L) \in A] \text{ uniformly in } A \in B([0, \infty)) \text{ as } n \to \infty,$$

where X and L are independent random variables with $L \overset{d}{=} L_n$.

Proof. The stationarity of $\{\tau_n\}$ follows from the fact that $\{X_n\}$ is stationary and the sequences $\{L_m; m \geq n\}$ and $\{Y_m; m \geq n\}$ are independent of X_{n-1}. Further, for arbitrary $y \geq 0$

$$\sup_{A \in B([0,\infty))} \left| P_y[\tau_n \in A] - P_\pi[\tau(X, L) \in A] \right|$$

$$\leq \sup_{A \in B([0,\infty))} \int_0^\infty \left| P_y[\tau(X_{n-1}, l) \in A] - P_\pi[\tau(X, l) \in A] \right| F_L(dl)$$

$$\leq \sup_{\{f : \|f\|_\infty \leq 1\}} \left| \int_0^\infty f(z) P^{n-1}(y, dz) - \int_0^\infty f(z) \pi(dz) \right|.$$

Finally, by (3.15) in Theorem 3.8 the right hand side converges to zero as $n \to \infty$ which finishes the proof. $\square$

Theorem 3.3 implies in particular that the tail of the stationary distribution of τ is completely specified by a subexponential load input L and the input rate r. For definition, properties and examples of subexponential distributions see for instance [5] or [10].

Theorem 4.2. *Let $\{X_n\}$ be the stationary process defined in (2.19) with L having a subexponential distribution F_L (i.e. $\overline{F}_L^{2*}(x) \sim 2\overline{F}_L(x)$, as $x \to \infty$). Then*

$$(4.2) \qquad\qquad P_\pi[\tau > x] \sim \overline{F}_L(rx), \qquad as\ x \to \infty.$$

Proof. Let $\tau = \tau(X, L)$. Note that at the end of transmission $X(\tau) = X + L - r\tau$. Clearly $X + L - r\tau \geq 0$ which yields

$$\tau \leq \frac{X + L}{r}.$$

On the other hand considering the cases of starting above γ or below γ

$$X + L - r\tau \leq \max\{X + \delta, \gamma + \delta^{(\alpha)}\} \leq \max\{X, \gamma\} + \delta^{(\alpha)},$$

which yields

$$\tau \geq \frac{X + L - \max\{X, \gamma\} - \delta^{(\alpha)}}{r}.$$

The two inequalities written together give

$$(4.3) \qquad\qquad \frac{L + X - \max\{X, \gamma\} - \delta^{(\alpha)}}{r} \leq \tau \leq \frac{L + X}{r}.$$

From (4.3) we see that τ is bounded above and below by random variables of the form $(L/r) + V$ where V is independent of L and the tail of its distribution is exponentially decreasing. By Proposition 2.1 (ii) in [26], the tail of V is lighter than the tail of L/r and from part (iv) of the same proposition the asymptotic behavior of the distribution of τ in (4.2) follows. $\square$

Theorem 4.2 seems to be very surprising since for heavy-tailed inputs the tail of the distribution of τ is independent of d (as a first order approximation). One might wonder the extent to which other characteristic quantities depend on d. Besides the tail of the distribution of τ we consider therefore the expectation of τ. Assuming again that the process is stationary the next theorem shows bounds on the expectation of the transmission time τ.

Theorem 4.3. *Let $\{X_n\}$ be the stationary process defined in (2.19) with L satisfying Assumption 3.1. Then, for every $0 < c_0, c_1 \leq \gamma$ we have*

$$\max\left(\alpha^{-1}\left(-1 + E\sqrt{1 + 2\alpha L}\right), \frac{EL}{r} - EY\right)$$

(4.4)
$$\leq \; E_\pi\{\tau\} \; \leq \; \frac{EL}{r} - P\left[L > t(\gamma - c_0) + \alpha t(\gamma - c_0)^2/2\right] \int_0^{c_0/r} y\, dF(y)$$

$$- P[L > \beta_{c_1}]\frac{c_1}{r}P[Y > (\gamma + \delta^{(\alpha)})/r],$$

where $\beta_{c_1} = t(\gamma - c_1) + \alpha t(\gamma - c_1)^2/2 \vee d^{-1}$ and $t(x)$ is defined in (2.2).

