stochastic processes and their applications

ELSEVIER Stochastic Processes and their Applications 65 (1996) 139–146

Conservation laws and reflection mappings with an application to multiclass mean value analysis for stochastic fluid queues

Takis Konstantopoulos [a,1], Michael Zazanis [b,*,2], Gustavo De Veciana [a,3]

[a] *ECE Department, University of Texas, Austin, TX 78712, USA*

[b] *IEOR Department, University of Massachusetts, Amherst, MA 01003, USA*

Received April 1995; revised July 1996

Abstract

In this paper we derive an alternative representation for the reflection of a continuous, bounded variation process. Under stationarity assumptions we prove a continuous counterpart of Little's law of classical queueing theory. These results, together with formulas from Palm calculus, are used to explain the method for the derivation of the mean value of a buffer fed by a special type stochastic fluid arrival process.

Keywords: Fluid queues; Reflection mapping; Conservation laws; Palm probabilities

1. Introduction

Delay systems with stochastic fluid input processes are currently being used as models of components of high speed communication networks. Motivated by such applications, we derive a version of Little's law valid for reflected processes with stationary increments and continuous, bounded variation paths. In doing so, we find a new representation of the Skorokhod reflection mapping. Using this representation together with a generalized Campbell's formula from Palm calculus we interpret Little's law as a conservation law for a multiclass stochastic fluid queue. As an application, we consider the system of Dupuis and Hajek (1994) and give a rigorous proof of the formula for the mean buffer content. The latter system models the shared buffer in an asynchronous transfer mode multiplexer.

The alternative representation for the reflected process is derived in Section 2 and is a result of independent interest. We emphasize that the representation depends on the

* Corresponding author. Fax: 413 545 0724; e-mail: zazanis@ecs.umass.edu.

[1] This work was supported in part by NSF Research Initiation Award NCR-9211343, by NSF Faculty Career Development Award NCR-9502582, and by Grant ARP-224 of the Texas Higher Education Coordinating Board.

[2] Supported in part by NSF Grant EEC-9527519.

[3] Supported in part by NSF Research Initiation Award NCR-9409722.

PII S0304-4149(96)00103-2

continuity of the free process. Little's law is obtained from this result in Section 3. Finally, computations and formulas for the derivation of the mean buffer content are reported in Section 4.

We note that previous work on fluid queues and Little's laws can be found in Rolski and Stidham (1983), Glynn and Whitt (1989), and Miyazawa (1994). The connection with Palm Calculus is pointed out by Miyazawa (1994). In particular, Miyazawa introduces the concept of detailed Palm probability that helps to explain the notion of conditioning with respect to a "typical point" of a general stationary measure.

2. An integral representation for a reflected process

Consider a buffer fed by an arrival process $\{A(t),\ t \geqslant 0\}$ and served at rate $c > 0$ in a work-conserving fashion. In communications applications we think of $A(t)$ as the total number of "bits" arriving on $[0,t]$. The arrival process is assumed to be right continuous. The buffer content (or load) $Q(t)$ is found by *reflecting* the *free process*

$$X(t) := Q(0-) + A(t) - ct, \tag{1}$$

where $Q(0-)$ is the load just before time 0. The notion of reflection is classical: it requires that Q be obtained from X by adding to it a right continuous, increasing process L, with $L(0) = 0$, such that the points of increase of L be a subset of the zeros of Q and $Q(t) \geqslant 0$ for all t.

Given any right continuous X with left limits process, such as, for example, (1), there is a unique L (and hence a unique $Q = X + L$) satisfying the above requirements, namely,

$$L(t) = - \inf_{0 \leqslant s \leqslant t} X(s) \wedge 0.$$

The mapping $X \mapsto Q$ is referred to as the *Skorokhod reflection mapping*, a basic property of which is its *causality*, expressed by the fact that, for any $s \geqslant 0$, the reflection of $\{Q(s) + X(s+t) - X(s),\ t \geqslant 0\}$ equals $\{Q(s+t),\ t \geqslant 0\}$. Causality can be verified by the definitions above and can be explicitly expressed as

$$Q(t) = \sup_{s \leqslant u \leqslant t} X(u,t] \vee \{Q(s) + X(s,t]\}. \tag{2}$$

Here and in what follows we shall often use the convention $X(u,t] := X(t) - X(u)$.

