# Monitoring of Stochastic Particle Systems: Analysis and Optimization

Iddo Eliazar* Uri Yechiali†

October 9, 2007

**Abstract**

Consider a system to which particles with random lifetimes flow stochastically. The system is monitored at discrete time epochs following a renewal process. When the system is detected non-empty (upon monitoring), a service procedure is initiated – clearing the system off particles. Once the service procedure is concluded, the system's evolution regenerates.

Particles may represent customers or jobs in a queueing system, contaminants in a physical system, hazardous chemical or biological agents in an environmental system, standing buy/sell orders at a brokerage center, etc.

This class of stochastic systems is modeled and analyzed. We **(i)** derive the joint transform of the time-to-first-detection and the number of particles present in the system at that epoch, and compute their statistics; **(ii)** define and calculate various path-functionals and performance measures of the system; and, **(iii)** study the issue of optimal monitoring schemes.

**Keywords**: stochastic particle systems; queueing theory; discrete-time monitoring; optimal monitoring.

## 1 Introduction

A multitude of physical or 'real world' systems can be characterized, schematically, as follows:

Independent particles flow stochastically into a system. The particles remain in the system for a random duration of time – their lifetime – and then exit (or vanish). While in the system, the particles need to be attended and

*Department of Technology Management, Holon Institute of Technology, P.O. Box 305, Holon 58102, Israel. E-mail: *eliazar@post.tau.ac.il*

†Department of Statistics & Operations Research, School of Mathematical Sciences, Raymond & Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: *uriy@post.tau.ac.il*

processed. Examples of such systems include: impatient customers arriving to a service center; pollutants flowing into a physical system; hazardous chemical or biological agents contaminating an environmental system; standing buy/sell orders arriving to a brokerage center; etc. Often, service in such systems is not provided continuously. Rather, the systems are monitored discretely (i.e., at discrete time epochs), and when detected non-empty a service procedure is initiated.

The case of impatient customers, arriving to a service center with vacationing server(s), was studied in the realm of queueing theory. Queueing systems in which the server(s) is (are) unavailable while on vacation are treated in Levy & Yechiali [5], Takagi [7], Yechiali [9], and Levy & Yechiali [6]. Impatient customers, on the other hand, who abandon the queue if the server does not return from its vacation before their impatience time runs out, are considered in Altman & Yechiali [1]. In such systems the server's vacationing regime can be looked upon as a stochastic monitoring process.

Our aim in this work is to introduce and study a general model for the type of discretely-monitored stochastic particle systems described above. We consider a generic particle system with the following features: **(i)** particle lifetimes are governed by a common arbitrary probability law; **(ii)** monitoring takes place following an arbitrary renewal process; **(iii)** the service procedure is general (e.g., sequential, parallel, etc.). Upon the completion of a service procedure the system is set back to its empty state (i.e., with no particle present), and the monitoring process resumes anew. We study this model focusing on its analysis and optimization.

The paper is organized as follows. Section 2 introduces the generic system-model and describes, in detail, the system's three underlying processes: inflow, monitoring, and service. In Section 3 we conduct an analysis of the system, computing the joint statistics of the following pair of random variables: the time-to-first-detection and the number of particles present in the system at that time epoch. Section 4 introduces various path-functionals of the systems and studies their statistics. Section 5 introduces various types of service procedures and studies their statistics. Last, in Section 6, we define several performance measures (i.e., quantitative indicators measuring the systems' performance), and use them in order to devise optimal monitoring schemes.

**A note about notation**

Throughout the manuscript: **(i)** the sign $\stackrel{d}{=}$ will denote equality in law (of random variables); and, **(ii)** the function $\overline{F}(s) = 1 - F(s)$ will denote the tail probability of a given probability distribution function $F(s)$ (defined on the non-negative half-line $s \geq 0$).

## 2 The model

Consider a system, initiated at time $t = 0$, to which there is a random inflow of particles. The system is monitored at discrete time epochs. If the monitor finds the system clear of particles, no action is taken. However, when the monitor finds the system non-empty a service procedure is initiated. The service procedure clears off particles from the system, at the end of which the system is set back to its 'empty state' (with no particles in it).

Let us specify in further detail the system's three underlying processes: *inflow*, *monitoring*, and *service*:

**The inflow process**

Particles arrive to the system according to a Poisson process with rate $\lambda$. A particle arriving to the system, while not in service, remains in it for a random duration of time – the particle's 'lifetime' – and then departs. The particles are assumed independent and identically distributed (IID), and their common lifetime distribution function is denoted $F(s)$ $(s \geq 0)$.

**The monitoring process**

The system is monitored at a random sequence of discrete time epochs denoted $0 < T_1 < T_2 < T_3 < \cdots$. The sequence of monitoring epochs $\{T_k\}_{k=1}^{\infty}$ is assumed a renewal process. That is, $T_k = \Delta_1 + \cdots + \Delta_k$ where $\{\Delta_k\}_{k=1}^{\infty}$ is an IID sequence of positive-valued random variables, all distributed like the generic random variable $\Delta$ – the 'inter-monitoring period'.

**The service process**

Once the monitor finds the system non-empty, a service procedure commences. We denote by $S_n$ $(n = 1, 2, 3, \cdots)$ the random duration of time it takes the service procedure to clear the system and restore it to the 'empty state' – given that $n$ particles were found in the system when monitored.

Particles arriving to the system while it is in service can be either rejected, or accepted and treated within the service procedure in process – pending on the specific features of the service regime.

The three processes – inflow, monitoring, and service – are assumed independent. Once the system is restored to its 'empty state' the entire process *regenerates*: the system's clock is reset to $t = 0$, and a new monitoring sequence (independent of the past) is initiated.

# 3 Analysis

In this Section we conduct a basic analysis of the system described in Section 2. We begin with an analysis of the inflow process, and then turn to study issues regarding the *time-to-first-detection*: **(i)** how long does it take till a system is first detected non-empty? and, **(ii)** how many particles are in the system at that time epoch?

