A Literature Study on Dynamic Bandwidth Allocation Mechanisms for Tandem Communication Networks

1CH.V. RAGHAVENDRAN, 2G. Naga SATISH, 3M.V. Rama SUNDARI, 4P. Suresh VARMA
1,2Adikavi Nannaya University, Rajahmundry, India
3Dept. of IT, GIFT, Rajahmundry, India
4Dept. of Computer Science, Adikavi Nannaya University, Rajahmundry, India
1raghuchv@yahoo.com

Abstract—Modeling and performance prediction are becoming increasingly important issues in the design and operation of computer communications systems. In this paper a review is carried out on how Tandem queuing models with Dynamic Bandwidth Allocation have been applied so far into the performance evaluation of Communication Networks. Queuing network models with finite/infinite capacity buffers and blocking have been applied as more realistic models of systems with finite capacity resources. First we review basic properties of exponential queuing systems, and then give an overview of recent progress made in the areas of dynamic bandwidth allocation for tandem queuing network models and performance measures.

Keywords: Queuing models, Tandem Communication Networks, Dynamic Bandwidth, Feedback.

I. INTRODUCTION

Queuing theory history goes back more than 100 years. Johannsen’s “Waiting Times and Number of Calls” published in 1907 is the first paper on the subject. Another paper that has historic importance is by A.K. Erlang’s “The Theory of Probabilities and Telephone Conversations” in 1909. This laid the foundation for the Poisson distribution in queuing theory. In the next two decades several theoreticians became interested in these problems and developed general models for complex situations. A detailed account of the investigations made by these authors is found in books by Syski (1960) and Saaty (1961).

Traffic processes in computers and computer networks have necessitated the development of mathematical techniques to analyze them. The first article on queuing networks is by J. Jackson in 1957. Since around 1968 queuing network models (QNM’s) have been used as models of computer and communications systems. This new use has stimulated research into enlarging the class of models which can be solved and into developing new solution techniques [1]. From the beginning of 1970’s, complex queuing network problems have been investigated and several survey papers and books summarized the major contributions in this area.

In the context of computer networks, customers represent packets and service represents the assignment of packets to communication links. If a link has a bandwidth of B then the service time for a packet assigned to that link will be L/B, where L is the size of the packet in bits. Queuing theory is responsible for the study of such systems. This theory is helping us to gain insight about buffer space, packet delays and network utilization. This helps us in design of switching strategies and congestion and mechanisms. The communication systems have been dominated by voice transmission with circuit switching technology [2] for a long time. However, the demand for data communication increased dramatically and drove the development of the Internet, which uses packet switching; an efficient technology for data communication rather than circuit switching. Currently, the mobile communication system is facing the same evolution as the Internet. The network operators are migrating from circuit-switched GSM systems to GPRS and 3G networks [3], which have the packet switched functionality.

II. QUEUING NETWORKS

Mathematicians are studying queues from long time and one of their first applications was to telephone exchanges (Erlang’s loss formula). However they gained popularity with Computer Scientists about 50 years ago when it was realized that single queues and networks of queues could be used as performance models of computer systems. Queueing theory was first used in the late 1960s to model time-sharing computer systems and single queues were used to study allocation policies for CPUs. Queueing networks remain a useful performance modeling paradigm.

The development of queueing network theory, which provided application areas with solutions, formulas, and algorithms, commenced around 1950’s in the area of Operations Research, with importance to production, inventory, and transportation with the works of Jackson [4]. The second breakthrough in queueing network theory was already connected with Computer Science: the celebrated papers of Baskett, Chandy, Muntz, and Palacios [5], and of Kelly [6]. The two volumes of Kleinrock’s book [7, 8] appeared at the same time as [6]. From that time on, queueing network theory and its applications have become intimately connected with performance analysis of complex systems in
Computer and Communications at the hardware and software levels.

