M/M/C/K/N systems: part II

Lecturer: Dmitri A. Moltchanov
E-mail: moltchan@cs.tut.fi

OUTLINE:

- M/M/1 queuing system with state-dependent arrivals;
- M/M/m queuing system;
- M/M/m/m queuing system;
- M/M/1/K queuing system;
- M/M/1/$\infty$/K queuing system;
- M/M/$\infty$/-/K queuing system.
- Service time variation in M/-/C/K/N systems.
1. M/M/1 with state dependent arrivals

What we assume here:

- rate of the arrival is slowing down when the state goes up;
- where it may occur: some feedback mechanism controlling arrivals.

Parameters are given as follows:

\[
\begin{align*}
\lambda_k &= \frac{\lambda}{k + 1}, & k &= 0, 1, \ldots, \\
\mu_k &= \mu, & k &= 1, 2, \ldots, \\
\mu_0 &= 0.
\end{align*}
\]

Figure 1: Birth-death process of M/M/1 queuing system with state dependent arrivals.
1.1. Interesting notes

What is interesting about this queue:

• PASTA property does not hold for such queuing system:
  – arrival process is not **homogenous** Poisson!

• the mean arrival rate to the system is given by follows:

\[
E[\lambda] = \sum_{k=0}^{\infty} \lambda_k p_k = \sum_{k=0}^{\infty} \frac{\lambda}{k+1} p_k.
\]

  – depends on the state of the system;
  – we do not know mean arrival rate in advance;
  – Little’s result can still be applied with proper \( E[\lambda] \)!

What else we can define:

• state-dependent service times:
  – intensity of the service gets smaller when more customers in the system.
1.2. Steady-state distribution

Existence of steady-state distribution (capacity is infinite!):

- $E[\lambda]/\mu$ is always limited irrespective of initial $\lambda$

\[
\frac{E[\lambda]}{\mu} < \infty. \quad (3)
\]

- even when $\lambda >>> \mu$!

Solution for steady-state probabilities:

\[
p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} = p_0 \prod_{i=0}^{k-1} \frac{\lambda}{(i + 1) \mu} = p_0 \left( \frac{\lambda}{\mu} \right)^k \frac{1}{k!}, \quad k = 1, 2, \ldots, \quad (4)
\]

- where we can find $p_0$ from normalizing condition:

\[
p_0 = 1 - \sum_{i=1}^{\infty} p_k = 1 - \sum_{i=1}^{\infty} p_0 \left( \frac{\lambda}{\mu} \right)^k \frac{1}{k!} = e^{-(\lambda/\mu)}. \quad (5)
\]

- $\lambda$ is the arrival rate that decreases as the state number increases.
1.3. Mean number of customers

Mean number of customers in the system: $E[N]$

$$E[N] = \sum_{k=0}^{\infty} kp_k = \frac{\lambda}{\mu}. \quad (6)$$

Arrival rate to the system:

$$E[\lambda] = \sum_{k=0}^{\infty} \lambda kp_k = \mu(1 - e^{-\lambda/\mu}). \quad (7)$$

Mean time spent by customer in the system: $E[W]$

- apply Little’s result to get:

$$E[W] = \frac{E[N]}{E[\lambda]} = \frac{\lambda}{\mu^2(1 - e^{-\lambda/\mu})}. \quad (8)$$
2. M/M/m queuing system

Note the following:

- the system has $m$ servers each of which requires $1/\mu$ time to serve customer;
- the birth-death process has the following parameters:

\[
\begin{align*}
\lambda_k &= \lambda, & k &= 0, 1, \ldots, \\
\mu_k &= k\mu, & k &= 1, 2, \ldots, m, \\
\mu_k &= m\mu, & k &= m + 1, \ldots.
\end{align*}
\]  

Note: steady-state distribution exists when $\rho = \lambda/(m\mu) < 1$.

![Birth-death process associated with M/M/m queuing system.](image)

Figure 2: Birth-death process associated with M/M/m queuing system.
2.1. Steady-state distribution

The equilibrium state distribution is given by product form solution:

\[ p_k = p_0 \frac{\rho^k}{k!}, \quad k = 1, 2, \ldots, m, \]

\[ p_k = p_0 \frac{\rho^k}{m!m^{k-m}}, \quad k = m + 1, \ldots, \]

(10)

where \( p_0 \) can be derived from normalizing condition:

\[ p_0 = \left( \sum_{k=0}^{m-1} \frac{\rho^k}{k!} + \frac{m\rho^m}{m!(m-\rho)} \right)^{-1}. \]

(11)

Note the following:

- M/M/m queues was (is) extensively used in teletraffic theory;
- example: waiting time to get the free line.
2.2. Performance measures

The most important measure:

- probability that arriving customer will wait which is given by:

\[
C(m, \rho) = \sum_{k=m}^{\infty} p_k = \frac{m\rho^m}{m!(m-\rho)} \sum_{k=0}^{m-1} \frac{\rho^k}{k!} + \frac{m\rho^m}{m!(m-\rho)}.
\] (12)

- this formula is extensively tabulated in literature.

