Military Technical College
Kobry Elkobbah,
Cairo, Egypt
May 27-29,2008

4th International Conference on Mathematics and Engineering Physics (ICMEP-4)

EM-6

ATM NETWORKS CONGESTION CONTROL OPTIMIZATION USING STOCHASTIC PROGRAMMING

OSAMA I. SHAARAWY, ELIWA M. ROSHDY, EL-SAYED E. SOROUR

ABSTRACT.

The Asynchronous Transfer Mode (ATM) is a major transport technology to support the Broadband Integrated Services Digital Networks (B-ISDN). It supports variety of network services such as World Wide Web, videoconference, e-commerce, network multimedia, traditional file transfer protocol (ftp), email, newsgroup and etc. Many network applications not only ask for high speed, broadband, but also require high quality of services (QoS). More and more research works are focused on ATM networks. A number of researchers have demonstrated the virtual path concept. The advantage of this concept is that it allows a large group of virtual circuits to be handled and switched together, resulting in a lower computing complexity, faster processing speed, and an efficient use of network resources. The ATM network provides multiple services, a significant amount of traffic flows have stochastic characteristics; that makes it difficult to solve congestion control problem.

Stochastic programming (SP) is introduced to solve the essential problems of ATM network congestion control, Scenario Tracking (ST) approach of stochastic programming employed for ATM network optimization. Three optimization models: Capacity Assignment (CA), Capacity and Flow Assignment (CFA), and Flow Assignment (FA) are discussed; a proposed virtual path based CFA model is then simulated.

1. INTRODUCTION

Asynchronous Transfer Mode (ATM) is an ITU-T standard for cell relay, which provides multiple services, such as voice, video, and data, all that is conveyed in small, fixed-size cells. ATM is a cell-switching and multiplexing technology that combines the benefits of circuit switching (guaranteed capacity and constant transmission delay) with those of packet switching (flexibility and efficiency for intermittent traffic)[1]. It provides scalable bandwidth from a few megabits per second (Mbps) to many gigabits per second (Gbps).

ATM networks are fundamentally connection-oriented, which means that a virtual channel (VC) must be set up across the ATM network prior to any data transfer [5]. Two types of ATM connections exist: virtual paths (VP), which are identified by virtual path identifiers (VPI), and virtual channels (VC), which are identified by the combination of a VPI and a virtual channel identifier (VCI). A virtual path is a bundle of virtual channels, all of which are switched transparently across the ATM network on the basis of the common VPI.

Figure 1 Virtual paths (VP) and Virtual channels (VC)[1].

The basic operation of an ATM switch is straightforward: The cell is received across a link on a known VCI or VPI value. The switch looks up the connection value in a local translation table to determine the outgoing port (or ports) of the connection and the new VPI/VCI value of the connection on that link. The switch then retransmits the cell on that outgoing link with the appropriate connection identifiers. Because all VCIs and VPIs have only local significance across a particular link, these values are remapped, as necessary, at each switch.

Congestion control is important in any packet communication system. In ATM networks, congestion is the state that the network is unable to provide the guaranteed QoS to established connections [4][6].

Since ATM networks typically work with virtual paths (VP) and virtual channels (VC), conventional based on link-node models are not feasible to ATM network optimization any longer. As mentioned before, after virtual path connection (VPC) established, virtual channel connections (VCC) are established upon requests, the procedures just like a phone call connection. So in this paper, we pay more attention to build a simple and theoretical model, and the detail VPC/VCC operation are too complex, for convenience, we just induce virtual path (VP) conception. Consequently, the concepts of path flow and path capacity are proposed to characterize the path-oriented networks. And in order to deal with the probabilistic infeasibility cases occurring in network operation, the stochastic programming SP methodology is introduced.[2]