Remark 4.4. (a) For $X > 0$ a.s. the above statement can be sharpened. The left inequality in Theorem 4.3 becomes an equality, more precisely

$$E_\pi\{\tau\} = \frac{EL}{r} - EY\,.$$

This follows immediately from the fixpoint equation

$$X \stackrel{d}{=} (X + L - r\tau - rY)_+ = X + L - r\tau - rY$$

and taking expectations on both sides.
(b) For $\alpha = 0$, $\alpha^{-1}\left(-1 + E\sqrt{1 + 2\alpha L}\right)$, $t(\gamma - c_0) + \alpha t(\gamma - c_0)^2/2$ and β_{c_1} simplify to EL, $c_0/(1 - r)$ and $c_1/(1 - r) \vee d^{-1}$, respectively.
(c) The right bound in (4.4) depends on the choice of c_0 and c_1. The optimal choice of these constants (in terms of minimizing the right bound) can not be specified in general without knowledge of the underlying distributions of L and Y. However, if

(4.5)
$$s_\alpha := (2\alpha)^{-1}\left(r^2 + 1 + \alpha d^{-1} - 2r\sqrt{1 + \alpha d^{-1}} - (1 - r)^2\right) \geq \gamma$$

(note $s_0 = d^{-1}(1 - r)$) the optimal choice of c_1 is γ. This can be easily seen by noting that $f(c) := t(\gamma - c) + \alpha t(\gamma - c)^2/2$ is strictly monotone increasing in c and $s_\alpha \geq \gamma$ is the only positive root of $f(c) = d^{-1}$. Thus, for every $c \in (0, \gamma]$

$$P[L > \beta_c]\frac{c}{r} \;=\; P[L > d^{-1}]\frac{c}{r}$$
$$\leq \; P[L > d^{-1}]\frac{\gamma}{r} = P[L > \beta_c]\frac{\gamma}{r}$$

and $c_1 = \gamma$. In the case when s_α is strictly smaller than γ the above argument does not work anymore and we can only conclude that the optimal choice of c_1 lies in between s_α and γ.

In what follows we consider in more numerical detail the dependence of the expectation of τ on d for a special case which is analytically tractable. The expectation of τ is estimated by Monte-Carlo simulation. The estimated value of $E_\pi\{\tau\}$ can be also used for checking how good the bounds in Theorem 4.3 are. Recall that only the right bound in Theorem 4.3 depends on d. Further, by Remark 4.4(a), the expectation of τ is always independent of d whenever the support of the stationary distribution does not include $\{0\}$.

Example 4.5. Assume $\alpha = 0$ and L has the Pareto-distribution, i.e. $P[L > x] = (1 + x)^{-\kappa}$, $x \geq 0$ with $\kappa > 2$. Moreover, let Y be uniformly distributed in $[0, y_0]$.

By Theorem 4.3, we conclude that

$$\max\left(\frac{1}{\kappa - 1}, \frac{1}{(\kappa - 1)r} - \frac{y_0}{2}\right)$$

$$\leq \quad E_\pi\{\tau\} \quad \leq \quad \frac{1}{(\kappa - 1)r} - (1 + c_0/(1 - r))^{-\kappa}\frac{(c_0/r \wedge y_0)^2}{2y_0}$$

(4.6)
$$- (1 + (c_1/(1 - r) \vee d^{-1}))^{-\kappa}\frac{c_1\left(y_0 - (y_0 \wedge (\gamma + \delta^{(\alpha)})/r)\right)}{ry_0}.$$

Straightforward calculations yield for the optimal choice of c_0 and c_1:

(a) if $y_0 \geq \gamma/r$ then $c_0 = \gamma \wedge 2(1 - r)/(\kappa - 2)$,
(b) if $y_0 < \gamma/r$ then $c_0 = y_0 r \wedge 2(1 - r)/(\kappa - 2)$,
(c) if $d^{-1}(1 - r) \geq \gamma$ then $c_1 = \gamma$,
(d) if $d^{-1}(1 - r) < \gamma$ and $d < \kappa - 1$ then $c_1 = d^{-1}(1 - r)$,
(e) if $d^{-1}(1 - r) < \gamma$ and $d \geq \kappa - 1$ then $c_1 = \gamma \wedge (1 - r)/(\kappa - 1)$.

