**IEOR 6711: Stochastic Models I**

**Professor Whitt, Thursday, December 6, 2012**

# More CTMC's: Loss Models

## 1 The Classical Erlang $M/M/N/0$ Loss Model

### 1.1 A special birth-and-death process

This model is a finite-state BD process. The birth rates are $\lambda_k \equiv \lambda$ and the death rates are $\mu_k = k\mu$ for an overall arrival rate of $\lambda$ and an individual service rate of $\mu$. The detailed balance equations give

$$\alpha_j \lambda = \alpha_{j+1}(j+1)\mu \tag{1}$$

and

$$\alpha_j = C\frac{\nu^j}{j!}, \tag{2}$$

where $\nu \equiv \lambda/\mu$ is the offered load and the normalization constant $C$ is

$$C \equiv G(N) = \sum_{j=0}^{N} \frac{\nu^j}{j!} \tag{3}$$

### 1.2 The Generating Function of the Normalization Constant

To understand the application of numerical inversion to calculate normalization constants (partition functions) in more general models, it is good to see how to do that calculation here. The idea is to change the order of summation. In this elementary case, for $z$ a complex number with $|z| < 1$, we get the simple expression

$$\begin{aligned}
\hat{G}(z) &\equiv \sum_{N=0}^{\infty} G(N) z^N = \sum_{N=0}^{\infty} \left( \sum_{j=0}^{N} \frac{\nu^j}{j!} \right) z^N = \sum_{j=0}^{\infty} \sum_{N=j}^{\infty} \frac{\nu^j}{j!} z^N \\
&= \sum_{j=0}^{\infty} \frac{(\nu z)^j}{j!} \sum_{N=j}^{\infty} z^{N-j} = \sum_{j=0}^{\infty} \frac{(\nu z)^j}{j!} \sum_{N=0}^{\infty} z^N \\
&= \left( \frac{1}{1-z} \right) \sum_{j=0}^{\infty} \frac{(\nu z)^j}{j!} = \frac{e^{\nu z}}{1-z}.
\end{aligned} \tag{4}$$

## 2 Insensitivity in the Erlang $M/GI/N/0$ Loss Model

This model has the insensitivity property: The steady-state distribution does not depend on the general service-time distribution beyond its mean. If the mean service time is $ES = 1/\mu$, then the steady state distribution is the same as in the case of exponential service times.

The three concepts reversibility, insensitivity and product form are intimately related. There is by now a huge literature, but some issues remain complicated.

We will follow the discussion in §5.7.2 of Ross. We apply the generalization of Kelly's lemma in Theorem 6.8 of the CTMC notes. See Ethier and Kurtz [6], especially §4.9, for supporting

theory on Markov processes with general state space, as used by Ross without justification. See Burman [2]. See $W^2$ [10] for continuity argument to extend results.

The final story is simple: The process describing the system state can be made a continuous-time Markov process (but with uncountable state space) if we include the ages of the customers in service. A state of this Markov process can be written ias

$$(n; x) \equiv (n; (x_1, x_2, \ldots, x_n)), \tag{5}$$

where, without loss of generality, we take the ages to be expressed in increasing order, so that $x_1 < x_2 < \cdots x_n$.

The steady-state number in system is the same as for the $M/M/N/0$ model. Given that there are $n$ customers in service, the ages of these service times in process (and the residual lifetimes) in steady state are independently distributed, each with cdf $G_e$, the equilibrium excess distribution associated with the service-time cdf $G$, which has density

$$g_e(x) \equiv \frac{G^c(x)}{ES}. \tag{6}$$

Taking account of the order requirement on the ages, which introduces the factor $n!$, we obtain the steady state distribution of the Markov process.

