RESEARCH PAPER | Management | Volume : 3 | Issue : 7 | July 2013 | ISSN - 2249-555X

Stochastic Control Optimization Technique on Multi-Server Markovian Queueing System

KEYWORDS: Markov chain ,Quality of Service, Multiclass Queueing Network, Scheduling, Transition

K. Sivaselvan
Deparment of Mathematics, Jeppiaar Engineering College, Chennai, India.

C. Vijayalakshmi
VIT University, SAS Maths Division,Chennai,India.

ABSTRACT *Queueing network has widely applied to signify and investigate the resource sharing the system such as computer system, communication network. In many applications servers, like Web servers file servers, database servers, a huge amount of transaction has to be handled properly in a specified time limit. Each transaction typically consists of several sub-transactions that have to be processed in a fixed sequential order. One of the most important quality of service parameters for these applications is expected response time, which is the total time it takes a users request to be processed. In multi class queueing network, a job moves from a queue to another queue with some probability after getting a service. The portrayal of the queue stability is obtained based on the following parameters: general arrival, service time distributions, and multiple classes with specific arrival rate. A multiple class of customer could be open or closed where each class has its own set of queueing parameters. In distributed multi-server network in which the customer transitions have exemplified by more than one closed Markov chain. The main objective is to maximize the total source utility by posing the optimization network control which shown that the rate control problem has been solved completely. Numerical calculations and Graphical representation shows that the new method improves the performance measure in terms of reduction computational effort.*

1.0 INTRODUCTION

Recently communication infrastructure plays vital role in applications that share parts of the infrastructure. Consider a multi server queueing network with M servers and N classes of customers. Time is slotted to continue a fixed entity length to serve jobs. Consider the system has an infinite buffer in which assuming that the jobs are assumed to arrive for the duration of slot and depart at the end of the slot. In the class of queuing systems the problem of optimal control is maximize the optimization criterion has to be chosen in order to prevent degradation of services in computer communication networks. The optimization criterion depends on two factors that maximize the average network time delay constraint. In many applications servers, like Web servers file servers, database servers, a huge amount of transaction has to be handled properly in a specified time limit. Each transaction typically consists of several sub-transactions have to be processed in a fixed sequential order. For the distributed computer network model, an algorithm is designed in which the customer transitions have characterized by more than one closed Markov chain. A solution of product form algorithm is derived in the case of multiple closed sub chains and computational algorithm is presented for general class of queueing networks. The result is generalized to a queueing network in which the customer routing transitions are characterized by a Markov chain decomposable into multiple sub chains. Several aggregate states and their marginal distributions are discussed in conclusion.

2.0 LITERATURE SURVEY

M.Andrews (2004).et.al., had developed the concept of the Scheduling in a Queueing system with asynchronously varying service rates. S.Andradottir,et.al.,(2003) has explained efficiency of time segmentation parallel simulation of finite Markovian queueing networks. R.Artalejoet.al.,(2012) has envisaged Markovian retrial queues with two way communication system. S.Balsamo (2001), et.al., had explained the analysis of queueing networks with blocking. Parallel scheduling of multiclass M/M/m queues: approximate and heavy-traffic optimization of achievable performance by Glazebrooket.al., (2001).P.Cremonesi et.al.,(2002) had enlightened the approximate solution of closed multiclass queuing networks. M.Harchol-Baltere et.al.,(2005) has evidently envisaged the Multi-server queueing systems with multiple priority classes. W.K. Ehrlich.et.al.,(2001), had clearly explained the performance of web servers in a distributed computing environment. J.R.Ramos,et.al.,(2003)has approached an improved computational algorithm for Round Robin Service. The concept of node decomposition based approaches for multi-class closed queuing networks has envisioned by K.Satyamet.al.,(2005). M,Reiser et.al.,(1980) had analyzed multi chain closed queueing networks. R.Srinivaset.,al(2007) had analyzed a multi-server queueing model with Markovian arrivals and multiple thresholds., S.Stidham (2002) has clearly explained the methodology to design and control of queueing system. L.Tadj et.al.,(2005)had explained Optimal design and control of queues.