Note that the result (c) follows from Remark 4.4(c), whereas, here, (d) and (e) are precise results in contrast to Remark 4.4(c).

Figure 3 displays the numerically estimated expectation of τ as a function of d and the corresponding bounds for $\kappa = 2.5$ and 10, respectively, $\gamma = 1$ and different values of y_0 and r. Although the numbers of *on/off* periods were chosen very large (20,000 *on/off* periods) the numerical value of $E_\pi\{\tau\}$ is still not always stable. Note also that for d small the estimated value of $E_\pi\{\tau\}$ is not in all cases between our theoretical bounds. In some plots the upper and lower theoretical bounds are close to each other. However (unfortunately) this is not always the case. One conclusion of our empirical study is that the bounds might be for some parameters quite good and in other cases not. In general one might say that the theoretical bounds converge when r is increasing. Another striking feature is that $E_\pi\{\tau\}$ stays in the most cases closer to the lower bound. Last but not least, all pictures show that the expectation of τ seems hardly to depend on d. This observation together with Theorem 4.2 and Theorem 4.3 indicates that the value d used for the control mechanism is only minimally relevant for the transmission time τ.

Proof of Theorem 4.3. Note first that from (3.17)

(4.7)
$$E_\pi\{\tau\} \geq \alpha^{-1}\left(-1 + E\sqrt{1 + 2\alpha L}\right).$$

To deal with $\max(0, X + L - r\tau - rY)$ in a convenient way we define

(4.8)
$$Z = Y1_{\{Y \leq \frac{X+L-r\tau}{r}\}} + \frac{(X + L - r\tau)}{r}1_{\{Y > \frac{X+L-r\tau}{r}\}}.$$

Then it is seen that $(0, X + L - r\tau - rY)_+ = X + L - r\tau - rZ$. If X has the stationary distribution then the state at the next step, i.e. $X + L - r\tau - rZ$ also has the same distribution and from $E_\pi\{X\} = E_\pi\{X + L - r\tau - rZ\}$ we get that under the stationary distribution on X

(4.9)
$$E_\pi\{\tau\} = \frac{EL}{r} - E_\pi\{Z\}.$$

Since $Z \leq Y$ the left inequality of (4.4) follows from (4.7) and (4.9). We now work towards the right inequality.

From the definition of Z we have

(4.10) $\quad E_\pi\{Z\} = E_\pi\left\{\int_0^{\frac{X+L-r\tau}{r}} y\,dF(y)\right\} + E_\pi\left\{\frac{(X + L - r\tau)}{r}P_\pi\left(Y > \frac{X + L - r\tau}{r} \,\Big|\, X, L\right)\right\}.$

We examine how the values of L affect $(X + L - r\tau)$. Throughout we assume $0 < c \leq \gamma$ and $L > t(\gamma - c) + \alpha t(\gamma - c)^2/2$. Under these conditions we want to show that (4.11) holds.

First consider the case $X = 0$. Since $c \leq \gamma$ and $L > t(\gamma - c) + \alpha t(\gamma - c)^2/2$ transmission continues for a time greater than or equal to $t(\gamma - c)$. Thus $X + L - r\tau$ is greater than or equal to the buffer level at time $t(\gamma - c)$ which is $t(\gamma - c) + \alpha t(\gamma - c)^2/2 - rt(\gamma - c) = c$.

Next consider the case $0 < X < \gamma$. As before $0 < c \leq \gamma$ and $L > t(\gamma - c) + \alpha t(\gamma - c)^2/2$. Starting from X either the buffer level hits γ before time $t(\gamma - c)$ (which means at the end of transmission $c \leq \gamma \leq X + L - r\tau$) or by time $t(\gamma - c)$ the buffer level increases to $X + c$ so that again at the end of transmission $c \leq X + c \leq X + L - r\tau$. On the other hand for $X \geq \gamma$ clearly $X + L - r\tau \geq \gamma \geq c$.