2. STOCHASTIC PROGRAMMING IN ATM NETWORKS

In a conventional communication paradigm, the traffic passing through a network usually does not change significantly within a period, so a generic model can be used for a common problem. In a network model, the parameters needed in such formulations are also presented in terms of mean values such as the average arrival rate, the average interarrival time, or the average cost. However, this method is now greatly challenged by new communication modes represented by asynchronous transfer mode (ATM). An ATM network can support diverse rates applications each with a different requirement of QoS.
The major variety of ATM services proposed by ITU and ATM forum are[3]:

- Constant Bit Rate (CBR), also known as “Deterministic” Bit Rate
- Variable Bit Rate real time traffic (RT-VBR), Variable Bit Rate non-real time traffic (NRT-VBR), also known as “Statistical” Bit Rate
- Available Bit Rate (ABR)
- Unspecified Bit Rate (UBR).

The QoS can be specified by the following parameters: Cell Delay Variation (CDV), Maximum Cell Transfer Delay (MCTD), Mean Cell Transfer Delay (Mean CTD), Cell Loss Rate (CLR)

In general, traffic flows in modern communication networks are not only dynamic but also stochastic. In these circumstances, the approach that uses mean values of parameters may be no longer valid in many cases. The SP theory has been developed to primarily deal with that type of problems assuming that the probability density function (PDF) or the cumulative distribution function (CDF) of anticipation is known with certainty

All of the generic network models can be abstracted as a mathematical programming problem:

$$\begin{aligned} &\text{minimize} \quad f(X, a) \\ &\text{subject to} \quad g(X, b) \geq 0 \end{aligned}$$

where f: $\mathrm{IR}^n \rightarrow \mathrm{IR}$, g: $\mathrm{IR}^n \rightarrow \mathrm{IR}^m$. X is called the decision variable. $f(X)$ is called the objective function and $g(X)$ the constraint function. The feasible region of X is represented by mixed set of the solutions of both $f(X)$ and $g(X)$, expressed as a parameter set $\Omega = \{a, b\}$.
In SP formulation, subsets of the parameters are considered being random variables:

$$\begin{aligned} &\text{minimize} \quad f(X, a_d, a_u) \\ &\text{subject to} \quad g(X, b_d, b_u) \geq 0 \end{aligned}$$

where the subscripts d, u stand for deterministic and uncertain, respectively. A scenario is defined to be a particular joint realization of a set of uncertain parameters.

In practice, most network problems are analyzed with simulation techniques rather than analytical models. Consequently, the *wait-and-see* approach seems more flexible than the *here-and-now* one, since the latter needs analytical formulas of distributions. Another alternative approach is the *Scenario Tracking* (ST) [10], in which the stochastic programming problem is converted to a number of standard mathematical programming sub-problems. It takes scenarios (samples) of stochastic parameters, and based on these scenario values, a number of scenario sub-problems are constructed. Optimization then is performed for these scenario sub-problems. Finally, based on all scenario solutions, a coordinating model is constructed, and another time optimization is performed to track the scenario solutions to get single optimal solution.

Let a_s and b_s represent a particular joint realization of uncertain parameters, a_u and b_u, respectively. For each scenario $s \in S \equiv$ {set of all scenarios}, the SP formulations, expressed above, reduces to a deterministic problem shown below, which is referred to as the *scenario sub-problem*:

$$\begin{aligned} &\text{minimize} \quad f(X, a_d, a_s) = y_s \\ &\text{subject to} \quad g(X, b_d, b_s) \geq 0 \end{aligned}$$

For each scenario, there is a corresponding probability p_s. In such an environment, a solution x to a stochastic nonlinear system $g(X, b_d, b_s) \geq 0$ is said to be feasible if it minimizes:

$$\sum_s \mathrm{p_s} \left[\left\| \min(0, g(X, b_d, b_s \right\| \right]$$

Accordingly, a coordination model can be constructed as:

$$\text{minimize} \sum_s \mathrm{p_s} \left[\left\| f(X, a_d, a_s) - y_s \right\| + \left\| \min(0, g(X, b_d, b_s)) \right\| \right]$$

The coordinating model tracks the scenario solutions as closely as possible while still maintaining feasibility and this model is referred as tracking model.

