J. Indian Inst. Sci., July–Aug. 2005, **85**, 183–193

A dynamic bandwidth allocation scheme for VSAT-based networks

S. S. MANVI*, ANEES FLEXWALA AND ABHINAV SINGH
Department of Electronics and Communication Engineering, Basaveshwar Engineering College,
Bagalkot 587 102, Karnataka, India.
email: agentsun2002@yahoo.com; Phone: (91) 08354-234485; Fax: (91) 08354-234204.

Received on August 4, 2004; Revised on March 28, 2005.

Abstract

The limited bandwidth available to an organization through an expensive gateway like VSAT (very small aperture terminal) requires proper and effective utilization. This paper proposes a scheme, which allocates bandwidth to applications based on priorities, application metric weights coupled with max-min fairness approach. The applications requiring less bandwidth are fulfilled before the applications requesting for more bandwidth that come under same priority classes. The scheme is devised such that applications within the same priority are allocated bandwidth by using max-min fairness technique. The scheme is simulated and its performance is evaluated in terms of several performance parameters like bandwidth utilization, application rejection and bandwidth allocation. The simulation result shows that the proposed scheme can be used to cater to the needs of prioritized applications effectively. The flexibility, robustness and least mathematical complexity of the scheme make it suitable for networking applications over VSAT.

Keywords: VSAT, dynamic bandwidth, media access schemes.

1. Introduction

Very small aperture terminal (VSAT) emerged in 1970s as a satellite-based low-cost approach to connecting multiple locations in a private network for data transaction. Maral [1] states that VSAT has better features than terrestrial networks such as: ubiquitous availability, network reliability, uniform service levels, timely deployment and installations, multicast content distribution capability, easy site relocation and addition, network capacity expansion and emerging application support. Figure 1 depicts VSAT-based corporate networks. It comprises satellite as transponder, earth stations connecting local area networks via a router/switch, PABX and public networks. The communication between two nodes in any network takes place via the satellite through the earth stations [2, 3].

Earth station of a VSAT unit comprises the following: indoor unit, outdoor unit and antenna system. Outdoor unit is completely controlled by the indoor unit (Fig. 2). If the indoor unit transmits a signal, then the outdoor unit converts the signal up to the radio frequency band, amplifies it and transmits. The system network node computer (router/

---

*Author for correspondence.

switch) connected to the indoor unit controls the user interfaces of the VSAT station. It handles the data to the modulator card and from the demodulator card, manages queues, and performs protocol handling and management tasks.

VSAT facilitates both voice and data services. A router (gateway) connected to the VSAT allows Internet connectivity to either a single workstation or a LAN. Underutilized voice channels of an organization can be used for data services. A portion of VSAT bandwidth can be allocated to support video applications (such as teleconferencing, telemedicine, distance learning, etc.) as well. Data rates for video applications can be set at 64/128 kbps, or higher, depending upon specific site requirements. Whenever video is not required, the allocated bandwidth will be utilized to support data applications.

For a VSAT network to be an effective backbone network, it must handle a variety of traffic types and have efficient media access schemes. An intelligent VSAT uses DAMA (demand assignment multiple access) networking to permit the maximum utilization of the satellite capacity. Although DAMA may allow more VSATs to access the network, it is not robust enough to support low-duty cycle or bursty traffic of interactive transactions having short response time. This necessitates the use of random access protocols such as Aloha. Two such main protocols are Pure Aloha and Slotted Aloha [4].

