Queuing Models

Name

Institutional Affiliation

Queuing Models

Queuing Theory

Queuing theory is a mathematical approach to predicting the amount of time the clientele is likely to spend on a waiting line through examining the waiting time versus the queue lengths. Queuing theory factors in the possibility of congestion through scrutinizing all facets of queues which include the number of the available servers, the arrival of clients as well as the resources that have to be put in place to decongest a waiting line. Queuing models have typically been applied in business establishments which deal with commodities. However, in contemporary times, enterprises which deal in services have begun applying queuing theory to make decisions which would improve the quality of the services they render to their clients through the fundamental premise that customers are likely to become disappointed when they have to continually wait for too long before being attended.

There are varied queuing models which could suffice in different situations depending on the urgency of service, the rate of arrivals, and the number of servers. Firstly, is the M/M/1 queue model; whereby, the server exponentially distributes the service times. The distributing takes the form of $1/\mu$ using a variance of $1/\mu^2$ (Render, Stair, Hanna & Hale, 2014). Secondly, is the M/D/1 queue model; in which there are fixed service times (Render, Stair, Hanna & Hale, 2014). Thirdly, is the M/G/1 queue; in which there is unlimited capacity regardless of the Poisson-style arrivals (Render, Stair, Hanna & Hale, 2014). Fourthly, is the G/G/1 queue model; whereby, service times are distributed based on inter-arrival times (Render, Stair, Hanna & Hale, 2014). Other models include the M/D/k queue model and the M/D/c queue model which distributes the servers exponentially.

The M/D/1 Queue Model

The M/D/1 queue model denotes a stochastic process in which the arrival of clients that require services follows a Poisson random measure. From a peripheral standpoint, the M notation represents a Markovian input process. A Markovian arrival process is distributed exponentially. The second notation describes the distribution of services; in this case, they are deterministic. The third notation represents a single server which serves all entities which arrive randomly on the waiting line.

The discussion which Render, Stair, Hanna, and Hale (2014) developed encapsulates vital facets of the M/D/1 queue model. Firstly, is the presence of a single server; meaning, the waiting line in an M/D/1 queue is served by one server which could be an automated computer, a bank teller, or a customer service representative. Secondly, the service times are deterministic in an M/D/1 queue; meaning, they adhere to a fixed approach. According to Zukerman (2013), the job service times could be said to be deterministic since the service times follow the rate $\mu=1/D$. In this notation, D denotes the deterministic time. Thirdly, the buffer could contain, or store, an infinite number of entities. The entities, in an M/D/1 queue, refer to the clients that require the service that the single server provides.

The fourth element in an M/D/1 queue is the first-come, first-out service approach; whereby, the only server serves the clients based on their positions on the waiting line. Zukerman (2013) argued that the first-in-first-out (FIFO) approach does suffice for the single server system in an M/D/1 queue model since other decongestion and servicing mechanisms such as the preemptive shortest job first and the shortest job first would otherwise slow down the single server approach since they entail calculating the service times need for individual entities.

In blunt terms, the single server begins serving each arrival from the front; thus, reducing the number of queuing entities by one each time the server serves the front entity successfully. The single server system, therefore, has constant service times.

Notably, however, a Poisson process determines the arrival of all entities. A Poisson process, according to Budhiraja, Chen, and Dupuis (2013) is a random location of objects, or points, on a predetermined probability space which could be said to have substantial statistical properties. A Poisson point process, in queuing theory, strategically models the arrival of consumers who, on average, arrive on waiting line randomly. This implies that there are periods in which the customers could arrive overwhelmingly at once, and there are periods in which their arrival could be distributed differently. The Poisson random measure, therefore, is relevant in relevant in the M/D/1 queue model since there is infinite storage capacity for the number of entity arrivals.

The M/D/1 queue model, as such, is directly applicable in situations in which the arrival of entities takes an exponential fashion yet are best served by one server. Perry, Ousterhout, Balakrishnan, Shah, and Fugal (2015) studied how the M/D/1 queue model fits in network design. In wide area computer networks (WANs), there exists one central server that serves the packets which arrive to request its services. A cyclic redundancy check (Perry, Ousterhout, Balakrishnan, Shah & Fugal, 2015) in the central server independently computes the arrival of individual packets. Also, since WANs have infinite storage for the packet arrivals, the central server could serve many packets without necessarily depleting its resources.

Importance of Queuing Models for Business Operations

Queuing models are vital, in varied ways, for business operations. At the most elemental level, queuing models, according to Van Der Aalst (2013), serve the purpose of both timeliness and orderliness. A fast-food drive-through window, for example, would be chaotic if the service providers fail to institute appropriate mechanisms for queuing. In other words, queuing models in the service industry creates the logic for controlling the flow of customers. Since staffing resources are usually limited, a fast-food drive-through window has to encompass time as a key factor in ascertaining that consumers do not have to wait for prolonged periods before accessing the services they require. Since time has a critical effect on the profitability, public image as well as operational efficiency in the fast-food industry, queuing models are important in devising the practical mechanisms for serving the most considerable number of customers within short periods.

Queuing models are also significant in tracking the progress of business operations. Tracking how an enterprise serves its customers, according to Van Der Aalst (2013), is necessary for enabling the executive board to stay in control of all their operations, which have a critical effect on how the clientele perceives the services they provide. More specifically, for a fast-food drive-through window, the managers could efficiently apply a queuing model to devise how the technical staff might schedule the business' computers to speed up and serve multiple customers at a time.

The other way in which queuing models are vital for business operations is in making iterative assessments. Queuing models simplify how fast-food drive-through windows assess the efficacy of the services rendered to customers (Van Der Aalst, 2013). For instance, if the fast-food drive-through window keeps the consumer too long on the line for them to get their fast-food fix, it would imply that the efficiency of the fast-food enterprise would be in question.

References

Budhiraja, A., Chen, J., & Dupuis, P. (2013). Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastic Processes and their Applications*, *123*(2), 523-560.

Perry, J., Ousterhout, A., Balakrishnan, H., Shah, D., & Fugal, H. (2015). Fastpass: A centralized zero-queue datacenter network. *ACM SIGCOMM Computer Communication Review*, *44*(4), 307-318.

Render, B., Stair R., Hanna, M. & Hale, T. (2014). *Quantitative analysis for management, (12th ed.)*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Van Der Aalst, W. M. (2013). Business process management: a comprehensive survey. *ISRN Software Engineering*, *2013*.

Zukerman, M. (2013). Introduction to queueing theory and stochastic teletraffic models. *arXiv preprint arXiv:1307.2968*.