Analysis of an $M/M/1$ Queue Using Fixed Order of Search for Arrivals and Service

Patrick Eschenfeldt  
peschenfeldt@hmc.edu  
Ben Gross  
bgross@hmc.edu  
Nicholas Pippenger  
njp@math.hmc.edu  

Department of Mathematics  
Harvey Mudd College  
1250 Dartmouth Avenue  
Claremont, CA 91711

Abstract: We analyze an $M/M/1$ queue with a service discipline in which customers, upon arriving when the server is busy, search a sequence of stations for a vacant station at which to wait, and in which the server, upon becoming free when one or more customers are waiting, searches the stations in the same order for a station occupied by a customer to serve. We show how to find complete asymptotic expansions for all the moments of the waiting time in the heavy traffic limit. We show in particular that the variance of the waiting time for this discipline is more similar to that of last-come-first-served (which has a pole of order three as the arrival rate approaches the service rate) than that of first-come-first-served (which has pole of order two).

Keywords: Queueing theory, Lambert series, asymptotic expansions.

Subject Classification: 60K26, 90B22
1. Introduction

We consider the $M/M/1$ queue (with independent exponentially distributed interarrival times, independent exponentially distributed service times, and a single server) with various service disciplines. We shall be interested mainly in the “heavy traffic” limit, \( \lambda \to 1 \), where \( \lambda \) is the arrival rate (measured in units of the service rate), and all asymptotic statements in this paper refer to this limit.

It is well known (see Little [L]) that the average waiting time \( \text{Ex}[W] \) (\( W \) is the length of interval from arrival to commencement of service) does not depend on the service discipline (the rule used to determine which waiting customer is served next when the server becomes free). We have

\[
\text{Ex}[W] = \frac{\lambda}{1 - \lambda} \sim \frac{1}{1 - \lambda}.
\]

The variance of \( W \), however, (and more generally its higher moments) does depend on the service discipline. Kingman [K] has shown that, among all service disciplines, “first-come-first-served” (FCFS) minimizes the variance of \( W \). We have

\[
\text{Var}[W_{\text{FCFS}}] = \frac{\lambda(2 - \lambda)}{(1 - \lambda)^2} \sim \frac{1}{(1 - \lambda)^2}
\]

(see for example Riordan [R2, pp.102–103]). Tambouratzis [T] has shown that “last-come-first-served” (LCFS) maximizes this variance. We have

\[
\text{Var}[W_{\text{LCFS}}] = \frac{\lambda(2 - \lambda + \lambda^2)}{(1 - \lambda)^3} \sim \frac{2}{(1 - \lambda)^3}
\]

(see for example Riordan [R2, pp. 106–109]; LCFS was first analyzed by Vaulot [V2]). We note that the difference between FCFS and LCFS is qualitative, in that \( \text{Var}[W_{\text{LCFS}}] \) has a pole of order three at \( \lambda = 1 \), whereas \( \text{Var}[W_{\text{FCFS}}] \) has a pole of order only two there.

Another service discipline that has been studied is “random-order-of-service” (ROS), first successfully analyzed by Vaulot [V1]. One of the motivations for studying ROS was stated by Riordan [R1]:

“In many switching systems it is not feasible to fully realize this ethical ideal of first come, first served, and it has long been of interest to determine delays on another basis. The contrasting assumption is of calls picked at random, which is again an idealization but in large offices appears to be called for, as a bound for the service actually given.”