Edgar Reich [9] presented property of stationary birth and death processes, which implies that for certain queues; the output process is closely related to the input process. This is applied to a situation where customers proceed to a second queue after having been processed at a first queue. Later L. Takács [10] assumed that customers arrive at a counter in accordance with a Poisson process of density, and arrived by a single server in first come first server manner. Suppose that after being served each customer either immediately joins the queue again with probability p or departs permanently with probability q (p + q = 1).

Michael Shalmon et al. [11] presented a complete analysis of delays for a tandem network of queues with deterministic service and multiple, interfering sources. Packets enter the system at every station, are handed from station to station in store-and-forward fashion, and exit at the end; there are no intermediate departures. Jonathan Brandon et al. [12] considered an N-node tandem queueing network with Bernoulli feedback to the end of the queue of the first node. They derived a formula \( EL(n) \) for the expected queue length seen by a customer at his \( n \)th feedback. Balaji Prabhakar et al. [13] studied the flow of jobs on an infinite series of first-come-first-served queues. Jobs are placed in the buffer of the first queue and allowed to flow through the infinite tandem of queues. The asymptotic flow of jobs on an infinite queueing tandem has been studied under stationary and ergodic service times, given that the jobs are simultaneously released at the first queue. B. Krishna Kumar et al. [14] discussed a retrial queue with Bernoulli feedback and the server is subjected to starting failures. If an arriving customer finds the server busy or down, the customer leaves the service area and enters a group of blocked customer called orbit. Performance measures such as mean system size, the server utilization, mean orbit size, the probability of the server being under repair, the probability of the orbit being empty, etc. are obtained.

Haghighi A.M. et al. [15] considered a novel two-station single-processing tandem queueing system with task splitting, feedback, and blocking – a finite buffer at each station and only the intermediate buffer finite cases. After completing processing at each station, a task may leave the system, join the other station, go for a split, go to the other station or return to its own station. In [16] they considered a two-node single-processor Markovian tandem queueing system with task splitting and feedback. Blocking is not possible because of infinite buffer for each node. Haghighi A M et al. [17] considered a Poisson arrival with exponential single-processor model with splitter, feedback and three buffers. Feedback occurs exponentially with delay and in bounded random batch size 1. The steady-state system of differential-difference equations is solved analytically that reduces to a functional equation.

Walenty Oniszczuk [18] considered an open queueing model with a single task class and three stations – source, station A and B. After service completion at station A, the task proceeds to station B and once it finishes at B, it sent back to A for re-processing with probability \( \sigma \) and leave the network with probability \( 1-\sigma \). The model describes the behaviour of a computer network exposed to an open Markovian queueing model with blocking. Yong Wan Lee [19] considered an M/G/1 single server retrial queueing system with two types of customers and feedback. The customers arrive according to independent Poisson streams with rates \( \lambda_1 \) and \( \lambda_2 \). Josh Reed et al. [20] considered queues in tandem with customer deadlines and retrials. They studied a Markovian system of two queues with blocking at the second queue and used probability generating functions and matrix analytic techniques. Second a non-Markovian setting and derived the stability condition for an approximating diffusion by considering expectations of the first hitting times of compact sets.

### III. COMMUNICATION NETWORKS

The inauguration of commercial telegraph service by William Cooke and Charles Wheatstone in England in 1839 was the first technology to sever the transmission of information from the physical movement of people. In 1896 Guglielmo Marconi introduced wireless signaling for transmission of information from one place to another. In 1948, Shockley, Bardeen and Brattain were invented transistor and published a paper “A mathematical theory of Communication”. This was the foundation for communication satellites and the beginning of digital and data communications.

Traditional communication networks are designed to offer a single type of service, whereas an integrated-services network use packet-switching technology and offer multiple services, including data, voice, video and others. A proactive network control is needed to provide performance guarantees in terms of delay, throughput, delay jitter and loss rate. This kind of network architecture needs service discipline at the switching nodes.