Even though (2) can be taken as a definition of the reflected process, there is another representation (we call it *integral representation*), valid if X is continuous and of bounded variation, which is useful in connection with Palm calculus for stochastic fluids. The representation is given by Theorem 1 below, whose proof requires the following:

Lemma 1. *Let Q be the reflection of a process $X : \mathbb{R}_+ \to \mathbb{R}$ which is right continuous with left limits. Suppose there is $\lambda \geqslant 0$ such that $X(t) + \lambda t$ is an increasing function.*

Then, for any $t_1 < t_2$, the following sets are equal:

$$F := \{s \in [t_1, t_2] : Q(s) > \lambda(t_2 - s)\},$$

$$G := \{s \in [t_1, t_2] : X(t_2) - X(s) + \lambda(t_2 - s) < Q(t_2)\}.$$

Proof. Set $t = t_2$ in (2). Observe that if $s \in F$ then $Q(t_2) \geqslant Q(s) + X(s, t_2] > \lambda(t_2 - s) + X(s, t_2]$, that is, $s \in G$. Next assume that $s \in G$. Then $X(s, t_2] + \lambda(t_2 - s)$ is strictly smaller than $Q(s) + X(s, t_2]$, in which case $\lambda(t_2 - s) < X(s, t_2]$, i.e., $s \in F$, or $X(s, t_2] + \lambda(t_2 - s)$ is strictly smaller than $X(u, t_2]$ for some $u \in [s, t_2]$. But the latter case is impossible: Indeed, it is equivalent to $[X(s, u] + \lambda(u - s)] + \lambda(t_2 - u) < 0$; however, $X(s, u] + \lambda(u - s) \geqslant 0$ and $\lambda(t_2 - u) \geqslant 0$. □

Theorem 1 (Integral relation for a reflected process). *Suppose that $X : \mathbb{R}_+ \to \mathbb{R}$ is continuous with bounded variation, and $X(0) = 0$. Let Q be the reflection of X. If $X(t) + \lambda t$ is increasing, for some $\lambda \geqslant 0$, the following relation holds*:

$$Q(t) = \int_0^\infty 1[s \leqslant t < s + \lambda^{-1} Q(s)][X(\mathrm{d}s) + \lambda\,\mathrm{d}s]. \tag{3}$$

Proof. We will use the change of variables formula

$$\int_0^t g(Y(u)) Y(\mathrm{d}u) = \int_{Y(0)}^{Y(t)} g(y)\,\mathrm{d}y, \tag{4}$$

valid for any integrable function g, provided that Y is increasing and continuous (whence the inevitability of our continuity assumption). See, for instance, Brémaud (1981, p. 301). To arrive at the form (4), observe first that the right-hand side of (3) equals

$$\int_0^\infty 1[s \leqslant t : X(t) - X(s) + \lambda(t - s) < Q(t)][X(\mathrm{d}s) + \lambda\,\mathrm{d}s]. \tag{5}$$

This follows from Lemma 1. Define then $Y(u) := X(t) - X(t - u) + \lambda u$, $u \in [0, t]$, observe that it is increasing, and apply the change of variables formula (4) to (5):

$$\int_0^t 1[Y(u) < Q(t)] Y(\mathrm{d}u) = \int_0^{Y(t)} 1[y < Q(t)]\,\mathrm{d}y = Y(t) \wedge Q(t).$$

To conclude the proof, we show that $Q(t) \leqslant Y(t)$. For $0 \leqslant u \leqslant t$ we have $0 \leqslant X(u) + \lambda u \leqslant X(u) + \lambda t$, and so $\sup_{0 \leqslant u \leqslant t}[X(t) - X(u)] \leqslant X(t) + \lambda t = Y(t)$. But $Q(0) = 0$, by assumption, hence, from (2), $Q(t) = \sup_{0 \leqslant u \leqslant t}[X(t) - X(u)] \vee X(t) \leqslant Y(t)$. □

Special case. Consider now $X(t) = A(t) - ct$, as in the beginning of the section, and let $\lambda \geqslant c$. Assume that $A(t)$ is continuous, and that $A(0) = Q(0) = 0$. Then (3) holds. In particular, for $\lambda = c$, we have

$$Q(t) = \int_0^\infty 1[s \leqslant t < s + c^{-1} Q(s)] A(\mathrm{d}s). \tag{6}$$

The interpretation of (6) should be clear: it expresses the load at time t as the sum of all bits that have arrived before t but are still in the buffer at time t. Note that we have assumed that A is a continuous function. If A is only right continuous then (4), and hence Theorem 1, may fail to hold. The following case exhibits this problem.