Mathematically, a network optimization problem with the path-node incidence can be described by the multi-commodity (MC) model, a representative paradigm developed in the theory of network flows [7]. For a B-ISDN paradigm, the concept of commodity classes associated with traffic classes. In modern communication technologies, the Poisson model had played a dominant role in for several decades until the 1990s. Then several studies argued that some applications in packet data networks might not follow the Poisson model very well [8][9].

3. NETWORK CONGESTION CONTROL MODELS

To archive a new optimization model of ATM network congestion control, a group of wide used network models are reviewed.
A communication network can be expressed by a graph $G = (V, A)$, where V is the set of nodes (vertices) and A is the set of arcs (or links, edges).
Most packet communication networks can be regarded as an augmented graph with two types of indices: the operating index, expressed by T, e.g. packets delay or the number of packets in the system, generally speaking, network operation cost; the capital index, expressed by D, e.g. capacities cost. In a communication network, two important entities must be considered: link flow λ_i and link transmission capacity C_i. The physical meaning of λ_i is the traffic arrival rate in link i, expressed as data units per second, and C_i has the same dimension as λ_i. Accordingly, to a given network topology, the external traffic requirements, and the constraints of λ_i and C_i, the following three models can be developed:[2]

Capacity Assignment (CA) problem:
- Given λ_i
- Minimize T
- Adjust C_i
- Constraint D

Flow Assignment (FA) problem:
- Given C_i
- Minimize T
- Adjust λ_i

Capacity Flow Assignment (CFA) problem:
- Minimize T
- Adjust C_i and λ_i
- Constraint D

Several important issues may be discovered with further analysis of these generic models as flow:
■ Whether a model is CA, FA, or CFA, depends on what the design variables are, rather than what variables are presented in the objective function.
■ An objective function should be able to reflect the effect of design variables; generally speaking, all design variables should be integrated into the objective function.
■ In the formulations of the CFA problems, the roles of objective function and constraint function can be exchanged, thus the terminology of primary and dual formulations is induced, the two forms can be called dual each other. However, the capital index cannot be utilized as the objective function for the FA model; otherwise adjusting λ_i has no effect to reduce the objective function.

4. PROPOSED FLOW ASSIGNMENT PATH-NODE MODEL

The parameters employed in path-node models are as follows:

W : the set of all OD pairs.
P_w: the set of all paths that connect a particular OD pair w.
L : the set of all links.
C_i: the capacity of link i.
Q_i: the set of all paths that pass link i.
d_p: the unit flow cost of path p.
x_p: the flow of path p.

b_p: the unit capacity cost of path p.
G_p: the capacity of path p.
x_p/G_p: the path utilization factor.
γ_w: External traffic demand which will be a random variable with uniform PDF.

Model PF-1
The objective function and constraint functions are illustrated below:

$$\text{minimize} \qquad y = \sum_{w \in W} \sum_{p \in P_w} d_p \left(\frac{x_p}{G_p}\right) x_p$$

$$\text{subject to} \qquad \sum_{p \in P_w} x_p = \gamma_w$$

$$x_p \leq G_p$$

$$\sum_{p \in Q_i} G_p \leq C_i$$

$$x_p \geq 0$$

$$\forall p \in P_w, w \in W, i \in L$$

In this model, the target function is the same to that of CFA model, but the fundamental difference is: that in FA, only x_p is the design variable, G_p is a parameter; while in CFA both x_p and G_p are design variables. By introducing VP flow and VP capacity the path-node FA model becomes a CFA model on logical network level, but on physical network level, it still remains an FA model.

