Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

ERMR: a generalised equivalent random method for overflow systems with repacking

S. C. Borst, R. J. Boucherie, O. J. Boxma

Probability, Networks and Algorithms (PNA)

PNA-R9817 December 1998

Report PNA-R9817
ISSN 1386-3711

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

ERMR: A Generalised Equivalent Random Method for Overflow Systems with Repacking[1]

Sem Borst[a,b]

Richard J. Boucherie[a,c]

Onno J. Boxma[a,b]

ABSTRACT

This note develops two-moment approximations for blocking probabilities in overflow systems with repacking. The approximations are based on fixed point schemes, iteratively relating the arrival rate at the primary channels to the mean number of calls at the overflow channels. The analysis focuses on layered cellular mobile communications networks, but can also be applied to multiple retrial queues sharing a common (finite) orbit. Numerical results show that our approximation is fairly accurate for systems in which the mean repacking rate does not exceed the call arrival rate.

1991 Mathematics Subject Classification: 60K25, 90B22.

Keywords and Phrases: Layered cellular mobile communications network, Overflow system, Repacking, Blocking probabilities, Two-moment approximation, Retrial queue.

Note: This work is carried out under the project MOBILECOM in PNA 2.1.

1. Introduction

The spectacular growth of wireless communications has forced service providers to re-use frequencies more efficiently. Layered cellular systems achieve a huge increase of available capacity. Such systems typically consist of an overlay or macrocellular network with underlying microcells. Microcells cover a geographically small part of the service area, and have limited capacity, so as to accommodate small areas with a larger traffic density ('hot spots'), which enables a tighter re-use of frequencies. Calls generated in a microcell without free capacity are handled by the macrocellular network. In addition, repacking calls from the overlay or macrocellular network to the microcellular network considerably increases the capacity of the network.

---

[1]The research of R.J. Boucherie is supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs, The Netherlands.
[a] CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.
[b] Eindhoven University of Technology, Department of Mathematics and Computing Science, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
[c] Universiteit van Amsterdam, Department of Econometrics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

Key performance measures in wireless communications networks include blocking probabilities. Unfortunately, exact expressions for such performance measures are usually unavailable. The goal of this note is therefore to derive approximations for blocking probabilities in layered networks under various repacking policies.

Layered networks can be represented as overflow systems such as investigated for classical telecommunications networks, with primary channels representing microcells, and secondary channels representing macrocells. Two-moment methods such as the Equivalent Random Method (ERM) developed by Wilkinson [11] have been successful tools for the analysis of overflow systems. Such methods are known to accurately approximate the blocking probability for systems with Poisson arrivals and negative exponential holding times, and identical service requirements at the primary and secondary (overflow) channels; see Wolff [12], chapter 7, for an overview of two-moment methods, and a discussion of their accuracy. Schehrer [10] has generalised the ERM to include distinct service requirements at primary and secondary channels, and to include arrivals directly to the secondary channels. The latter generalisation of Schehrer is particularly useful for also allowing fresh calls to originate at the macrocell.

With repacking from the macrocell to the microcell, Schehrer's generalisation of the ERM is no longer applicable as repacking influences mean and variance of the traffic overflowing the primary channels. This note proposes an approximative scheme, based on Schehrer's generalisation of the ERM, for obtaining blocking probabilities in an overflow system with repacking. The key element is our assumption that the rate at which repacked calls arrive at the primary cell is determined by the average number of calls overflowing the primary channels. This assumption implies that the joint arrival process of fresh calls and repacked calls is a Poisson process. Similar approximations have been proposed for retrial queues, e.g. the RTA (Retrials see Time Averages) approximation of [1, 2]. These approximations provide fairly good estimates of overflow probabilities for low retrial rates. This gives additional support for our approximation (although retrial queues differ from our model as we consider multiple microcells sharing a finite overflow buffer and also include a service rate at the overflow channels).

Here is the organisation of this note. Section 2 specifies the model and Section 3 discusses exact results and existing two-moment approximations for the mean and variance of overflow traffic from a single channel. Section 4 presents our approximation scheme. This scheme is evaluated in Section 5 through numerical experiments, where it is shown that our approximation provides accurate results for systems in which the repacking rate does not exceed the fresh call arrival rate. In the Appendix, an analytical expression is derived for an overflow system with a single primary channel and repacking.

2. The model

Consider a cellular communications network consisting of d primary cells Q_i, $i = 1, \dots, d$. Calls with negative exponential holding time with rate μ_i are generated

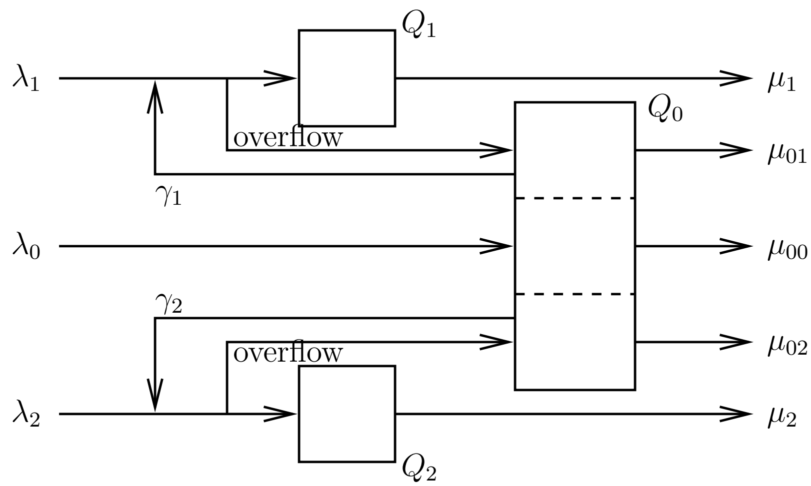

Figure 1: Overflow system with 2 primary cells.

in cell Q_i according to a Poisson process with rate λ_i. Cell Q_i has C_i channels reserved for calls arriving at that cell. Calls generated in cell Q_i finding all C_i channels occupied are handled by a secondary (overflow or macro) cell Q_0 containing C_0 channels. In addition, fresh calls are generated in the overflow cell with Poisson rate λ_0. The holding time of calls in the overflow cell is assumed to be negative exponential with rate μ_{0i} for calls originating in cell Q_i, $i = 0, \dots, d$. To improve efficiency, calls residing at the overflow channels originating from cell Q_i are repacked into cell Q_i at rate γ_i. Note that, due to the assumption of negative exponential holding times, a call that is repacked into cell Q_i has the same characteristics as a fresh call generated in cell Q_i. Figure 1 illustrates our notation for $d = 2$.

