A Differential Measurement(DM) of Congestion Level in ATM Network

Dr. Hamad Abu Zaied

Philadelphia University

Summary
A differential measurement of congestion level is introduced to enhance the cost function to be applied in routing strategy inorder to maximize the utilization of the installed resources as well as balancing the load in ATM networks is introduced. The differential measurement of congestion level is based on computing the number of in-progress calls within each reallocation period several times, which leads to find out the rate of loading. At each node (ATM switch) there is a controller for computing the occupied capacity, the residual capacity and the available capacity to prevent any call rejection if there is sufficient amount of bandwidth supporting the arriving calls. When a new call requests a connection, the bandwidth allocation controller assigns for it the required bandwidth, if there is enough resources, otherwise it will be rejected. After the call is completed, the assigned bandwidth will be released and returned back to the remaining amount of capacity, which we will call the channel reserve capacity. Once the call is accepted, the cost function will guide it to follow the least congested route taking into account the congestion level and the rate of loading at the next node. Some simulation results are presented.

Key words:
traffic, cost function, congestion level, residual bandwidth.

1. Introduction

The traffic on the ATM node (switch) is classified into a number H of classes. Each class is characterized by its own statistical parameters and QoS requirements. In our approach the ATM switch is considered as a set of queues of finite size. The cells are picked out by a scheduler, complying with a routing algorithm, which is governed by a switch control device.

The proposed connection admission control will play as follows: When a new call requests a connection, the bandwidth allocation controller assigns for it the required bandwidth from the channel available reserve capacity, if there is enough bandwidth [1,2,3]. Otherwise the call will be rejected. The described procedure will be repeated for all the outgoing links. If only one link can accept the new call, then it will be shipped over it. If none of the outgoing links can accept the new call, then the call will be rejected. This case can take place if and only if the total capacity is occupied, or the total reserves can not cope with the new call requirements or the set of the next nodes are congested. If more than one link can accept the new call, then we need to differentiate between the available links (routes). The differentiation process is based on the congestion level of all the outgoing links taking into account the channel capacity, the link traffic being in progress, the congestion level at the next node at the most close previous instants, and on the new call requirements[4-8].

One of the main aims of the introduced algorithm is to balance the bandwidth utilization among all of the outgoing links. By defining the residual capacity over the possible routes, the controller will select the one with the maximum residual capacity. As a result of this, the traffic will be distributed equally among the routes, which means that we will not have overuse resources once we have underused ones. Consequently we get a uniform distribution of the traffic over the network.

Node-by-node Call Admission Control (CAC) and routing can be joined and managed together. In our approach, the best route is not chosen beforehand and then verified, but at each node, we choose the least loaded link among the outgoing links by means of the cost (residual capacity) function associated with each link [9-12,17]. The cost function should take into account the link's local traffic, the traffic associated with the subsequent hops to the destination, as well as the traffic originated from different types. It should be considered only for those links that are not congested (in the sense that it is defined by the admission control rule). A bit-stream in the messenger packet is used to remember the nodes traversed along the route [18]. At each node, a table look-up is performed to find out the congested channels. The link with the lowest cost (lowest utilization) among the non-congested ones is chosen, and the packet is sent along the route. If all outgoing links become congested, at an intermediate node, then the connection can be saved from rejection by tracking back to a previously visited node and requests the connection again [19-21]. If the messenger reaches its destination, then the connection is accepted, a Virtual Connection (VC) is established, and the path that the cells of that connection will follow is fixed. After the connection is completed, the resources

Manuscript received December 5, 2007
Manuscript revised December 20, 2007

can be released, the assigned bandwidth becomes idle and so the bandwidth allocation controller will return it to the reserve capacity.

In the next section, we discuss the proposed differential measurement of the congestion level. Section 3 presents the network performance and simulation results. Section 4 contains the conclusion.