Note that objective function of model PF-1 incorporates the path utilization factor and the unit path flow cost d_p, in the VP flow model. The path utilization factor x_p/G_p introduces the degree of congestion into the objective function. Obviously, when x_p/G_p approaches 1, the VP flow is close to the VP capacity. Hence, the delay and the possibility of congestion in this VP increase. By incorporating the utilization factor x_p/G_p into the objective function, we distribute the traffic between an OD pair on all available VPs evenly, and decrease the possibility of congestion. In this model, because the path utilization factor x_p/G_p is kept small while minimizing the total flow delay, the resulting VP system can be resilient to input traffic changes. Therefore, possibility of congestion is likely to be decreased. Since traffic is distributed evenly on all available VPs between OD pairs, the highest link load can be minimized, and the traffic can be distributed evenly on all links.

In ATM networks, the higher the maximum load on any specific link, the more catastrophic may be the effect of the failure of the link. If traffic can be distributed evenly on all links and highest link load can be reduced, the network will have high robustness. Therefore, the resulting VP system has high robustness to physical link failure.

5. WEIGHTED SCENARIO TRACKING MODEL

The *Scenario Tracking* (ST) approach is used here to solve the VP distribution problem. A weighted scenario tracking scheme is proposed here. The ST method does not need analytical formulas of distribution of stochastic parameters and can find a single optimal solution. These features make ST method more feasible and flexible for VP optimization when comparing with other approaches.

In this thesis, we apply the ST scheme to proposed FA model PF-1 to solve the VP optimization problem. Here, scenario values of γ_w are taken, and for every scenario, solve the scenario sub-problem, which is indeed a deterministic problem. After we get solutions for all scenarios, construct a coordinating model based on all scenarios to track the scenario solutions as closely as possible while still maintaining feasibility. Finally, the optimal solution is calculated based on the tracking model. The tracking model consists of two parts: "*objective value*" and "*feasibility penalty*" cost functions.

Objective value

The objective value cost function is used to track all scenario solutions by taking summation of difference between the final objective value and every scenario objective value.

Let the final optimal objective value be denoted by $G = \sum_{w \in W} \sum_{p \in P_w} d_p(\frac{x_p}{G_p}) x_p$, then

$$Z_{obj} = \left| G - y_s \right|$$

where S: the set of all scenarios; y_s: objective function value at scenario s; the remaining notations are same as defined in subsection 3.5.3 Chapter 3.

This objective value part Z_{obj} shown above is used to track all scenario solutions to find the final single solution by minimizing summation of difference between the final solution and every scenario objective value. This part is to track the scenario solutions as closely as possible.

Feasibility Penalty

This cost function is used to reflect the feasibility into the tracking model by summating the negative distance between the final solution and each scenario input flow. Let the lowest difference between the final solution and scenario input flow is denoted by $D_s = \min[((\sum_{p \in P_w} x_p) - \gamma_w), w \in W]$, then

$$Z_{penalty} = \left| \min(0, D_s) \right|$$

Here D_s is used to determine the lowest difference between the final solution and every scenario input flow γ_w. Note that in networks, only when path flows between an OD pair cannot meet external input traffic flows $(\sum_{p \in P_w} x_p < \gamma_w)$, the feasibility of the flow assignment is violated. Hence, we count only these negative distances here. We then compare the lowest value D_s with 0 to get the summation of all negative distances. Finally, we get absolute value of this summation. By doing this, we reflect the feasibility into tracking model by minimizing the largest negative distance between final solution and every scenario input flow γ_w. By adding the feasibility penalty into the tracking model, we are trying to maintain the feasibility while tracking the scenario solutions.

Hence, the complete tracking model can be constructed as follows:

minimize $\sum_{s \in S} p_s (aZ_{obj} + bZ_{penalty})$

The fully extended form of the objective value is:

$$\sum_{s \in S} p_s [a \left| \sum_{w \in W} \sum_{p \in P_w} d_p(\frac{x_p}{G_p}) x_p - y_s \right| + b \left| \min(0, \min[(\sum_{p \in P_w} x_p) - \gamma_w]] \right|$$

subject to

$$x_p \le G_p$$

$$\sum_{p \in Q_i} G_p \le C_i$$

$$x_p \geq 0$$

$$\forall s \in S, p \in P_w, w \in W, i \in L$$

where p_s: probability of scenario s; a: weighting factor of objective value tracking; b: weighting factor of feasibility tracking.