This situation typically occurs in layered cellular networks with a macrocell overlaying d microcells, where the microcells cover only part of the macrocell. The holding time of calls in the macrocell might differ from that time in the microcells because the macrocell covers a larger area. Handovers between microcells are typically avoided in layered networks, and are also excluded from our model. Our model is restricted to a single macrocell, which is in accordance with models in the literature, where handovers from other macrocells are often considered as arrivals of fresh calls. In addition, under assumptions leading to a product form expression (see Section 3.1), a decomposition of the network into macrocells for obtaining blocking probabilities is justified (see [3]). The network formulation also applies to loss networks with d different trunk groups, and alternative routing via a set of C_0 trunks, where calls are repacked into the primary trunks upon availability (see [6]). Furthermore, a set of d retrial queues sharing a common (finite) orbit fits into our formulation (see [5]).

Let $X_{0i}(t)$ be the number of calls generated at cell Q_i present at cell Q_0 at time t,

$i = 0, \ldots, d$, and let $X_i(t)$ denote that number present at cell Q_i, $i = 1, \ldots, d$. With

$$S = \{\mathbf{n} = (n_{00}, n_{01}, \ldots, n_{0d}, n_1, \ldots, n_d) \; : \; n_i \le C_i, \quad \sum_{i=0}^{d} n_{0i} \le C_0\},$$

the $2d+1$-dimensional Markov chain $\mathbf{X} = ((X_{00}(t), \ldots, X_{0d}(t), X_1(t), \ldots, X_d(t)),\ t \ge 0)$ on S of the number of calls has transition rates $q = (q(\mathbf{n}, \mathbf{n}'),\ \mathbf{n}, \mathbf{n}' \in S)$ given by

$$q(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_0 & \mathbf{n}' = \mathbf{n} + \mathbf{e}_{00}, \\ \lambda_i & \mathbf{n}' = \mathbf{n} + \mathbf{e}_i, \text{ or } \mathbf{n}' = \mathbf{n} + \mathbf{e}_{0i}, \text{ if } n_i = C_i, \\ \mu_i & \mathbf{n}' = \mathbf{n} - \mathbf{e}_i, \\ \mu_{0i} & \mathbf{n}' = \mathbf{n} - \mathbf{e}_{0i}, \\ \gamma_i & \mathbf{n}' = \mathbf{n} - \mathbf{e}_{0i} + \mathbf{e}_i, \end{cases}$$

where $\mathbf{e}_i$, $\mathbf{e}_{0i}$ denote ith unit vectors. We are interested in obtaining the *overflow probability* (the probability that an arriving call finds all primary channels occupied), and the *call loss probability* (the probability that a fresh call finds all channels occupied in both its primary cell and in the overflow cell).

Let $\pi(\mathbf{n}) = \lim_{t\to\infty} \mathrm{P}(X(t) = \mathbf{n})$ denote the equilibrium distribution. This distribution is the unique solution of the system of global balance equations, with $I(A)$ denoting the indicator of event A, for $\mathbf{n} \in S$,

$$\begin{aligned} \pi(\mathbf{n}) \Bigg\{ & \lambda_0 I(\sum_{i=0}^{d} n_{0i} < C_0) + \sum_{i=1}^{d} \lambda_i \left[I(n_i < C_i) + I(n_i = C_i, \ \sum_{i=0}^{d} n_{0i} < C_0) \right] \\ & + \sum_{i=1}^{d} n_i \mu_i + \sum_{i=0}^{d} n_{0i}\mu_i + \sum_{i=1}^{d} n_{0i}\gamma_i I(n_i < C_i) \Bigg\} \\ = \; & \lambda_0 \pi(\mathbf{n} - \mathbf{e}_{00}) + \sum_{i=1}^{d} \lambda_i \left[\pi(\mathbf{n} - \mathbf{e}_i) + \pi(\mathbf{n} - \mathbf{e}_{0i}) I(n_i = C_i) \right] \end{aligned} \tag{2.1}$$

$$+ \sum_{i=1}^{d} (n_i + 1)\mu_i \pi(\mathbf{n} + \mathbf{e}_i) + \sum_{i=0}^{d} (n_{0i} + 1)\mu_{0i} \pi(\mathbf{n} + \mathbf{e}_{0i}) + \sum_{i=1}^{d} (n_{0i} + 1)\gamma_i \pi(\mathbf{n} + \mathbf{e}_{0i} - \mathbf{e}_i),$$

normalised such that $\sum_{\mathbf{n}\in S} \pi(\mathbf{n}) = 1$. From the equilibrium distribution the desired blocking probabilities can be computed. However, the global balance equations can only be explicitly solved for special cases of the network parameters.

3. Exact expressions for call blocking

Two extreme cases of retrial behaviour are well understood. On the one hand, an overflow system with instantaneous repacking can be modelled as a loss network, with resulting multidimensional Poisson distribution for π (Section 3.1). On the other hand, an overflow system without repacking can be suitably approximated using two-moment methods (Section 3.2). However, in this note we are interested in the intermediate case of a finite positive repacking rate. Section 3.3 (see also the Appendix) provides a result for a single channel with infinite overflow buffer and arbitrary repacking rates. The proof of this result suggests that such models with multiple primary channels cannot be analysed explicitly. Therefore, in Section 4 we provide an approximation for the blocking probabilities for general values of the repacking rates.

3.1 Immediate repacking

For $\gamma_i = \infty$, $i = 1, \dots, d$, repacking is immediate, i.e., as soon as a channel becomes available in cell Q_i, a call residing in the overflow cell, and originating in cell Q_i, is repacked into cell Q_i. Then, for $\mu_i = \mu_{0i}$, $i = 1, \dots, d$, equation (2.1) has an explicit solution (cf. Ross [9])

$$\pi(\mathbf{n}) = G^{-1} \prod_{i=0}^{d} \frac{\rho_i^{n_i+n_{0i}}}{(n_i+n_{0i})!}, \quad \mathbf{n} \in S, \quad G = \sum_{\mathbf{n}\in S} \prod_{i=0}^{d} \frac{\rho_i^{n_i+n_{0i}}}{(n_i+n_{0i})!},$$

where we have defined $\rho_i = \lambda_i/\mu_{0i}$, $i = 0, \dots, d$, and we have set $n_0 \equiv 0$ to simplify notation. This distribution is a truncated multivariate Poisson distribution, from which blocking probabilities can be easily obtained. Blocking of a fresh call in cell Q_i is computed as

$$B_i = \sum_{\mathbf{n}\in T_i} \pi(\mathbf{n}), \quad \text{where} \quad T_i = \{\mathbf{n} : \mathbf{n} \in S,\ \mathbf{n} + \mathbf{e}_{i0} \notin S\}.$$

The summations involved in this expression can be evaluated for small networks using exact methods, and for larger networks using Monte Carlo methods, or approximate methods, see e.g. [9] for an overview of such methods.

3.2 No repacking

We now consider the case of a single primary cell with C channels, offered Poisson traffic at rate λ, holding time with rate μ_{cell} at the primary cell, and μ_{over} at an overflow cell containing infinitely many channels. The mean, m_{prim}, and variance, v_{prim}, of the number of calls overflowing the primary queue are (see Wolff [12], p. 351)

$$\begin{aligned} m_{prim} &= \rho_{cell}\mathrm{Erl}(\rho_{cell}, C), \\ v_{prim} &= m_{prim}\left(1 - m_{prim} + \frac{\rho_{cell}}{C + 1 + m_{prim} - \rho_{cell}}\right), \end{aligned}$$

where we have abbreviated $\rho_{cell} = \lambda/\mu_{cell}$, and $\mathrm{Erl}(\rho, C)$ is the Erlang loss probability.