### 3.1 The distribution of the inflow process

Consider a system with inflow *alone* – that is, with no monitoring and service taking place. Let $N(t)$ $(t \geq 0)$ denote the number of particles in the system at time $t$, and let $D(t)$ $(t \geq 0)$ denote the number of particles that have departed the system up to time $t$. The following result, which is a generalization of Bartlett's theorem for Poisson Processes (see, for example, Kingman [4]), will serve us in the sequel:

**Proposition 1** *$N(t)$ and $D(t)$ are independent and Poisson-distributed random variables. The mean of $N(t)$ is $\Lambda(t)$, and the mean of $D(t)$ is $\lambda t - \Lambda(t)$, where:*

$$\Lambda(t) = \lambda \int_0^t \overline{F}(s)ds \ . \tag{1}$$

Readers familiar with the theory of queueing systems should identify $N(t)$ as the queue-size (at time $t$) of an $M/G/\infty$ queueing system with arrival rate $\lambda$ and job-size governed by the distribution $F$. The proof of Proposition 1 can be found in Takacs [8].

For example, if the particles' lifetime is Exponentially-distributed with parameter $\alpha$ – namely: $\overline{F}(s) = \exp\{-\alpha s\}$ – then

$$\Lambda(t) = \frac{\lambda}{\alpha}(1 - \exp\{-\alpha t\}) \ . \tag{2}$$

And, if the particles' lifetime is Pareto-distributed with exponent $\alpha$ – namely: $\overline{F}(s) = (1+s)^{-\alpha}$ – then

$$\Lambda(t) = \begin{cases} \frac{\lambda}{1-\alpha}\left\{(1+t)^{1-\alpha} - 1\right\} & \text{if } 0 < \alpha < 1 \ , \\ \lambda \ln(1+t) & \text{if } \alpha = 1 \ , \\ \frac{\lambda}{\alpha-1}\left\{1 - (1+t)^{-(\alpha-1)}\right\} & \text{if } \alpha > 1 \ . \end{cases} \tag{3}$$

### 3.2 The time-to-first-detection

Let $K \in \{1, 2, 3, \cdots\}$ be the number of the first monitoring epoch at which the system was found non-empty. Namely:

$$K = \inf\{k \geq 1 \mid N(T_k) > 0\} \ . \tag{4}$$

The random variable $K$ is geometrically-distributed with parameter $p = 1 - \mathbf{E}[\exp\{-\Lambda(\Delta)\}]$, that is:

$$\mathbf{P}\left(K > k\right) = \mathbf{E}[\exp\{-\Lambda(\Delta)\}]^k \tag{5}$$

$(k = 0, 1, 2, \cdots)$. We explain:

Observe the system at the first monitoring epoch $T_1 \stackrel{d}{=} \Delta$. If there are no particles in the system at time $T_1$ then the entire system process regenerates, and otherwise – a service procedure initiates. Hence, $K$ is geometrically-distributed with parameter $p = \mathbf{P}(N(T_1) > 0)$. However

$$\mathbf{P}\left(N(T_1) = 0\right) = \mathbf{P}\left(N(\Delta) = 0\right) = \mathbf{E}\left[\mathbf{P}\left(N(\Delta) = 0 \mid \Delta\right)\right] = \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right] , \tag{6}$$

and hence $\mathbf{P}(N(T_1) > 0)$ equals $1 - \mathbf{E}[\exp\{-\Lambda(\Delta)\}]$.

We now turn to analyze the random pair $(\tau, N(\tau))$ where: $\tau := T_K$ is the '*time-to-first-detection*' – the time till the first monitoring epoch at which a non-empty system was detected (and a service procedure was called upon); and, $N(\tau)$ is the number of particles present at the system at the time $\tau$. It should be emphasized that while the inflow process and the monitoring epochs are independent, the random time $\tau$ and the inflow process are, on the other hand, highly correlated.

The random pair $(\tau, N(\tau))$ satisfies the following regeneration equation:

$$(\tau, N(\tau)) \stackrel{d}{=} (T_1, N(T_1)) + (\tau', N'(\tau')) \cdot \mathbf{I}_{\{N(T_1)=0\}} , \tag{7}$$

where $(\tau', N'(\tau'))$ is an IID copy of $(\tau, N(\tau))$ which is independent of $(T_1, N(T_1))$. The explanation of equation (7) is identical to the explanation regarding the distribution of random variable $K$:

Observe the system at the first monitoring epoch $T_1 \stackrel{d}{=} \Delta$. If particles are present in the system at time $T_1$ then $(\tau, N(\tau)) = (T_1, N(T_1))$. However, if the system is empty at time $T_1$ (i.e., if $N(T_1) = 0$) then $\tau = T_1 + \tau'$ and $N(\tau) = N'(\tau')$. Writing these two scenarios in a single equation yields (7).

Equation (7), in turn, leads to:

**Proposition 2** *The joint transform of the random pair $(\tau, N(\tau))$ is*

$$\boldsymbol{E}\left[\exp\{-\omega\tau\} z^{N(\tau)}\right] = \frac{\boldsymbol{E}\left[\exp\{-\omega\Delta - (1-z)\Lambda(\Delta)\}\right] - \boldsymbol{E}\left[\exp\{-\omega\Delta - \Lambda(\Delta)\}\right]}{1 - \boldsymbol{E}\left[\exp\{-\omega\Delta - \Lambda(\Delta)\}\right]} . \tag{8}$$

*where $\omega \geq 0$ and $|z| \leq 1$.*

The proof of Proposition 2 is given in the Appendix. Since $\tau$ is positive valued, and since $N(\tau)$ is integer valued, we used a 'hybrid' transform (Laplace transform for the positive-valued random variable $\tau$, and $z$-transform for the integer-valued random variable $N(\tau)$).