This statement suggests that (1) in practical systems of that era (around 1953) it was not possible to keep track of the order of arrival, (2) ROS was analyzed as a substitute for the service discipline actually implemented, and (3) it was hoped that the performance of ROS would approximate that of the discipline actually implemented. For ROS, we have

\[
\text{Var}[W_{\text{ROS}}] = \frac{\lambda(4 - 2\lambda + \lambda^2)}{(2 - \lambda)(1 - \lambda)^2} \sim \frac{3}{(1 - \lambda)^2}
\]

(see for example Riordan [R2, pp. 103–106]). Comparing (1.4) with (1.2) and (1.3), we see that ROS is qualitatively “more like” FCFS than LCFS, in that \( \text{Var}[W_{\text{ROS}}] \) has a pole of order two rather than three, though its coefficient is larger than that of FCFS by a factor of three.
In the early 1950s, telephone switching systems were electromechanical and did not employ randomization beyond that present in the arrival and service processes. In this paper we shall analyze a service discipline that we shall call “fixed-order-of-search” (FOS). For this discipline, there is an infinite sequence \( W_1, W_2, \ldots \) of “waiting stations”, each of which can be either “vacant” or “occupied”. A customer arriving when the server is busy searches these stations in increasing order of their indices and occupies the first vacant station it finds. When the server becomes free and one or more customers are waiting, it searches the stations in the same order and serves the customer waiting at the first occupied station it finds, thereby vacating that station. This discipline was introduced by Eschenfeldt, Gross and Pippenger [E]. Like FCFS and LCFS (and unlike ROS), it does not employ randomization beyond that present in the arrival and service processes. If the server were to search the stations in the reverse order to that used by arriving customers (serving the customer waiting at the occupied station with the largest, rather than the smallest, index), the result would be LCFS, with its concomitant maximum variance for the waiting time. The choice of the same order of search for both customers and the server thus represents an attempt to improve upon LCFS, while still using a fixed order of search in each case.

We shall give an exact formula for \( \text{Var}[W_{FOS}] \):

\[
\text{Var}[W_{FOS}] = \frac{\lambda(6 + \lambda + \lambda^2)}{(1 - \lambda)^3} - \frac{4 \psi^{(1)}(1)}{(1 - \lambda) \log^2 \lambda},
\]

where \( \psi^{(1)}(x) \) is the \( q \)-trigamma function (defined below). We shall indicate how similar, but increasingly more complicated, formulas can be derived for the higher moments of \( W_{FOS} \). We shall also give an asymptotic formula for \( \text{Var}[W_{FOS}] \):

\[
\text{Var}[W_{FOS}] \sim \frac{8 - 4 \zeta(2)}{(1 - \lambda)^3},
\]

where \( \zeta(s) = \sum_{n \geq 1} 1/n^s \) is the Riemann zeta function and \( \zeta(2) = \pi^2 / 6 \) (see for example Whittaker and Watson [W2, pp. 265–280]). We shall also indicate how complete asymptotic expansions (with error terms of the form \( O((1 - \lambda)^R) \) for any \( R \)) can be derived for the variance of \( W_{FOS} \), as well as for the higher moments. Comparing (1.6) with (1.2) and (1.3), we see that FOS is qualitatively “more like” LCFS than FCFS, in that \( \text{Var}[W_{FOS}] \) has a pole of order three rather than two, though its coefficient is smaller than that of LCFS by a factor of \( 4 - 2 \zeta(2) = 4 - \pi^2 / 3 = 0.7101 \ldots \).

Eschenfeldt, Gross and Pippenger [E] initiated the study of FOS, determining the distribution of the index \( I \) of the station \( W_I \) at which a newly arriving customer waits (where \( I = 0 \) if the server is idle at the time of the arrival): we have \( \Pr[I \geq 0] = 1 \) and

\[
\Pr[I \geq i] = \frac{(1 - \lambda) \lambda^i}{1 - \lambda},
\]

for \( i \geq 1 \). The moments of \( I \) can be expressed in terms of the sums

\[
T_i(\lambda) = \sum_{i \geq 1} \frac{i^j \lambda^i}{1 - \lambda},
\]

for which we have the asymptotic formulas

\[
T_0(\lambda) \sim \frac{1}{1 - \lambda} \log \frac{1}{1 - \lambda},
\]
and

\[ T_l(\lambda) \sim \frac{l! \zeta(l+1)}{(1-\lambda)^{l+1}} \] (1.10)

for \( l \geq 1 \), and the exact formulas

\[ T_0(\lambda) = \frac{\psi(1) + \log(1-\lambda)}{\log \lambda} \] (1.11)