2. The Proposed Congestion Level Differential Measurement

2.1 The applied Cost Function

The Dynamic Least Congested Path (DLCP) algorithms' family decisions are based on the computation of a simple cost function, related to each outgoing link, given by the sum of a local cost and an aggregate cost [22-25]. Let i be the considered node; then the cost of link ij for class h at decision instant k (in slots) is defined as follows. At instant $\tilde{k}$ (in slots), a generic node i chooses the link through which it will forward a call request packet generated by a class h connection request, by minimizing (over all successor nodes j) the cost function as described below:

The bandwidth allocation controller periodically computes the occupied capacity by the current calls for each class $V_{ij,o}^{(h)}(\tilde{k})$, then calculates the sum of the occupied capacities by all classes $\sum_{h=1}^{H} V_{ij,o}^{(h)}(\tilde{k})$. This sum will be called the total occupied capacity $C_{ch,o(i)}$ and it can not exceed the total channel capacity C_T. Mathematically we can write the following inequality(1):

$$\sum_{h=1}^{H} V_{ij,o}^{(h)}(\tilde{k}) \le C_T \quad \text{...........(1)}$$

The difference between the total channel capacity and the total occupied capacity represents the channel residual capacity in hand of node i at instant $\tilde{k}$ and is represented by $C_{ch,r(i)}$ as expressed in equation (2).

$$C_{ch,r(i)} = C_T - \sum_{h=1}^{H} V_{ij,o}^{(h)}(\tilde{k}) \quad \text{.......(2)}$$

This residual capacity will be compared with the channel residual capacity at the next node to find the available capacity, which can be used to handle the new calls. CAC will admit new calls as long as there is sufficient available capacity (bandwidth) to carry the newly arrived calls on a certain link, under the QoS requirements. Obviously at the initial moment the reserve capacity will be the channel capacity C_T. While at the run regime, the reserve capacity will be as expressed in equation (2), in addition, the reserve capacity can not exceed the channel capacity, see inequality (3) below:

$$C_{ch,r(i)} \le C_T \quad \text{.................. (3)}$$

Any efficient cost function must take account for the next node; we have to introduce a new term, which we will call the available capacity C_e. The available capacity means how much the channel reserve capacity is affected by the congestion level at the next node. The effective capacity will be taken as the minimum value of both nodes j and I at instants $\tilde{k}$ and $\tilde{s}$ respectively, keeping in mind that $\tilde{s}$ always leads $\tilde{k}$ are measured in slots. Mathematically we can express:

$$C_e = \min\,(C_{ch,r(i)}, C_{ch,r(j)}) \quad \text{........... (4)}$$

Where C_e is the actual available bandwidth, which is able to transfer the traffic from node i through node j. The cost function represents the amount of the available capacity C_e related to the total installed link capacity C_T to give a precise view of the link congestion. Our cost function depends on the link capacity, the link current load and the congestion level at the next node (measured at the previous instant). The cost function takes into account the link instantaneous congestion information as well as the aggregate information that is passed along periodically, which reflects the congestion level of the adjacent nodes. Uniform utilization of the resources and load balancing among the network components is the main goal of the cost function. This cost function consists of two terms: the first term represents the instantaneous congestion state of the node i at instant $\tilde{k}$, while the second term represents the congestion level at the next adjacent nodes (j) at the instant $\tilde{s}$. It should be noted that $\tilde{s} < \tilde{k}$. Finally the cost function might be expressed as follows:

$$\text{COST}=\begin{cases}\dfrac{C_{ch,o(i)}}{C_{ch,r(i)}}+\alpha_i\cdot\dfrac{C_{ch,o(j)}}{C_{ch,r(j)}} & \text{if } C_{ch,r(i)} \neq 0.0 \,\&\, C_{ch,r(j)} \neq 0.0 \\ y & \text{if } C_{ch,r(i)} \,\&\, C_{ch,r(j)} \text{ are zeros}\end{cases} \quad \ldots\ldots\ldots\ldots\ldots (5)$$

Where $C_{ch,o(i)}, C_{ch,o(j)}$, $C_{ch,r(i)}$ and $C_{ch,r(j)}$ are the total occupied capacities and the channel residual capacities in hand of nodes i and j respectively and measured in cells. α is a weighting factor that ranges between zero and one. The threshold value y is chosen very high to prevent using the saturated link even if there is reserve capacity for transmitting only one cell on another link, this value was chosen for 150 Mbps working channel capacity (our simulations are carried out on 150 Mbps) and it must be increased proportionally if the channel capacity increased.