In particular, Proposition 2 implies the following corollaries:

- The Laplace transform of the time-to-first-detection $\tau$ is given by:

$$\mathbf{E}\left[\exp\{-\omega\tau\}\right] = \frac{\mathbf{E}\left[\exp\{-\omega\Delta\}\right] - \mathbf{E}\left[\exp\{-\omega\Delta - \Lambda(\Delta)\}\right]}{1 - \mathbf{E}\left[\exp\{-\omega\Delta - \Lambda(\Delta)\}\right]} . \tag{9}$$

- The $z$-transform of the random variable $N(\tau)$ is given by:

$$\mathbf{E}\left[z^{N(\tau)}\right] = \frac{\mathbf{E}\left[\exp\{-(1-z)\Lambda(\Delta)\}\right] - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]} \ . \tag{10}$$

  The Taylor expansion of the $z$-transform of Equation (10), in turn, implies that the probability distribution of the random variable $N(\tau)$ is given by:

$$\mathbf{P}\left(N(\tau) = n\right) = \frac{1}{n!} \frac{\mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\Lambda(\Delta)^n\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]} \quad \left( = \frac{\mathbf{P}\left(N(\Delta) = n\right)}{\mathbf{P}\left(N(\Delta) > 0\right)} \right) \tag{11}$$

  $(n = 1, 2, 3, \cdots)$.

- The covariance $\mathbf{Cov}\left(\tau, N(\tau)\right)$ of the pair $(\tau, N(\tau))$ equals

$$\frac{\mathbf{E}\left[\Delta \cdot \Lambda(\Delta)\right] \mathbf{E}\left[1 - \exp\{-\Lambda(\Delta)\}\right] - \mathbf{E}\left[\Lambda(\Delta)\right] \mathbf{E}\left[\Delta(1 - \exp\{-\Lambda(\Delta)\})\right]}{\left(1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]\right)^2} \ . \tag{12}$$

  Equation (12) is derived from equation (8) by differentiation (see the Appendix for the details).

# 4 Path functionals

A *path functional* $Y$ is a functional of the system's sample path trajectory along the random time interval $[0, \tau]$ (spanning from system initiation till the epoch of first-detection).

Various path functionals satisfy the following regeneration equation:

$$Y \stackrel{d}{=} X + Y' \cdot \mathbf{I}_{\{N(T_1)=0\}} \ , \tag{13}$$

where $X$ is the value of the functional $Y$ along the random time interval $[0, T_1]$ (spanning from system initiation till the first monitoring epoch), and where $Y'$ is an IID copy of $Y$ which is independent of the random pair $(T_1, N(T_1))$.

We shall henceforth refer to functionals admitting the regenerative representation of equation (13) as *additive path functionals*. The mean value of an additive path functional $Y$ is given by:

$$\mathbf{E}\left[Y\right] = \frac{\mathbf{E}\left[X\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]} \ . \tag{14}$$

We explain:

Taking expectation on both sides of equation (13), while using equation (6), gives

$$\mathbf{E}\left[Y\right] = \mathbf{E}\left[X\right] + \mathbf{E}\left[Y\right] \mathbf{P}\left(N(T_1) = 0\right) = \mathbf{E}\left[X\right] + \mathbf{E}\left[Y\right] \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right] \ . \tag{15}$$

Equation (15), in turn, yields equation (14).

We present six examples of additive path functionals (in examples $2-6$ we recall that $T_1 \overset{d}{=} \Delta$ and use Proposition 1):

1. The number of monitoring epochs is the additive path functional $Y = K$. In this case $X = 1$ and hence:
$$\mathbf{E}[X] = 1 \ .$$
2. The time-to-first-detection is the additive path functional $Y = \tau$. In this case $X = T_1$ and hence:
$$\mathbf{E}[X] = \mathbf{E}[\Delta] \ .$$
3. The number of particles present in the system at the epoch of first detection is the additive path functional $Y = N(\tau)$. In this case $X = N(T_1)$ and hence:
$$\mathbf{E}[X] = \mathbf{E}[\Lambda(\Delta)] \ .$$
4. The number of particles that have departed the system till the epoch of first detection is the additive path functional $Y = D(\tau)$. In this case $X = D(T_1)$ and hence:
$$\mathbf{E}[X] = \lambda\mathbf{E}[\Delta] - \mathbf{E}[\Lambda(\Delta)] \ .$$
5. The cumulative particle-load incurred by the system till the epoch of first detection is the additive path functional $Y = \int_0^\tau N(s)ds$. In this case $X = \int_0^{T_1} N(s)ds$ and hence:
$$\mathbf{E}[X] = \mathbf{E}\left[\int_0^\Delta \Lambda(s)ds\right] \ .$$
6. The system's 'particle-free time' till the epoch of first detection is the additive path functional $Y = \int_0^\tau \mathbf{I}\{N(s) = 0\}ds$. In this case $X = \int_0^{T_1} \mathbf{I}\{N(s) = 0\}ds$ and hence:
$$\mathbf{E}[X] = \mathbf{E}\left[\int_0^\Delta \exp\{-\Lambda(s)\}ds\right] \ .$$

**Remark**

It is tempting to use equation (14) in order to compute the joint transform of the random pair $(\tau, N(\tau))$ by setting $Y = \exp\{-\omega\tau\}z^{N(\tau)}$ (having fixed $\omega \geq 0$ and $|z| \leq 1$). However, $Y$ is *not* an *additive* functional of the system's sample-path trajectories, and hence it does *not* admit the regenerative representation of equation (13). Rather, the random variable $Y$ satisfies the regeneration equation $Y \overset{d}{=} X + (Y' - 1) \cdot \mathcal{E}_1$ where: **(i)** $X = \exp\{-\omega T_1\} z^{N(T_1)}$; **(ii)** $\mathcal{E}_1 = \exp\{-\omega T_1\} \mathbf{I}_{\{N(T_1)=0\}}$; and, **(iii)** $Y'$ is an IID copy of $Y$ (which is independent of the random pair $(T_1, N(T_1))$). This representation implies that $\mathbf{E}[Y] = (\mathbf{E}[X] - \mathbf{E}[\mathcal{E}_1]) / (1 - \mathbf{E}[\mathcal{E}_1])$. A straightforward computation of $\mathbf{E}[X]$ and $\mathbf{E}[\mathcal{E}_1]$ yields, in turn, Proposition 2.

