A Note on the Event Horizon for a Processor Sharing Queue *

Robert C. Hampshire

Heinz School of Public Policy and Management

Carnegie Mellon University

hamp@andrew.cmu.edu

William A. Massey

Department of Operations Research and Financial Engineering

Princeton University

wmassey@princeton.edu

July 20, 2008

Abstract

In this note we identify a phenomenon for processor sharing queues that is unique to ones with time-varying rates. This property was discovered while correcting a proof in Hampshire, Harchol-Balter and Massey [5]. If the arrival rate for processor sharing queue has unbounded growth over time, then it is possible for the number of customers in a processor sharing queue to grow so quickly that a newly entering job never finishes. We define the minimum size for such a job to be the *event horizon* for a processor sharing queue. We discuss the use of such a concept and develop some of its properties. This short article serves both as errata for [5] and as documentation of a characteristic feature for some processor sharing queues with time varying rates.

Keywords: processor sharing queues, fluid limits, diffusion limits, transient behavior, dynamical queueing systems, uniform acceleration, sojourn times, virtual customers.

1 Introduction

In recent years there has been an increased research interest in the analysis of processor sharing queues. Examples of this research include two papers of Gromoll, Puha and Williams [3] and Puha and Williams [7]. These two papers provide a general framework for studying the fluid limits of $GI/GI/1/PS$ processor sharing queues via a measure-valued state descriptor, where

---

the queue length and the residual service time process are modeled as measure-valued processes. Additionally, the papers by Puha and Williams [6] and Gromoll [2], develop diffusion limit results as well.

The work of Hampshire, Harchol-Balter and Massey [5] derives fluid and diffusion limit theorems for the sojourn time of a virtual customer in an $M_t/M/1$ processor sharing queue with time varying or non-homogeneous, Poisson arrival rates. This note, based on work in Hampshire [4], is an extension of [5]. Here we chronicle a phenomenon that can *only* be observed in a processor sharing system with time varying rates.

To begin, we let $Q = \{ Q(t) \mid t \geq 0 \}$ be a positive (including zero) stochastic process. As in Hampshire, Harchol-Balter and Massey [5] we construct an associated, positive process $\{ T(x) \mid x \geq 0 \}$ where $T(x) = \infty$ unless it solves the equation

$$x = \int_0^{T(x)} \frac{dt}{1 + Q(t)}. \tag{1.1}$$

When Q is the queue length process for a processor sharing queue, then the random variable $T(x)$ is the sojourn time for a virtual customer with constant job size x that arrives at time 0. By virtual, we mean that the processing time $T(x)$ of this job is affected by the presence of the current jobs and potentially by the new jobs that arrive in the future. However, this specific job has *no* effect on the processing times of any of the other jobs in the queue.

We now define the *event horizon* for a virtual job in a processor sharing queue to be a random variable X_∞, where

$$X_\infty \equiv \int_0^\infty \frac{dt}{1 + Q(t)}. \tag{1.2}$$

Unlike the first-in, first-out (FIF0) service discipline, it is possible for a job to arrive to a finite queue but the arrival rates for the subsequent jobs grow so quickly that this job never finishes under the processor sharing service discipline. For any job of size x, we then have

$$\{T(x) < \infty\} = \{x < X_\infty\}. \tag{1.3}$$

Below we give some simple bounds for a finite event horizon.

Theorem 1.1 *If our generalized queueing system process $\{ Q(t) \mid t \geq 0 \}$ has the lower bound of*

$$Y \cdot t^\alpha \leq Q(t) \tag{1.4}$$

for all $t \geq 0$, where $Y > 0$ is a random variable and $\alpha > 1$ is a constant, then the resulting event horizon is finite with

$$0 < X_\infty \leq \frac{\pi/\alpha}{Y^{1/\alpha} \cdot \sin(\pi/\alpha)}. \tag{1.5}$$

Moreover, equality holds for X_∞ and its upper bound if and only if $Q(t) = Y \cdot t^\alpha$ for almost all $t \geq 0$. Finally, if the queueing system process has a linear upper bound or

$$Q(t) \leq Z \cdot t \tag{1.6}$$

for all sufficiently large t, where $Z > 0$ is a random variable independent of t, then the event horizon is always infinite.