If $\mu_{over} = \mu_{cell}$, then m_{prim} and v_{prim} are also the mean and variance of the occupancy of the overflow cell. If $\mu_{over} \neq \mu_{cell}$, however, mean and variance of the channel occupancy of the overflow cell are different, and are given by (cf. Schehrer [10])

$$m_{sec}(\lambda, \mu_{cell}, \mu_{over}, C) = \frac{\mu_{cell}}{\mu_{over}} m_{prim}, \tag{3.1}$$

$$v_{sec}(\lambda, \mu_{cell}, \mu_{over}, C) = m_{sec}\left(1 - m_{sec} + \frac{\lambda}{\mu_{over}} \frac{g(C)}{\sum_{n=0}^{C} g(n)}\right), \tag{3.2}$$

where we have abbreviated $m_{sec} = m_{sec}(\lambda, \mu_{cell}, \mu_{over}, C)$ in the expression for the variance. The function $g(\cdot)$ can be recursively computed as (cf. [10])

$$\begin{aligned} g(0) &= 1, \quad g(-1) = 0, \\ g(n) &= \frac{1}{n}(\rho_{cell} + n - 1 + \frac{\mu_{over}}{\mu_{cell}})g(n-1) - \frac{\rho_{cell}}{n} g(n-2), \quad n = 1, \dots, C. \end{aligned}$$

For $\gamma_i = 0$ (no repacking), and $\mu_i = \mu_{cell}$, $i = 1, \dots, d$, $\mu_{0i} = \mu_{over}$, $i = 0, \dots, d$, Schehrer's expressions (3.1), (3.2) can be applied to *approximate* the fresh call blocking probabilities for the system with *finitely many overflow channels*. To this end, an *equivalent primary cell* replacing the primary cells is defined such that the mean and variance of the traffic overflowing the equivalent primary cell equal the mean and variance of the aggregated overflow traffic of the primary cells $Q_1, \dots, Q_d$. Let m and v be defined as

$$m = \sum_{i=0}^{d} m_{sec}(\lambda_i, \mu_{cell}, \mu_{over}, C_i), \quad v = \sum_{i=0}^{d} v_{sec}(\lambda_i, \mu_{cell}, \mu_{over}, C_i),$$

then the equivalent primary cell has mean holding time μ_{over}^{-1}, and offered traffic a, and number of channels C, determined from

$$m = a\mathrm{Erl}(a, C), \quad v = m\left(1 - m + \frac{a}{C + 1 + m - a}\right).$$

The blocking probability for calls arriving to cell Q_i is then approximated as

$$B_i \sim \mathrm{Erl}(\lambda_i/\mu_i, C_i)\frac{\mathrm{Erl}(a, C + C_0)}{\mathrm{Erl}(a, C)}.$$

More accurate approximations can be obtained when using the peakedness of the overflow traffic from the cells. We shall elaborate on improvements using peakedness in Section 4.

3.3 Single primary channel

Let us now study the special case of a single primary cell with one channel, offered Poisson traffic at rate λ, holding time with rate μ_{cell} at the primary cell, and μ_{over} at an overflow cell containing infinitely many channels. The mean and variance of the traffic overflowing the primary channel can be computed for all repacking rates $0 < \gamma < \infty$. In particular, we have obtained an explicit expression for the generating function of the joint distribution of the number of calls at the primary channel and at the overflow channels. Details of the analysis leading to these explicit expressions are provided in the Appendix. Here we only state the results.

For $C = 1$ an explicit expression for the mean and variance of the occupancy of the overflow channels can be obtained as

$$m = \frac{\lambda}{\mu}Q_0'(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}) + \frac{\lambda}{\mu}Q_1'(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}), \tag{3.3}$$

$$v = (\frac{\lambda}{\mu})^2 Q_0''(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}) + (\frac{\lambda}{\mu})^2 Q_1''(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}) + m - m^2, \tag{3.4}$$

where $Q_0(\cdot)$ and $Q_1(\cdot)$ are defined in the Appendix. This result is used in Section 5 to compare the approximation of Section 4 with exact results.

From the analysis presented in the Appendix it appears highly unlikely that exact expressions can be obtained for a network with a primary cell containing $C \geq 2$ channels. A similar observation is justified when comparing the model with repacking to retrial queues. Generally speaking, for such queues, analytical solutions have been obtained only for $C = 1$, and approximations are used for $C \geq 2$, see Falin and Templeton [5], or Wolff [12], for an overview of results for retrial queues.

4. Approximation of call blocking

This section presents a simple approximation for the loss probability of an overflow system with repacking. The approximation uses Schehrer's expressions (3.1), (3.2) for estimating the occupancy of the overflow channels, and then approximates the blocking probabilities of the network as if the joint arrival process of fresh calls and repacked calls forms a Poisson arrival process.

Invoking two-moment methods requires the call holding times to be identically distributed. Therefore, we restrict ourselves to the case that

$$\mu_i = \mu_{cell}, \quad i = 1, \ldots, d, \quad \text{and} \quad \mu_{0i} + \gamma_i = \mu_{over}, \quad i = 0, \ldots, d. \tag{4.1}$$

In the system with repacking, arrivals at primary cell Q_i are either fresh calls, or repacked calls. Let $\mathrm{E}X_{0i}$ denote the mean number of calls in the overflow cell that have overflowed cell Q_i. Then repacked calls arrive at primary cell Q_i at rate $\gamma_i \mathrm{E}X_{0i}$. Thus, the arrival rate of calls to primary cell Q_i is

$$\nu_i = \lambda_i + \gamma_i \mathrm{E}X_{0i}, \quad i = 1, \ldots, d. \tag{4.2}$$

Our *approximation* is that blocking occurs as if calls arrive at primary cell Q_i at *Poisson* rate ν_i, $i = 1, \ldots, d$, and that calls depart from the overflow cell at rate μ_{over}; see Figure 2 for an illustration of the approximate model. Thus the repacking process is decoupled into separate arrival and departure processes. This natural decomposition has been used for retrial queues, and is referred to as RTA (Retrials see Time Averages) approximation; see [1, 4, 12]. The RTA approximation is numerically shown [2] to result in a fairly accurate approximation for the overflow probability of an $M/M/C/C$ queue with infinite overflow buffer, where calls in the overflow buffer attempt retrials at low exponential rate (service in the overflow buffer is not included in the model).