# 5 Service procedures

In this Section we study service procedures of the monitoring system, induced by various types of *service regimes*.

Let $S$ denote the duration of a service procedure. Recall that the duration of a service procedure, initiated when the monitor detects $n$ particles present in the system, was denoted by $S_n$ ($n = 1, 2, 3 \cdots$). Hence:

$$\mathbf{E}[S \mid N(\tau) = n] = \mathbf{E}[S_n] \ . \tag{16}$$

Let $\xi$ denote the random time it takes to process a single particle, and let $\mu$ and $G(s)$ ($s \geq 0$) denote, respectively, the corresponding mean and distribution function. The particles – as noted already in Section 2 – are assumed IID.

We consider five types of service regimes: Gated-Sequential, Exhaustive-Sequential, Gated-Parallel, Exhaustive-Parallel, and Lévy-Structured. For these service regimes the following result holds:

**Proposition 3** *The mean duration $\mathbf{E}[S]$ of the service procedure admits the form:*

$$\boldsymbol{E}[S] = \frac{\boldsymbol{E}[\Psi(\Lambda(\Delta))]}{1 - \boldsymbol{E}[\exp\{-\Lambda(\Delta)\}]} \ , \tag{17}$$

*where $\Psi(\theta)$ ($\theta \geq 0$) is a function contingent on the type of the system's service regime.*

The proof of Proposition 3 is given in the Appendix. We turn now to specify the function $\Psi(\theta)$ for each of the five different service-regime types.

**Gated-Sequential**

In the Gated-Sequential service regime: **(i)** the system is 'gated' once the service procedure commences – rejecting particles arriving to the system while in service; and, **(ii)** the service is rendered in a sequential fashion – processing the particles serviced in a sequential order.

The service time $S_n$ is thus the *sum* of $n$ IID copies of the random time $\xi$ – implying that $\mathbf{E}[S_n] = \mu n$. In this case

$$\Psi_{\mathrm{GS}}(\theta) = \mu\theta \ . \tag{18}$$

**Exhaustive-Sequential**

In the Exhaustive-Sequential service regime: **(i)** the service procedure 'exhausts' the system – accepting and processing also particles arriving to the system while in service; and, **(ii)** the service is rendered in a sequential fashion – processing the particles serviced in a sequential order.

The service procedure thus operates in a form of an $M/G/1$ queueing system with arrival rate $\lambda$ and job-size governed by the distribution $G$. Consequently, the service time $S_n$ equals an $M/G/1$ *Busy Period* initiated with $n$ awaiting jobs – implying that $\mathbf{E}[S_n] = \frac{\mu}{1-\lambda\mu}n$ [8]. In this case

$$\Psi_{\mathrm{ES}}(\theta) = \frac{\mu}{1-\lambda\mu}\theta \ . \tag{19}$$

**Gated-Parallel**

In the Gated-Parallel service regime: **(i)** the system is 'gated' once the service procedure commences – rejecting particles arriving to the system while in service; and, **(ii)** the service is rendered in a parallel fashion – processing the particles serviced in a parallel order.

The service time $S_n$ is thus the *maximum* of $n$ IID copies of the random time $\xi$. In this case

$$\Psi_{\mathrm{GP}}(\theta) = \int_0^\infty \Big(1 - \exp\{-\theta\overline{G}(x)\}\Big)\, dx \ . \tag{20}$$

**Exhaustive-Parallel**

In the Exhaustive-Parallel service regime: **(i)** the service procedure 'exhausts' the system – accepting and processing also particles arriving to the system while in service; and, **(ii)** the service is rendered in a parallel fashion – processing the particles serviced in a parallel order.

The service procedure thus operates in a form of an $M/G/\infty$ queueing system with arrival rate $\lambda$ and job-size governed by the distribution $G$. Consequently, the service time $S_n$ equals an $M/G/\infty$ *Busy Period* initiated with $n$ awaiting jobs. In this case

$$\Psi_{\mathrm{EP}}(\theta) = \int_0^\infty \Big(1 - \exp\{-\theta\overline{G}(x)\}\Big) \exp\left\{\lambda \int_x^\infty \overline{G}(u)du\right\} dx \ . \tag{21}$$

**Lêvy-Structured**

In all the four service regime specified so far the mean service duration $\mathbf{E}[S_n]$ can be expressed in the general form

$$\mathbf{E}[S_n] = an + \int_0^\infty \Big(1 - \exp\{-nx\}\Big)\, \phi(x)dx \ , \tag{22}$$

where $a$ is a non-negative valued constant, and where $\phi(x)$ $(x > 0)$ is a non-negative valued function.

Readers familiar with the theory of Lévy processes should identify the function appearing on the right-hand-side of equation (22) as the Lévy characteristic of a Lévy subordinator with drift parameter $a$ and Lévy measure $\phi(x)dx$, evaluated at the point $n$ (see, for example, Bertoin [2]).[1] Hence, we refer to service regimes whose mean service durations are expressible in the form of equation (22) as "Lévy-Structured".