3.0 STOCHASTIC SCHEDULING CONTROL ON MULTI SERVER QUEUEING NETWORK

The behavior of multi server queueing network has been classified into three major pastures:

(1) Network Control,
(2) Control over networks, and
(3) Multistage scheme.

Networks Control has provides a certain level of performance to a network data stream, although attains proficient and fair utilization of network resources. Control over networks deals with the design of feedback policy to modify the control systems in which control data is swap over through unreliable communication links. Multistage scheme explain network structural design and communications between network components which describes behavior of the individual components agents of the networked system. Queueing network model is used to study the behavior of large system of different components and it is very complex for input and output from a component. Only limited case of queuing network models yielding a analytic solutions. Application of Markovian queuing network models with product form solution gives exactly solution. The stochastic nature of the input and output system represents the arrival and departures of packets in computer communication networks. The product form solution provides approximation to the actual behavior of packet switching network. Each user has to choose a

Volume : 3 | Issue : 7 | July 2013 | ISSN - 2249-555X

flow control strategy and the optimal choice for decentralized flow control depends on the strategies of the other users. Consider a queueing network that composed of three nodes, each modeled as a multiclass queueing network. Jobs arrive λ_1 at the first node in the servers of the node with the rate μ_0 ,μ_1 , *and* μ_2 . After completion of service at the first node jobs are forward to the second node with probability p and leave the node with probability $1-p$. The incoming jobs are server in the second node with rates μ_1 *and* μ_2 with $\gamma_1 = \lambda_1 - (\mu_0 + \mu_1 + \mu_2)$.

3.1 Steady State Distribution

The first step is to partition the most arriving job into several sub-states and aggregate the surrounding state into one. The partitioning strategy imposes a structure on the feasible state and is therefore very important for the rate of convergence of the algorithm. If the partitioning is such that most of the good solutions tend to be clustered together in the same sub states, it is likely that the algorithm quickly concentrates the search in these subsets of the feasible state. More efficient partitions could be constructed if the performance function is considered. This type of partitioning techniques is called data base group partitioning. Then the approach of moment fragmentation technique is appropriate to a class of queueing networks with thrashing and communication blocking stations. Let $\Re_i$ denote the number of servers and C_j denotes the buffer capacity at j. According to Poisson process jobs arrival rate is λ_i, $i=1,2,....,n$ to the station i and the service time is exponentially distributed with μ_i, $i=1,2,....,n$ .Let P_j denotes the probability that a job endeavor to join the queue at station j, immediately after a service completion at station i, for j = 1,… n, and let $P_{i,n+1}$ denotes the probability that a job will leave the system after being served at station i. When a job enters the system, which founds that the buffer station is full leaves the system immediately, unless it is arriving from another station in the network, in which case it endures another service and then rerouted. Let $\Psi(t)=\{\Psi_1(t)\,\Psi_2(t)\,\Psi_3(t)...,\Psi_n(t)\}$ denote the state of the system at time t, where $\Psi_j(t)$, $j=1,2,...,n$ denotes the number of jobs in station j at time t. Next determine the station at which the event has to be executed and whether the event is an arrival or a service completion. An arrival is a feasible event at station j in a routing if the buffer is not full. If a job is feasible for a routing, then execute the event and update the state of the system accordingly. If a job is not feasible for a routing, simply ignore that event and the state of the system in that sample path will remain unchanged. This procedure for parallel simulation of multiple routings of discrete event systems using a general cycle of latent measures is fundamentally based on standardized of the Markov chain.

3.2Round Robin

In the round robin scheduling, processes are dispatched in a FIFO manner but are given a limited amount of CPU time called a time-slice or a quantum. Round Robin fashion runs the algorithm based on the length of the quantum that will drastically decrease the waiting time compared with the other scheduling algorithm.