Counterexample. Suppose $A(t) = 2t1[0 \leqslant t < 1] + 1[t \geqslant 1]$ and $c = 1$. One can check that the reflection of $X(t) = A(t) - ct$ is given by $Q(t) = t1[0 \leqslant t < 1] + (3-t)1[1 \leqslant t < 3]$. On the other hand, it is seen that this function does not satisfy (3). Indeed, *the* solution to the integral equation (3) is given by $Q(t) = t1[0 \leqslant t < 1] + (3-t)1[1 \leqslant t < 2] + 1[2 \leqslant t < 3]$. Thus the integral equation does *not* represent the evolution of the fluid queueing model unless the arrival process is continuous.

3. Stochastic fluid queues and conservation laws

By stochastic fluid queue we mean a reflected stochastic process Q with the free process X having stationary ergodic integrable increments. Formally, as in Baccelli and Brémaud (1994), let $(\Omega, \mathscr{F}, P)$ be a probability space equipped with a measurable flow $\{\theta_t,\ t \in \mathbb{R}\}$ which is invariant under P. Assume that the pair $(P, \{\theta_t\})$ is ergodic. Given a process $\{X(t),\ t \in \mathbb{R}\}$ having right continuous paths with left limits, increments compatible with the flow, i.e., $X(t) - X(s) = [X(t-u) - X(s-u)] \circ \theta_u$, and such that $E[X(t) - X(s)] = \kappa(t-s) < 0$, we take (2) as the definition of the reflected process Q. Under the aforementioned conditions, there is only one stationary ergodic strong solution of (2):

Theorem 2 (Existence and uniqueness of stationary load process). *Under the above conditions, the fixed point equation* (2) *possesses a unique right continuous solution* $\{Q(t),\ t \in \mathbb{R}\}$, *which is compatible with the flow, i.e.,* $Q(t) \circ \theta_s = Q(t+s)$, *for all* $t, s \in \mathbb{R}$.

Due to space limitations we omit the proof of this theorem and refer the reader to [7]. A consequence of the proof is that there are infinitely many times t at which $Q(t)$ hits zero.

Assuming now that A is a *continuous stationary ergodic random measure* with $0 < \alpha = EA(0,1] < c < \infty$ we define $X(s,t] = A(s,t] - c(t-s)$ and consider Q to be the unique stationary solution of (2). Since $Q(T) = 0$ for some T, we can represent Q using our alternative integral representation formula (6):

$$Q(t) = \int_T^\infty 1[s \leqslant t < s + c^{-1}Q(s)]A(\mathrm{d}s) = \int_{-\infty}^\infty 1[s \leqslant t < s + c^{-1}Q(s)]A(\mathrm{d}s). \quad (7)$$

The range of integration was taken from $-\infty$ to $+\infty$ due to the fact that the integrand $1[s \leqslant t < s + c^{-1}Q(s)]$ is zero for all $s < T$.

We pass on to the derivation of Little's law and the conservation law. We take the expectation values on both sides of (7) and make use of a generalized Campbell's formula. To set up the framework, recall that the *Campbell measure* for the random measure A is defined as $E[A(S)1_F]$, where S is a Borel subset of $\mathbb{R}$ and $F \in \mathscr{F}$. The

Palm transformation P_A of P with respect to A is defined as the value of the Radon–Nikodým derivative $E[A(\mathrm{d}t)1_F]/E[A(\mathrm{d}t)]$ at $t = 0$. Denoting by E_A the expectation with respect to the probability measure P_A, we have, almost immediately from the definition, the *generalized Campbell's formula*

$$\alpha E_A \int_{\mathbb{R}} Z(s)\,\mathrm{d}s = E \int_{\mathbb{R}} Z(s) \circ \theta_s A(\mathrm{d}s), \tag{8}$$

true for any bounded measurable process Z. The reader is referred to the monograph of Daley and Vere-Jones (1988), and also to the recent report of Schmidt and Serfozo (1994) for a nice compact exposition of the concepts.