In the case of Lévy-Structured service regimes we have:

$$\Psi_{\text{LS}}(\theta) = a\theta + \int_0^1 \Big(1 - \exp\{-\theta y\}\Big) \frac{\phi(-\ln(1-y))}{1-y} dy \ . \tag{23}$$

The following examples illustrate the wide structural spectrum attainable by the class of Lévy-Structured service regimes (in all examples the drift parameter $a$ is set to zero):

- The power-law structure $\mathbf{E}[S_n] = n^\alpha$, with exponent $0 < \alpha < 1$, is obtained via $\phi(x) = (\alpha/\Gamma(1-\alpha))x^{-1-\alpha}$.
- The logarithmic structure $\mathbf{E}[S_n] = \ln(1+n)$ is obtained via $\phi(x) = \exp\{-x\}/x$.
- The power-law structure $\mathbf{E}[S_n] = 1 - (1+n)^{-\alpha}$, with exponent $\alpha > 0$, is obtained via $\phi(x) = (1/\Gamma(\alpha))\exp\{-x\}x^{\alpha-1}$.

# 6 Performance measures and monitoring optimization

In this last Section we introduce *performance measures* which, as their name implies, quantitatively measure the performance of the monitoring system. Equipped with these performance measures we address the issue of *monitoring optimization*: how to monitor a system so that to obtain optimal performance.

## 6.1 Performance measures

Given a path functional $Y$, consider the following *performance measure* $\Theta(Y)$ associated with it:

$$\Theta(Y) = \frac{\mathbf{E}[Y]}{\mathbf{E}[\tau] + \mathbf{E}[S]} \ . \tag{24}$$

The performance measure $\Theta(Y)$ measures the value of the path functional $Y$ relative to time. We explain:

Consider the first $m$ system cycles. Each cycle is composed of a period of length $\tau$ during which the system is occasionally monitored (but not serviced),

[1] The necessary and sufficient condition for the well-definiteness of the Lévy characteristic is that the integrability condition $\int_0^\infty \min\{x,1\}\,\phi(x)dx < \infty$ be satisfied.

and a period of length $S$ during which the system is serviced. Hence, the cumulative value of the path functional over all the first $m$ system cycles, relative the cumulative length of these cycles is given by

$$\frac{Y_1 + \cdots + Y_m}{(\tau_1 + S_1) + \cdots + (\tau_m + S_m)} \ . \tag{25}$$

Moreover, at the termination of each cycle the system regenerates, and the triplets $(\tau_1, S_1, Y_1), \cdots, (\tau_m, S_m, Y_m)$ are therefore IID. Hence, due to the Law of Large Numbers, the stochastic ratio (25) converges (almost surly), as $m \to \infty$, to the deterministic limit (24).

Now, if $Y$ is an additive path functional (i.e., if it satisfies equation (13)), then the mean value of $Y$ is given by equation (14). On the other hand, if the service procedure follows one of the five service regimes considered in Section 5, then the mean service duration $\mathbf{E}[S]$ admits the form of equation (17). Hence, using equations (14) and (17), we obtain that the performance measure $\Theta(Y)$ is given by

$$\Theta(Y) = \frac{\mathbf{E}[X]}{\mathbf{E}[\Delta] + \mathbf{E}[\Psi(\Lambda(\Delta))]} \ , \tag{26}$$

where: **(i)** $X$ is the value of the functional $Y$ along the random time interval $[0, T_1]$ (spanning from system initiation till the first monitoring epoch); and, **(ii)** the function $\Psi(\theta)$ is contingent on the *type* of the system's service regime.

For example, if we take the additive path functional to be the time-to-first-detection $Y = \tau$ then: $\Theta(Y)$ is the proportion of time in which the system *is* in service; and, conversely, $1 - \Theta(Y)$ is the proportion of time in which the system *is not* in service. These proportions are given, respectively, by:

$$\frac{\mathbf{E}[\Psi(\Lambda(\Delta))]}{\mathbf{E}[\Delta] + \mathbf{E}[\Psi(\Lambda(\Delta))]} \quad \text{and} \quad \frac{\mathbf{E}[\Delta]}{\mathbf{E}[\Delta] + \mathbf{E}[\Psi(\Lambda(\Delta))]} \ . \tag{27}$$

**Remark**

In some systems the length $S$ of the service procedure is "*load-independent*". That is, the distribution of $S$ is fixed and does not depend on the number of particles detected in the system. For such systems the counterpart of equation (26) is given by:

$$\Theta(Y) = \frac{\mathbf{E}[X]}{\mathbf{E}[\Delta] + \mathbf{E}[S]\left(1 - \mathbf{E}[\exp\{-\Lambda(\Delta)\}]\right)} \ . \tag{28}$$

## 6.2 Monitoring optimization

With performance measures at hand, we turn to study the issue of *optimal monitoring*.

On one hand monitoring is costly, and hence the system administrator would like to monitor the system as rarely as possible. On the other hand, costs are incurred also when there are particles present in the system and are not being processed (by the service procedure). Thus, the system administrator has to balance between these 'opposing' costs by designing an optimal monitoring policy.

To formulate this mathematically, let: **(i)** $C_{mon}$ denote the cost of a single monitoring scan; and, **(ii)** $Y_{cost}$ denote the costs incurred during a service cycle. Hence, the overall cost during a service cycle is $C_{mon}K + Y_{cost}$. If $Y_{cost}$ is an additive path functional (i.e., if it satisfies equation (13)), then the performance measure (26) associated with $Y = C_{mon}K + Y_{cost}$ is

$$\frac{C_{mon} + \mathbf{E}\left[X_{cost}\right]}{\mathbf{E}\left[\Delta\right] + \mathbf{E}\left[\Psi(\Lambda(\Delta))\right]} , \tag{29}$$

where $X_{cost}$ is the value of the functional $Y_{cost}$ along the random time interval $[0, T_1]$ (spanning from system initiation till the first monitoring epoch).