2

Proof: Given the bound (1.4) and the integral formula in Gradshteyn and Ryzhik [1] (see 3.241 on page 340 for the case of $\nu = 2$), we have a finite event horizon since

$$\int_0^\infty \frac{dt}{1 + Q(t)} \leq \int_0^\infty \frac{dt}{1 + Y \cdot t^\alpha} = \frac{1}{Y^{1/\alpha}} \cdot \frac{2}{\alpha} \cdot \int_0^\infty \frac{s^{(2/\alpha)-1} ds}{1 + s^2} = \frac{1}{Y^{1/\alpha}} \cdot \frac{\pi/\alpha}{\sin(\pi/\alpha)}. \tag{1.7}$$

Moreover, if we are given the bound (1.6), then we have

$$\int_0^\infty \frac{dt}{1 + Z \cdot t} \leq \int_0^\infty \frac{dt}{1 + Q(t)}. \tag{1.8}$$

and now the event horizon is infinite. ∎

Observe that queues with bounded arrival rates satisfy this linear growth rate. This means that the phenomenon of a finite event horizon only occurs in queues with growth rates that are time-varying, not constant. As an example, consider the case of a deterministic model with input rate $\lambda(t) \equiv 2at$ and output rate $\mu(t) \equiv 2bt$ with $a > b$. We then have

$$Q(t) \equiv \int_0^t \left(\lambda(s) - \mu(s) \right) ds = (a - b) \cdot t^2 \quad \text{and} \quad X_\infty = \frac{\pi}{2 \cdot \sqrt{a - b}}. \tag{1.9}$$

Any job arriving at time 0 that is the size of X_∞ or larger takes an infinite amount of time to finish under the processor sharing service discipline.

Now consider a family of scaled queueing processes $\{ Q^\eta / \eta \mid \eta > 0 \}$ as constructed in [5]. For a given job size x, the corresponding sojourn times are denoted by $\{ T^\eta(x) \mid \eta > 0 \}$. Moreover, assume that this family of queueing processes has a *functional strong law of large numbers* limit where

$$\lim_{\eta \to \infty} \sup_{0 \leq t \leq \tau} \left| Q^{(0)}(t) - \frac{1}{\eta} Q^\eta(t) \right| = 0 \quad \text{a.s.} \tag{1.10}$$

for all $\tau > 0$, where $\{ Q^{(0)}(t) \mid t \geq 0 \}$ is some positive deterministic process. In [5] we asserted that such a family of queueing processes has a corresponding limit theorem for their processor sharing sojourn times. We restate the theorem here and correct the proof given there. Now our result extends to the cases of finite and infinite event horizons for the limiting fluid model which we denote by $X_\infty^{(0)}$.

Theorem 1.2 *If x is any job size, then we have*

$$\lim_{\eta \to \infty} T^\eta(x) = T^{(0)}(x) \quad \text{a.s.} \quad \text{where} \quad x = \int_0^{T^{(0)}(x)} \frac{dt}{1 + Q^{(0)}(t)}. \tag{1.11}$$

Proof: First, we assume that

$$\mathsf{P}\left\{ \liminf_{\eta \to \infty} T^\eta(x) < \infty \right\} > 0 \tag{1.12}$$

This implies that there exists a random variable T and a random subsequence $\{ \eta(k) \mid k \geq 0 \}$ where

$$\mathsf{P}\left\{ \lim_{k \to \infty} T^{\eta(k)}(x) = T < \infty \right\} > 0 \quad \text{and} \quad \lim_{k \to \infty} \frac{1}{\eta(k)} Q^{\eta(k)}(t) = Q^{(0)}(t) \quad \text{a.s.} \tag{1.13}$$

3

The a.s. convergence follows from (1.10). By (1.1) and dominated convergence, we have for all k

$$x = \int_0^{T^{\eta(k)}(x)} \frac{dt}{1 + Q^{\eta(k)}(t)/\eta(k)} = \int_0^T \frac{dt}{1 + Q^{(0)}(t)}. \tag{1.14}$$

Hence $T^{(0)}(x) = T < \infty$ and so $x < X_\infty^{(0)}$.

Now for any $\delta > 0$, we have

$$x < \int_0^{T^{(0)}(x)+\delta} \frac{dt}{1 + Q^{(0)}(t)}. \tag{1.15}$$

Hence, there exists an $\epsilon > 0$ such that

$$x = \int_0^{T^{(0)}(x)+\delta} \frac{dt}{1 + \epsilon + Q^{(0)}(t)}. \tag{1.16}$$

Using (1.10),we have for all sufficiently large η

$$\sup_{0 \le t \le T^{(0)}(x)+\delta} \left| Q^{(0)}(t) - \frac{1}{\eta} Q^\eta(t) \right| \le \epsilon \quad \text{a.s.} \tag{1.17}$$