From the above discussion, regardless of the initial X, if $0 < c \leq \gamma$ and $L > t(\gamma - c) + \alpha t(\gamma - c)^2/2$, then

$$(4.11) \qquad\qquad c \leq X + L - r\tau.$$

Finally, note that if $L > 1/d$ then

$$(4.12) \qquad\qquad X + L - r\tau \leq \gamma + \delta^{(\alpha)}$$

independent on the state of X.

Thus (4.10), (4.11) and (4.12) imply for every $0 < c_0, c_1 \leq \gamma$

$$E_\pi\{Z\} \geq E1_{\{L > t(\gamma - c_0) + \alpha t(\gamma - c_0)^2/2\}} \int_0^{c_0/r} y dF(y) + E1_{\{L > \beta_{c_1}\}} \frac{c_1}{r} P[Y > \frac{\gamma + \delta^{(\alpha)}}{r}]$$

$$(4.13) \qquad = P[L > t(\gamma - c_0) + \alpha t(\gamma - c_0)^2/2] \int_0^{c_0/r} y dF(y) + P[L > \beta_{c_1}] \frac{c_1}{r} P[Y > \frac{\gamma + \delta^{(\alpha)}}{r}],$$

establishing the right inequality of (4.4). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Knowing the approximative expectation of the transmission time of an input load gives a first idea about the effectiveness of the system. Another quantity which can be considered for the same reason is the probability that the system is empty. A system which is likely to be empty is not used to capacity and hence not economically efficient. The next result gives the asymptotic percentage of time that the buffer is empty. The result is an immediate consequence of Theorem 4.3 and its proof. We use the fact that $\{X(t)\}$ is regenerative and thus has a stationary limit distribution.

Corollary 4.6. *Let $X(t)$ be the continuous time process in (2.21) and (2.22) and assume that the conditions in Theorem 3.8 are satisfied. Then, for every $x \geq 0$*

$$(4.14) \qquad\qquad \lim_{t \to \infty} P_x[X(t) = 0] = 1 - \frac{E\{L/r\}}{E_\pi\{\tau + Y\}}.$$

Proof. From (4.9) we have

$$(4.15) \qquad\qquad \frac{E\{L/r\}}{E_\pi\{\tau + Y\}} = \frac{E_\pi\{\tau + Z\}}{E_\pi\{\tau + Y\}}.$$

By Theorem 3.8 and Lemma 4.1, the sequence (Z_i, Y_i, τ_i) is P_π-stationary and ergodic and from the ergodic theorem the right hand side of (4.15) is the P_π-almost sure limit of

$$(4.16) \qquad\qquad \frac{\sum_{i=1}^n (\tau_i + Z_i)}{\sum_{i=1}^n (\tau_i + Y_i)}.$$

Let $\{N_n\}$ be the successive return times to 0 by $\{X_n\}$ and suppose

$$k = \sup\{j : N_j \leq n\}.$$

The numerator in (4.16) equals

$$
\begin{aligned}
(\tau_1 + Y_1 &+ \cdots + \tau_{N_1-1} + Y_{N_1-1} + \tau_{N_1} + Z_{N_1}) \\
&+ (\tau_{N_1+1} + Y_{N_1+1} + \cdots + \tau_{N_2-1} + Y_{N_2-1} + \tau_{N_2} + Z_{N_2}) \\
&+ \cdots \\
&+ (\tau_{N_k+1} + Y_{N_k+1} + \cdots + \tau_n + Y_n)
\end{aligned}
$$

(4.17)

and the denominator of (4.16) is

$$
\begin{aligned}
(\tau_1 + Y_1 &+ \cdots + \tau_{N_1-1} + Y_{N_1-1} + \tau_{N_1} + Y_{N_1}) \\
&+ (\tau_{N_1+1} + Y_{N_1+1} + \cdots + \tau_{N_2-1} + Y_{N_2-1} + \tau_{N_2} + Y_{N_2}) \\
&+ \cdots \\
&+ (\tau_{N_k+1} + Y_{N_k+1} + \cdots + \tau_n + Y_n).
\end{aligned}
$$