Taking the inter-monitoring period $\Delta$ to be *deterministic*, the optimal monitoring policy is given by the optimization problem

$$\inf_{\Delta>0} \frac{C_{mon} + \mathbf{E}\left[X_{cost}\right]}{\Delta + \Psi(\Lambda(\Delta))} . \tag{30}$$

Analogously, for systems with "load-independent" service procedures, the performance measure (28) leads us to the optimization problem

$$\inf_{\Delta>0} \frac{C_{mon} + \mathbf{E}\left[X_{cost}\right]}{\Delta + \mathbf{E}\left[S\right]\left(1 - \exp\{-\Lambda(\Delta)\}\right)} . \tag{31}$$

We present four examples of optimal monitoring problems:

**Server utilization**

Assume that the system has a server which executes the service procedures. The system administrator wishes to utilize the server optimally. Let $C_{idle}$ denote the cost, per unit time, of keeping the server idle. Therefore, $Y_{cost} = C_{idle}\tau$ and hence

$$\mathbf{E}\left[X_{cost}\right] = C_{idle}\Delta . \tag{32}$$

**Restoration**

Assume that the system administrator's objective is to restore the system back to the 'particle-free' state as quick as possible. Let $C_{pen}$ denote the penalty cost, per unit time, of service. Therefore, $Y_{cost} = C_{pen}S$ and hence

$$\mathbf{E}\left[X_{cost}\right] = C_{pen}\Psi(\Lambda(\Delta)) . \tag{33}$$

**De-contamination**

Assume that the particles are hazardous contaminants. Once the monitor detects hazardous particles, the system is quarantined and a de-contamination procedure (=service procedure) initiates. Let $C_{con}$ denote the cost of the damage, per unit time, incurred by a single hazardous particle present in the system (while not quarantined). Also, let $C_{decon}$ denote the cost, per unit time, of the de-contamination procedure. Therefore, $Y_{cost} = C_{con} \int_0^\tau N(s)ds + C_{decon}S$ and hence

$$\mathbf{E}\left[X_{cost}\right] = C_{con} \int_0^\Delta \Lambda(s)ds + C_{decon}\Psi(\Lambda(\Delta)) \ . \tag{34}$$

**Loss minimization**

Assume that the particle system represents jobs arriving to a service center. A job that does not begin to receive service during its 'lifetime' is lost. The system administrator's objective is to minimize the number of particles lost. Let $C_{loss}$ denote the cost per particle lost, and let $C_{ser}$ denote the cost, per unit time, of the service procedure. Therefore, $Y_{cost} = C_{loss}D(\tau) + C_{ser}S$ and hence

$$\mathbf{E}\left[X_{cost}\right] = C_{loss}\left(\lambda\Delta - \Lambda(\Delta)\right) + C_{ser}\Psi(\Lambda(\Delta)) \ . \tag{35}$$

# 7 Appendix

## 7.1 Proof of Proposition 2

For an inter-monitoring period $\Delta$ set

$$U(\omega; z) = \mathbf{E}\left[\exp\{-\omega\Delta - (1-z)\Lambda(\Delta)\}\right] \ , \tag{36}$$

where $\omega \geq 0$ and $|z| \leq 1$. Also set $V(\omega; z) = \mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\right]$, and note that

$$V(\omega; z) = \mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\mathbf{I}_{\{N(T_1)=0\}}\right] + \mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\mathbf{I}_{\{N(T_1)>0\}}\right] \ . \tag{37}$$

Using equation (7) we have

$$\begin{aligned} &\mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\mathbf{I}_{\{N(T_1)=0\}}\right] \\ &= \mathbf{E}\left[\exp\{-\omega(T_1+\tau')\}z^{N'(\tau')}\mathbf{I}_{\{N(T_1)=0\}}\right] \\ &= \mathbf{E}\left[\exp\{-\omega T_1\}\mathbf{I}_{\{N(T_1)=0\}}\right] \cdot V(\omega; z) \ , \end{aligned} \tag{38}$$

and

$$\begin{aligned} &\mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\mathbf{I}_{\{N(T_1)>0\}}\right] \\ &= \mathbf{E}\left[\exp\{-\omega T_1\}z^{N(T_1)}\mathbf{I}_{\{N(T_1)>0\}}\right] \\ &= \mathbf{E}\left[\exp\{-\omega T_1\}z^{N(T_1)}\right] - \mathbf{E}\left[\exp\{-\omega T_1\}\mathbf{I}_{\{N(T_1)=0\}}\right] \ . \end{aligned} \tag{39}$$

Due to Proposition 1 the random variable $N(t)$ is Poisson-distributed with mean $\Lambda(t)$ $(t \geq 0)$. Hence, using equation (37) and the fact that inter-monitoring periods are independent of the inflow process, we have

$$\begin{aligned} &\mathbf{E}\left[\exp\{-\omega\Delta\}\mathbf{I}_{\{N(\Delta)=0\}}\right] \\ &= \mathbf{E}\left[\exp\{-\omega\Delta\}\mathbf{P}\left(N(\Delta)=0 \mid \Delta\right)\right] \\ &= \mathbf{E}\left[\exp\{-\omega\Delta\}\exp\{-\Lambda(\Delta)\}\right] \\ &= U(\omega;0)\ , \end{aligned} \tag{40}$$

and

$$\begin{aligned} &\mathbf{E}\left[\exp\{-\omega\Delta\}z^{N(\Delta)}\right] \\ &= \mathbf{E}\left[\exp\{-\omega\Delta\}\mathbf{E}\left[z^{N(\Delta)} \mid \Delta\right]\right] \\ &= \mathbf{E}\left[\exp\{-\omega\Delta\}\exp\{-(1-z)\Lambda(\Delta)\}\right] \\ &= U(\omega;z)\ . \end{aligned} \tag{41}$$

Now, since $T_1 \stackrel{d}{=} \Delta$, combining equations (38) and (40) together yields

$$\mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\mathbf{I}_{\{N(T_1)=0\}}\right] = U(\omega;0)\cdot V(\omega;z)\ ; \tag{42}$$

and, from equations (39)-(41) we have

$$\mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\mathbf{I}_{\{N(T_1)>0\}}\right] = U(\omega;z) - U(\omega;0)\ . \tag{43}$$