and

\[ T_l(\lambda) = \frac{\psi^{(l)}(1)}{\log^{l+1} \lambda} \] (1.12)

for \( l \geq 1 \). Here \( \psi_q(x) = \partial \log \Gamma_q(x)/\partial x \) is the \( q \)-digamma function, the logarithmic derivative of the \( q \)-gamma function \( \Gamma_q(x) = (1-q)^{1-x} \prod_{n \geq 0} ((1-q^n)+1)/(1-q^{n+x}) \) (see for example Gasper and Rahman [G, p. 16]), and \( \psi^{(l)}_q(x) = \partial^l \psi_q(x)/\partial x^l \) is the \( l \)-th \( q \)-polygamma function. (The sums \( T_l(\lambda) \), which are called Lambert series (see for example Hardy and Wright [H, p. 257]), are the generating functions \( T_l(\lambda) = \sigma_l(n) \lambda^n \) for the sums \( \sigma_l(n) = \sum_{d|n} d^l \) of the \( l \)-th powers of the divisors of of \( n \) (see for example Hardy and Wright [H, p. 239]).) In terms of the \( T_l(\lambda) \), we have

\[ \text{Ex}[I^m] = (1-\lambda) \sum_{0 \leq l \leq m-1} \binom{m}{l} (-1)^{m-1-l} T_l(\lambda). \]

In Section 2 we shall determine the moment generating \( M_W(s) \) function for \( W_{\text{FOS}} \) (which in what follows we shall denote simply \( W \)). In Section 3, we shall derive the exact formula (1.5) and the asymptotic formula (1.6). We shall also indicate how similar exact and asymptotic formulas can be found for the higher moments of \( W \). Finally, we shall indicate how these asymptotic formulas can be extended to complete asymptotic expansions (with error terms of the form \( O((1-\lambda)^R) \) for any \( R \)) for these quantities.

2. The Generating Functions

Consider the random process whose state variable \( J \) denotes the number of customers in the system. The random variable \( J \) is zero during an idle period (interval of time during which the server is idle). It is incremented whenever a customer arrives, and decremented whenever a customer departs (that is, at the termination of a service interval). Arrivals occur in a Poission process with rate \( \lambda \). During a busy period (interval of time during which \( J \geq 1 \)), departures occur in an independent Poisson process at rate 1. Thus, during a busy period, “transitions”, by which we mean arrivals and departures together, occur in a Poisson process at rate \( 1+\lambda \). Furthermore, during a busy period, the probability that the next transition will be an arrival is \( p = \lambda/(1+\lambda) \), and the probability that it will be a departure is \( q = 1/(1+\lambda) \). In this section we shall study the distribution of the random variable \( N \), defined as the number of transitions that occur between the arrival of a customer (excluded) and the departure that initiates its service interval (included). Specifically, we shall determine the probability generating function \( g(t) = \sum_{n \geq 0} \Pr[N = n] t^n \) for \( N \). We have \( N = 0 \) if the arrival initiates a busy period, and \( N \geq 1 \) if it occurs during a busy period.

The index \( J \) of the station \( W_t \) at which a newly arriving customer waits has the distribution given by (1.7). As a first step to determining \( g(t) \), we shall determine the conditional generating function \( g_i(t) = \sum_{n \geq 0} \Pr[N = n \mid I = i] t^n \). We begin with two special cases. If \( i = 0 \), the arrival initiates a busy period, so \( N = 0 \) and \( g_0(t) = 1 \). If \( i = 1 \), the customer waits at \( W_t \) and will be served as soon as the next departure occurs. The number of transitions preceding and including this next departure has a geometric distribution, and