(4.18)

For the purpose of taking limit of the ratio of (4.17) and (4.18) the first and last terms can be ignored and then we see that the denominator adds independent continuous time regeneration cycle lengths and the numerator adds busy periods within the cycles. The limit is the ratio of the expectations and from regenerative process theory (see e.g. [25], p.265) this is $\lim_{t\to\infty} P_\pi[X(t) > 0]$. The conclusion of the corollary is an immediate consequence of Theorem 4.3. □

References

[1] Rajeev Agrawal, Armand M. Makowski, and Philippe Nain. On a reduced load equivalence for fluid queues under subexponentiality. *Queueing Systems Theory Appl.*, 33(1-3):5–41, 1999. Queues with heavy-tailed distributions.

[2] E. Altman, K. Avratchenkov, and C. Barakat. A stochastic model for tcp/ip with stationary random losses, 2000. ACM SIGCOMM, September 2000.

[3] S. Asmussen. *Applied Probability and Queues*. Wiley, Chichester, West Sussex, UK, 1987.

[4] J. Beran, R. Sherman, W. Willinger, and M.S. Taqqu. Long range dependence in variable-bit-rate video. *IEEE Transactions on Communications*, 43:1566–1579, 1995.

[5] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

[6] O.J. Boxma and V. Dumas. Fluid queues with long-tailed activity period distributions. *Computer Communications*, 21:1509–1529, 1998. Special issue on "Stochastic Analysis and Optimization of Communication Systems".

[7] F. Brichet, J. Roberts, A. Simonian, and D. Veitch. Heavy traffic analysis of a storage model with long range dependent on/off sources. *Queueing Systems*, 23:197–215, 1996.

[8] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.

[9] M. Crovella, A. Bestavros, and M. Taqqu. Heavy-tailed probability distributions in the world wide web. In Murad S. Taqq Robert Adler, Raisa Feldman, editor, *A PRACTICAL GUIDE TO HEAVY TAILS: Statistical Techniques for Analysing Heavy Tailed Distributions*. Birkhäuser, Boston, 1999.

[10] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extreme Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997.

[11] A. Gilbert, Y. Joo, and N. McKeown. Congestion control and periodic behavior. Preprint: AT&T Labs–Research, 180 Park Ave, Florham Park, NJ 07932-0971, 2000.

[12] D. Heath, S. Resnick, and G. Samorodnitsky. Patterns of buffer overflow in a class of queues with long memory in the input stream. *Ann. Appl. Probab.*, 7(4):1021–1057, 1997.

[13] D. Heath, S. Resnick, and G. Samorodnitsky. Heavy tails and long range dependence in on/off processes and associated fluid models. *Math. Oper. Res.*, 23(1):145–165, 1998.

[14] D. Heath, S. Resnick, and G. Samorodnitsky. How system performance is affected by the interplay of averages in a fluid queue with long range dependence induced by heavy tails. *Ann. Appl. Probab.*, 9:352–375, 1999.

[15] P. Jelenković and A. Lazar. Asymptotic results for multiplexing subexponential on-off processes. *Advances in Applied Probability*, 31:394–421, 1999.

[16] S. Keshav. *An Engineering Approach to Computer Networking; ATM Networks, the Internet, and the Telephone network*. Addison-Wesley, Reading, Mass., 1997.

[17] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1–15, 1994.

[18] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag London Ltd., London, 1993.

[19] T. J. Ott, J. H. B. Kemperman, and M. Mathis. The stationary behavior of ideal tcp congestion avoidance. In *Proceedings of IEEE INFOCOM'99, New York*, 1999.

[20] T. J. Ott and A. Misra. The window distribution of idealized tcp congestion avoidance with variable packet loss. *http://www.argreenhouse.com/papers/tjo/*, 2000.

[21] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling tcp throughput: a simple model and its empirical validation, 1998. ACMSIGCOMM, September 1998.