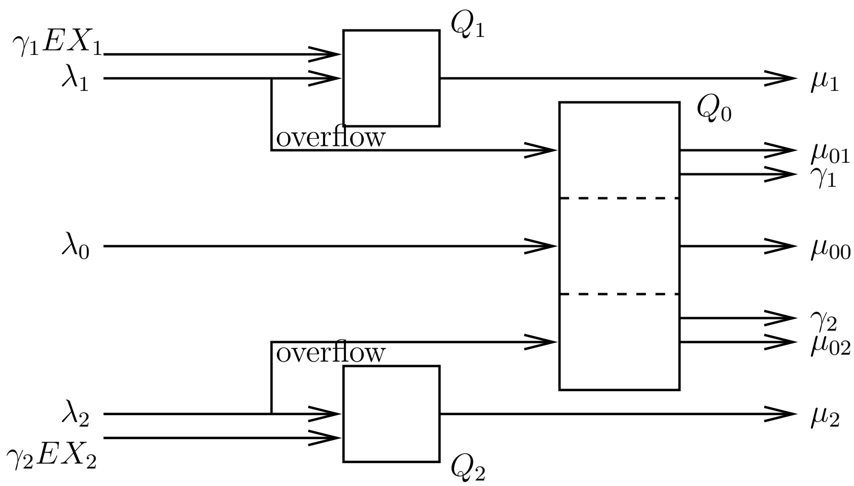

Figure 2: Approximate overflow system with 2 primary cells.

The approximation (4.2) requires an estimate of $\mathrm{E}X_{0i}$. The mean number of calls in the overflow cell can be approximated using Schehrer's two-moment method. The mean number of calls overflowing cell Q_i is $\nu_i \mathrm{Erl}(\nu_i/\mu_{cell}, C_i)(1 - K_i)$, where K_i is the probability that a call overflowing cell Q_i is blocked at the overflow cell. Thus, incorporating the holding time of calls in the overflow cell,

$$\mathrm{E}X_{0i} = \nu_i \mathrm{Erl}(\nu_i/\mu_{cell}, C_i)(1 - K_i)/\mu_{over}, \quad i = 1, \dots, d. \tag{4.3}$$

Notice that the right-hand side of this expression for $\mathrm{E}X_{0i}$ also includes $\mathrm{E}X_{0i}$ via ν_i, and also requires computation of K_i.

The blocking probability K_i can be approximated using the Katz approximation (see Wolff [12], p. 357) which takes into account the peakedness of the traffic overflowing the primary cells. To this end, first determine an equivalent primary cell with offered load a and C channels from (recall (3.1), (3.2))

$$a\mathrm{Erl}(a, C) = \sum_{i=0}^{d} m_{sec}(\nu_i, \mu_{cell}, \mu_{over}, C_i), \tag{4.4}$$

$$a\mathrm{Erl}(a, C)\left(1 - a\mathrm{Erl}(a, C) + \frac{a}{C + 1 + a\mathrm{Erl}(a, C) - a}\right) = \sum_{i=0}^{d} v_{sec}(\nu_i, \mu_{cell}, \mu_{over}, C_i). \tag{4.5}$$

Assume that K_i is affine with the peakedness z_i of the traffic overflowing cell Q_i. This assumption is justified as the overflow cell has fresh call arrivals at Poisson rate λ_0, see Wolff [12], p. 357, for a discussion. Approximating the total blocking probability in the overflow cell using Schehrer's generalisation of the ERM as $a\mathrm{Erl}(a, C + C_0)/a\mathrm{Erl}(a, C)$, the *Katz approximation* of the blocking probability at the overflow cell is

$$K_i = \frac{a\mathrm{Erl}(a, C + C_0)}{a\mathrm{Erl}(a, C)}\left(v(C, C_0)^{-1} + \frac{z_i - 1}{z - 1}(1 - v(C, C_0)^{-1})\right), \tag{4.6}$$

where $v(C, C_0)$ is determined from the recursion, with initial condition $v(C, 0) = 1$,

$$v(C, j) = \frac{aj}{a\text{Erl}(a, C)[C + j - a - a\text{Erl}(a, C)v(C, j-1)]}, \quad j = 1, 2, \dots,$$

and the peakedness, z_i, of the traffic overflowing primary cell Q_i, and z, of the total traffic arriving at the overflow cell are respectively given by

$$z_i = \frac{v_{sec}(\nu_i, \mu_{cell}, \mu_{over}, C_i)}{m_{sec}(\nu_i, \mu_{cell}, \mu_{over}, C_i)}, \quad z = \frac{\sum_{i=0}^{d} v_{sec}(\nu_i, \mu_{cell}, \mu_{over}, C_i)}{\sum_{i=0}^{d} m_{sec}(\nu_i, \mu_{cell}, \mu_{over}, C_i)}.$$

We have now established all ingredients for our approximation. The *loss probability for arrivals at primary cell Q_i is approximated* by

$$B_i = \text{Erl}(\nu_i/\mu_{cell}, C_i)K_i. \tag{4.7}$$

This blocking probability B_i approximates the fraction of *time* that 'stream' i is blocked. As arrivals of stream i occur according to a Poisson process with rate λ_i, B_i (which is estimated using ν_i) is the probability that a fresh call arriving to cell i is blocked. We will refer to the approximation (4.7), using the fixed point equations (4.2), (4.3), and (4.6) to determine ν_i, $i = 1, \dots, d$, as the ERMR approximation.

It is important to observe that, to approximate the call loss probability, ERMR is based on only two assumptions:

1. Blocking of calls occurs as if combined fresh call and repacked arrivals to the primary cells form Poisson processes.

2. Blocking of calls at the overflow cell is approximated via the ERM.

Both approximations turn out to be fairly accurate in special cases: Assumption 1 is accurate for a retrial queue with infinite overflow buffer and low retrial rate (Falin and Artelejo [2]), and Assumption 2 is an established approximation for the resulting system of Figure 2 that does not include retrials.

Approximation (4.3) is call conserving at the overflow cell. On the one hand, the mean number of calls, $\mathrm{E}X_0$, in the overflow cell is approximated as $\mathrm{E}X_0 = a[\text{Erl}(a, C) - \text{Erl}(a, C+C_0)]$, using Schehrer's approximation. On the other hand, the average number of calls at the overflow cell via the primary cells and as fresh calls is $\frac{\lambda_0}{\mu_{00}}(1 - K_0) + \sum_{i=1}^{d} \mathrm{E}X_{0i}$. It is straightforward to verify that the approximation (4.3) of $\mathrm{E}X_{0i}$ using peakedness of the overflow traffic satisfies

$$\frac{\lambda_0}{\mu_{00}}(1 - K_0) + \sum_{i=1}^{d} \mathrm{E}X_{0i} = \mathrm{E}X_0,$$

which provides a sanity check of our approximation.

Computation of the approximate blocking probabilities, B_i, of (4.7) requires determining a fixed point ν_i, $i = 1, \dots, d$, of the system of equations (4.2), (4.3), (4.6). For obtaining a and C we will use a continuous approximation of the Erlang loss probability $\mathrm{Erl}(\rho, C)$, so that this loss probability can be obtained for each real number C. Similarly, the function $v(C, C_0)$ that is used in the Katz approximation (4.6) is a continuous approximation of $v(C, C_0)$ and can be obtained for each real C. For computing this continuous approximation v it must be noted that the recursion for v is unstable for non-integer values of C. We therefore compute values for integer C first, and approximate after completion of the recursion. The resulting recursions appear to be stable.