In 1950’s, the United States formed the Advanced Research Project Agency (ARPA) within the Department of Defense and is well-known as ARPANET. Initially this network is constructed with four nodes – University of California at Los Angeles, SRI in Stanford, University of California at Santa Barbara, and University of Utah. Jon Postel, Steve Crocker and Vinton Cerf developed the packet switching principle of the ARPANET. The Network Control Protocol (NCP) was the first protocol that connects all hosts in ARPANET. By the mid 1970s, the research networking has grown rapidly and that ARPANET would need to connect to other networks. As other networks uses a different protocol, Robert Khan and Vinton Cerf started to design a network overlapping protocol. In May 1974 a workgroup
with the name “A Protocol for Packet Network Intercommunication” published, and in December the first RFC (RFC675) “Specification of Internet Transmission Control Protocol” edited. TCP was created and enabled the expansion to a World Wide Internet. In 1978 TCP split into a pair of protocols called Transport Control Protocol (TCP) and Internet Protocol (IP).

Comitate Consultative International Telephonique Telegraphs (CCITT) has defined Integrated Services Digital Network (ISDN) in 1988 in its red book [21]. This is support a wide range of voice (telephone calls) and non-voice (data exchange) applications in the same network. ISDN is a set of communication standards for simultaneous digital transmission of voice, video, data and other network services over the traditional circuits of the public switched telephone network.

J.C.R. Licklider & W. Clark [22] proposed directions in which advances can be made and described on-going programs that seek to improve man-machine interaction in teaching and learning, in planning and design, and in visualizing the internal processes of computers. They discussed basic problems involved in improving man-computer communication. Leonard Kleinrock et al. [23] studied the specification, analysis and evaluation of some hierarchical routing procedures which are effective for large store-and-forward packet-switched computer networks. Optimal clustering structures are determined to minimize the length of the routing tables.

According to Caglan M. Aras et al. [24] the dramatically increased bandwidths and processing capabilities of future high-speed networks make possible many distributed real-time applications, such as sensor-based applications and multimedia services. Since these applications will have traffic characteristics and performance requirements, new communication network architectures and protocols will be required. Lefelhocz et al. [25] claim that packet scheduling, buffer management, feedback, and source algorithms are four necessary and sufficient components for providing better best-effort services. They proposed a general design principle: the network should manage and distribute its resources through packet scheduling and buffer management and give the best possible explicit feedback. Dimitrios Miras [26] focused on audio and video application and presented a detailed analysis of the end-to-end performance requirements of applications like audio-video conferencing, voice over IP and streaming of high quality audio and video, and gives an overview of the adaptation choices available to these applications so that they can operate within a wider range of network conditions.

B. Filipowicz and J. Kwiecień [27] describes queuing systems and queuing networks which are successfully used for performance analysis of different systems such as computer, communications, transportation networks and manufacturing. They incorporated classical Markovian systems with exponential service times and a Poisson arrival process, and queuing systems with individual service. Aderemi A. Atayero et al. [28] studied a Markov model describing the duration of error intervals and error-free reception for streaming video transmission and was developed based on the experimental data obtained as a result of streaming video from a mobile source on IEEE 802.16 standard network. Koorosh Firouzbakht et al. [29] used a game theoretic approach to formulate communication between two nodes over a wireless link in the presence of an adversary. In this model, the transmitter’s goal is to maximize the achievable expected performance of the communication link, defined by a utility function, while the jammer’s goal is to minimize the same utility function.

IV. DYNAMIC BANDWIDTH ALLOCATION IN COMMUNICATION NETWORKS

Very little work has been reported in literature regarding load dependent transmission and Transient Analysis of Tandem Communication Networks which are very useful for accurate predictions of the performance measures. Network resources such as bandwidth of physical link, buffer space and processing time of each node should be allocated in cost-effective manner.

Each application expects the network to provide a desired Quality of Service (QoS). In a load dependent communication network the transmission depends on the content of buffer at that instant, i.e. on the number of packets in the buffer. This type of queuing system is analyzed by Conway R.W and Maxwell W.L. [30]. They considered a single server queueing system with Poisson arrival and service times having a state dependent exponential distribution. Carl M. Harris, [31] considered a Poisson bulk-arrival single-channel first-come-first-served queuing system with service times are conditioned on the length of the queue at the moment service is begin. Gilles Davignon et al. [32] presented some results for queues with a state dependent feedback decision maker. They studied some of the properties of a single server queue of M/G/1 type with a feedback mechanism that may depend on the “state” of the system and, in a Markov manner, the “history” of previous feedback decisions. They discussed the queue length process, the busy period process, the output process and the departure process.

