# DOWNLOAD PDF QUEUING THEORY IN OPERATION RESEARCH

## Chapter 1 : Queuing Theory in Operation Research - GATE l GATExplore

*Discussion Slide -1 Define Queuing Model or Queuing Theory Queuing theory is the mathematical study of waiting lines (or queues) that enables mathematical analysis of several related processes, including arriving at the (back of the) queue, waiting in the queue, and being served by the Service Channels at the front of the queue.*

Queueing Theory Video Transcript Instructor: And then the things that need work to be done queue up behind it. This could be a cashier at a supermarket and these are the people online. This could be a processor in a computer system. But the theory will end up being exactly the same. Then we need to know the speed of the processor or the cashier at the supermarket. At what speed can they get customers satisfied or done and leave the system? T ends up been the mean service time for each arrival. So the mean service time will be the reciprocal. If it can do three per second, then is one third of a second per person or per process. Another interesting thing we like to keep track of is what percentage of the time is this server in use? And then this â€¦ the processor is still working. Other things that we might want to know, what is the mean number of customers in the system like at any given time, if something is running and what one is in the running state and the ready state and the total rate. We might want to analyze how many are in the queue at a time. That would be the response time or the turnaround time. The through pit of the system is the amount of work it can get through per time so if it can do four jobs per minute that its through put. So if you think about this, suppose the amount of work coming into a queuing system is more than the rate at which it can get out, what would happen? Not only theoretically but what would happen in actuality, if we had a queue? I mean yeah, it will like build off to infinity. And that queuing system would be the bottleneck. The second one could be the probability distribution for the service time. The third one could be the number of servers. The fourth one could be the capacity of the queue. Just the last example, there was an infinite queue but we could have a capacity on the queue. The N will end up being the number of customers in the system. And then Z would be the queuing discipline like first come first serve. But there could be other queuing disciplines. So, have you guys taken a probability class? So you know how likeâ€¦ you could have a completely random distribution like maybeâ€¦ sometimes called the Poisson distribution, something happens at a rate of one per second. But howâ€¦ when the next event from the previous one is completely unrelated. So you could have, one comes exactly every second. This is an arrival. So they both have an average of one per second but one of them comes every second and the other one comes totally random with an average of one per second. So just because we know the rate, one per second or 10 per second, we still need more information about how the distance between each one is. So this would be the distance between the arrivals, the distribution for the service time too. We could have a service time where like for example, a processor a lot of times, it runs for a fixed amount of time and then kicks somebody out. So every time somebody goes into the processor, we could say it runs for a fixed amount of time and runs out so that might be a deterministic service time whatâ€¦ things thatâ€¦ processes that arrive to the queue come in randomly. So we could end up having something like a Markov distribution of where they come in and then maybe a deterministic distribution of when they go out. So just for example of using this notation, suppose we have a system like this. In this case our queue has exactly four seats in it. We have three servers. If we think of this as three cashiers at a supermarket, and then we have a finite number. This is a closed system. So we could say the arrival of markovian distribution of the service times are deterministic. There are three servers. The queue has four seats in it. And when they arrive at the queue, they are processed in a first come first served. We could later do things like in the queue, we can say we could have priorities like high priority ones get pushed to front and little priorities gets pushed to the back. There are other things besides first come first served. But first come first served is like the intuitive one. And the service time is also random. So once we have hereâ€¦ lambda is our arrival rate. W of the queue is the mean timeâ€¦ process spent sitting in the queue. The number of items in the queue could be N and so queueâ€¦ and then we could do the same thing for the whole systemâ€¦ the wait time and the number in our system. So hopefully, this is an intuitive example. But if we have 10 people walk into a supermarket every minute and on average they stay there for 30 minutes. Then at a random point in time, how many people are in the supermarket? But that will be true if they stay for one minute. And if on average they stay for 30 minutes, then it will be 10 times 30 which will be people on average. Then there might be only 10 people in at a time but if they can come in and stay random amounts of time some, some people can stay an hour some people leave after two minutes, but on average they stay for 30 minutes. Then this is how many people would be in the system on average. And the principle applies to the whole system, which is the queue and the processor. And if we wanted to just focus on the queue, same principle applies for the queue. Okay, so take a look at this MM1 queue. So we left out a lot in that Kendall notation. We left out a lot of stuff. The queue size is going to be infinite. The inter-arrival times are markovian. The service time distribution is markovian. One is for one server. If the rate into the system is one per second and this thing can get the processes served at a rate of one per second. But the cashier can get the people out at a rate of one per second whatâ€¦ how big would the queue be? Coming in at one per second. And it takes one second to process it. Yeah the processor or the cashier whatever you want to think of it as can move at a speed of one per second. If you give it no work, it obviously does no work but if you gave it like tons of work, it can get the work done at one per second. So on average is like maybe oneâ€¦ Instructor: Or one being processed. In that case, at time equals zero, one arrives and starts being served. At time equals one, the first one leaves as the second one arrives. And this will go on all day. And the same thing is true with the server, with that circumstance make it different. So you can see the difference that even though the rate how many come in per second and how many can get done per second, the distribution between the events actually will affect the queue. Therefore, every arrival is served in the exactly one second. And the pervious customer leaves when the new one comes in. So you might think that the queue is of size zero. But if you think about it, it really depends on these letters up hereâ€¦ could cause the queue increase in size and decrease in size. But what about the utilization? And the queue might build up to like three of four, then go back to zero, then build up to three or four and go back down to zero. So the utilization ends up beingâ€¦ we can think of a formula now as the rate in divided by the rate out. And that would be independent of the input distribution and output distribution. If one comes in per second and it can move at a speed of getting two out per second, this will be in use half the time and half the time, it will be sitting and waiting. Okay, so then what we want to do little later on down the road is we want to start taking inputs from different areas and running them into the same queue. I just want to work on a little intuition here. But this could be an ND3 system. So the, three means three servers. So, if we had for some reason two different inputs coming into the same queue, what would be the rate into the queue? Hopefully, this is a pretty simple intuition. But if one source that was sending inputs into a queue was coming in at a rate of two per second, and another source was sending processes into the same queue at a rate of one per second, then from the point of this queue, how many are coming in?