Theorem 3 (Little's law). *Suppose that A is a continuous stationary ergodic random measure. Under the assumptions of Theorem 2, the stationary solution Q of the reflection equation (2) satisfies*

$$E[Q(0)] = \frac{\alpha}{c} E_A[Q(0)]. \tag{9}$$

Proof. Let $Z(s) := 1[s \leqslant 0 < s + c^{-1}Q(0)]$ in (8) and use (7). □

The terminology "Little's law" should be clear by analogy to the classical queueing result. One interprets (9) as: "the mean buffer load $E[Q(0)]$ equals the bit arrival rate α times the mean delay $E_A[Q(0)/c]$ experienced by the typical bit". The relation is "clear for physical reasons"; however, its mathematical proof requires the aforementioned setup. Considering next A as the superposition of M jointly stationary random measures A_i, $i = 1, \ldots, M$, and using the fact that P_A is a mixture of P_{A_i}, namely, $P_A = \sum_{i=1}^{M} (\alpha_i/\alpha) P_{A_i}$, we write Little's law (9) as

$$EQ(0) = \sum_{i=1}^{M} \frac{\alpha_i}{c} E_{A_i} Q(0), \tag{10}$$

and refer to this as *conservation law* in harmony with namesake relations for traditional multiclass queues: it does not depend on the service discipline employed for individual streams.

4. Mean value analysis for bursty fluid sources

We now apply the Palm-calculus-based methods to the problem considered by Dupuis and Hajek (1994). It concerns a multi-class fluid queue served at rate c and fed by M independent ON/OFF sources. The ith source initiates *sessions* at times $\cdots < T^i_{-1} < T^i_0 \leqslant 0 < T^i_1 < \cdots$ at rate $1/m_i$. The time interval between two successive sessions n and $n+1$ consists of an initial *active period* of duration L^i_n and a remaining *silent period* of duration S^i_n. The number of bits transmitted by the ith source on the first t time units of its active period is denoted by $F^i_n(t)$. The random functions $\{F^i_n,\ n \in \mathbb{Z}\}$ are assumed to be i.i.d. and independent of $\{T^i_n,\ n \in \mathbb{Z}\}$ and $\{S^i_n,\ n \in \mathbb{Z}\}$. Naturally, we take $F^i_n(t) = 0$ for $t < 0$, and $F^i_n(t) = F^i_n(L^i_n)$ for $t > L^i_n$, and assume that F^i_n is continuous.

Furthermore, it is assumed that a *burstiness assumption* is satisfied, namely, $F_n^i(t) \geqslant ct$ for $0 \leqslant t \leqslant L_n^i$. The number of bits transmitted on the set $B \subseteq \mathbb{R}$ by the ith source is thus given by $A_i(B) := \sum_n \int_{B-T_n^i} F_n^i(\mathrm{d}t)$. The total arrival process is $A = \sum_i A_i$. The service policy is *first bit-first served.*

Let α_i be the mean bit rate of A_i, and $\alpha = \sum_i \alpha_i$. It is useful to introduce the indicator process $\zeta_i(t)$ taking value 0 if source i is on its silent period and 1 otherwise. Let $p_i = P(\zeta_i(t) = 1)$. Denote by P the underlying probability measure, constructed as outlined above, under which the arrival process is stationary, let P_{N_i} be the Palm transformation of P with respect to the point process N_i with points $\{T_n^i,\ n \in \mathbb{Z}\}$, and let P_{A_i} be the Palm transformation of P with respect to the random measure A_i. Under the assumption $\alpha < c$ there is a unique stationary solution Q (the load process) to the fixed point equation (2) with $X(s,t] := A(s,t] - c(t-s)$, constructed on the same probability space.