FCFS and SJF are examples for non-primitive algorithms. Scheduling primitive algorithms like MLFQ and RR, which provides response time and fair dispatching of CPU time. Round Robin function technique will provide a proper response time and no starvation with low overhead. Feedback scheduling algorithm has not possibility for starvation since this algorithm gives better results of I/O bound processes and there is no importance for throughput and response time.

4.0 OPTIMIZATION OF QUANTUM QUEUE REPEATED NETWORK

Repeated Network is the best model that reserves the optimization that recognize the trend information of time series data. The network leaves a trace of its behavior and keeps a memory of its previous states. The inputs of repeated networks are the quantum of queues and the average response time. Average response time enters as an input in neural network, and then the network obtains the relation between the quantum change of a specified queue with the average response time and the quantum of other queues. According to the quantum change in a specified queue, it is possible to optimize the average response time. The network finds the relation between increase or decrease of quantum of a queue with the average response time and tries to reduce the average response time. Then the input of neural network updates the quantum changes of the queues and specifies a new quantum for queues. By entering the new quantum of queue to the intelligent multi class feedback queue function and pre-assumed processes are fed to this function, which leads to obtain the average response time. The optimized quantum for lower queues is found in the previous stage and these quantum amounts are not optimized further and the network may fail to achieve the desired result. This can be prevented by replacing the new quantum with previous ones and the new amount of specified queue with the former amounts.

The quantum of queues are returned to the input in a recursive way, it means that only the new quantum of specified queue is returned to the input and the other queues receive the former amounts as inputs. The new average response time is obtained by replacing the new quantum of a specified queue in intelligent multi class feedback queue function. When a change is applied in the quantum of a specified queue, the number of queues can be changed. It is possible that reducing the quantum caused more processes are moved to the lower queues or a new queue is added to the number of required queues. On the other hand increasing the quantum of a queue may cause no process is moved to the lower queues and as a result the lower queues are eliminated. Applying the changes in the specified queue, by decreasing the quantum it causes more processes are moved to the lower queues or a new queue is added to the required number of queues. By increasing the quantum of a queue may cause no process is moved to the lower queues which are eliminated. The effect of these changes into the network by eliminating or adding a queue would be recognized by the new quantum for a specified queue. The multi class feedback queue outputs are used to calculate the average response time.

5.0 OPTIMIZATION OF QUANTUM QUEUE REPEATED NETWORK

A interconnect queue that describes a multi-class queueing network, where a job moves from a queue to another queue with some probability after getting a service. A multiple class of customer could be open or closed where each class has its own set of queueing parameters. A closed queuing network model is suitable for large number of job arrivals. These parameters are obtained by analyzing each station in isolation under the assumption that the arrival process of each class is a state-dependent Markovian process. These jobs are served at different sub queues with different service time distributions. The entire sub-queues have been served completely, then only the customer leave the system. Queueing network with finite capacity queues to represent the system with finite capacity resources and population constraints.

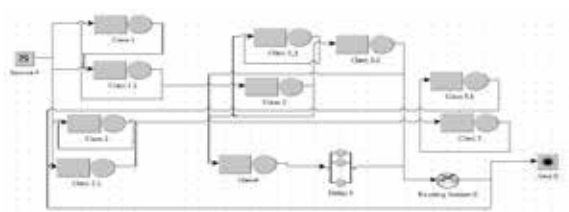


Fig. 1: Design of Multiclass Closed Queueing Network

Consider a queueing network system consisting of K service centers as $1 \le i \le K$. There exist R different classes of custom-

Volume : 3 | Issue : 7 | July 2013 | ISSN - 2249-555X

ers and their transition are presiding over from one state to another by a first-order Markov chain M. $P_{r(i\sigma r)}$ represent the probability that a customer of class r ν ch completes service at centre i will go to service centre i¢ and changes to class r¢ in M. Then the Markov chain M is decomposable into L sub chains $M_1,M_2,M_3,....M_L$ which are all irreducible. Assume that without loss of generality R ≥ L ≥ 1. The sub chains M_l's are either open are closed. The sub chains are driven by L independent Poisson arrival streams with rate $\lambda_l(n_l)$, where n_l is the total number of customers in sub chain M_l at a given system. Representation of general service time distribution in the method of stages is, only one stage can accommodate a job at a given time .In FCFS discipline, a customer waiting at the head of the line is not allowed to enter the first stage until the job currently in service completes its last stage and departs from this centre. That is the entrance stage is blocked as longs as a job exists in the some stage. The steady state distribution provides a solution in a product form when it is not blocking. In the situation of blocking, the solution is complicated for the queueing system. Hence the service centre is assumed to be a queue dependent exponential server in FCFS discipline.