Mean number of customers in the buffer: \(E[N_Q]\)

- we obtain it directly from steady-state distribution:

\[
E[N_Q] = \sum_{k=m}^{\infty} kp_k = \frac{p_m}{1-\rho} \sum_{k=0}^{\infty} k(1-\rho)\rho^n = C(m, \rho) \frac{\rho}{1-\rho}.
\] (13)

Mean waiting time in the buffer: \(E[W_Q]\)

- using Little’s result we get:

\[
E[W_Q] = \frac{E[N_Q]}{\lambda} = C(m, \rho) \frac{\rho}{(1-\rho)\lambda}.
\] (14)
3. M/M/m/m queuing system

Note the following:

- there are no waiting positions in this queue;
- capacity of the system is limited and equal to the number of servers.

Parameters of birth-death process are as follows:

\[
\begin{align*}
\lambda_k &= \lambda, & k &= 0, 1, \ldots, m - 1 \\
\lambda_k &= 0, & k &\geq m - 1 \\
\mu_k &= k\mu, & k &= 1, 2, \ldots, m, \\
\mu_k &= 0, & k &> m.
\end{align*}
\]

Figure 3: Birth-death process associated with M/M/m/m queuing system.
3.1. Steady-state distribution

Note the following:

• customers enters the queue if at least one position is free!

• otherwise, the customer is immediately rejected: loss system!

Condition of existence of steady-state distribution:

• state-space is limited;

• system is always stable irrespective of \( \lambda/m\mu \)!

Defining \( \rho = \lambda/\mu \), the steady-state distribution is given by:

\[
p_k = p_0 \frac{\rho^k}{k!}, \quad k = 1, 2, \ldots, m, \\
p_k = 0, \quad k = m + 1, \ldots,
\]  

(16)

• where from the normalizing conditions we get \( p_0 \):

\[
p_0 = \left( \sum_{k=0}^{m} \frac{\rho^k}{k!} \right)^{-1}.
\]  

(17)
3.2. Performance measures

Application of the M/M/m/m queue:

- call service process between telephone exchanges.

![Diagram of M/M/m/m queue](image_url)

Figure 4: Using M/M/m/m as a model in telephone network.

**Most important parameter:** probability of blocking:

- probability that an arrival finds all server busy forcing customer to leave without service.

**Note:** blocking is often used to describe the situation when the system is full:

- we say this system is with blocking;
- we say customer is blocked, etc.
To derive probability of blocking:

- arrival process is Poisson: PASTA property holds;
- probability that arrival sees all servers busy = fraction of time all servers busy:

\[
B(m, \rho) = \frac{\rho^m}{m! \sum_{k=0}^{m} \frac{\rho^k}{k!}},
\]

which is known as **Erlang-C** formula.

Using the following recursion it is easy to estimate probabilities \( B(m, \rho) \):

\[
B(0, \rho) = 1, \quad B(m, \rho) = \frac{\rho B(m-1, \rho)}{1 + \frac{\rho B(m-1, \rho)}{m}}.
\]

Other performance parameters:

- can be obtained using Little’s result (first, you have to get \( E[N] \));
- note that the arrival rate at which customers enter the queue is:

\[
\lambda_A = \lambda(1 - B(m, \rho)).
\]
4. M/M/1/K queuing system

Note the following:

- frequently used in performance evaluation of packet networks;
- birth-death process associated with M/M/1/K has the following parameters:

\[
\begin{align*}
\lambda_k &= \lambda, & k &= 0, 1, \ldots, K - 1 \\
\lambda_k &= 0, & k &\geq K \\
\mu_k &= \mu, & k &= 1, 2, \ldots, K, \\
\mu_k &= 0, & k &> k.
\end{align*}
\]

(21)

Figure 5: Birth-death process associated with M/M/1/K queuing system.
4.1. Steady-state distribution

Existence of steady-state distribution:

- state space of the system is limited: \( \{0, 1, \ldots, K\} \);
- steady-state distribution exist for all \( \lambda \) and \( \mu \) irrespective of \( \lambda/\mu \)!