To compute $E[Q(0)]$ we first condition on $\zeta_i(0) = 1$:

$$E[Q(0)] = E[Q(0)|\zeta_i(0) = 1]p_i + E[Q(0)|\zeta_i(0) = 0](1 - p_i). \tag{11}$$

Due to the burstiness assumption, on the event $\{\zeta_i(0) = 1\}$, we can write $Q(0) = Q(T_0^i) + A(T_0^i, 0] - c(0 - T_0^i)$. Taking expectations, we obtain

$$\begin{aligned} E[Q(0)|\zeta_i(0) = 1] &= E[Q(T_0^i)|\zeta_i(0) = 1] + E[A(T_0^i,0] + cT_0^i|\zeta_i(0) = 1] \\ &= E_{N_i}[Q(0)] + C_1^i. \end{aligned} \tag{12}$$

The first term follows from $E[Q(T_0^i)1(\zeta_i(0) = 1)] = E[Q(T_0^i)]p_i$, by independence, and $E[Q(T_0^i)] = E_{N_i}[Q(0)]$, by the Palm inversion formula (19) and independence. The second term

$$C_1^i := E[A(T_0^i, 0] + cT_0^i|\zeta_i(0) = 1] \tag{13}$$

depends only on the sources' statistics and is computed in Appendix A.

Assume now that S_n^i is *exponentially distributed.* Then (see Appendix B),

$$E[Q(0)|\zeta_i(0) = 0] = E_{N_i}[Q(0)]. \tag{14}$$

Substituting (12) and (14) into (11) we have

$$E[Q(0)] = E_{N_i}[Q(0)] + p_i C_1^i. \tag{15}$$

The final step is to relate P_{N_i} to P_{A_i} by an *exchange formula.* In Appendix A it is shown that

$$E_{N_i}[Q(0)] = E_{A_i}[Q(0)] - C_2^i, \tag{16}$$

where C_2^i is a positive expression, computable from the sources' characteristics. Substitute now (16) into (15) and use the conservation law (10) to obtain

$$E[Q(0)] = \frac{\sum_i \frac{\alpha_i}{c}(C_2^i - p_i C_1^i)}{1 - \alpha/c} =: \frac{C_3}{1 - \alpha/c}. \tag{17}$$

The values of C_1^i, C_2^i, and $C_3 := \sum_i (\alpha_i/c)(C_2^i - p_i C_1^i)$ are computed in Appendix A, as formulas (A.3), (A.4), and (A.5), respectively. Substituting them in (17) we obtain the explicit formula

$$E[Q(0)] = \frac{1}{2(c-\alpha)} \sum_i m_i^{-1} E_{N_i}[F_i^0(L_0^i) - \alpha_i L_0^i]^2 - \sum_i m_i^{-1} E_{N_i} \int_0^{L_0^i} t[F_0^i(\mathrm{d}t) - \alpha_i \,\mathrm{d}t]. \tag{18}$$

Appendix A. Computations involving the statistics of the sources

Everything depends on a proper use of the exchange formulas

$$E[Z(0)] = \frac{E_{N_i} \int_{T_0^i}^{T_1^i} Z(t)\,\mathrm{d}t}{E_{N_i}[T_1^i - T_0^i]}, \tag{A.1}$$

$$E_{A_i}[Z(0)] = \frac{E_{N_i} \int_{(T_0^i, T_1^i]} Z(t) A_i(\mathrm{d}t)}{E_{N_i} A_i (T_1^i - T_0^i]}. \tag{A.2}$$

They both follow from the integration formula (8), and hold for any integrable process Z, jointly stationary with the data of the problem.

Regarding C_1^i, defined in (13), we use the first exchange formula (A.1) to obtain

$$C_1^i = (p_i m_i)^{-1} E_{N_i} \int_0^{L_0^i} [A(0,t] - ct]\,\mathrm{d}t. \tag{A.3}$$

Regarding C_2^i, appearing in (16), we relate P_{A_i} to P_{N_i} using the second exchange formula (A.2) to obtain

$$\begin{aligned} E_{A_i}[Q(0)] &= (\alpha_i m_i)^{-1} E_{N_i} \int_0^{L_0^i} Q(t) A_i(\mathrm{d}t) \\ &= (\alpha_i m_i)^{-1} E_{N_i} \int_0^{L_0^i} [Q(0) + A(0,t] - ct] A_i(\mathrm{d}t) \\ &= E_{N_i}[Q(0)] + (\alpha_i m_i)^{-1} E_{N_i} \int_0^{L_0^i} [A(0,t] - ct] A_i(\mathrm{d}t), \end{aligned}$$

and hence,

$$C_2^i = (\alpha_i m_i)^{-1} E_{N_i} \int_0^{L_0^i} [A(0,t] - ct] A_i(\mathrm{d}t). \tag{A.4}$$

