

The probability that an arriving customer has to wait on arrival

$$= \sum_{n \geq c} P_n = \frac{\rho_c}{1-\rho} \quad \left(\text{Erg. } C(c, \frac{\lambda}{\mu}) \right)$$

This is known as Erlang's C formula.

$C(c, \frac{\lambda}{\mu})$: fraction of time all the c servers are busy.

Let N_q denote the number of customers in the queue.

$$P[N_q = j, W > 0] = P[N = c+j] = \rho^j P_c \\ = \rho^j (1-\rho) C(c, \frac{\lambda}{\mu})$$



Other than the steady-state probability we can get some more [Indiscernible] [00:00:03]. The first one is in the probability that the arriving customer has to wait on arrival. What is the probability that the arriving customer has to wait on arrival. So that means the number of customers in the system is a greater or equal to c then only the customer has to wait.

So the probability you had the probability of P_c sorry P_n sorry here is some mistake P suffix n where n is running from c to infinity if you had all those probabilities that is going to be P_c divided by $1 - \rho$ and this probability is known as a Erlang's C formula for a multi server infinite capacity model that I am denoting with the letter c of c , λ by μ because you need a number of servers in the system and you need a λ as well as μ . If I know this quantity I can find out what is a Erlang's C formula. This is very important formula using that you can find out what is the optimal c such a way that and the probability has to be minimum.

You can find out what is optimal number of servers is needed to have a some upper bound probability of arriving customer has to wait. Therefore this Erlang's C formula is very useful in performance analysis of any system.

Thus,

$$P\{N_q = j_0 / W > 0\} = \frac{P\{N_q = j, W > 0\}}{P\{W > 0\}}$$
$$= (1-\rho) \rho^j, \quad j = 0, 1, \dots$$

Expected number of busy servers

$$E(B) = \sum_{n=0}^{c-1} n P_n + \sum_{n=c}^{\infty} c P_n = c \rho$$

Expected number of idle servers

$$E(I) = E(c - B) = c - c \rho = c(1 - \rho)$$



The next quantity is N_q denotes the number of customers in the queue. So either I use the letter N and suffix q earlier I use the letter Q itself. So for that I am finding the joint distribution of what is the probability that the number of customers in the queue is j and the waiting time is going to be greater than 0. W is used for the waiting time. So the waiting time is going to be greater than 0. That is same as the number of customers in the system that is a c plus j. What is a probability that j customers in the queue as well as the waiting time is greater than 0 that is same as what is a probability that c plus j customers in the system. You do the little simplification so you will get this joint probability in terms of Erlang's C formula. So using that I am finding the conditional probability what is the conditional probability that j customers in the queue given that the waiting time is greater than 0. If I do little simplification I will get a 1 minus Rho times Rho power j where Rho is lambda divided by cmu. This is nothing but the probability mass function of a geometric distribution. This is the probability mass function of a geometric distribution therefore this conditional probability is geometrically distributed with the parameter Rho.