It follows that

$$\frac{1}{\eta} Q^\eta(t) \le Q^{(0)}(t) + \sup_{0 \le t \le T^{(0)}(x)+\delta} \left| Q^{(0)}(t) - \frac{1}{\eta} Q^\eta(t) \right| \le Q^{(0)}(t) + \epsilon \quad \text{a.s.} \tag{1.18}$$

for all $0 \le t \le T^{(0)}(x) + \delta$, which gives us

$$x = \int_0^T \frac{dt}{1 + \epsilon + Q^{(0)}(t)} \le \int_0^T \frac{dt}{1 + Q^\eta(t)/\eta} \quad \text{a.s.} \tag{1.19}$$

This implies that $T^\eta(x) \le T^{(0)}(x) + \delta < \infty$ a.s. for all sufficiently large η and so

$$\limsup_{\eta \to \infty} T^\eta(x) \le T^{(0)}(x) + \delta < \infty \quad \text{a.s.} \tag{1.20}$$

Now all subsequences of $\{ T^\eta(x) \mid \eta > 0 \}$ have convergent subsequences with the same limit $T^{(0)}(x)$, so (1.11) holds.

Finally, what we first proved can be restated as $x \ge X_\infty^{(0)}$ implies

$$\mathsf{P} \left\{ \liminf_{\eta \to \infty} T^\eta(x) < \infty \right\} = 0, \tag{1.21}$$

which is equivalent to saying

$$\lim_{\eta \to \infty} T^\eta(x) = \infty \quad \text{a.s.} \tag{1.22}$$

Since $x \ge X_\infty^{(0)}$ means that $T^{(0)}(x) = \infty$, then (1.11) holds. ∎

4

This theorem does not necessarily imply a convergence result for event horizons. Consider a scaled family of deterministic processes $\{\, Q^\eta \mid \eta > 0 \,\}$ where

$$Q^\eta(t) \equiv \eta \cdot \left(Q^{(0)}(t) - \frac{\left(Q^{(0)}(t) + 1\right)^2}{Q^{(0)}(t) + 1 + t + \eta} \right), \qquad (1.23)$$

These are positive processes as long as $\eta \geq 2$ and $Q^{(0)}(t) \geq 1$ for all $t \geq 0$. We then have

$$\sup_{0 \leq t \leq T} \left| Q^{(0)}(t) - \frac{1}{\eta} Q^\eta(t) \right| \leq \frac{1}{\eta} \cdot \sup_{0 \leq t \leq T} \left(Q^{(0)}(t) + 1 \right)^2, \qquad (1.24)$$

which means that (1.10) holds when $Q^{(0)}$ is bounded on compact intervals of time. Moreover, we have the relation

$$\frac{1}{1 + Q^\eta(t)/\eta} = \frac{1}{1 + Q^{(0)}(t)} + \frac{1}{t + \eta}, \qquad (1.25)$$

so with the proper choice of the process $Q^{(0)}$, we have

$$X_\infty^{(0)} = \int_0^\infty \frac{dt}{1 + Q^{(0)}(t)} < \infty \quad \text{but} \quad X_\infty^\eta = \int_0^\infty \frac{dt}{1 + Q^\eta(t)/\eta} = \infty \quad \text{for all } \eta \geq 2. \qquad (1.26)$$

However using Fatou's lemma, we can say in general that

$$X_\infty^{(0)} \leq \liminf_{\eta \to \infty} X_\infty^\eta. \qquad (1.27)$$

Acknowledgements

References

[1] Gradshteyn, I. S. and Ryzhik, I. M. *Table of Integrals, Series and Products (5th Edition)*, Academic Press, 1994.

[2] Gromoll, H. C. "Diffusion approximation for a processor sharing queue in heavy traffic", *Annals of Applied Probability*, 14 (2004), 555–611.

[3] Gromoll, H. C., Puha, A. and Williams, R. "The Fluid Limit of a Heavily Loaded Processor Sharing Queue", *Annals of Applied Probability*, 12 (2002), 797–859.

[4] Hampshire, R. C. "Dynamic Queueing Models for the Operations Management of Communication Services.". *Ph.D. Dissertation*, Princeton University, March 2007.

[5] Hampshire, R. C., Harchol-Balter, M. and Massey, W. A. "Fluid and Diffusion Limits for Transient Sojourn Times of Processor Sharing Queues with Time Varying Rates.". *Special issue of Queueing Systems and Their Applications on Fair Resource Sharing*, Issue 1/2, 53 (2006), 19–30.

[6] Puha, A. and Williams, R. "Invariant States and Rates of Convergence for a Critical Fluid Model of a Processor Sharing Queue", *Annals of Applied Probability*, 14 (2004), 517–554.

[7] Puha, A. and Williams, R. "The Fluid Limit of an Overloaded Processor Sharing Queue", *Mathematics of Operations Research*, 31 (2006), 316–350.