## Chapter 2 : Queuing Theory and Practice: A Source of Competitive Advantage

*Queueing theory is the mathematical study of waiting lines, or queues. A queueing model is constructed so that queue lengths and waiting time can be predicted. Queueing theory is generally considered a branch of operations research because the results are often used when making business decisions about the resources needed to provide a service.*

This waiting problem leads the Danish engineer A. He developed in the Queuing theory. Queuing theory analyze the shared facility needs to be accesed for service by a large number of jobs or customers. Examples for the queuing theory are waiting lines in cafeterias, hospitals, banks, airports and so on. In the following you can find more detailled informations for this topic. Definition In computer science, queueing theory is the study of queues as a technique for managing processes and objects in a computer. A queue can be studied in terms of: The queues that a computer manages are sometimes viewed as being in stacks. In most systems, an item is always added to the top of a stack. A process that handles queued items from the bottom of the stack first is known as a first-in first-out FIFO process. A process that handles the item at the top of the stack first is known as a last-in first-out LIFO process. System of Queuing theory Elements of Queuing Systems Figure 1 shows the elements of a single queue queuing system: Population of Customers can be considered either limited closed systems or unlimited open systems. Unlimited population represents a theoretical model of systems with a large number of possible customers a bank on a busy street, a motorway petrol station. Example of a limited population may be a number of processes to be run served by a computer or a certain number of machines to be repaired by a service man. It is necessary to take the term "customer" very generally. Customers may be people, machines of various nature, computer processes, telephone calls, etc. Arrival defines the way customers enter the system. Mostly the arrivals are random with random intervals between two adjacent arrivals. Typically the arrival is described by a random distribution of intervals also called Arrival Pattern. Queue represents a certain number of customers waiting for service of course the queue may be empty. Typically the customer being served is considered not to be in the queue. Sometimes the customers form a queue literally people waiting in a line for a bank teller. Sometimes the queue is an abstraction planes waiting for a runway to land. There are two important properties of a queue: Maximum Size and Queuing Discipline. Maximum Queue Size also called System capacity is the maximum number of customers that may wait in the queue plus the one s being served. Queue is always limited, but some theoretical models assume an unlimited queue length. If the queue length is limited, some customers are forced to renounce without being served. There are these ways: These methods are typical for computer multi-access systems. Most quantitative parameters like average queue length, average time spent in the system do not depend on the queuing discipline. In fact the only parameter that depends on the queuing discipline is the variance or standard deviation of the waiting time. There is this important rule that may be used for example to verify results of a simulation experiment: Theoretical models without priorities assume only one queue. This is not considered as a limiting factor because practical systems with more queues bank with several tellers with separate queues may be viewed as a system with one queue, because the customers always select the shortest queue. Of course, it is assumed that the customers leave after being served. Systems with more queues and more servers where the customers may be served more times are called Queuing Networks. Service represents some activity that takes time and that the customers are waiting for. Again take it very generally. It may be a real service carried on persons or machines, but it may be a CPU time slice, connection created for a telephone call, being shot down for an enemy plane, etc. Typically a service takes random time. Theoretical models are based on random distribution of service duration also called Service Pattern. Another important parameter is the number of servers. Systems with one server only are called Single Channel Systems, systems with more servers are called Multi Channel Systems. Output represents the way customers leave the system. Output is mostly ignored by theoretical models, but sometimes the customers leaving the server enter the queue again "round robin" time-sharing systems. Queuing Theory is a collection of mathematical models of various queuing systems that take as inputs parameters of the above elements and that provide quantitative parameters describing the system performance. Because of random nature of the processes involved the queuing theory is rather demanding and all models are based on very strong assumptions not always satisfied in practice. Many systems especially queuing networks are not soluble at all, so the only technique that may be applied is simulation. Nevertheless queuing systems are practically very important because of the typical trade-off between the various costs of providing service and the costs associated with waiting for the service or leaving the system without being served. High quality fast service is expensive, but costs caused by customers waiting in the queue are minimum. On the other hand long queues may cost a lot because customers machines e. So a typical problem is to find an optimum system configuration e. The solution may be found by applying queuing theory or by simulation. Applications and limitations Applications The queueing theorie is used to analyze computer, telecommunication systems, traffic systems traffic flue , logistic and manufacturing systems. A queueing model is characterized by: The model is the most elementary of queueing models[1] and an attractive object of study as closed-form expressions can be obtained for many metrics of interest in this model. Here is the arrival population unlimited and all arrivals wait to be served.