[22] V. Paxton and S. Floyd. Wide-area traffic: The failure of poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[23] S. Resnick and G. Samorodnitsky. Activity periods of an infinite server queue and performance of certain heavy tailed fluid queues. *Queueing Systems*, 33:43–71, 1999. Available as TR1201.ps.Z at http://www.orie.cornell.edu /trlist/trlist.html.

[24] S. Resnick and G. Samorodnitsky. Steady state distribution of the buffer content for $m/g/\infty$ input fluid queues. Technical report, available at www.orie.cornell.edu/trlist/trlist.html, 1999.

[25] S.I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.

[26] J. Rosiński and G. Samorodnitsky. Distributions of subadditive functionals of sample paths of infinitely divisible processes. *Ann. Probab.*, 21:996–1014, 1993.

[27] R. Srikant. Control of communication networks. In Tariq Samad, editor, *Perspectives in Control Engineering: Technologies, Applications, New Directions*, pages 462–488. IEEE Press, 2000.

[28] M. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. *Computer Communications Review*, 27:5–23, 1997.

[29] M.S. Taqqu and J. Levy. Using renewal processes to generate long-range dependence and high variability. In E. Eberlein and M.S. Taqqu, editors, *Dependence in Probability and Statistics*, pages 73–89, Boston, 1986. Birkhäuser.

[30] R. L. Tweedie. Criteria for classifying general Markov chains. *Advances in Appl. Probability*, 8(4):737–771, 1976.

[31] R. L. Tweedie. Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability, statistics and analysis*, pages 260–276. Cambridge Univ. Press, Cambridge, 1983.

[32] R. L. Tweedie. Invariant measures for Markov chains with no irreducibility assumptions. *J. Appl. Probab.*, (Special Vol. 25A):275–285, 1988. A celebration of applied probability.

[33] V. van Jacobson. Congestion avoidance and control. *Proceedings of ACM SIGCOMM*, 1988.

[34] W. Willinger, M.S. Taqqu, M. Leland, and D. Wilson. Self–similarity in high–speed packet traffic: analysis and modelling of ethernet traffic measurements. *Statistical Science*, 10:67–85, 1995.

[35] W. Willinger, M.S. Taqqu, M. Leland, and D. Wilson. Self–similarity through high variability: statistical analysis of ethernet lan traffic at the source level. *Computer Communications Review*, 25:100–113, 1995. Proceedings of the ACM/SIGCOMM'95, Cambridge, MA.

[36] A. P. Zwart. A fluid queue with a finite buffer and subexponential input. *Adv. in Appl. Probab.*, 32(1):221–243, 2000.

School of Operations Research and Industrial Engineering, Rhodes Hall, Cornell University, Ithaca, New York 14853, USA
  E-mail address: {borkovec, sid, gennady}@orie.cornell.edu