Expected number in the system

$$E(N) = E(B) + E(Q)$$

$$E(Q) = \sum_{n=c}^{\infty} (n-c) P_n$$

$$= \sum_{n=c}^{\infty} (n-c) \frac{(\frac{\lambda}{\mu})^n}{c! c^{n-c}} P_0$$

$$= \frac{\rho}{1-\rho} c(c, \frac{\lambda}{\mu})$$

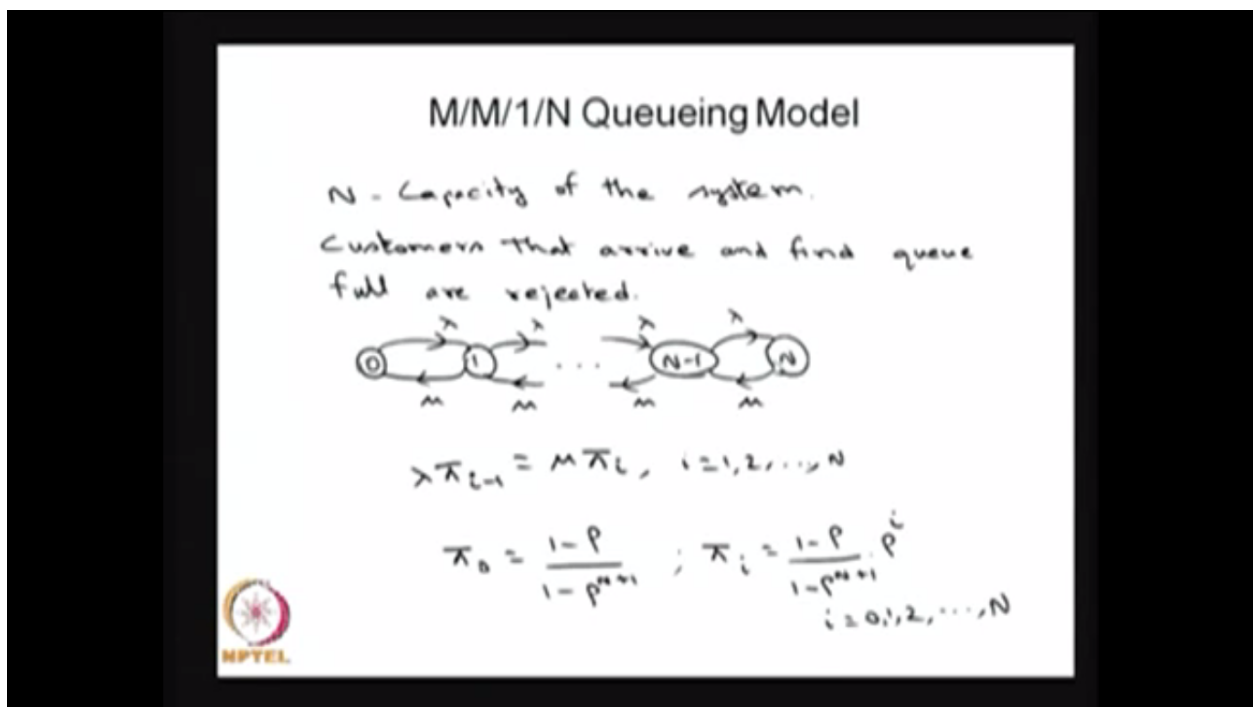
$$E(N) = c\rho + \frac{\rho}{1-\rho} c(c, \frac{\lambda}{\mu})$$


From these you can find out the expected number of the next measure is the expected number of busy servers. What is the average number of busy servers that is nothing but the summation of n equal to 0 to c minus 1 n times P_n that means whenever the system size is less than c only those many servers are busy and with the probability. Whenever n customer are more than n customers in the system all the c servers are going to be busy therefore c times P . If you simplify you will get a c times ρ so that is the expected number of busy servers. Once I know the expected number of busy servers I can find out what is the expected number of idle servers also. It's [Indiscernible] [00:04:18] that is an expected number of ideal server is nothing but expectation of it's a random variable so ideal number is nothing but there are totally c servers in the system therefore c minus busy servers are B therefore c minus B is same as I . So the expectation satisfies the linear property therefore expectation of I is same as expectation of c minus B . C is a constant and B 's are random variable therefore it is a c minus expectation of B . Expectation of B just now we got c times ρ therefore the expected number of ideal server is c times 1 minus ρ .

So other than stationary distribution for the MMC model we are getting what is the probability that arriving customer has to wait and we are getting the conditional probability of a j customers in the queue given that waiting time is greater than 0 as well as this expected quantities we are getting. Also we can find out what is the expected number of customers in the system. That is nothing but expected number is nothing but expected of the busy, busy service plus expected number in the queue. Earlier I use the notation N suffix quarter. N suffix q and Q are both one and the same so I can compute what is the expectation of Q do the little simplification then I can substitute expectation of Q here therefore I will get expected number of customers in the system that involves the Erlang's C formula. So this Erlang's C formula is used to get the expected number of customers in the system and then later we can do some optimization over the probability expected number with the specified c and λ by μ . So using Little's formula I

can find out the expected time spent in the system because I know what is arrival rate and from the stationary distribution I got expected number in the system in a steady state therefore since I know lambda and expectation of IN I can get expectation of R where R is the response time or sojourn time or total time spent in the system. So that expectation is going to be expectation of EN divided by lambda, do little simplification you will get expectation of R. You can apply the Little's formula in the queue level also. So this is the system level and you can apply the queue level also. So lambda times expectation of waiting time is same as expectation of a number of customers in the queue.

So expectation of waiting time or average waiting time is same as expectation of Q divided by lambda. So using since the MMC infinity queue the underlying stochastic process is a birth-death process therefore we are getting all the measures using the birth-death logic.



Next I am going for the finite capacity. So the N is a capacity of the system that means whenever the customers arrives and find a queue full that customer will be rejected. Therefore at any time the number of customers in the system if you make it as a random variable and that random variable takes the possible values from 0 to N therefore the state space is finite the number of customers in the system make any time t that a random variable and you will have a stochastic process and since the inter-arrival time is exponentially distributed service is exponentially distributed only one server finite capacity, therefore, the underlying stochastic process is a birth-death process with the birth rates lambda, the death rates mu. If you see the queue matrix for this one infinite decimal generator matrix that's a tri-diagonal matrix with all the off diagonals or lambdas as well as mu and the diagonals are minus lambda plus mu except the first determinant and the last determinant. The [Indiscernible] [00:09:20] in the first row and the last row.