Using Brouwer's fixed point theorem, it can be easily shown that the thus obtained continuous approximation of the system of equations (4.2), (4.3), (4.6), has a fixed point $(\nu_1, \dots, \nu_d)$. To this end, first observe that $m_{sec}(\lambda, \mu_{cell}, \mu_{over}, C)$, and $v_{sec}(\lambda, \mu_{cell}, \mu_{over}, C)$ of (3.1), (3.2) are continuous functions of λ. (In these equations C is the number of channels of the primary cells, and not the number of channels obtained from Schehrer's approximation.) As a consequence, the right-hand sides of (4.4), (4.5) are continuous functions of ν_i, $i = 1, \dots, d$. Denote these right-hand sides by $m(\nu)$ and $v(\nu)$, respectively. Then, the number of channels of the equivalent primary cell, say $C(\nu)$, is easily obtained from (4.5) as $C(\nu) = a((v(\nu)/m(\nu)) + m(\nu))\,/((v(\nu)/m(\nu)) + m(\nu) - 1) - m(\nu) - 1$. Insertion of this expression into (4.4) yields that the equivalent load a is a continuous function of $\nu_1, \dots, \nu_d$, say $a(\nu)$ that is determined by $m(\nu) = a(\nu)\mathrm{Erl}(a(\nu), C(\nu))$. It is not too difficult to see that this equation has a fixed point $a(\nu)$ that is a continuous function of ν.

Observe that K_i of (4.6) is a continuous function of $\nu_1, \dots, \nu_d$, say $K_i(\nu)$, and that $0 \le K_i(\nu) \le 1$. Equation (4.2) now reads

$$\nu_i = \lambda_i + \nu_i \frac{\gamma_i}{\gamma_i + \mu_{0i}} \mathrm{Erl}(\nu_i/\mu_{cell}, C_i)(1 - K_i(\nu)), \quad i = 1, \dots, d,$$

or in functional form $\nu_i = f_i(\nu_1, \dots, \nu_d)$, $i = 1, \dots, d$,
where $f(\nu_1, \dots, \nu_d) = (f_1(\nu_1, \dots, \nu_d), \dots, f_d(\nu_1, \dots, \nu_d))$ is a continuous function

$$f : [0, \tfrac{\lambda_1(\mu_{01}+\gamma_1)}{\mu_{01}}] \times \dots \times [0, \tfrac{\lambda_d(\mu_{0d}+\gamma_d)}{\mu_{0d}}] \to [0, \tfrac{\lambda_1(\mu_{01}+\gamma_1)}{\mu_{01}}] \times \dots \times [0, \tfrac{\lambda_d(\mu_{0d}+\gamma_d)}{\mu_{0d}}],$$

that has a fixed point according to Brouwer's fixed point theorem.

The system of equations (4.2), (4.3), (4.6) might have multiple fixed points, a problem similar to that involved in the reduced load methods for loss networks, see e.g. [6]. Our numerical experiments, however, indicate that, in the light load situation such as typically occurring for cellular networks, the apparently unique fixed point of (4.2), (4.3), (4.6) can be easily found via a successive iteration method.

Intuitively, the blocking probabilities B_i, $i = 0, \dots, d$, are bounded from above by the blocking probability in the system without repacking, and bounded from below by the blocking probability in the system with immediate repacking. Therefore, interpolation between these blocking probabilities to obtain an estimate for B_i at first sight seems an alternative for our approximation method. This interpolation approximation generally does not lead to a suitable approximation as the ratio of the bounding probabilities can easily exceed a factor of 1000.

5. Numerical results

This section presents numerical results illustrating the accuracy of our approximation via a comparison with simulation results (estimated probabilities have 95% confidence and 10% relative precision). First, Section 5.1 considers the special case of Section 3.3: a single channel with infinite overflow buffer. The results indicate that the approximation is fairly good for low retrial rates. Then, in Section 5.2 it is shown that ERMR coincides with the RTA approximation, and provides good estimates of blocking probabilities at the primary channels. Section 5.3 applies our approximation to layered mobile communications networks that generally operate under light load. Finally, the behaviour of ERMR for moderately loaded networks is investigated.

5.1 Single channel with infinite overflow buffer

Consider the network of Section 3.3: a single channel with infinite overflow buffer. For this system, mean and variance of the occupancy of the overflow buffer can be obtained in closed form, see (3.3), (3.4). From ERMR, m and v can also be obtained explicitly. The fixed point ν of (4.2), (4.3), (4.6) satisfies

$$\nu = \lambda + \gamma\nu \mathrm{Erl}(\nu/\mu_{cell}, 1)/(\gamma + \mu_{over}).$$

For $\lambda = \mu_{cell} = \mu_{over} = 1$ this yields $\nu = \sqrt{1+\gamma}$, so that m and v can be obtained from (3.1), (3.2) in analytical form as

$$m = \mathrm{E}X_{01} = \frac{1}{1+\sqrt{1+\gamma}}, \qquad v = m\left(1 - m + \frac{\nu}{1+\gamma}\frac{1+\nu+\gamma}{2+\nu+\gamma}\right).$$

Table 5.1 lists m and v as well as the blocking probability, B_{loc}, at the primary buffer from ERMR, and from (3.3), (3.4). Comparison of these results shows that m and v are only well approximated if the retrial rate is smaller than the arrival rate. Further observe that the probability that an arriving call overflows the primary channel is approximated fairly well.

	Exact			ERMR		
γ	$\mathrm{E}X_0$	$\mathrm{var}(X_0)$	B_{loc}	$\mathrm{E}X_0$	$\mathrm{var}(X_0)$	B_{loc}
0.1	0.492	0.578	0.508	0.488	0.567	0.512
0.5	0.469	0.563	0.532	0.449	0.516	0.551
1	0.450	0.550	0.551	0.414	0.469	0.586
5	0.400	0.518	0.599	0.290	0.312	0.710
10	0.387	0.509	0.612	0.231	0.243	0.768
$d=1,\ C_0=\infty,\ C_1=1,\ \lambda_0=0,\ \lambda_1=1,\ \mu_{cell}=\mu_{over}=1$						

Table 5.1: Single channel with infinite overflow buffer.

5.2 Retrial queues

A basic $M/M/C/C$ retrial queue is a single primary buffer with infinite overflow buffer called orbit. Customers arriving to the system that find the primary buffer occupied wait *in orbit*. These customers reattempt to join the primary buffer at exponential rate. Thus, a retrial queue corresponds to the setting of Section 2 with $d = 1$, $C_0 = \infty$, $\lambda_0 = 0$, and $\mu_{01} = 0$. For a retrial queue, it is not the total blocking probability that is of interest, but the probability that an arriving customer overflows the primary buffer (cannot be served immediately). For an overview of results for $M/M/C/C$ retrial queues see Falin and Templeton [5], Chapter 2.