We apply the ST method to model PF-1, and then compare the resulting VP system with the VP system obtained from mean values.

In our research, we assume that the external traffic demands; expressed in terms of bandwidth, change randomly. A random number generator is used to generate sets of traffic demands with Poisson distribution. Each group corresponds to external traffic demand of an OD pair, and each group includes numbers corresponding to scenarios.

In ST approach, the more scenarios we use the higher accuracy we obtain. However, if we choose a large number of scenarios, the computation will take long time and the tracking model will be difficult to solve.

6. SIMULATION AND ANALYSIS

A prototype network with 4 nodes, 4 links and 6 OD pairs is shown in Fig. (2).

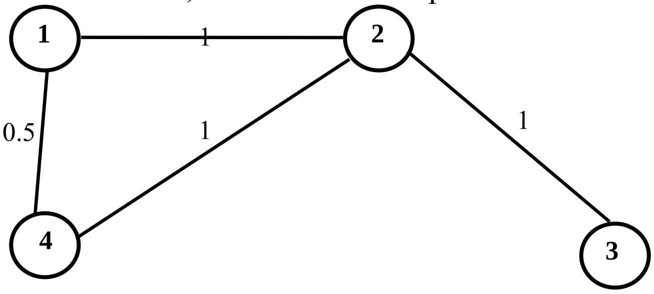

Figure 2 Small size prototype network

In the network, we consider at most two VPs for each OD pair, so there are 11 paths, which are shown in Table (1). Table (1) shows the all paths associated with OD pairs and corresponding path cost. Table (2) shows the path / link incidence. Table (3) shows a few selected network parameters that we are using in the simulation.

Table 1 OD Pairs – Paths and costs (small prototype network)

OD Pairs		Paths		Path cost (distance)
1	(1,2)	P1	{1,2}	1
		P2	{1,4,2}	1.5
2	(1,3)	P3	{1,2,3}	2
		P4	{1,4,2,3}	2.5
3	(1,4)	P5	{1,4}	0.5
		P6	{1,2,4}	2
4	(2,3)	P7	{2,3}	1
5	(2,4)	P8	{2,4}	1
		P9	{2,1,4}	1.5
6	(3,4)	P10	{3,2,4}	2
		P11	{3,2,1,4)	2.5

Table 2 Links – Paths (small prototype network)

	Path										
Link	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Link1 (1,2)	1		1			1			1		1
Link2 (2,3)			1	1			1			1	1
Link3 (2,4)		1		1		1		1		1	
Link4 (1,4)		1		1	1				1		1

Table 3 Network model parameters (small prototype network)

d_p, $p = 1,\ldots, 11$				C_i, $i = 1,\ldots, 4$		γ_w, $w = 1,\ldots, 6$	
d_1	1	d_8	1	C_1	35	γ_1	13
d_2	1.5	d_9	1.5	C_2	38	γ_2	10.2
d_3	2	d_{10}	2	C_3	20	γ_3	13
d_4	2.5	d_{11}	2.5	C_4	40	γ_4	8
d_5	0.5					γ_5	10.2
d_6	2					γ_6	10.2
d_7	1						