It is worth noting that the proposed policy is distributed, based on a combination of the congestion situation of the node and its successors in two consecutive instants $\tilde{k}$ and $\tilde{s}$, and does not require the presence of a real time supervisory controller, which would be questionable in a wide area network. Moreover it can be observed that the characteristics of the strategy mentioned above are substantially independent to the specific access control and bandwidth allocation scheme.

2.2 The Congestion rate computation

By the congestion rate hereandafter we mean the growth speed of the number of in-progress calls that had been accommodated in the link under consideration during a specified period. To find out the congestion rate of the link over a considered period first of all it is required to know the number of in-progress calls in several instances during this period. For simplicity we substitute the number of in-progress calls by the occupied bandwidth by these calls. Suppose we have a set of occupied bandwidth values such as $V_{ij,o}^{(h)}(\tilde{k1})$, $V_{ij,o}^{(h)}(\tilde{k2})$ .... $V_{ij,o}^{(h)}(\tilde{kn})$ in different instances k1, k2…..kn of the considered link during a specified period. The following step is to find the average of these values, which will give us the congestion rate of the link μ_{ij} during this period. Consequently we can say

$$\mu_{ij} = \frac{\sum_{z=1}^{n} V_{ij,o}^{(h)}(\tilde{kz})}{n} \quad \ldots\ldots\ldots\ldots (6)$$

It is worth noting that the congestion here is represented in terms of occupied bandwidth instead of the number of in-progress calls, which gives us a direct indication for choosing the least congested link.

2.3 The routing rule

After accepting the incoming calls arises the following problem: through which link the admitted call will be carried towards its destination. In previous works [2, 3] several techniques were applied, among these techniques the shortest path, which leads to early congestion of a set of links that connects the intermediate nodes causing a complete blocking of the network. Another technique recommends selecting the link with minimum number of in-progress calls, which causes wasting the network bandwidth [6]. The equivalent bandwidth-based strategy is a good technique for saving bandwidth but it does not consider the congestion acceleration phenomena obviously[9]. In this approach we will enhance one of the equivalent strategies that use the cost function, which is addressed in section 2.1. The enhancement is carried out as follows:

First step: Finding the link with the highest free bandwidth by applying the equation (2) in order to choose it primarily but not finally. It must be checked again for rate congestion build up, which is carried out in the next step. The first step is a major check because it is useless to check the congestion rate of an overloaded link.

Second step: Defining the dynamic congestion behavior of the considered link during the specified period the least congested links are to be examined for the acceleration of congestion build up during the studied period.

4. Numerical Results

The performance of the proposed new cost function has been tested by simulation on a simple network. We will refer to traffic flow generated by the above data as an offered load 1; an offered "x" corresponds to the same data, except for the traffic intensities $N_a^{(h)}$,h=1, 2, 3 which are multiplied by x. The behavior of the DLCP was tested extensively in [2] and some initial simulation results of the performance of the routing scheme are reported in [3]. In this paper we aim at evaluating the established enhancement of the DLCP algorithm as a result of applying the Differential Measurements (DM) of congestion level. Figure 1 shows the obtained improvement of DLCP in terms of output calls after applying the introduced algorithm. Figure 2 shows the decrease in blocked calls as a result of applying the current method.
Figure 3 is dedicated to show the established gain over the network in terms of residual capacity in terms of Kilocells after loading the network with the same number of input calls.