Our interest is to get the stationary distribution. Later I'm going to explain the time dependent solution also. So to get the stationary distribution either you write the P_{i0} is equal to 0 and the summation of P_{i0} is equal to 1 and solve that or you write the balance equation the P_{i0} is equal to 0 that will land up a balanced equation so some books writes this as a balanced equation. What is the inflow rate and what is the outflow rate both are going to be same whenever the system reaches the equilibrium solution, equilibrium state. Therefore, the outflow is a λ times this the inflow is μ times λ^{-1} ; like that you can go for understanding the balance equation for this state and second and so on. And this also satisfies the time reversibility – this is also called satisfying the time reversible equation therefore one can use the time reversible property of a birth-death process so you can find out the P_{i0} 's easily using the time reversible equation itself. You don't want to use a P_{i0} 's is equal to 0 instead of that you can write the time reversible equation since it is have abstract by all the states now we can use the summation of P_{i0} 's equal to 1 i starting from 0 to the N therefore you will get P_{i0} naught and here the birth-death process in the finite state space therefore the P_{i0} naught will be 1 divided by the denominator series that's a finite series, finite terms in it therefore it is always converges immaterial of the value of λ and μ . Therefore you will get a P_{i0} naught without any restriction over λ and μ . So once you get the P_{i0} naught you can get P_{i0} 's in terms of P_{i0} naught therefore that is a $1 - \rho$ divided by $1 - \rho^{n+1}$ times ρ^i where ρ is λ/μ . So this is the birth – the underlying stochastic process, the birth-death process with the birth rates λ and death rates μ . So you can use all the concepts of the birth-death process and you can analyze the system in ECV.

. Effective arrival rate
 $\lambda_{eff} = \lambda (1 - \pi_N)$

. Throughput
 $\mu (1 - \pi_0)$

. Blocking probability
 π_N

. $E(R) = \frac{E(N)}{\lambda_{eff}}$



So this is a steady state probability. Once you know the steady-state probability you can get the other measures also. Here the important thing is called the effective arrival rate that means the system, the queuing system is a finite capacity. So maximum N customers can wait in the system

and the service rate is μ the arrival rate is λ from the infinite population. So whenever the system size is full the customer is rejected. Therefore there is a rejection. After the service is completed the system leaves the system. So the effective arrival rate is nothing but what is the rate in which the system the customers are entering into the system. So there is a partition here. So the effective arrival rate is a λF . That rate will be what is the probability that the system is not full multiplied by the arrival rate λ that is going to be the λ effective. Whenever the system is not full that proportion of the time or the probability is $1 - P_n$ where P_n is the steady-state probability just now we got it. From here you can get a P_n suffix n that is the probability that the system is full and $1 - P_n$ is the probability that the system is not full and multiplied by the arrival rate and that is going to be the λ effective. And you can also find out the throughput. Throughput is nothing but what is the rate in which the customers are served per unit of time. The service rate is μ and this is the probability that the system is not empty $1 - P_0$ therefore $1 - P_0 \mu$ that is the rate in which the customers are served in the MM1 n system. Whenever the system is not full not empty sorry whenever the system is not empty multiplied by that probability multiplied by μ that is going to be the throughput. By using the time reversible equation the μ times $1 - P_0$ you can get in terms of λ s equivalent also but the throughput is the service rate multiplied by what is the probability that the system is not empty. Since it is a finite capacity system one can find out the blocking probability also. Blocking probability is nothing but the probability that the customers are blocked. The customers are blocked whenever the system is full. Therefore, the blocking probability same as the probability that the system is full that is P_n .

Once we know the steady-state probabilities you can find out the average number of customers in the system and using the Little's formula you can get expected time spent in the system by any custom divided by not λ it is λ effective, because the effective arrival rate is a used in the Little's formula not the arrival rate. For a moment infinity system the effective arrival rate and arrival rate are one and the same because there is no blocking therefore the probability of $1 - P_n$ that is equal to 1 only. Therefore the effective arrival rate and arrival rate are same for a infinite capacity system because there is no blocking. For a finite capacity system the effective arrival rate has to be computed. Similarly we have to go for finding the MM1 λ effective or the MMCK model also. So other than stationary distribution or equilibrium probabilities we are getting the other performance measures using the birth-death process concepts.