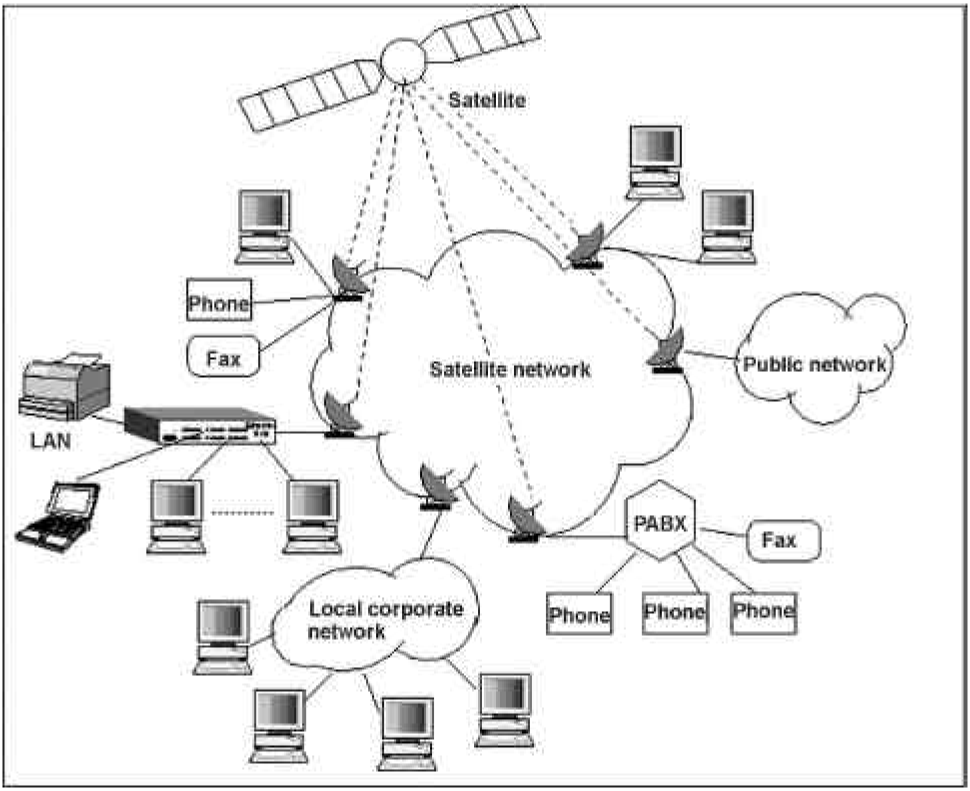

FIG. 1. VSAT-based corporate network.

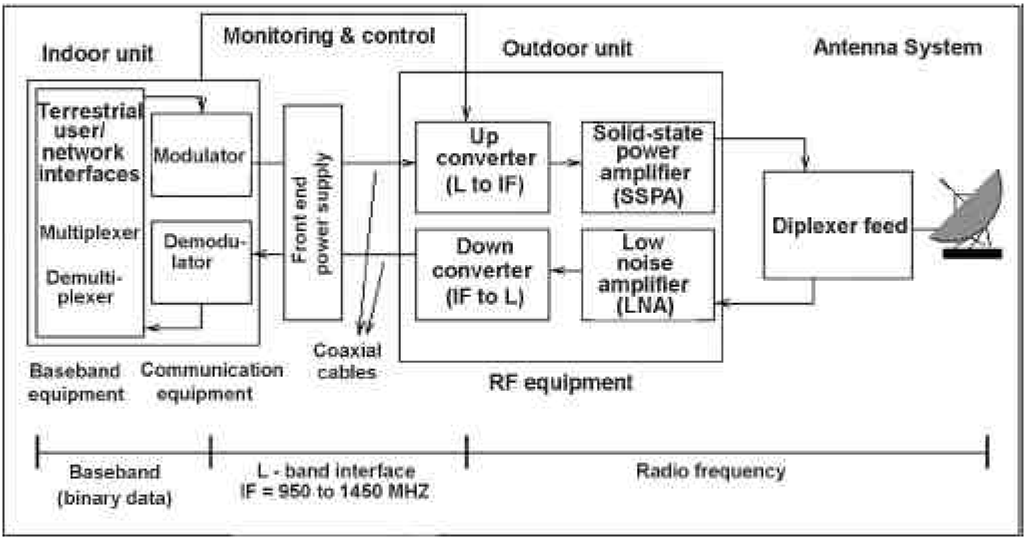

FIG. 2. Basic structure of VSAT.

Multimedia applications running in a VSAT-based network require certain quality of service (QoS). The goal of QoS is to provide guarantees on the ability of a network to deliver predictable results. Qian presents elements of network performance that are within the scope of QoS, which often include *availability* (uptime), *bandwidth* (throughput), *latency* (delay), *error rate* and *jitter* [5]. Providing QoS generally involves prioritization of network traffic based on the user/application requirements.

Satellite bandwidth is expensive. In any corporate network, bandwidth is fixed and limited. As more and more connection requests arrive in a network, distribution of fixed amount of bandwidth becomes a difficult and challenging job. Therefore, allocating the bandwidth from VSAT modules effectively and dynamically within the networks of an organization that are using the VSAT services is an important issue.