Steady-state distribution is given by:

\[
p_k = p_0 \rho^k, \quad k = 1, 2, \ldots, K, \tag{22}
\]
\[
p_k = 0, \quad k = K + 1, \ldots, \tag{23}
\]

where \( p_0 \) can be found from the normalizing condition:

\[
p_0 = \frac{(1 - \rho)}{(1 - \rho^{K+1})} \tag{24}
\]

Note: from equilibrium state distribution all mean parameters can be found.
5. M/M/1/$\infty$/K queuing system

Note the following:

- the first case when we consider queuing system with finite population;
- parameters of birth-death process are as follows:

\[
\begin{align*}
\lambda_k &= \lambda(K - k), & k &= 0, 1, \ldots, K - 1 \\
\lambda_k &= 0, & k &\geq K \\
\mu_k &= \mu, & k &= 1, 2, \ldots, K, \\
\mu_k &= 0, & k &> k.
\end{align*}
\]

(25)

Figure 6: Birth-death process associated with M/M/1/$\infty$/K queuing system.
5.1. Steady-state distribution

**Existence of steady-state distribution:**

- state space of the system is limited: \{0, 1, …, K\};
- arrival rate decreases when state increases;
- steady-state distribution exist for all \( \lambda \) and \( \mu \) irrespective of \( \lambda/\mu \)!

**Defining** \( \rho = \lambda/\mu \) **the steady-state distribution is:**

\[
p_k = p_0 \rho^k \frac{K!}{(K - k)!}, \quad k = 1, 2, \ldots, K,
\]

\[
p_k = 0, \quad k > K,
\]

- where \( p_0 \) can be found from normalizing condition:

\[
p_0 = \left( \sum_{k=0}^{K} \rho^k \frac{K!}{(K - k)!} \right)^{-1}
\]

**Note:** performance parameters can be found immediately.
6. M/M/∞/-/K

Note the following: number of servers need not be more $K$

- limited population of customers.

Parameters of birth-death process are given:

\[
\lambda_k = \lambda (K - k), \quad k = 0, 1, \ldots, K - 1 \\
\lambda_k = 0, \quad k \geq K \\
\mu_k = k \mu, \quad k = 1, 2, \ldots, K, \\
\mu_k = 0, \quad k > k.
\]

(28)

Figure 7: Birth-death process associated with M/M/∞/-/K queuing system.
6.1. Steady-state distribution

**Existence of steady-state distribution:**

- state space of the system is limited: \( \{0, 1, \ldots, K\} \);
- steady-state distribution exist for all \( \lambda \) and \( \mu \) irrespective of \( \lambda/\mu \)!

**Defining \( \rho = \lambda/\mu \), steady-state distribution is given by:**

\[
p_k = p_0 \rho^k \frac{K!}{k!(K-k)!}, \quad k = 1, 2, \ldots, K, \]
\[
p_k = 0, \quad k > K,
\]

where \( p_0 \) can be found from normalizing condition:

\[
p_0 = \left( \sum_{k=0}^{K} \rho^k \frac{K!}{k!(K-k)!} \right)^{-1}
\]  

(29)

The mean arrival rate to the system is:

\[
E[\lambda] = \sum_{k=0}^{K} \lambda_k p_k = K \lambda \frac{1}{1 + \rho}
\]  

(31)
6.2. Performance measures

Mean number of customers in the system: $E[N]$

$$E[N] = \frac{K \rho}{1 + \rho}. \quad (32)$$

Mean number of customers in the buffer: $E[N_Q]$

$$E[N_Q] = 0. \quad (33)$$

- since the number of servers is infinite.

Mean time spent by customer in the system: $E[W]$

$$E[W] = \frac{E[N]}{E[\lambda]} = \frac{1}{\mu}. \quad (34)$$

- since any arriving customer immediately enters the server.

Mean time spent by customer in the buffer – $E[W_b]$

$$E[W_b] = 0. \quad (35)$$

- since the number of servers is infinite.
7. Service time variation in M/-/C/K/N system

Why service time variations:

• exponential distribution gives poor approximation;

• which other distribution we can use:
  – Erlang distribution;
  – hyperexponential distribution;
  – Cox distribution;
  – phase type distribution.

How we can analyze these systems:

• analyzing M/G/- queuing system;
  – complete lack of memoryless property in service process.