Hence, substituting equations (42)-(43) back into equation (37) gives

$$V(\omega;z) = U(\omega;0)\cdot V(\omega;z) + \left(U(\omega;z) - U(\omega;0)\right)\ . \tag{44}$$

Equation (44) implies that

$$V(\omega;z) = \frac{U(\omega;z) - U(\omega;0)}{1 - U(\omega;0)}\ , \tag{45}$$

which, in turn, yields

$$\mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\right] = \frac{\mathbf{E}\left[\exp\{-\omega\Delta - (1-z)\Lambda(\Delta)\}\right] - \mathbf{E}\left[\exp\{-\omega\Delta - \Lambda(\Delta)\}\right]}{1 - \mathbf{E}\left[\exp\{-\omega\Delta - \Lambda(\Delta)\}\right]}\ . \tag{46}$$

**Proof of equation (12)**

Differentiating the function $V(\omega;z) = \mathbf{E}\left[\exp\{-\omega\tau\}z^{N(\tau)}\right]$ gives:

$$\mathbf{E}[\tau] = -\frac{\partial V}{\partial \omega}(0;1)\ ;\ \mathbf{E}[N(\tau)] = \frac{\partial V}{\partial z}(0;1)\ ;\ \mathbf{E}[\tau N(\tau)] = -\frac{\partial^2 V}{\partial \omega \partial z}(0;1)\ . \quad (47)$$

Computing the derivatives of the function $V(\omega;z)$, the covariance between the random variables $\tau$ and $N(\tau)$ is thus given by:

$$\begin{aligned} &\mathbf{Cov}(\tau, N(\tau)) = \mathbf{E}[\tau N(\tau)] - \mathbf{E}[\tau]\,\mathbf{E}[N(\tau)] \\ &= -\tfrac{\partial^2 V}{\partial\omega\partial z}(0;1) + \tfrac{\partial V}{\partial \omega}(0;1)\tfrac{\partial V}{\partial z}(0;1) \\ &= \tfrac{\mathbf{E}[\Delta\cdot\Lambda(\Delta)]\mathbf{E}[1-\exp\{-\Lambda(\Delta)\}]-\mathbf{E}[\Lambda(\Delta)]\mathbf{E}[\Delta(1-\exp\{-\Lambda(\Delta)\})]}{(1-\mathbf{E}[\exp\{-\Lambda(\Delta)\}])^2}\ . \end{aligned} \quad (48)$$

## 7.2 Proof of Proposition 3

We split the proof to three parts: **(i)** the Gated-Sequential & Exhaustive-Sequential service regimes; **(ii)** the Gated-Parallel & Exhaustive-Parallel service regimes; and, **(iii)** the Lévy-Structured service regime.

**(i) Gated-Sequential & Exhaustive-Sequential**

Assume that $\mathbf{E}[S_n] = an$ $(n = 1,2,3\cdots)$ where $a$ is a positive constant.

Equation (14) implies that the mean of the random variable $N(\tau)$ is given by

$$\mathbf{E}[N(\tau)] = \frac{\mathbf{E}[\Lambda(\Delta)]}{1-\mathbf{E}[\exp\{-\Lambda(\Delta)\}]}\ . \quad (49)$$

Combining equations (49) and (16) together we obtain that

$$\begin{aligned} &\mathbf{E}[S] = \mathbf{E}[\mathbf{E}[S \mid N(\tau)]] = \mathbf{E}[aN(\tau)] \\ &= a\tfrac{\mathbf{E}[\Lambda(\Delta)]}{1-\mathbf{E}[\exp\{-\Lambda(\Delta)\}]} = \tfrac{\mathbf{E}[\Psi(\Lambda(\Delta))]}{1-\mathbf{E}[\exp\{-\Lambda(\Delta)\}]}\ , \end{aligned} \quad (50)$$

where

$$\Psi(\theta) = a\theta \quad (51)$$

$(\theta \geq 0)$.

In the Gated-Sequential service regime $a = \mu$ and hence equation (51) implies that $\Psi_{\mathrm{GS}}(\theta) = \mu\theta$ $(\theta \geq 0)$. In the Exhaustive-Sequential service regime $a = \frac{\mu}{1-\lambda\mu}$ and hence equation (51) implies that $\Psi_{\mathrm{ES}}(\theta) = \frac{\mu}{1-\lambda\mu}\theta$ $(\theta \geq 0)$.

**(ii) Gated-Parallel & Exhaustive-Parallel**

Assume that

$$\mathbf{E}\left[S_n\right] = \int_0^\infty (1 - \psi(x)^n)\phi(x)dx \tag{52}$$

($n = 1, 2, 3 \cdots$), where $0 \le \psi(x) \le 1$ and $\phi(x) \ge 0$ ($x \ge 0$).

Equation (10) implies that

$$1 - \mathbf{E}\left[z^{N(\tau)}\right] = \frac{1 - \mathbf{E}\left[\exp\{-(1-z)\Lambda(\Delta)\}\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]} \tag{53}$$

($|z| \le 1$).

Combining equations (52), (53), and (16) together we obtain that

$$\begin{aligned}
\mathbf{E}\left[S\right] &= \mathbf{E}\left[\mathbf{E}\left[S \mid N(\tau)\right]\right] \\
&= \mathbf{E}\left[\int_0^\infty \left(1 - \psi(x)^{N(\tau)}\right)\phi(x)dx\right] \\
&= \int_0^\infty \left(1 - \mathbf{E}\left[\psi(x)^{N(\tau)}\right]\right)\phi(x)dx \\
&= \int_0^\infty \frac{1 - \mathbf{E}\left[\exp\{-(1-\psi(x))\Lambda(\Delta)\}\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]}\phi(x)dx \\
&= \frac{\mathbf{E}\left[\int_0^\infty \left(1 - \exp\left\{-(1-\psi(x))\Lambda(\Delta)\right\}\right)\phi(x)dx\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]} \\
&= \frac{\mathbf{E}\left[\Psi(\Lambda(\Delta))\right]}{1 - \mathbf{E}\left[\exp\{-\Lambda(\Delta)\}\right]} \ ,
\end{aligned} \tag{54}$$

where

$$\Psi(\theta) = \int_0^\infty \left(1 - \exp\left\{-\theta(1-\psi(x))\right\}\right)\phi(x)dx \tag{55}$$

($\theta \ge 0$).