\[ g_1(t) = \sum_{n \geq 1} p^{n-1} q t^n = qt/(1-qt). \]
For the general case, we consider the random process whose state variable \( K \) denotes the number of vacant stations among \( W_1, \ldots, W_i \). Since the customer in question waits at station \( W_i \), we have \( K = 0 \) immediately after the arrival of that customer. Furthermore, the first time thereafter at which \( K = i \) coincides with the beginning of the service interval for that customer, and thus occurs after exactly \( N \) transitions have occurred. When \( 1 \leq K \leq i - 1 \), the random variable \( K \) is incremented by a departure (because the next customer served will be waiting at one of the stations under consideration) and decremented by an arrival (because the arriving customer will wait at one of these stations). If however \( K = 0 \), a departure will increment \( K \), but an arrival will leave \( K \) unchanged (because it will have to wait at a station beyond \( W_i \)). Thus the process determining \( N \) given \( I = i \) is a random walk with one “reflecting barrier” (at \( K = 0 \)) and one “absorbing barrier” (at \( K = i \)), as shown in Figure 1.

![Figure 1](image)

Figure 1. The process determining \( N \), given that \( I = i \).

Non-terminal states are labeled with values of the random variable \( K \).

The number \( N \) of steps to absorption in this process has been studied by Weesakul [W1], who shows that the generating function is given by

\[
g_i(t) = \frac{q^i t^i (Q(t) - P(t))}{Q(t)^{i+1} - P(t)^{i+1} - pt(Q(t)^i - P(t)^i)},
\]

(2.1)

where

\[
P(t) = \frac{1 - \sqrt{1 - 4pqt^2}}{2}
\]

and

\[
Q(t) = \frac{1 + \sqrt{1 - 4pqt^2}}{2}.
\]

We note that for \( t = 1 \) we have \( P(1) = p, Q(1) = q \) and \( g_i(1) = 1 \).

We can now express the unconditional generating function \( g(t) \) by using summation by parts:

\[
g(t) = \sum_{i \geq 0} g_i(t) \Pr[I = i]
\]

\[
= \sum_{i \geq 0} g_i(t) (\Pr[I \geq i] - \Pr[I \geq i + 1])
\]

\[
= g_0(t) \Pr[I \geq 0] + \sum_{i \geq 1} (g_i(t) - g_{i-1}(t)) \Pr[I \geq i]
\]

\[
= 1 + \sum_{i \geq 1} (g_i(t) - g_{i-1}(t)) \Pr[I \geq i],
\]

because \( g_0(t) = \Pr[I \geq 0] = 1 \). Thus, using (1.7), we have

\[
g(t) = 1 + (1 - \lambda) \sum_{i \geq 1} \frac{(\nabla g)_i(t) \lambda^i}{1 - \lambda^i},
\]

(2.2)
where \((\nabla g)_i(t) = g_i(t) - g_{i-1}(t)\) denotes the backward difference of \(g_i(t)\). We note that for \(t = 1\) we have \((\nabla g)_i(1) = 0\), so \(g(1) = 1\).

We are now ready to drive the moment generating function \(M_W(s) = \mathbb{E}[e^{sW}]\) for \(W\). Each intertransition time \(X\) is exponentially distributed with mean \(1/(1 + \lambda)\), so the moment generating function for \(X\) is \(M_X(s) = \frac{1 + \lambda}{1 + \lambda - s}\). The moment generating function for the sum \(\sum_{1 \leq k \leq n} X_k\) of \(n\) independent intertransition times \(X_1, \ldots, X_n\) is \(M_X(s)^n = \left(\frac{(1 + \lambda)}{(1 + \lambda - s)}\right)^n\). Thus the waiting time \(W\), which is the sum of the random number \(N\) of independent intertransition times has the moment generating function

\[
M_W(s) = \sum_{n \geq 0} \Pr[N = n] M_X(s)^n \\
= g(M_X(s)) \\
= g \left( \frac{1 + \lambda}{1 + \lambda - s} \right). \tag{2.3}
\]

3. The Moments

In this section we shall derive the mean and variance for \(N\) and \(W\), and indicate how to derive the higher moments as well.