In Processor sharing (PS) queue discipline , the problem of blocking in the exponential server could not exist. In a multi server queue, there are many servers available than jobs and no waiting line is formed, thus blocking is nonexistent. In an infinite server queue, where the service rate is lowered according to the number of jobs in the centre at a given time.In FCFS, when a new job enters, the first stage of the server is provided to it. If a new job is entered prior this has been served by its own server. If a job is stayed at any stage restart the service among those remaining in the system. When a new job entered the service centre without blocking, this leads to provide a product form solution. In Processor Sharing(PS) and Last Come First Serve(LCFS),the service stages are specified in the system.

6.0 NUMERICAL CALCULATIONS

Number of Customers

Average number of customers for each class at each station.

*	Aggregate	Class0	Class1	Class2	Class3	Class4	Class5
Aggregate	54.000000	2.000000	10.000000	15.000000	5.000000	10.000000	12.000000
Class 1	0.598934	0.011280	0.002423	0.321567	0.108101	0.154789	0.000770
Class 1.1	2.079728	0.040533	0.005230	0.024852	0.002176	0.063025	1.934002
Class 2	0.557316	0.011320	0.000261	0.183289	0.128501	0.233417	0.000528
Class 2.1	0.503948	0.000045	0.039745	0.102072	0.048697	0.000792	0.311598
Class 3	7.912940	0.000520	4.721682	0.160429	0.004096	3.028089	0.002113
Class 3_1	3.577932	0.145821	0.355906	0.008343	3.044073	0.021356	0.002433
Class4	8.786273	0.005419	3.928286	0.012342	0.848724	3.915795	0.025756
Class 3.2	10.774813	1.329461	0.480935	0.668911	0.136846	0.006868	7.951791
Delay 0	3.477350	0.050377	0.015157	0.005849	0.657561	1.001321	1.747085
Class 5	3.728603	0.392450	0.025479	1.718888	0.002505	1.570834	0.018447
Class 5.1	12.052162	0.003762	0.224937	11.793457	0.017719	0.007714	0.004572

Residence Times

Total time spent by each customer class at each station.

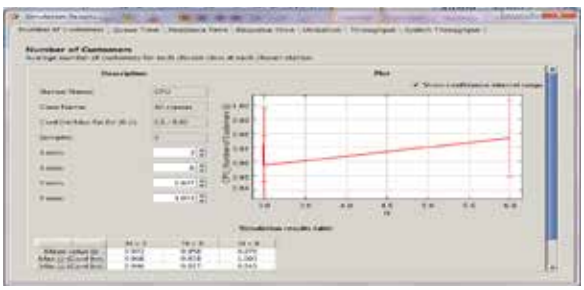
*	Aggregate	Class0	Class1	Class2	Class3	Class4	Class5
Aggregate	367.999715	330.674831	1071.5360...	439.693797	190.458175	337.832505	289.918039
Class 1	4.081623	1.864931	0.259587	9.426082	4.117728	5.229263	0.018736
Class 1.1	14.172947	8.189739	0.561332	0.728497	0.082888	2.129203	46.746907
Class 2	3.798001	1.871654	0.027948	5.372727	4.894822	7.885581	0.012756
Class 2.1	3.434308	0.007458	4.258786	2.992019	1.893039	0.026746	7.528148
Class 3	53.925181	0.087693	505.945267	4.702652	0.156037	102.163545	0.051055
Class 3_1	24.382926	24.109732	38.136654	0.244556	115.953699	0.721473	0.058780
Class4	59.536037	0.896022	420.924636	0.361782	32.329295	132.288294	0.622261
Class 3.2	73.428300	219.809724	72.964645	19.607734	5.212697	0.232033	192.113974
Delay 0	23.697480	8.329128	1.624117	0.171464	25.047589	33.827885	42.209283
Class 5	25.409722	64.886677	2.730219	50.385618	0.095431	53.067875	0.445675
Class 5.1	82.133189	0.622075	24.102823	345.700665	0.674950	0.260606	0.110464