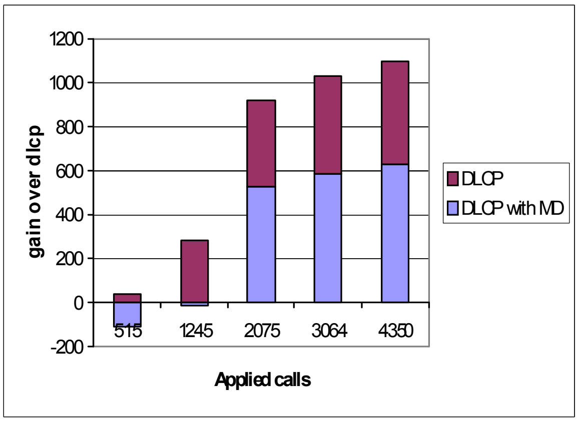

Fig. 1 The gain obtained of applying MD

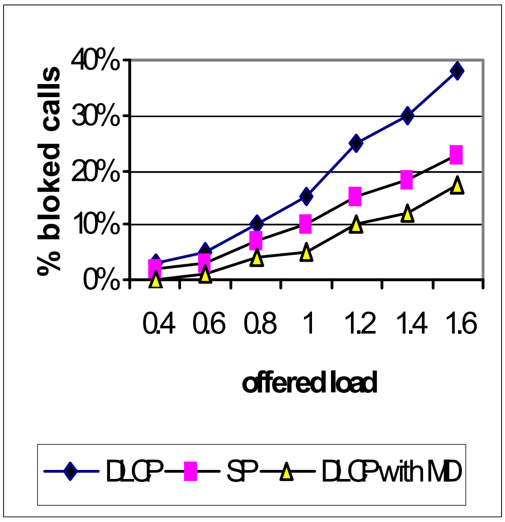

Fig. 2 Reduction in blocked calls of DLCP algorithm after applying the introduced algorithm in comparison with other algorithms

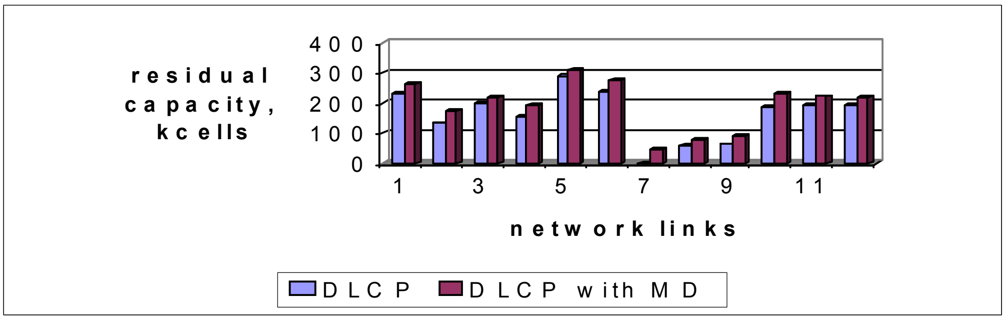

Fig. 3 The increase of residual capacity after applying MD

5. Conclusion

The proposed algorithm was introduced. The results confirmed the significant improvement of the cost function after applying Differential Measurements (DM) of congestion level. Simulation results show a sensible enhancement of DLCP when applying the proposed method.

References

[1] H. Abu Zaied. A Generic Connection Admission Control in ATM Networks . The 8th International Conference on Laser and Fiber-Optical Networks Modeling. Kharkov, Ukraine, 29th June -1st July, 2006;

[2] H. Abu Zaied.An Early Congestion Sensing Probe for ATM Networks.Proc. of International Conference on Information Technology and Sciences, ICITNS 2003,October 19th - 21st , Amman, Jordan, p.110-113.

[3] H. Abu Zaied and A. Dalala'h. A New Call Admission Control Function for an ATM Backbone. Proc. of The 2000 Arab Conference on Information Technology, ACIT"2000, Oct. 31st – Nov. 2nd, Zarka, Jordan, p.88 - 93.

[4] H. Abu Zaied, Victor Besliu, A. Dalala'h and M. S. Obaidat. Dynamic Near Shortest Path routing algorithm. – The Proceedings of the 2001 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2001, July 15-19, Orlando, Florida, p. 415-422.

[5] H. Abu Zaied, Victor Besliu and A. Dalala'h. Virtual link based connection Admission Control in ATM Networks (VCAC). The Proceedings of The 2001 International Arab Conference of Information Technology (ACIT 2001), Irbid, 16 - 19 November, 2001, p. 341-349.

[6] H. Abu Zaied, Victor Besliu and A. Dalala'h. A New Cost Function for integrated dynamic routing algorithm in ATM networks. The Proceedings of The 2001 International Arab Conference of Information Technology (ACIT 2001), Irbid, 16 - 19 November, 2001, p. 356 - 363.

[7] H. Abu Zaied. A distributed look-around with minimum Number of hops routing algorithm in ATM networks. BIT+, I International Conference on Information Technologies, Chişinău, Moldova, 2001, p.52.