For the mean of \(N\), we use the formula \(\mathbb{E}[N] = g'(1)\). We have

\[
g'_i(1) = \frac{(1 + \lambda) (i(1 - \lambda) - \lambda (1 - \lambda^i))}{(1 - \lambda)^2},
\]

so

\[
(\nabla g')_i(1) = \frac{(1 + \lambda)(1 - \lambda^i)}{1 - \lambda}.
\]

Thus we have

\[
\mathbb{E}[N] = g'(1) \\
= (1 - \lambda) \sum_{i \geq 1} \frac{(\nabla g')_i(1) \lambda^i}{1 - \lambda^i} \\
= (1 - \lambda) \sum_{i \geq 1} \frac{(1 + \lambda)(1 - \lambda^i)}{1 - \lambda} \frac{\lambda^i}{1 - \lambda^i} \\
= \frac{(1 + \lambda) \lambda}{1 - \lambda} \frac{1}{1 + \lambda}. \tag{3.1}
\]

For the mean of \(W\), we use the formula for the expectation of the sum of a random number \(N\) of independent, identically distributed random variables \(X, X_1, X_2, \ldots\): \(\mathbb{E}[\sum_{1 \leq k \leq N} X_k] = \mathbb{E}[N] \mathbb{E}[X]\). Since the intertransition time \(X\) satisfies \(\mathbb{E}[X] = 1/(1 + \lambda)\), we have

\[
\mathbb{E}[W] = \mathbb{E}[N] \mathbb{E}[X] \\
= \frac{(1 + \lambda) \lambda}{1 - \lambda} \frac{1}{1 + \lambda} \\
= \frac{\lambda}{1 - \lambda},
\]

in accordance with (1.1).
For the variance of $N$, we begin by using the formula for the second factorial moment: $\text{Ex}[N(N-1)] = g''(1)$. We have

$$g''(1) = \frac{(1 - \lambda)^2 (1 + \lambda) i^2 - (1 - \lambda)(1 + \lambda)(1 - 10\lambda - 3\lambda^2) i}{(1 - \lambda)^4}$$

$$- \frac{(6\lambda(1 - \lambda)(1 + \lambda)^2 i^2 + 2\lambda(1 + \lambda)(1 + 4\lambda - \lambda^2) - 2\lambda^2 (1 + \lambda)^2 \lambda^i (1 - \lambda^i)}{(1 - \lambda)^4},$$

so

$$(\nabla g'')(1) = \frac{2(1 - \lambda)(1 + \lambda)^3 \lambda^i + 6i(1 - \lambda)^2 (1 + \lambda)^2 - 2(1 - \lambda)^3 (1 + \lambda) (1 - \lambda^i) - 4i(1 - \lambda)^2 (1 + \lambda)^2}{(1 - \lambda)^4}.$$

Thus we have

$$\text{Ex}[N(N - 1)] = \frac{2(1 + \lambda)^3}{(1 - \lambda)^2} \sum_{i \geq 1} \lambda^{2i} + \frac{6i(1 + \lambda)^2}{1 - \lambda} \sum_{i \geq 1} i \lambda^i - 2(1 + \lambda) \sum_{i \geq 1} \lambda^i - \frac{4(1 + \lambda)^2}{1 - \lambda} \sum_{i \geq 1} \frac{1}{1 - \lambda^i}$$

$$\begin{align*}
&= \frac{2\lambda^2 (1 + \lambda)^2}{(1 - \lambda)^3} + \frac{6\lambda(1 + \lambda)^2}{(1 - \lambda)^3} - \frac{2\lambda (1 + \lambda)}{1 - \lambda} - \frac{4(1 + \lambda)^2}{1 - \lambda} T_1(\lambda),
\end{align*}$$

where we have used the definition (1.8) to evaluate the last sum. It follows that

$$\text{Var}[N] = \text{Ex}[N(N - 1)] + \text{Ex}[N] - \text{Ex}[N]^2$$

$$= \frac{2\lambda^2 (1 + \lambda)^2}{(1 - \lambda)^3} + \frac{6\lambda(1 + \lambda)^2}{(1 - \lambda)^3} - \frac{\lambda (1 + \lambda)}{1 - \lambda} - \frac{\lambda^2 (1 + \lambda)^2}{(1 - \lambda)^2} - \frac{4(1 + \lambda)^2}{1 - \lambda} T_1(\lambda),$$