Finally, the constant $C_3 = \sum_i (\alpha_i/c)(C_2^i - p_i C_1^i)$ follows by a straightforward substitution and manipulation of the integrals involved, as in [7]. We thus obtain

$$C_3 = \sum_i \frac{\alpha_i}{c}(C_2^i - p_i C_1^i) = \sum_i (cm_i)^{-1} E_{N_i} \int_0^{L_0^i} [A(0,t] - ct][A_i(\mathrm{d}t) - \alpha_i \,\mathrm{d}t]. \tag{A.5}$$

Appendix B. Papangelou's theorem and exponential silent times

A rigorous derivation of relation (14) can be provided by means of Papangelou's theorem (cf. Baccelli and Brémaud, 1994, p. 54, Section 3.2). Let $\mathscr{F}_t$ be a history for the point process N_i. In our case we choose $\mathscr{F}_t$ to be the right-continuous σ-field generated by the history of the arrival processes up to time t. According to Papangelou's theorem, P_{N_i} is absolutely continuous with respect to P on $\mathscr{F}_{t-}$ if and only if the $\mathscr{F}_t$-stochastic intensity λ_t^i of N_i exists. But it is easy to verify that

$$\lambda_t^i = ((1 - p_i)m_i)^{-1} 1(\zeta_i(0) = 0).$$

Indeed, silent times of source i are exponential random variables with mean $(1 - p_i)m_i$. We also have $E\lambda_t^i = 1/m_i$. The Radon–Nikodým derivative of P_{N_i} with respect to P on $\mathscr{F}_{t-}$ is given by

$$\left.\frac{\mathrm{d}P_{N_i}}{\mathrm{d}P}\right|_{\mathscr{F}_{t-}} = \frac{\lambda_t^i}{E\lambda_t^i} = \frac{1}{1 - p_i} 1(\zeta_i(0) = 0).$$

Note that, by the continuity of the process Q, the random variable $Q(0)$ is measurable with respect to $\mathscr{F}_{0-}$ and thus

$$E_{N_i}[Q(0)] = E\left[Q(0) \left.\frac{\mathrm{d}P_{N_i}}{\mathrm{d}P}\right|_{\mathscr{F}_{0-}}\right] = \frac{E[Q(0)1(\zeta_i(0) = 0)]}{1 - p_i} = E[Q(0)|\zeta_i(0) = 0].$$

This explains the relation (14).

References

P. Brémaud, Point Processes, and Queues. Martingale Dynamics (Springer, New York, 1981).

F. Baccelli and P. Brémaud, Elements of Queueing Theory. Palm-Martingale Calculus and Stochastic Recurrences (Springer, Berlin, 1994).

D.J. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes (Springer, New York, 1988).

H. Dupuis and B. Hajek, A simple formula for mean multiplexing delay for independent regenerative sources, Queueing Systems 16 (1994).

P. Glynn and W. Whitt, Extensions of the queueing relations $L = \lambda W$, Oper. Res. 37 (1989) 634–644.

T. Konstantopoulos and V. Anantharam, Optimal flow control schemes that regulate the burstiness of traffic, IEEE Trans. Networking 3 (1993) 423–432.

T. Konstantopoulos, M. Zezanis and G. de Veciana, Conservation laws and reflection mappings with an application to multiclass mean value analysis for stochastic fluid queues, Tech. Report SCC-95-02, Univ. of Texas at Austin (1995).

M. Miyazawa, Palm calculus for a process with a stationary random measure and its applications to fluid queues, Queueing Systems 17 (1994) 183–211.

T. Rolski and S. Stidham, Continuous versions of the queueing formulas $L = \lambda W$ and $H = \lambda G$, Oper. Res. Lett. 2 (1983) 211–215.

V. Schmidt and R. Serfozo, Campbell's formula and applications to queueing, in: J.H. Dshalalow, ed., Advances in Queueing; Theory, Methods and Open Problems (CRC Press, Boca Raton, 1994).