Brown T.C. and Pollett P.K. [33] considered a single server Markovian queueing networks with state dependent service rates. The distance of customer flows from Poisson process was estimated in both open and closed queues. I.F. Akyildiz et al. [34] presented a method for approximate solutions to load dependent closed queueing networks containing general service time distributions and FCFS scheduling disciplines. A new formula for the conditional throughputs is derived. Parthasarathy, P.R. and Selvaraju, N [35] obtained a transient solution analytically using continued fractions for a state dependent birth-death queue in which potential customers are discouraged by the queue length.
They compare this model with the infinite server queueing system.

Zuyuan Fang et al. [36] studied the problem and theoretical aspects of fair bandwidth sharing between nodes at the link level in mobile ad-hoc networks through a game theoretic approach. They presented some applications to show how such a framework can be invoked to design efficient media access control protocols in a non-cooperative, self-organized, topology-blind environment as well as in environments where the competing nodes share some basic information to guide their choice of channel access policies. F. R. B. Cruz et al. [37] analyzed the problem of service and capacity allocation in state dependent M/G/c/c queueing networks and developed algorithms to compute the optimal allocation c. They evaluated the measurements like, the blocking probability, throughput, and the expected number of customers in the system and expected travel (service) time for pedestrian planning evacuation problems in building.

P. Suresh Varma et al. [38] developed a load dependent Tandem Queueing Model with two nodes to the communication network with the integrated data/voice packetized statistical multiplexing. They assumed that the messages are packetized at source and stored in buffer for transmission. After transmitted in the first node, it is being transmitted through the second node and the transmission is carried with load dependent strategy. It was observed that the load dependent transmission can reduce the delays in transmission and enhance the channel capacity.

Rama Subramanian S et al. [39] developed a framework for fair bandwidth sharing called Access Mechanism for Efficient Sharing (AMES), where every core router in the network employs link-specific queuing, round-robin scheduling, and reacts to congestion by restricting the adjacent routers from transmitting, rather than dropping packets. According to S. Vijay Bhanu et al. [40] less bandwidth utilization is the key reasons for reduced number of channel accesses in VOIP. But the free bandwidth of at least 1-5% will improve the voice quality in VOIP. Their proposal utilizes the maximum bandwidth. They proposed a Bandwidth Data rate Moderation (BDM) algorithm that correlates the data rate specified in IEEE802.11b with the free bandwidth. Before sending the packet, the BDM algorithm will calculate the bandwidth utilization to improve performance and voice quality of VoIP. Erlang and VOIP bandwidth calculator is used for calculating the bandwidth.

K. Srinivasa Rao et al. [41] developed and analyzed a two node tandem communication network with dynamic bandwidth allocation (DBA) having two stage direct bulk arrivals. The arrival and transmission processes at each node are characterized through compound Poisson and Poisson processes such that several of the statistical characteristics of communication networks identically matches. K. Srinivasa Rao et al. [42] introduced a method for performance evaluation and stochastic control of the two node tandem communication system with phase type transmissions. The massages arrive to the communication network are converted into a random number of packets and stored in buffers for forward transmission. The arrival process of the communication network is well characterized by compound binomial Poisson processes such that the bulk arriving nature is induced in the analysis. They observed that the arrival distribution parameters and Dynamic Bandwidth Allocation strategy have significant influence on the performance measures of the network.