$$= \frac{\lambda(1 + \lambda)^2 (6 + \lambda + \lambda^2)}{(1 - \lambda)^3} - \frac{\lambda (1 + \lambda)}{1 - \lambda} - \frac{4(1 + \lambda)^2}{1 - \lambda} T_1(\lambda),$$

where we have used (3.1), then combined the first, second and fourth terms. For the variance of $W$, we use the formula for the variance of a random number $N$ of independent, identically distributed random variables $X, X_1, X_2, \ldots$: $\text{Var} \left[ \sum_{1 \leq k \leq N} X_k \right] = \text{Var}[N] \text{Ex}[X]^2 + \text{Ex}[N] \text{Var}[X]$. Since the intertransition time $X$ satisfies $\text{Ex}[X] = 1/(1 + \lambda)$ and $\text{Var}[X] = 1/(1 + \lambda)^2$, we have

$$\text{Var}[W] = \text{Var}[N] \text{Ex}[X]^2 + \text{Ex}[N] \text{Var}[X]$$

$$= \left( \frac{\lambda(1 + \lambda)^2}{(1 - \lambda)^3} - \frac{\lambda (1 + \lambda)}{1 - \lambda} - \frac{4(1 + \lambda)^2}{1 - \lambda} T_1(\lambda) \right) \frac{1}{(1 + \lambda)^2} + \frac{\lambda (1 + \lambda)}{1 - \lambda} \frac{1}{(1 + \lambda)^2}$$

$$= \frac{\lambda(6 + \lambda + \lambda^2)}{(1 - \lambda)^3} - \frac{4}{1 - \lambda} T_1(\lambda).$$

Evaluating $T_1(\lambda)$ using (1.12) yields

$$\text{Var}[W] = \frac{\lambda(6 + \lambda + \lambda^2)}{(1 - \lambda)^3} - \frac{4 \psi^{(1)}(1)}{(1 - \lambda) \log^2 \lambda},$$

confirming (1.5), whereas using (1.10) yields

$$\text{Var}[W] \sim \frac{\lambda(6 + \lambda + \lambda^2)}{(1 - \lambda)^3} - \frac{4 \zeta(2)}{(1 - \lambda)^3},$$
confirming (1.6).

It is straightforward to generalize the derivations of the mean and variance of \( W \) given above to the higher moments. We begin by indicating how to derive the higher factorial moments of \( l \) for \( \geq 1 \), we have

\[
\text{Ex}[N(N-1) \cdots (N-m+1)] = g^{(m)}(1) = (1-\lambda) \sum_{i \geq 1} \frac{(\nabla g^{(m)})_i(1) \lambda^i}{1-\lambda^i}.
\]

After differentiating \( g_i(t) \) with respect to \( t \) \( (m \) times), then evaluating the result at \( t = 1 \), and finally differencing with respect to \( i \), the result is a bivariate polynomial \( P(i, u) \) (with coefficients that are rational functions of \( \lambda \) in the variables \( i \) and \( u = \lambda^i \). Dividing this polynomial by \( 1-u = 1-\lambda^i \), we obtain \( P(i, u) = Q(i, u)(1-u) + R(i) \), with quotient \( Q(i, u) \) and remainder \( R(i) \). We then have

\[
\text{Ex}[N(N-1) \cdots (N-m+1)] = (1-\lambda) \sum_{i \geq 1} Q(i, \lambda^i) \lambda^i + (1-\lambda) \sum_{i \geq 1} R(i) \lambda^i.
\]

The first sum in (3.2) can be expressed as a linear combination (with coefficients that are rational functions of \( \lambda \) of sums of the form

\[
S_{l,k}(\lambda) = \sum_{i \geq 1} i^l \lambda^{ki}.
\]