For the prototype network in Fig.(2), the model PF-1 can be formulated as follows:

$$\text{minimize} \quad y = \sum_{p=1}^{11} d_p \left(\frac{x_p}{G_p}\right) x_p$$

subject to

$$x_1 + x_2 = \gamma_1$$

$$x_3 + x_4 = \gamma_2$$

$$x_5 + x_6 = \gamma_3$$

$$x_7 = \gamma_4$$

$$x_8 + x_9 = \gamma_3$$

$$x_{10} + x_{11} = \gamma_3$$

$$x_p \le G_p,\ p = 1, \ldots, 11$$

$$G_1 + G_3 + G_6 + G_9 + G_{11} \le C_1$$

$$G_3 + G_4 + G_7 + G_{10} + G_{11} \le C_2$$

$$G_2 + G_4 + G_6 + G_8 + G_{10} \le C_3$$

$$G_2 + G_4 + G_5 + G_9 + G_{11} \le C_4$$

$$x_p \ge 0,\ p = 1,\ldots 11$$

The above model is optimized using the parameters given in Table (3). After converting the optimization problem into GAMS software (General algebraic modeling system) and running, we get the optimized VP flows and VP capacities shown in Table (4).

Table 4 Optimized VP flows and capacities (small prototype network).

Path flow x_p		Path capacity G_p	
x_1	13	G_1	17.44
x_2	0	G_2	0
x_3	10.2	G_3	13.96
x_4	0	G_4	0
x_5	13	G_5	36.41
x_6	0	G_6	0
x_7	8	G_7	11.18
x_8	10.2	G_8	10.74
x_9	0	G_9	0
x_{10}	7.78	G_{10}	9.26
x_{11}	2.42	G_{11}	3.59

The minimized objective function $y = \sum_{p=1}^{11} d_p\left(\frac{x_p}{G_p}\right) x_p = 59.47$

It is observed that in most cases, the path utilization factor x_p / G_p is low for high traffic load VPs. In this way, the traffic can be distributed in the network evenly, and each VP is assigned some redundant capacity. Therefore, the assigned VP is resilient to external traffic changes.
For the ST tracking method, firstly, we perform 64 times optimization for 64 scenarios. Secondly, based on the tracking model, we perform another time of optimization to get the optimal solution. After converting the optimization problem into GAMS and running, we get the optimized VP flows and VP capacities. The optimal solution is compared with the solution obtained from applying mean values of external traffic demands.

7. CONCLUSION

The idea of ST method is to construct a tracking model to track the scenario solutions as closely as possible while still maintaining feasibility.

Based on the above simulation, it is observed that the ST approach shows great flexibility on solving stochastic programming problem. Based on different objectives, we can construct different tracking model. By adding two weighing factor *a* and *b* into the tracking model, we provide flexibility for users to change the weight of objective value part and feasibility part in the tracking model to meet their objectives.

The simulation results show that the stochastic programming solution for the VP optimization problem, by applying the ST approach to our proposed FA model, has better performance than mean value approach in terms of more evenly distributed traffic in the network and fewer failures to accommodate external traffic changes.

REFERENCES

[1] K. Sato, S. Ohta, and I. Torizwa, Broadband ATM network architecture based on virtual path. IEEE Transactions on communications, 1990.

[2] I. Chlamtac, A. Farago and T. Zhang, Optimizing the system on virtual paths. IEEE/ACM Transactions on networking. pp 581-587, Vol.2, December 1994.

[3] A. Alleles, ATM internetworking. Cisco Systems, Inc. report, 1995.

[4] G. Kesidis, ATM Network Performance. (second edition), pp 2, 2000.

[5] P.S. Neellakanta, ATM Telecommunication's principles and implementation, pp 335-400, 2000.

[6] O. kyas, ATM Networks (second edition), pp 151-177, 1996.

[7] M. Bizaar, and J. Jarvis, Linear programming and network flows (Wiley, New York 1977)

[8] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, On the self-similar nature of Ethernet traffic (Extended Version), IEEE/ACM Transactions on networking, pp. 1-15, 1994.

[9] M. Crovella and A. Bestavros, Self-similarity in World Wide Web traffic: evidence and possible causes. IEEE/ACM Transactions on networking, pp. 835-846, 1997.

[10] Y. Xiong, L. G. Mason, Restoration strategies and spare capacity requirements in self-Healing ATM Networks. IEEE/ACM Transactions on networking, vol.7, Feb, 1999.