## Chapter 3 : Queueing Theory Applications, Articles, and Video Tutorials

*Queuing Theory is a collection of mathematical models of various queuing systems that take as inputs parameters of the above elements and that provide quantitative parameters describing the system performance.*

Queuing Theory and Practice: Sometimes, it is a pleasant experience, but many times it can be extremely frustrating for both the customer and the store manager. Given the intensity of competition today, a customer waiting too long in line is potentially a lost customer. Queues are basic to both external customer-facing and internal business processes, which include staffing, scheduling and inventory levels. For this reason, businesses often utilize queuing theory as a competitive advantage. Fortunately, Six Sigma professionals â€” through their knowledge of probability distributions, process mapping and basic process improvement techniques â€” can help organizations design and implement robust queuing models to create this competitive advantage. The Cost of Waiting in Line The problem in virtually every queuing situation is a trade-off decision. The manager must weigh the added cost of providing more rapid service i. For example, if employees are spending their time manually entering data, a business manager or process improvement expert could compare the cost of investing in bar-code scanners against the benefits of increased productivity. Likewise, if customers are walking away disgusted because of insufficient customer support personnel, the business could compare the cost of hiring more staff to the value of increased revenues and maintaining customer loyalty. The relationship between service capacity and queuing cost can be expressed graphically Figure 1. Initially, the cost of waiting in line is at a maximum when the organization is at minimal service capacity. As service capacity increases, there is a reduction in the number of customers in the line and in their wait times, which decreases queuing cost. The optimal total cost is found at the intersection between the service capacity and waiting line curves. Chase and Nicholas J. Aquilano, Production and Operations Management, , page  Queuing Theory Queuing theory, the mathematical study of waiting in lines, is a branch of operations research because the results often are used when making business decisions about the resources needed to provide service. At its most basic level, queuing theory involves arrivals at a facility i. The number of arrivals generally fluctuates over the course of the hours that the facility is available for business Figure 2. Number of Arrivals at Facility Customers demand varying degrees of service, some of which can exceed normal capacity Figure 3. The store manager or business owner can exercise some control over arrivals. For example, the simplest arrival-control mechanism is the posting of business hours. Other common techniques include lowering prices on typically slow days to balance customer traffic throughout the week and establishing appointments with specific times for customers. The point is that queues are within the control of the system management and design. Service Requirements Queuing management consists of three major components: How customers arrive The condition of the customer exiting the system Arrivals: Arrivals are divided into two types: Constant â€” exactly the same time period between successive arrivals i. Variable â€” random arrival distributions, which is a much more common form of arrival. A good rule of thumb to remember the two distributions is that time between arrivals is exponentially distributed and the numbers of arrivals per unit of time is Poisson distributed. The Servicing or Queuing System: The servicing or queuing system consists of the line s and the available number of servers. Factors to consider include the line length, number of lines and the queue discipline. Queue discipline is the priority rule, or rules, for determining the order of service to customers in a waiting line. Others include a reservations first, treatment via triage i. An important feature of the waiting structure is the time the customer spends with the server once the service has started. This is referred to as the service rate: Another important aspect of the servicing system is the line structure. There are four types: The simplest type of waiting line structure is the single-channel, single-phase. Here, there is only one channel for arriving customers and one phase of the service system. An example is the drive-through window of a dry-cleaning store or bank. There are two possible outcomes after a customer is served. The customer is either satisfied or not satisfied and requires re-service. Waiting Line Models and Equations Table 1 shows the four types of commonly used waiting line models, along with key properties and examples.

## Chapter 4 : Queuing theory 3 â€“ Operations-Research-Wiki

*[Hindi] Queuing Theory in Operation Research l GATE l M/M/1 Queuing Model Operation Research #1 - Duration: GATE Lectures by Dishank 53, views*

## Chapter 5 : Queuing theory | mathematics | racedaydvl.com

*Queuing theory, subject in operations research that deals with the problem of providing adequate but economical service facilities involving unpredictable numbers and times or similar sequences. In queuing theory the term customers is used, whether referring to people or things, in correlating such.*

## Chapter 6 : Various characteristics of queuing system in Operations Research - IIBM Institute LMS

*This web portal is specially for candidates who are preparing GATE, IES, SSC JE,IIT JAM, IIT JEE, BARC and others competitive examination. Here we are providing all latest updates about examination, strategy, previous year papers, syllabus and many more.*

## Chapter 7 : Queueing theory - Wikipedia

*Definition: Queuing Theory or Waiting Line Theory. Queuing Theory is a branch of operations research which is used to predict the length of queues and waiting times in order to decide the amount of resources required to provide any service.*