M.V.Rama Sundari et al. [43] introduced a three node communication network model with non homogeneous Poisson arrivals having dynamic bandwidth allocation under modified phase type transmission is introduced for performance evaluation and monitoring of several tele and satellite communications. They showed that the dynamic bandwidth allocation can reduce congestion in buffers and delay in transmission by utilizing idle bandwidth. M.V.Rama Sundari et al. [44] developed and analyzed a three node communication network model with the assumption that the arrivals are characterized by non homogenous Poisson process. It was assumed that transmission time required by each packet at each node is dependent on the content of the buffer connected to it.

A.V.S. Suhasini et al. [45] developed and analyzed a two-node tandem queueing model with the assumption that the arrivals follows a non-homogeneous compound Poisson process, and the service rates of each node depends on the number of customers in the queue connected it. Using the difference-differential equations and a probability generating function of the number of customers in the queue and in the system is analyzed. Remco Germs et al. [46] considered a general state-dependent finite-buffer bulk queue in which the rates and batch sizes of arrivals and services are allowed to depend on the number of customers in queue and service batch sizes. They developed a unifying method to study the performance of this system. Based on the limiting probabilities, they presented various performance measures for evaluating admission control and batch service policies, such as the loss probability for an arriving group of customers and for individual customers within a group.

N. Thirupathi Rao et al. [47] developed and analyzed a two node communication network where nodes in series make the model applicable to any of the modern communication networks. They assumed that the input data arrival is in batch and the batch size distribution is zero truncated binomial distribution. N. Thirupathi Rao et al. [48] developed and analyzed the optimal operating policies of a two-transmitter communication network. In this the arrivals of the network are characterized by compound binomial Poisson process and transmission of both the transmitters is characterized by Poisson process. With suitable cost considerations, the expected total revenue function is derived and maximized with respect to the mean arrival rate, mean transmission rates. They observed that the optimal policies are highly influenced by the input parameters and costs.
Communication networks are developed to evaluate the performance of the communication system in particular with DBA is difficult and complicated. Hence, mathematical models of communication networks are developed to evaluate the performance of the newly proposed communication network model under transient conditions.

V. CONCLUSION

Communication network modeling is one of the thrust areas of computer science and systems engineering which utilizes the resources more efficiently and effectively. The wide development of broadband, reliable and cost efficient networks is pushing towards a major paradigm shift in telecommunication world. The wired as well as wireless telecom industries are both heading towards modern methods of evaluation of their traffic. The IP networks are packet switched rather than circuit switched. The resources of IP networks are different from those of circuit switched networks. The non blocking nature of packet networks requires adding a call admission control (CAC) component. The introduction of IP Telephony and wide spread use of wireless technology has significantly affected the communication network usage and traffic behavior.

Time dependent arrivals and dynamic bandwidth allocation strategies can reduce the burstiness in the buffers and mean delay in transmission through congestion control by utilizing the vacant bandwidth. Models based on this approach are much helpful for maintaining and monitoring the communication systems by utilizing resources more optimally. It is highly desirable for developing many more communication network models with credible conditions in order to utilize the resources more optimally and to understand the natural phenomenon of the systems more close to the reality.

REFERENCES


CH. V. Raghavendra is a research scholar in Computer Science Department of Adikavi Nannaya University, Rajahmundry, AP, India. His is working as an Associate Professor in Ideal College, Kakinada, AP, India. He has presented several papers in various National and International Conferences. His areas of interest are Communication Networks, Mobile Ad hoc Networks, Swarm Intelligence and Data Mining.

G. Naga Satish is a research scholar in Computer Science Department of Adikavi Nannaya University, Rajahmundry, AP. At present he is working as an Associate Professor in Ideal College, Kakinada, AP, India. He has presented and published papers in National and International Conferences. His areas of interest include Computer Networks and Network Security.

Dr. M.V. Rama Sundari is working as an Associate Professor, Department of IT, Godavari Institute of Engineering and Technology, Rajahmundry, AP, India. She published several research papers in national and international journals.

Dr. P. Suresh Varma is a Professor in the Department of Computer Science, Adikavi Nannaya University, Rajahmundry, India. He published several research papers in national and international journals and presented papers at various conferences, seminars and workshops. His current research interests are Communication Networks, Data Mining and Cloud Computing.