These sums are themselves rational functions of \( \lambda \):

\[
S_{l,k}(\lambda) = \frac{A_l(\lambda^k)}{(1-\lambda^k)^{l+1}},
\]

where \( A_l(x) = \sum_{0 \leq k \leq l} a(l, k) x^k \) is the \( l \)-th Eulerian polynomial and the \( a(l, k) \) are the Eulerian numbers, with generating function \( \sum_{l,k \geq 0} a(l, k) z^l y^l/l! = z(1-z)/(e^{y(1-z)} - z) \) (see Comtet \([C, p. 245]\)). The second sum in (3.2) can be expressed as a linear combination (again with coefficients that are rational functions of \( \lambda \) of the sums \( T_l(\lambda) \) given by (1.8), with asymptotic formulas given by (1.9) and (1.10), and with exact formulas given by (1.11) and (1.12).

We are now ready to obtain the moments of \( W \). Differentiating the identity (2.3) \( m \) times, we obtain

\[
M_{W}^{(m)}(s) = F_m(M_X^{(1)}(s), \ldots, M_X^{(m)}(s); g^{(1)}(M_X(s)), \ldots, g^{(m)}(M_X(s))),
\]

where \( F_m(x_1, \ldots, x_m; y_1, \ldots, y_m) \) is the polynomial

\[
F_m(x_1, \ldots, x_m; y_1, \ldots, y_m) = m! \sum_{1 \leq l \leq m} x_l \sum_{e_1, e_2, \ldots, e_m \geq 1} \prod_{1 \leq k \leq m} \left( \frac{y_k}{k!} \right)^{e_k},
\]

and the inner sum is over all \( e_1, e_2, \ldots, e_m \) such that \( e_1 + e_2 + \cdots + e_m = l \) and \( e_1 + 2 e_2 + \cdots + m e_m = m \) (see for example Comtet \([C, p. 137]\]). Evaluating at \( s = 0 \) and using \( M_X(0) = 1 \) and \( M_X^{(0)}(0) = 1/(1+\lambda)^l \) for \( l \geq 1 \), we have

\[
M_{W}^{(m)}(0) = F_m \left( g^{(1)}(1), \ldots, g^{(m)}(1); \frac{1}{1+\lambda}, \ldots, \frac{1}{(1+\lambda)^m} \right)
\]

\[
= \frac{1}{(1+\lambda)^m} \sum_{1 \leq l \leq m} g^{(l)}(1) \sum_{e_1, e_2, \ldots, e_m} \left( \frac{m}{l} \right)_{e_1, e_2, \ldots, e_m}
\]

\[
= \frac{1}{(1+\lambda)^m} \sum_{1 \leq l \leq m} g^{(l)}(1) \left\{ \frac{m}{l} \right\},
\]
where the \( \{ \frac{m}{l} \} \) are the Stirling numbers of the second kind, with the generating function
\[
\sum_{m \geq l \geq 0} \{ \frac{m}{l} \} y^l z^m / l! = e^{y(e^z-1)}
\]
(see for example Comtet [C, pp. 206–207]). Thus we have
\[
\text{Ex}[W^m] = M_W^{(m)}(0)
\]
\[
= \frac{1}{(1 + \lambda)^m} \sum_{1 \leq l \leq m} \{ \frac{m}{l} \} g^{(l)}(1),
\]
where the \( g^{(l)}(1) \) are the factorial moments of \( N \) determined in the preceding paragraph.