For the numerical analysis of an $M/M/C/C$ retrial queue, Falin and Artelejo [2] investigate the RTA approximation. This approximation is based on limiting arguments for retrial models. For the setting presented above, RTA approximates the retrial rate δ as the unique fixed point of (see [2, 4]):

$$\delta = (\lambda + \delta)\mathrm{Erl}(\lambda + \delta, C).$$

It can easily be seen that the fixed point ν of (4.2), (4.3), (4.6) satisfies

$$\nu = \lambda + \nu\mathrm{Erl}(\nu, C).$$

As $\nu = \lambda + \delta$, ERMR coincides with the RTA approximation. Comparison with simulation (see also [2]) shows that ERMR provides fairly accurate results for low retrial rates (see Table 5.2). The RTA approximation does not involve the actual retrial rate γ, and is therefore insensitive to the value of γ. Only for low retrial rates RTA provides a good estimate of the overflow probability at the primary buffer. As will be shown below, ERMR extends the results of the RTA approximation to finite orbits and to multiple buffers sharing a finite orbit. In those cases ERMR does involve the actual value γ of the retrial rates.

	$C = 2$		$C = 3$		$C = 5$	
γ	RTA	simul	RTA	simul	RTA	simul
0.1	0.293	0.278	0.0729	0.0726	0.00311	0.00301
0.2		0.286		0.0739		0.00300
0.5		0.296		0.0762		0.00307
1		0.304		0.0784		0.00313
5		0.321		0.0847		0.00342
10		0.324		0.0867		0.00351
$d = 1,\ \lambda_0 = 0,\ \lambda_1 = 1,\ \mu_{01} = 0,\ \mu_1 = 1$						

Table 5.2: Blocking probabilities of an $M/M/C/C$ retrial queue with infinite orbit.

5.3 Light traffic: layered mobile communications networks

Cellular mobile communications networks generally operate under light load; blocking probabilities not exceeding 1% are desirable. As capacity is severely limited, microcells are added to the macrocells to accommodate areas with a high traffic density.

Channels can be re-used in multiple microcells within a macrocell, but channels assigned to microcells cannot be used in the remaining part of the macrocell. As a microcell physically restricts part of the capacity to a small part of the coverage area, there is a trade-off between capacity allocation to the microcells and to the macrocell. This results in microcells being heavily loaded (with high local blocking probabilities). The macrocell is lightly loaded so that the total blocking probability remains within bounds. Obviously, capacity increases when repacking of calls from the macrocell to the microcells is taken into account.

Table 5.3 presents a comparison of ERMR with simulation results for a macrocell with two underlying microcells in a homogeneous situation (load is equally shared by the cells). As the results for cells Q_1 and Q_2 are identical, blocking probabilities for cell Q_1 are listed only. Here B_{loc} is the probability that a call overflows the microcell, and B_{tot} is the probability that an arriving call is rejected by both the microcell and the macrocell. The results presented in Table 5.3 indicate that ERMR is reasonably accurate for lower repacking rates. In the regime $\gamma \mathrm{E} X_0 < \lambda$ the results are within engineering accuracy. Observe that the approximation for B_{loc} is extremely accurate. This indicates that ERMR performs excellently for retrial models, where multiple queues share a common finite orbit.

Similarly, Table 5.4 lists results of a comparison of our approximation with simulation for an inhomogeneous model, and shows that ERMR remains accurate for low retrial rates.

	Simulation			ERMR			
cell	γ_i	B_{loc}	B_{tot}	ν_i	B_{loc}	B_{tot}	$\mathrm{E}X_0$
Q_0	0	1	0.0926	14	1	0.0937	16.04
Q_1	0	0.146	0.0242	14	0.148	0.0278	
Q_0	0	1	0.0655	14	1	0.0679	15.28
Q_1	0.1	0.152	0.0182	14.16	0.153	0.0204	
Q_0	0	1	0.0472	14	1	0.0479	14.53
Q_1	0.2	0.158	0.0147	14.34	0.159	0.0157	
Q_0	0	1	0.0189	14	1	0.0168	12.51
Q_1	0.5	0.172	0.0068	14.83	0.175	0.0062	
Q_0	0	1	0.0058	14	1	0.0030	10.02
Q_1	1	0.193	0.0026	15.52	0.197	0.0022	
Q_0	0	1	0.0015	14	1	0.0001	7.18
Q_1	2	0.225	0.0009	16.52	0.229	0.00006	
$d=2,\ \lambda_i=14,\ \forall i,\ \mu_i=\mu_{0i}=1,\ \forall i,\ C_0=21,\ C_1=C_2=15$							

Table 5.3: Homogeneous model.

			Simulation		ERMR			
cell	λ_i	γ_i	B_{loc}	B_{tot}	ν_i	B_{loc}	B_{tot}	EX_0
Q_0	10	0	1	0.268	10	1	0.251	12.60
Q_1	3	0.1	0.021	0.0061	3.00	0.022	0.0071	
Q_2	4	0.1	0.062	0.019	4.02	0.064	0.022	
Q_3	5	0.1	0.123	0.040	5.04	0.123	0.044	
Q_4	6	0.1	0.192	0.062	6.07	0.189	0.069	
Q_5	7	0.1	0.251	0.082	7.11	0.255	0.092	
Q_6	8	0.1	0.314	0.103	8.15	0.317	0.113	
Q_7	9	0.1	0.370	0.120	9.20	0.372	0.130	
Q_0	10	0	1	0.238	10	1	0.218	12.36
Q_1	3	0.2	0.021	0.0056	3.00	0.022	0.0064	
Q_2	4	0.2	0.064	0.017	4.03	0.064	0.020	
Q_3	5	0.2	0.124	0.036	5.07	0.125	0.040	
Q_4	6	0.2	0.193	0.057	6.13	0.194	0.064	
Q_5	7	0.2	0.260	0.076	7.21	0.262	0.086	
Q_6	8	0.2	0.325	0.097	8.31	0.325	0.105	
Q_7	9	0.2	0.376	0.111	9.41	0.381	0.121	
Q_0	10	0	1	0.179	10	1	0.158	11.64
Q_1	3	0.5	0.023	0.0050	3.02	0.022	0.0046	
Q_2	4	0.5	0.070	0.015	4.07	0.066	0.015	
Q_3	5	0.5	0.132	0.029	5.17	0.132	0.030	
Q_4	6	0.5	0.206	0.046	6.33	0.207	0.048	
Q_5	7	0.5	0.280	0.063	7.54	0.282	0.065	
Q_6	8	0.5	0.344	0.078	8.80	0.351	0.080	
Q_7	9	0.5	0.408	0.092	10.08	0.412	0.092	
Q_0	10	0	1	0.117	10	1	0.093	10.61
Q_1	3	1	0.023	0.0034	3.03	0.023	0.0028	
Q_2	4	1	0.070	0.010	4.12	0.069	0.0093	
Q_3	5	1	0.143	0.022	5.32	0.141	0.020	
Q_4	6	1	0.220	0.034	6.65	0.227	0.032	
Q_5	7	1	0.307	0.046	8.09	0.313	0.044	
Q_6	8	1	0.379	0.058	9.62	0.392	0.054	
Q_7	9	1	0.445	0.068	11.24	0.460	0.062	
Q_0	10	0	1	0.068	10	1	0.041	9.14
Q_1	3	2	0.025	0.0019	3.05	0.023	0.0012	
Q_2	4	2	0.078	0.0066	4.19	0.073	0.0042	
Q_3	5	2	0.150	0.013	5.53	0.155	0.0093	
Q_4	6	2	0.250	0.024	7.16	0.259	0.016	
Q_5	7	2	0.341	0.031	9.08	0.366	0.022	
Q_6	8	2	0.432	0.039	11.26	0.461	0.027	
Q_7	9	2	0.506	0.045	13.62	0.540	0.030	
$d = 7,\ \mu_i = 1,\ \mu_{0i} = 1,\ \forall i,\ \mu_{00} = 1 + \gamma_1,\ C_0 = 15,\ C_i = 7,\ \forall i$								