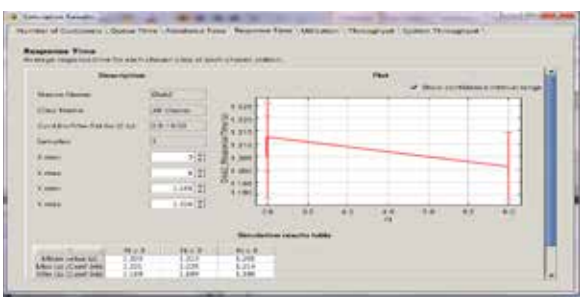
Utilization

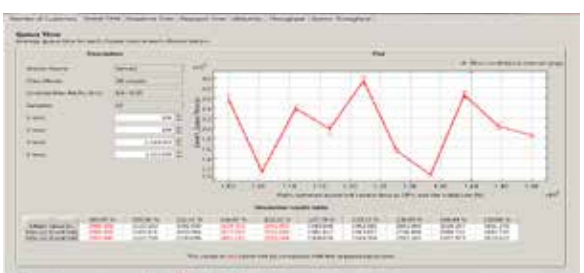
Utilization of a customer class at the selected station.

*	Aggregate	Class0	Class1	Class2	Class3	Class4	Class5
-	-	-	-	-	-	-	-
Class 1	0.375873	0.007068	0.001506	0.201157	0.068051	0.097606	0.000485
Class 1.1	0.695039	0.014468	0.001688	0.008071	0.000708	0.020661	0.649443
Class 2	0.360434	0.007278	0.000166	0.117717	0.083157	0.151777	0.000339
Class 2.1	0.336152	0.000029	0.026429	0.067872	0.033157	0.000526	0.208138
Class 3	0.946420	0.000059	0.584520	0.017970	0.000449	0.343185	0.000236
Class 3_1	0.957019	0.032609	0.077676	0.001822	0.839802	0.004581	0.000528
Class4	0.946974	0.000544	0.417982	0.001267	0.089244	0.435290	0.002647
Class 3.2	0.996074	0.127534	0.058222	0.056799	0.011589	0.000579	0.741351
Delay 0	3.477350	0.050377	0.015157	0.005849	0.657561	1.001321	1.747085
Class 5	0.828745	0.095288	0.005131	0.364256	0.000527	0.359705	0.003839
Class 5.1	0.999598	0.000282	0.017343	0.979681	0.001359	0.000584	0.000349

7.0 GRAPHICAL REPRESENTATION







CONCLUSION

This paper shows that the interactive behaviors in multi server queueing network which has been modeling by a stochastic technique. The stochastic isolation decomposition technique is to maximizing the scalability and minimizing the complexity in large networks. In distributed multi-server network in which the customer transitions have exemplified by more than one closed Markov chain. Generating function has implemented to derive closed form of solutions and product form solution with the parameters such as stability, normalizations constant and marginal distributions

Acknowledgment

I wish to express my gratitude and thanks to my parents, family members and my guide Dr.C. Vijayalakshmi for their valuable support and cooperation extended to design this model in a successful way.