The asymptotic formulas given above for the moments of \( W \) can be extended to complete asymptotic expansions, with error terms of the form \( O((1 - \lambda)^R) \) for any \( R \). Any rational function of \( \lambda \) has a Laurent series around \( \lambda = 1 \), which will serve an an asymptotic expansion as \( \lambda \to 1 \) as well. Thus the only remaining problem is to find asymptotic expansions for the sums \( T_l(\lambda) \). These expansions have been given by Eschenfeldt, Gross and Pippenger [E]. We have
\[
T_0(\lambda) \sim \frac{1}{h} \log \frac{1}{1 - \lambda} + \frac{\gamma}{h} + \sum_{r \geq 0} \frac{(-1)^r B_{r+1}(B_{r+1} - (-1)^{r+1}) h^r}{(r+1)(r+1)!},
\]
where \( \gamma = 0.5772 \ldots \) is Euler’s constant, \( B_r \) is the \( r \)-th Bernoulli number, defined by \( t/(e^t - 1) = \sum_{r \geq 0} B_r t^r / r! \) (see for example Roman [R3, p. 94], and \( h = -\log \lambda \) has the expansion
\[
h = -\log(1 - (1 - \lambda)) = \sum_{r \geq 1} \frac{(1 - \lambda)^r}{r},
\]
so that its reciprocal has the expansion
\[
\frac{1}{h} = \frac{1}{1 - \lambda} \sum_{r \geq 0} \frac{(-1)^r C_r (1 - \lambda)^r}{r!},
\]
where \( C_r \) is the \( r \)-th Bernoulli number of the second kind, defined by \( t / \log(1 + t) = \sum_{r \geq 0} C_r t^r / r! \) (see for example Roman [R3, p. 116]). (These numbers are also called the Cauchy numbers of the first kind, and are given by \( C_r = \int_0^1 x(x-1) \cdots (x-r+1) \, dx \); see for example Comtet [C, pp. 293–294].) For \( l \geq 1 \), we have
\[
T_l(\lambda) \sim \sum_{r \geq 0} \frac{(-1)^{r+l-1} B_r B_{r+l} h^{r-l}}{r! (r+l)},
\]
We note that, if \( l \) is odd, then this expansion has only finitely many terms (because \( B_r = 0 \) for odd \( r \geq 3 \)).

5. Acknowledgment

The research reported here was supported by Grant CCF 0917026 from the National Science Foundation.
6. References

[C] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansion*, D. Reidel Publishing Co., Dortrecht, 1974.

[E] P. Eschenfeldt, B. Gross and N. Pippenger, “Stochastic Service Systems, Random Interval Graphs and Search Algorithms”, arXiv:1107.4113v2 [math.PR].

[G] G. Gasper and M. Rahman, *Basic Hypergeometric Series*, Cambridge University Press, Cambridge, 1990.

[H] G. H. Hardy and E. M. Wright, *Introduction to the Theory of Numbers (5th edition)*, Clarendon Press, Oxford, 1979.

[K] J. F. C. Kingman, “The Effect of Queue Discipline on Waiting Time Variance”, *Math. Proc. Cambridge Phil. Soc.*, 58:1 (1962) 163–164.

[L] J. D. C. Little, “A Proof for the Queuing Formula: \( L = \lambda W \)”, *Oper. Res.*, 9 (1961) 383–387.

[R1] J. Riordan, “Delay Curves for Calls Served at Random”, *Bell System Technical Journal*, 32 (1953) 100-119.

[R2] J. Riordan, *Stochastic Service Systems*, John Wiley and Sons, New York, 1962.

[R2] S. Roman, *The Umbral Calculus*, Academic Press, New York, 1984.

[T] D. G. Tambouratzis, “On a Property of the Variance of the Waiting Time of a Queue”, *J. Appl. Prob.*, 5:3 (1968) 702–703.

[V1] É. Vaulot, “Délais d’attente des appels téléphoniques traités au hasard”, *Comptes Rendus Acad. Sci. Paris*, 222 (1946) 268–269.

[V2] É. Vaulot, “Délais d’attente des appels téléphoniques dans l’ordre inverse de leurs arrivée”, *Comptes Rendus Acad. Sci. Paris*, 238 (1954) 1188–1189.

[W1] B. Weesakul, “The Random Walk between a Reflecting and an Absorbing Barrier”, *Ann. Math. Statist.*, 32:3 (1961) 765–769.

[W2] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis (4th edition)*, Cambridge University Press, London, 1927.