Let $M_n = \max\{\xi_1, \cdots, \xi_n\}$ denote the maximum of $n$ IID copies of the random time $\xi$. Note that the tail probability of the maximum $M_n$ is given by:

$$\mathbf{P}\left(M_n > x\right) = 1 - G(x)^n \tag{56}$$

($x \ge 0$).

In the Gated-Parallel service regime the service time $S_n$ is the maximum of $n$ IID copies of the random time $\xi$. Hence $S_n \overset{d}{=} M_n$ which, in turn, implies that

$$\mathbf{E}\left[S_n\right] = \int_0^\infty \mathbf{P}\left(S_n > x\right)dx = \int_0^\infty \left(1 - G(x)^n\right)dx \ . \tag{57}$$

Thus, in the Gated-Parallel service regime $\psi(x) = G(x)$ and $\phi(x) = 1$ ($x \ge 0$). Using equation (55) we conclude that

$$\Psi_{\mathrm{GP}}(\theta) = \int_0^\infty \left(1 - \exp\left\{-\theta\overline{G}(x)\right\}\right)dx \tag{58}$$

$(\theta \geq 0)$.

In the Exhaustive-Parallel service regime the service time $S_n$ is the Busy Period of an $M/G/\infty$ queueing system, with arrival rate $\lambda$ and job-size governed by the distribution $G$, initiated with $n$ awaiting jobs.

Let $\beta(x)$ denote the mean length of an $M/G/\infty$ Busy Period initiated by jobs whose maximal size is $x$ ($x > 0$). In [3] it is proven that

$$\beta(x) = \int_0^x \exp\left\{\lambda \int_s^\infty \overline{G}(u)du\right\} ds \tag{59}$$

$(x > 0)$.

If the $M/G/\infty$ Busy Period is initiated with $n$ awaiting jobs then

$$\begin{aligned}
&\mathbf{E}\left[S_n\right] = \mathbf{E}\left[\beta\left(M_n\right)\right] \\
&= \textstyle\int_0^\infty \beta(x)\mathbf{P}\left(M_n \in dx\right) = \int_0^\infty \beta'(x)\mathbf{P}\left(M_n > x\right)dx \\
&= \textstyle\int_0^\infty \exp\left\{\lambda \int_x^\infty \overline{G}(u)du\right\}\left(1 - G(x)^n\right)dx
\end{aligned} \tag{60}$$

Thus, in the Exhaustive-Parallel service regime $\psi(x) = G(x)$ and $\phi(x) = \exp\left\{\lambda \int_x^\infty \overline{G}(u)du\right\}$ $(x \geq 0)$. Using equation (55) we conclude that

$$\Psi_{\mathrm{EP}}(\theta) = \int_0^\infty \left(1 - \exp\left\{-\theta\overline{G}(x)\right\}\right)\exp\left\{\lambda \int_x^\infty \overline{G}(u)du\right\}dx \tag{61}$$

$(\theta \geq 0)$.

**(iii) Lévy-Structured**

Assume that

$$\mathbf{E}\left[S_n\right] = an + \int_0^\infty (1 - \exp\{-nx\})\phi(x)dx \tag{62}$$

$(n = 1, 2, 3\cdots)$, where $a \geq 0$ and where $\phi(x) \geq 0$ $(x \geq 0)$.

The linear component $(an)$ of equation (62) was treated in part **(i)** of the proof; the integral component $(\int_0^\infty (1 - \exp\{-nx\})\phi(x)dx)$ of equation (62) was treated in part **(ii)** of the proof. Hence, using equation (51) and equation (55) (with $\psi(x) = \exp\{-x\}$), we obtain that:

$$\Psi_{\mathrm{LS}}(\theta) = a\theta + \int_0^\infty \left(1 - \exp\left\{-\theta(1 - \exp\{-x\})\right\}\right)\phi(x)dx \tag{63}$$

$(\theta \geq 0)$.

Using the change of variables $y = 1 - \exp\{-x\}$ in the integral component of equation (63) we conclude that:

$$\Psi_{\mathrm{LS}}(\theta) = a\theta + \int_0^1 \left(1 - \exp\{-\theta y\}\right)\frac{\phi(-\ln(1-y))}{1-y}dy\ . \tag{64}$$

## References


[1] Altman, E.; Yechiali, U. Analysis of impatience in queues with server vacations. Queueing Systems **2006**, *52*, 261-279.

[2] Bertoin, J. *Lévy Processes*; Cambridge University Press: Cambridge, 1996.

[3] Eliazar, I.; Klafter, J. On the active periods of nonlinear shot noise. Physica A **2006**, *363*, 237-259.

[4] Kingman, J.F.C. *Poisson Processes*; Oxford University Press: Oxford, 1993.

[5] Levy, Y.; Yechiali, U. Utilization of idle time in an $M/G/1$ queueing system. Management Science **1975**, *22*, 202-211.

[6] Levy, Y.; Yechiali, U. An $M/M/s$ queue with servers' vacations. Canadian Journal of Operational Research **1976**, *14*, 153-163.

[7] Takagi, H. *Queueing Analysis, vol. 1: Vacation and Priority Systems*; North Holland: Amsterdam, 1991.

[8] Takacs, L. *Introduction to the Theory of Queues*; Oxford University Press: Oxford, 1962.

[9] Yechiali, U. On the $M^X/G/1$ queue with a waiting server and vacations. Sankhya **2004**, *66*, part 2, 1-17.