Volume : 3 | Issue : 7 | July 2013 | ISSN - 2249-555X

REFERENCE

[1] Andrews,M.,Kumaran.K.,Ramanan.K.,Stolyar.A.,Vijayakumar.R and Whiting.P: Scheduling in a Queueing system with asynchronously varying service rates, Probability in the Engineering and Informational Sciences, Vol.18,no.02,(2004),pp.191-217. | [2] Andrad´ottir, S., Hosseini-Nasab, M.: Efficiency of time segmentation parallel simulation of finite Markovian queueing networks. Operations Research 51 (2003) 272–280. | [3] Artalejo, J.R. and Resing J.A.C. Mean value analysis of single server retrial queues. Asia-Pacific Journal of Operational Research 27 (2010), 335-345. | [4] Artalejo, Jesús R. and Tuan, P.D. (2012) Markovian retrial queues with two way communication. Journal of industrial and management optimization,8 (4) pp.781-206. ISSN 1547-5816. | [5] Balsamo.S, De Nitto Personè.V, Onvural.R, Analysis of Queueuing Networks with Blocking, Kluwer Academic Publishers, Dordrecht, (2001). | [6] Boxma.O.J. and Daduna.H, Sojourn times in queueing networks, In : Stochastic Analysis of Computer and Communication Systems, Elsevier Science Publishers, 1990. | [7] Glazebrook, K. and Niño-Mora, J. (2001). Parallel scheduling of multiclass M/M/m queues: approximate and heavy-traffic optimization of achievable performance. Operations Research, vol. 49 no. 4, 609-623. | [8] Cremonesi.P, Schweitzer.P.J, and Serazzi.G A unifying framework for the approximate solution of closed multiclass queuing networks. IEEE Trans. Comp.,51:1423 -1434,(2002). | [9] Ehrlich.W.K, Hariharan.R, Reeser.P.K and Van der Mei.R.D., Performance of Web servers in a distributed computing environment, In : Teletraffic Engineering in the Internet Era, Proceedings ITC-17 (Salvador-de Balia,Brazil), 137-148, 2001. | [10] Fujimoto, R.M.: Parallel and Distributed Simulation Systems. John Wiley & Sons,New York (2000). | [11] Katsaros, P. and Lazos C. 2000. "A technique for determiningqueuing network simulation length based on desired accuracy" International Journal of Computer Systems Science &Engineering, Vol. 15, 399-404. | [12] Leonardi.E, Mellia.M, Ajmone MarsanM, and Neri,F "On thethroughput achievable by isolated and interconnected input-queueing switches under multiclass traffic," Dipartimento di Elettronica, Politecnico di Torino, Tech. Rep. 16-02-2004,2004. | [13] Paganini. F, Wang, Doyle. J. C., and Low. S. H."Congestion control for high performance, stability, andfairness in general networks". IEEE/ACM Trans. on Networking, 13:43–56, 2005. | [14] Srinivas R. Chakravarthy, "A Multi-Server Queueing Model with Markovian arrivals and multiple thresholds" , Asia Pac. J. Oper. Res., 24, 223 (2007). | [15] Ramos,J.R., Rego,V., and Sang. J : An improved computational algorithm for Round Robin Service. Proceeding of the 2003 Winter Simulation Conference,Dec.7-10,IEEE – Xplore Press, USA ,pp : 721-728. | [16] Satyam.K, A. Krishnamurthy.A, and Kamath.M, "Node decomposition based approaches for multi-class closed queuing networks,"Technical Report Decision Sciences and Engineering Systems, (2005). | [17] Stidham. S., "Analysis, design and control of queueing systems",Operations Research 50 ,197–216 (2002). | [18] Siva Selvan .K and Vijayalakshmi C.: Design and Analysis of Multi server queueing model networks for webbased system, Proceedings of International Conference – ICOREM, Anna University, Thiruchirapalli (2009), PP.1296-1313. | [19] Siva Selvan .K and Vijayalakshmi C.: Algorithmic Approach For the Design Of Markovian Queueing Network with Multiple Closed Chains International Conference on TRENDZ information Sciences and Computing. Proceedings IEEE xplore, Sathyabama University ,TISC-2010. | [20] Sleptchenko, A. Harten and M. Heijden, An exact solution for the state probabilities of the multi-class, multi-server queue with preemptive priorities Queuing systems, Theory and Applications 50, (2005)81-107. | [21] Tadj, L. and Choudhury, G. (2005). Optimal design and control of queues. Top, 13, 359–412.