Stat-Math Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta 700035, India
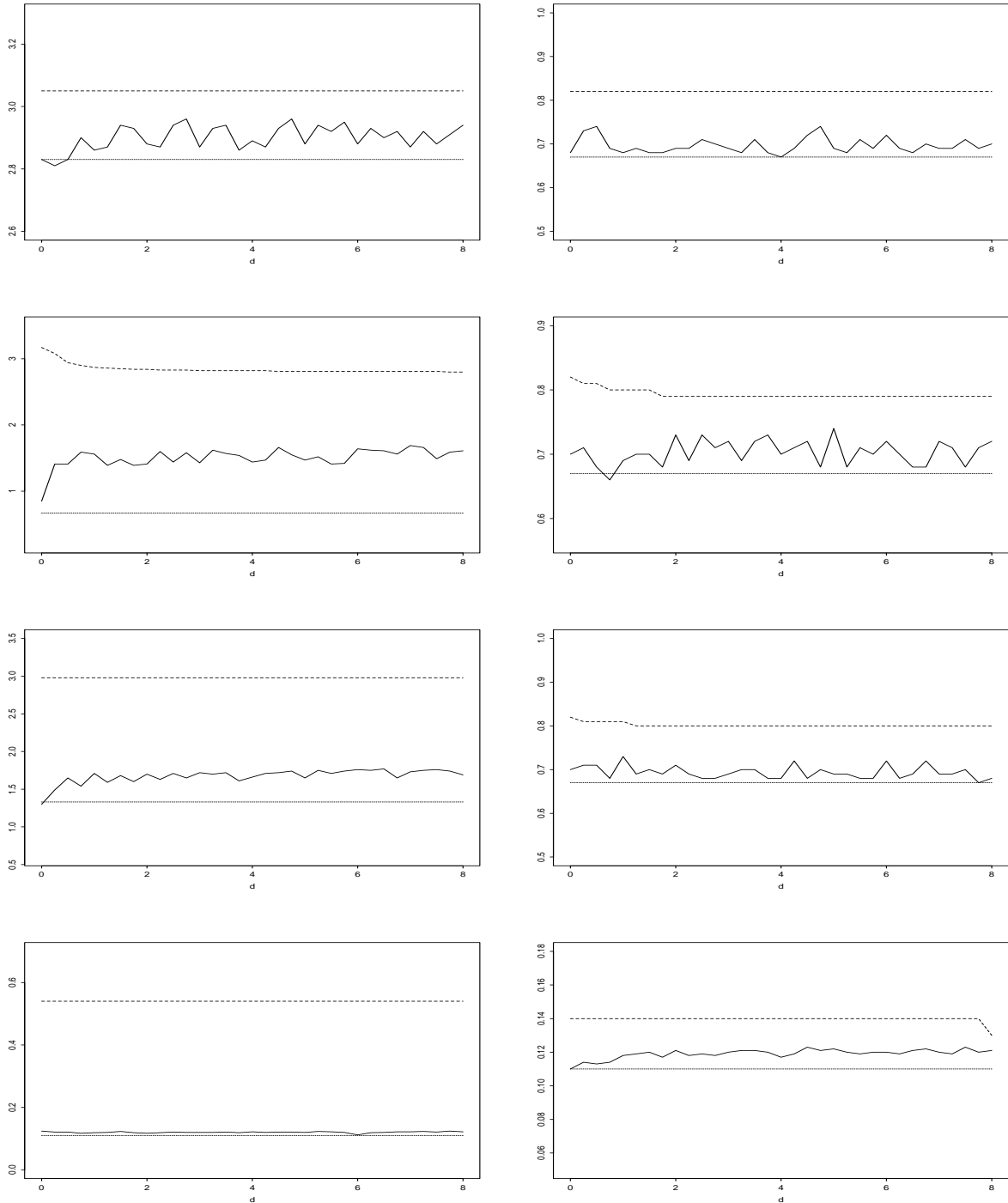 *E-mail address*: amites_v@www.isical.ac.in