• analyzing M/PH/- queuing systems and its special cases:
  – we may still benefit from memoryless property of components.
7.1. Method of stages for $M/E_r/1$

When this method is useful:

- service time is not exponentially distributed;
- service time is a combination of exponentials.

Consider the example of queuing system:

- Kendall’s notation: $M/E_r/1$;
- Poisson arrivals, single server, infinite waiting room, Erlang service times.

![Illustration of the queue of $M/E_r/1$ type.](image)

Figure 8: Illustration of the queue of $M/E_r/1$ type.
How we can represent the service in this system:

- customer first starts getting service at stage 1:
  - it is served for exponentially distributed time with mean $1/2\mu$.

- after completion it enters the stage 2:
  - it is served for exponentially distributed time with mean $1/2\mu$.

- the service time is the sum of two exponentials:
  - the result: Erlang distribution with mean $1/\mu$.

![Diagram of server with two stages](image)

Figure 9: Illustration of server with two stages.

Note: new customer enters the service after full service completion.
7.2. State of the system

State of the system:

- pair \((n, j)\):
  - \(n\): total number of customers in the system;
  - \(j\): stage at which the current customer is served.

Do the following:

- \(S_S = \{0, 1, \ldots\}\): number of customers in the system;
- \(S_B = \{1, 2\}\): phase of the service process;
- state space of the system is given by Cartesian product:
  \[ S_S \times S_B = \{0, 1, \ldots\} \times \{1, 2\}, \tag{36} \]
7.3. Example: generalized Erlang service times

Assume: $\mu_1 \neq \mu_2$: generalized Erlang distribution!

![Transition diagram of M/E_r/1 system with state description given by $(n,j)$](image)

Figure 10: Transition diagram of M/E_r/1 system with state description given by $(n,j)$.

Write balance equations for this diagram as:

\[ \lambda p_{00} = p_{12} \mu_2, \]
\[ p_{11}(\lambda + \mu_1) = p_{00} \lambda + p_{22} \mu_2, \]
\[ p_{12}(\lambda + \mu_2) = p_{11} \mu_1, \]
\[ p_{21}(\lambda + \mu_1) = p_{11} \lambda + p_{32} \mu_2, \]
\[ \ldots \]

(37)
Rewrite it as follows:

\[ p_{12} = \frac{\lambda}{\mu_2} p_{00}, \]
\[ p_{11} = \frac{\lambda(\lambda + \mu_2)}{\mu_1 \mu_2} p_{00}, \]

\[ \ldots \]

(38)

Analyze as follows:

- express all \( p_{ii}, i = 1, 2, \ldots \) as a function of \( p_{00} \);
- use normalizing condition to find \( p_{00} \).

This approach can be extended to:

- \( r \) stages of service time;
- finite number of waiting positions in the buffer;
- more general service time: hyperexponential, Cox, phase-type.
7.4. Alternative state description

What is bad about \((n, j)\) description:

- \((n, j)\) description is two-dimensional;
- when \(n\) or \(j\) are large it is getting complicated!

Another state description:

- number of uncompleted phases of work in the system:
  - each waiting customer adds \(r\) phases of work;
  - served customer adds \(j \leq r\) phases of work:

- there is one-to-one correspondence between \((n, j)\) and uncompleted phases of work:

\[
(n, j) = (n - 1)r + j,
\]

- \(j\) is the number of phases left for customer currently being served;
- \(r\) is the total number of phases in service time;
- \(n\) is the number of customers in the system.
7.5. Example: Erlang service times

**Assume:** \( \mu_1 = \mu_2 = \mu \): Erlang distribution.

![Transition diagram of M/E\(_r\)/1 system with a single state descriptor.](image)

**Note:** any arrival adds \( r \) phases of work to the system.

**Denote:** \( p_n \) be the steady-state probability that \( n \) phases of work are in the system:

- use global balance principle to get:

\[
\begin{align*}
    p_0 \lambda &= p_1 \mu, \\
    p_n (\lambda + \mu) &= p_{n+1} \mu, & n &= 1, 2, \ldots, r - 1, \\
    p_n (\lambda + \mu) &= p_{n-r} \lambda + p_{n+1} \mu, & n &= r, r + 1, r + 2, \ldots.
\end{align*}
\]
Do not forget normalizing condition:

\[ \sum_{n=0}^{\infty} p_n = 1, \quad (41) \]

How to get the solution:

- one can solve equations to get uncompleted phases of work;
- we do not directly get any interesting quantity;
- after obtaining steady-state distribution we have to find distribution of states.