**Theorem 2.1** (*steady-state of $M/GI/N/0$ Markov process*) *For the $M/GI/N/0$ model, if the service-time cdf has a pdf $g$, then the steady-state distribution is*

$$\alpha(n; x) \equiv \alpha(n; (x_1, x_2, \ldots, x_n)) = \alpha_{M/M/N/0}(n) \times n! \prod_{i=1}^{n} g_{ie}(x_i). \tag{7}$$

**Proof.** We guess that the claim is correct and identify the reverse-time Markov process and apply the generalization of Kelly's lemma. The reverse-time Markov process is also an $M/GI/N/0$ Markov process with Poisson arrival process, but in reverse time the forward-time ages become residual lifetimes or excess variables. There are two possible events: (i) an arrival or (ii) a service completion. An arrival takes the state $(n; x)$ into $(n+1; (0, x))$ for $x$ of dimension $n$. A service completion takes the state $(n; x)$ into $(n-1; e_i(x))$, where $x$ is of dimension $n$ and $e_i(x)$ is the $n-1$-dimension vector with the $i^{\text{th}}$ component of $x$ removed. Let the hazard rate of the service-time cdf be

$$h(x) \equiv \frac{g(x)}{G^c(x)}, \quad x > 0. \tag{8}$$

The two detailed flow equations are

$$\begin{aligned} \alpha(n; x)\lambda &= \alpha(n+1; (0, x)) & (9) \\ \alpha(n; x)h(x_i) &= \alpha(n-1; e_i(x))\lambda g(x_i) & (10) \end{aligned}$$

These can be seen to hold, as we now show. First, for (9),

$$\alpha(n+1; (0, x)) = \left(\frac{\lambda}{(n+1)\mu}\right)(n+1)G_e(0) \times \alpha(n; x), \tag{11}$$

but

$$\left(\frac{\lambda}{(n+1)\mu}\right)(n+1)G_e(0) = \lambda, \tag{12}$$

so that (9) holds. We no turn to (10).

The second formula in (10) becomes

$$\alpha(n;x)h(x_i) = \alpha(n;x)\frac{g(x_i)}{G^c(x_i)} \tag{13}$$

while

$$\frac{\alpha(n;x)}{\alpha(n-1;e_i(x))} = \left(\frac{\lambda}{n\mu}\right) n g_e(x_i) = \frac{\lambda g_e(x_i)}{\mu} = \lambda G^c(x_i). \tag{14}$$

Applying (13) and (14), we see that indeed (10) holds. ∎

# 3 Stochastic Loss Networks

This is a generalization of the Erlang loss model model with multiple classes and multiple facilities. Customers require and use subsets of resources from each facility and hold them all for the duration of their time in the system. See Kelly [8] and $W^2$ [11].

For numerical inversion of generating functions of the normalization constant, see Choudhury, Leung and $W^2$ [4]. For applications, see [3].

## 3.1 The Basic Model

Consider a **network** with $J$ **links** indexed by $j$, $1 \le j \le J$. Suppose that link $j$ contains $C_j$ **circuits**. Let there by $R$ customer **classes** (often called routes, **classes** $\equiv$ **routes**), indexed by $r$, $1 \le r \le R$. Customers of class $r$ **arrive** according to a Poisson process with rate $\lambda_r$. If the customer can get all its required resources in the network, the class-$r$ customer stays in the network for an exponential length of time (**service or holding time** with mean $1/\mu_r$. When a customer leaves, it releases all its resources. A class-$r$ customer requires $A_{j,r}$ circuits (units of resource) on link $j$, $1 \le j \le J$. Thus customers may use more than one circuit on a link and require circuits on more than one link. If a customer can not be given all its required resources immediately upon arrival, then this customer is blocked without affecting the future arrivals.

## 3.2 The Steady-State Behavior

Let $X(t) = (X_1(t), \ldots, X_R(t))$ be the number of customers from each class in the system at time $t$. The $R$-dimensional stochastic process is a CTMC. It is easily seen to be a **reversible CTMC**. The local balance equation is

$$\alpha_n \lambda_r = \alpha_{n+e_r} \mu_r (n_r + 1), \tag{15}$$

where $n \equiv (n_1, n_2, \ldots, n_R)$ and $e_i$ is an $R$-dimensional vector with a 1 in the $i^{\text{th}}$ place and 0 elsewhere. The system has a steady-state distribution that is a natural generalization of the Erlang loss model, in particular,

$$\alpha_n = G(C)^{-1} f(n), \quad n \in S(C), \tag{16}$$

where

$$f(n) = f(n_1, n_2, \ldots, n_R) \equiv \prod_{r=1}^{R} \frac{\nu_r^{n_r}}{n_r!} \tag{17}$$