[8] H. Abu Zaied, Victor Besliu. The Minimum Number of Passes - based Routing Strategy in ATM Networks(MNP). – Proceedings of the 3rd International Conference on "Microelectronics and Computer Science", (ICMCS-02), Volume II, Chişinău, 26 - 28 September, 2002, p. 138-144

[9] H. Abu Zaied, Victor Besliu and A. Dalala'h. The Minimum Number of Links-Based Routing Algorithm in ATM Networks (MNL)- The Proceedings of The 2002 International Arab Conference on Information Technology (ACIT'2002), Vol. 1 Qatar, 2002, p.268-273

[10] H. Abu Zaied, Victor Besliu and M. Al-Amaireh. Bandwidth Space Allocation Algorithm IN ATM Networks - The Proceedings of The 2002 International Arab Conference on Information Technology (ACIT'2002), Vol.2 Qatar, 2002, p. 794-799

[11] H. Abu Zaied. Optimum Bandwidth Allocation Algorithm in ATM Network- BIT+, II International Conference on Information Technologies, Chişinău, Moldova, 2002, p.95

[12] R. Bolla, A. Dalal'ah, F. Davoli and M. Marchese, "Dynamic Route Selection at Call Set-Up level in ATM Network", in D.D. Kouvatsos, Ed., ATM Networks Performance Modelling and Analysis, Vol. 3, Chapman & Hall, London, 1997, pp. 121-140.

[13] R. Bolla, A. Dalal'ah, F. Davoli, M. Marchese, M.S. Obaidat, "Integrated Dynamic Distributed Routing and Admission Control in ATM Networks", Internat. J. of Communication Systems, 1997.

[14] R. Bolla, F.Davoli, M.Marchese "Performance of Hop-by-Hop Distributed Routing and Resource Allocation in an ATM Network" . Proc.INFOCOM'93,San Francisco, CA, Vol. 3, pp. 1289- 1297, March 1993

[15] K. Murakami, H.S. Kim, "Virtual Path Routing for Survivable ATM Networks", IEEE/ACM Trans. Networking, vol. 4. No. 1, Feb. 1996, p. 22-39.

[16] V.J. Friesen, J.J. Harms, J.W. Wong, "Resource Management with Virtual Paths in ATM Networks", IEEE Network Mag., vol. 10, no. 5, Sept./Oct. 1996, p. 10-20.

[17] Z. Dziong, M. Juda, L.G. Mason, "A Framework for Bandwidth Management in ATM Networks - Aggregate Equivalent Bandwidth Estimation Approach", IEEE/ACM Trans. Networking, vol. 5, no.1, Feb. 1997.

[18] R.G. Akerley, "Hysteresis-Type Behavior in Networks with Extensive Overflows", British Telecom Tech. J., vol. 5, no. 4, Oct 1987.

[19] A. Girard, Routing and Dimensioning in Circuit-Switched Networks, Addison Wesley,1990.

[20] R. Bolla, F. Davoli, M. Marchese, "A Distributed Routing and Access Control Scheme for ATM Networks",proc.IEEE. Conf .on commun.,ICC'94,New Orleans,LA,Vol. 1, p.4-50,May 1994.

[21] R.O. Onvural, Asynchronous Transfer Mode Networks: Performance Issues, Artech House, Inc. Norwood, MA, 1994.

[22] Michael G. Hluchyj, Mark J. Karol. Queueing in high-performance packet switching. -IEEE journal on selected areas in communications, vol. № 9, 1988, -p.1587-1597.

[23] Miltos D. Grammatikakis and Miro Kraetzl and Eric Fleury. Shortest-path and hot-potato routing on unbuffered 2-D Tori. -GLOBECOM'96, London, 18-22 November, 1996, -p.1-6.

[24] Eric W. M. Wong, Andy K. M. Chain. Sammy Chan and K.T. Ko VP formulation and dynamic routing in ATM network. -IEEE Global Telecommunication Conference, 18-22 November, 1996, London, -p. 1715-1720.

[25] Jochen Beharens , J .J. Garcia-Luna-Aceves. Distributed scalable routing based on link-state vectors. -SIGCOMM 94, 1994, ACM, -p. 136- 147.