Table 5.4: Inhomogeneous model.

Besides the results for lightly loaded networks ($B_{tot} \sim 0.01 - 0.1$), we have also tested our approximation for moderately loaded networks ($B_{tot} \sim 0.3$). As is shown in Table 5.5, where the homogeneous network of Table 5.3 is investigated under moderate load, in that range the accuracy of the approximation is similar to the accuracy listed in the tables above: for $\gamma \mathrm{E}X_0 < \lambda$ the blocking probabilities B_{tot} are approximated reasonably well, and the approximation of the blocking probabilities B_{loc} is excellent. Furthermore, our results support the observation that typically the call loss probability decreases with increasing repacking rates.

	Simulation			ERMR			
cell	γ_i	B_{loc}	B_{tot}	ν_i	B_{loc}	B_{tot}	$\mathrm{E}X_0$
Q_0	0	1	0.911	10	1	0.880	1.87
Q_1	0.1	0.396	0.370	21.03	0.395	0.377	
Q_0	0	1	0.896	10	1	0.857	1.84
Q_1	1	0.403	0.373	21.21	0.399	0.380	
Q_0	0	1	0.870	10	1	0.836	1.81
Q_1	5	0.418	0.376	21.43	0.404	0.380	
Q_0	0	1	0.860	10	1	0.822	1.80
Q_1	10	0.428	0.380	21.15	0.398	0.390	
$d = 2,\ \lambda_0 = 10,\ C_0 = 2,\ \lambda_i = 21,\ C_i = 14,\ \mu_i = \mu_{0i} = 1$							

Table 5.5: Homogeneous model under moderate load.

References

1. J.R. Artelejo, A queueing system with returning customers and waiting line, *Operations Research Letters*, 17:191-199, 1995.
2. G.I. Falin and J.R. Artelejo, Approximations for multiserver queues with balking/retrial discipline, *OR Spektrum*, 17:239-244, 1995.
3. R.J. Boucherie and M. Mandjes, Estimation of performance measures for product form cellular mobile communiations networks, *To appear: Telecommunication Systems*.
4. J.W. Cohen, Basic problems of telephone traffic theory and the influence of repeated calls, *Philips Telecommunication Review*, 18:49-100, 1957.
5. G.I. Falin and J.G.C. Templeton, *Retrial Queues*, Chapman and Hall, 1997.
6. F.P. Kelly, Loss networks, *The Annals of Applied Probability*, 1:319-378, 1991.
7. A. Kuczura, The interrupted Poisson process as an overflow process, *The Bell System Technical Journal*, 52:437-448, 1973.
8. N.N. Lebedev, *Special Functions and their Applications*, Dover, New York, 1972.
9. K.W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer, 1995.

10. R.G. Schehrer, A two moment method for overflow systems with different mean holding times, *Proceedings ITC-15*, 1303-1314.

11. R.I. Wilkinson, Theories for toll traffic engineering in the U.S.A., *The Bell Systems Technical Journal*, 35:421-514, 1956.

12. R.W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice-Hall, 1989.

6. Appendix

Consider the model of Section 3.3: A single primary cell Q_1 with C channels is offered Poisson traffic at rate λ. The holding times at the primary cell are exponentially distributed with rate μ. The overflow traffic from the primary cell is directed towards an overflow cell Q_0 containing infinitely many channels. The holding times at the overflow cell are exponentially distributed with rate ζ. Calls that leave the overflow cell return to the primary cell with probability p (and are treated there in the same way as newly offered traffic), and leave the system with probability $1-p$. The choice $\zeta = \mu + \gamma$, $p = \gamma/(\mu+\gamma)$ is of particular interest to us, as it corresponds to the case in which the actual holding time at the overflow cell is exponential with rate μ, while repacking occurs at rate γ.

The purpose of this appendix is to determine the equilibrium distribution of the two-dimensional Markov chain $\mathbf{X} = ((X_0(t), X_1(t)),\ t \geq 0)$ of numbers of calls at Q_1 and Q_0, for the case $C=1$. Below we set up the balance equations for this equilibrium distribution, for general positive integer C, but we are only able to solve them explicitly for $C=1$.

Let $\pi(j,i) = \lim_{t\to\infty} \mathrm{P}(X(t) = (j,i))$, $i = 0,\ldots,C$, $j \geq 0$, denote the equilibrium distribution. Then, π is the solution of the following global balance equations, that read with $\pi(j,-1) = 0$, $j \geq 0$, for ease of notation

$$\begin{aligned}(\lambda + i\mu + j\zeta)\pi(j,i) &= \lambda\pi(j,i-1) + (i+1)\mu\pi(j,i+1)\\ &+ (j+1)\zeta(1-p)\pi(j+1,i) + (j+1)\zeta p\pi(j+1,i-1), \quad i = 0,\ldots,C-1,\end{aligned}$$

$$\begin{aligned}(\lambda + C\mu + j\zeta(1-p))\pi(j,C) &= \lambda\pi(j,C-1) + \lambda\pi(j-1,C)\\ &+ (j+1)\zeta(1-p)\pi(j+1,C) + (j+1)\zeta p\pi(j+1,C-1).\end{aligned}$$

Introducing, for $|z| \leq 1$, $i = 0,1,\ldots,C$, the generating function $P_i(z) = \sum_{j=0}^{\infty} \pi(j,i) z^j$ of the number of calls in the overflow cell for given occupancy of the primary channels, the following equations are obtained:

$$\begin{aligned}(\lambda + i\mu)P_i(z) + \zeta z P_i'(z) &= \lambda I(i \geq 1) P_{i-1}(z) + (i+1)\mu P_{i+1}(z)\\ &+ \zeta(1-p)P_i'(z) + \zeta p I(i \geq 1) P_{i-1}'(z), \quad i = 0,1,\ldots,C-1, \qquad (6.1)\end{aligned}$$

$$\begin{aligned}(\lambda + C\mu)P_C(z) + \zeta(1-p)zP_C'(z) &= \lambda P_{C-1}(z) + \lambda z P_C(z)\\ &+ \zeta(1-p)P_C'(z) + \zeta p P_{C-1}'(z). \qquad (6.2)\end{aligned}$$