$\nu \equiv \lambda_r/\mu_r$,

$$S(C) \equiv S(C_1, C_2, \ldots, C_J) \equiv \{n \in \mathbb{R}_+^R : An \le C\} \tag{18}$$

and

$$G(C) \equiv G(C_1, C_2, \ldots, C_J) \equiv \sum_{n \in S(C)} f(n). \tag{19}$$

The set $S(C)$ contains the possible states of the process, while $G(C)$ in (19) is the normalization constant or partition function.

The blocking probability of an arbitrary class-$r$ request is

$$B_r = 1 - \frac{G(C - Ae_j)}{G(C)}, \tag{20}$$

where $e_j$ is a vector that is 1 in the $j^{\text{th}}$ place and 0 otherwise.

### 3.3 The Generating Function of the Partition Function

The generating function of the normalization constant is a generalization of (4), namely,

$$\begin{aligned} \hat{G}(z) &\equiv \hat{G}(z_1, \ldots, z_J) \equiv \sum_{C} G(C) \prod_{r=1}^{R} z_i^{C_i} \\ &= \sum_{C_1=0}^{\infty} \cdots \sum_{C_J=0}^{\infty} G(C_1, \ldots, C_J) z_1^{C_1} \times \cdots \times z_R^{C_R} \\ &= \frac{e^{\left(\sum_{r=1}^{R} \nu_r \prod_{j=1}^{J} z_j^{A_{r,j}}\right)}}{\prod_{j=1}^{J}(1 - z_j)}. \end{aligned} \tag{21}$$

(There is a typo in (2.12) on p. 1115 of [4]; the second sum in the numerator there should be a product.)

### 3.4 Interesting Asymptotics and Approximations

See [8] and references there.

- Erlang fixed point approximation
- heavy-traffic limits as the link capacities $C_j$ grow
- limits as the number of nodes grows (in a controlled symmetric way)
- phase transitions

## References


[1] Abate, J., G. L. Choudhury, G. L., W. Whitt. 1999. An introduction to numerical transform inversion and its application to probability models. in *Computational Probability*, W. Grassman (ed.), Kluwer, Boston, 257–323.

[2] Burman, D. Y. 1981. Insensitivity in Queueing Systems. Advances in Appl. Prob. 13, 846–859.

[3] Choudhury, G. L., K. K. Leung and W. Whitt. 1995. Efficiently Providing Multiple Grades of Service with Protection Against Overloads in Shared Resources. *AT&T Technical Journal* 74, 50–63.

[4] Choudhury, G. L., K. K. Leung and W. Whitt. 1995b. An Algorithm to Compute Blocking Probabilities in Multi-Rate Multi-Class Multi-Resource Loss Models. Advances in Appl. Prob. 27, 1104–1143.

[5] Choudhury, G. L., D. L. Lucantomi and W. Whitt. 1994. Multi-Dimensional Transform Inversion with Applications to the Transient M/G/1 Queue. *Annals of Applied Probability* 4, 719–740.

[6] Ethier, S. N. and T. G. Kurtz. 1986. *Markov Processes: Characterization and Convergence*, Wiley.

[7] Kelly, F. P. 1979. *Reversibility and Stochastic Networks*, Wiley.

[8] Kelly, F. P. 1991. Loss networks. *Ann. Appl. Prob.* 1, 319–378.

[9] Serfozo, R. F. 1999. *Introduction to Stochastic Networks*, Springer.

[10] Whitt, W. 1980. Continuity of Generalized Semi-Markov Processes.

[11] Whitt, W. 1985. Blocking When Service is Required from Several Facilities Simultaneously. *AT&T Technical Journal* 64, 1807–1856.

[12] Whittle, P. 1986. *Systems in Stochastic Equilbirium*, Wiley.