FIGURE 3.   Estimated mean transmission time τ as a function of d for 20 000 on/off periods with parameters $r = 0.2, \kappa = 2.5, y_0 = 1, \gamma = 1, \alpha = 0$ (top,left), $r = 0.8, \kappa = 2.5, y_0 = 1, \gamma = 1, \alpha = 0$ (top, right), $r = 0.2, \kappa = 2.5, y_0 = 10, \gamma = 1, \alpha = 0$ (second line, left), $r = 0.8, \kappa = 2.5, y_0 = 10, \gamma = 1, \alpha = 0$ (second line, right), $r = 0.2, \kappa = 2.5, y_0 = 4, \gamma = 1, \alpha = 0$ (third line, left), $r = 0.8, \kappa = 2.5, y_0 = 4, \gamma = 1, \alpha = 0$ (third line, right), $r = 0.2, \kappa = 10, y_0 = 4, \gamma = 1, \alpha = 0$ (bottom, left) and $r = 0.8, \kappa = 10, y_0 = 4, \gamma = 1, \alpha = 0$ (bottom, right). The dotted lines represent the upper and lower bounds for $E_\pi\{\tau\}$ using the optimal constants c_0 and c_1 in Example 4.5.
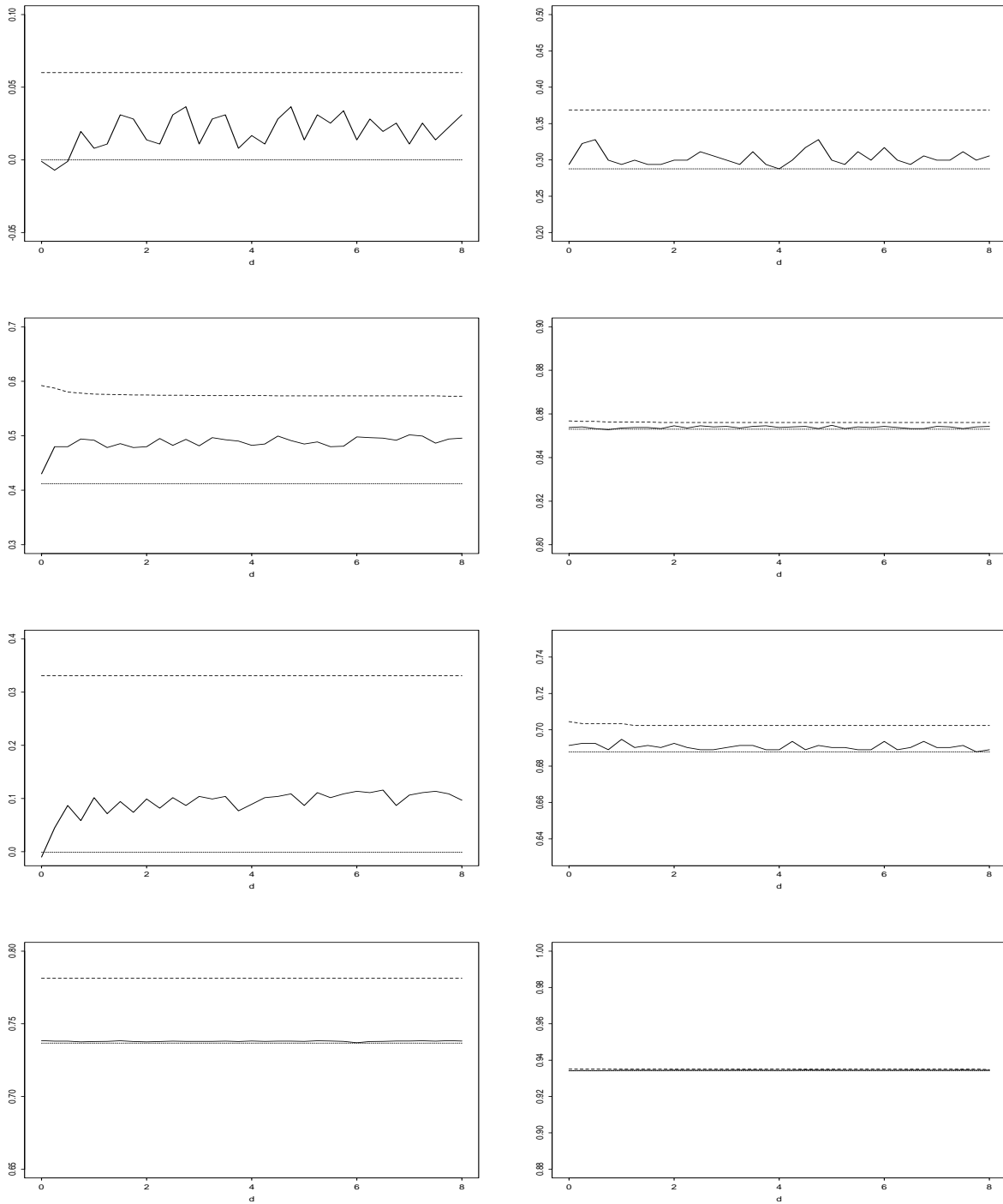
FIGURE 4. Estimated percentage of time the buffer is empty as a function of d. The chosen parameters are the same as in Figure 3. The dotted lines represent theoretical upper and lower bounds for the probability. The bounds can be easily derived by using Corollary 4.6 and (4.6).