This set of equations can be solved explicitly in the case $C = 1$, as will now be demonstrated. For $C = 1$ the equations (6.1), (6.2) reduce to a set of two first-order differential equations:

$$\lambda P_0(z) + \zeta z P_0'(z) \;=\; \mu P_1(z) + \zeta(1-p)P_0'(z), \tag{6.3}$$

$$(\lambda(1-z)+\mu)P_1(z) + \zeta(1-p)(z-1)P_1'(z) \;=\; \lambda P_0(z) + \zeta p P_0'(z). \tag{6.4}$$

Elementary addition-, differentiation- and substitution operations lead to two very similar second-order differential equations in $P_0(z)$ respectively $P_1(z)$. For $|z| \le 1$,

$$\zeta(1-p-z)P_0''(z) - [\zeta + \lambda + \frac{\mu}{1-p} + \frac{\lambda(1-p-z)}{1-p}]P_0'(z) + \frac{\lambda^2}{\zeta(1-p)}P_0(z) = 0, \tag{6.5}$$

$$\zeta(1-p-z)P_1''(z) - [\zeta + \lambda + \frac{\mu}{1-p} + \frac{\lambda(1-p-z)}{1-p}]P_1'(z) + \frac{\lambda(\lambda+\zeta)}{\zeta(1-p)}P_1(z) = 0. \tag{6.6}$$

Putting

$$\xi \;=\; \frac{-\lambda(1-p-z)}{\zeta(1-p)}, \quad \text{and} \quad Q_i(\xi) \;=\; P_i(z), \quad i = 0, 1, \tag{6.7}$$

formulas (6.5) and (6.6) transform into the following differential equations, which are well-known to be differential equations for the confluent hypergeometric function ([8]):

$$\xi Q_0''(\xi) + (1 + \frac{\lambda}{\zeta} + \frac{\mu}{\zeta(1-p)} - \xi)Q_0'(\xi) - \frac{\lambda}{\zeta}Q_0(\xi) \;=\; 0,$$

$$\xi Q_1''(\xi) + (1 + \frac{\lambda}{\zeta} + \frac{\mu}{\zeta(1-p)} - \xi)Q_1'(\xi) - \frac{\lambda+\zeta}{\zeta}Q_1(\xi) \;=\; 0.$$

It now follows immediately from [8] that

$$Q_0(\xi) \;=\; D_0 \; {}_1F_1(\frac{\lambda}{\zeta}; 1 + \frac{\lambda}{\zeta} + \frac{\mu}{\zeta(1-p)}; \xi),$$

$$Q_1(\xi) \;=\; D_1 \; {}_1F_1(\frac{\lambda+\zeta}{\zeta}; 1 + \frac{\lambda}{\zeta} + \frac{\mu}{\zeta(1-p)}; \xi),$$

where the confluent hypergeometric function ${}_1F_1(a; b; c)$ is defined as follows:

$${}_1F_1(a; b; c) \;=\; \sum_{n=0}^{\infty} \frac{(a)_n c^n}{(b)_n n!},$$

with

$$(y)_n \;=\; y(y+1)\ldots(y+n-1), \quad (y)_0 \;=\; 1.$$

D_0 and D_1 are constants that are yet to be determined. A first equation for this purpose follows from the normalisation equation $P_0(1) + P_1(1) = 1$, or

$$Q_0(\frac{\lambda p}{\zeta(1-p)}) + Q_1(\frac{\lambda p}{\zeta(1-p)}) = 1. \tag{6.8}$$

The second equation is obtained from the following two flow balance equations. With Λ_i the total flow (including overflow traffic) through Q_i, $i = 0, 1$,

$$\Lambda_1 = \lambda + p\Lambda_0, \quad \Lambda_0 = \Lambda_1 - \mu \mathrm{P}(X_1 = 1). \tag{6.9}$$

Eliminating Λ_1, and then observing that in the infinite server queue Q_0 one has (using Little's formula) that $\mathrm{E}X_0 = \Lambda_0/\zeta$, it follows that

$$\mu \mathrm{P}(X_1 = 1) + \zeta(1-p)\mathrm{E}X_0 = \lambda. \tag{6.10}$$

Of course, in the above equation $\mathrm{P}(X_1 = 1)$ can be replaced by $\mathrm{E}X_1 = Q_1(\frac{\lambda p}{\zeta(1-p)})$. Further note that $\mathrm{E}X_0 = P_0'(1) + P_1'(1)$, so that (6.10) reads

$$\lambda Q_0'(\frac{\lambda p}{\zeta(1-p)}) + \mu Q_1(\frac{\lambda p}{\zeta(1-p)}) + \lambda Q_1'(\frac{\lambda p}{\zeta(1-p)}) = \lambda, \tag{6.11}$$

and D_0 and D_1 immediately follow from (6.8), (6.11).

In case $\zeta = \mu + \gamma$ and $p = \gamma/(\mu + \gamma)$, the case of repacking at rate γ as considered in Section 3.3, Equation (6.10) yields $\mathrm{E}X_1 + \mathrm{E}X_0 = \lambda/\mu$. Indeed, in this case the whole system behaves like an $M/M/\infty$ system with service rate μ. In that case, mean and variance of the occupancy of the overflow buffer are

$$\begin{aligned}
\mathrm{E}X_0 &= P_0'(1) + P_1'(1) = \frac{\lambda}{\mu} Q_0'(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}) + \frac{\lambda}{\mu} Q_1'(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}), \\
\mathrm{Var}(X_0) &= P_0''(1) + P_1''(1) + \mathrm{E}X_0 - (\mathrm{E}X_0)^2 \\
&= (\frac{\lambda}{\mu})^2 Q_0''(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}) + (\frac{\lambda}{\mu})^2 Q_1''(\frac{\lambda}{\mu}\frac{\gamma}{\mu+\gamma}) + \mathrm{E}X_0 - (\mathrm{E}X_0)^2.
\end{aligned}$$

Remark (Case $p = 0$). For $p = 0$ the overflow process from Q_1 is an IPP (Interrupted Poisson Process) with "on"-periods $\exp(\mu)$ and "off"-periods $\exp(\lambda)$. Kuczura [7] has analyzed an $\cdot/M/\infty$ queue with as input process an IPP. Our results for $p = 0$ reduce to those obtained by Kuczura.

Remark (Case $p = 1$). If $p = 1$ then (6.5) and (6.6) reduce to simple first-order differential equations that yield

$$P_0(z) = D(\frac{1}{\mu - \lambda z})^{\frac{\lambda}{\zeta}}, \quad P_1(z) = \lambda D(\frac{1}{\mu - \lambda z})^{\frac{\lambda+\zeta}{\zeta}},$$

D being determined via the normalisation equation. Note that this case $p = 1$ corresponds to a retrial queue, with retrial rate ζ.