Bandwidth allocation has been a classical problem since long time in communication networks. Dynamic allocation schemes effectively utilize the bandwidth as compared to static bandwidth allocation schemes. Allocation of fixed amount of satellite bandwidth to different nodes of a VSAT network based on the various changing parameters and different constraints instantly with time is dynamic allocation. The constraints could be real-time application requests, error rate, pricing, priority, node's allocation state information, congestion status, etc. Thus, system software must be developed that will be placed in the router to allocate bandwidth dynamically (to nodes/applications) within an organization based on various constraints/parameters.

1.1. *Related work*

A detailed review of static and dynamic bandwidth allocation schemes based on buffers, congestion levels, queue lengths, feedback information, etc. is presented by Manvi and Venkataram [6]. Bertsekas and Gallager [7] present max-min fair scheme, in which, it allocates resources in order of increasing demand, no source gets a resource share larger than its demand, and sources with unsatisfied demands get an equal share of the resource. Fuzzy-based dynamic bandwidth allocation that uses bandwidth as a fuzzy parameter but allocates slightly less than ideal bandwidth is presented by Chandramathi and Shanmugavel [8]. Zhai *et al.* [9] present a queue length-based fair queuing in networks that deals with the calculation of dropping threshold. Arora and Brinkman [10] present a randomized online algorithm for bandwidth utilization, which lead to instability as well as unfair sharing of resources.

The work of Gunes *et al.* [11] and Arabshahi [12] provides a swarm-based algorithm that uses principle of ants-colony in finding paths and bandwidth allocation. Marbach [13] presents a scheme where max-min fair algorithm coupled with pricing is used for bandwidth allocation. Kelly [14] proposes a rate-based flow control algorithm closely related to where prices are interpreted as control signals to determine user transmission rate. Other proposed bandwidth allocation schemes use weighted max-min fairness, weighted fair queuing, dense graph method, distributed congestion control for max-min fair allocation, and max-min fair scheduling in wireless networks [15–20].

1.2. *Our contribution*

Our work differs from the other works in the following respects: 1) maintains certain state information about the nodes recent connectivity status and utilized bandwidth of the nodes to help in allocation decision making, 2) divides the applications into different classes based on the priorities computed using different application metrics (bandwidth, delays, jitter, reliability, etc.) and its weight; and 3) uses max-min fairness for same priority applications. The various priorities that are considered while allocating the bandwidth make the scheme very useful. The simplicity in which the priorities can be changed and the scheme be made to suit the users' need makes the scheme highly effective.

In the proposed scheme, applications requiring less bandwidth are fulfilled before the applications requesting for more bandwidth that come under the same priority classes. The scheme is devised such that applications within the same priority are allocated bandwidth by using max-min fairness technique.

1.3. *Paper organization*

The rest of the paper is organized as follows. Section 2 presents the proposed work. Simulation and results are presented in Sections 3 and 4, respectively. Section 5 conclude the work.

2. Proposed work

There is no problem if the desired bandwidth is less than the available bandwidth, but, when it exceeds the available bandwidth, there is a need to dynamically allocate it taking into consideration the various parameters (such as priority and criteria), i.e.

$$\sum_{i=1}^{n} x_i < \text{D bps} \;\text{———}\; \begin{array}{c}\text{Bandwidth}\\ \text{distributed}\\ \text{as requested}\end{array}$$

$$\sum_{\text{i}=1}^{n} x_i > \text{D bps} \;\text{———}\; ?$$

where *n* is the number of nodes and $x_{i,}$ the bandwidth requested by i^{th} application and *D*, the maximum capacity. Clearly, in such a case, some sort of prioritization is to be implemented, i.e. based on the various application metrics, there should be a scheme that allocates bandwidth slots on a priority basis to various nodes within a network and decides which node should transmit first.

The proposed dynamic bandwidth allocation scheme is devised around the following techniques: priority, application metric weights and max-min fairness, and allows many combinations of priorities. In this section, we describe max-min fairness scheme and our proposed scheme that uses the max-min fairness scheme along with priority of applications.

2.1. Max-min fairness

To illustrate the max-min fairness scheme, consider a network that has a total of 20 units of bandwidth (Fig. 3). At some point, there are four active connections A, B, C, and D using 5, 4, 3, and 2 units of bandwidth, respectively.

Referring to Fig. 3, assume that a new connection E, requesting seven units, has been accepted into the network. Clearly, the total amount of bandwidth requested by these connections is 21 units, exceeding the capacity of the network. How should the bandwidth be partitioned between the applications in a fair manner? The key idea in max-min fairness is that if *n* connections need to partition *b* units of bandwidth, then each is guaranteed its equal share of *b*/*n* units. Connections that require at most their equal share are granted their desired bandwidth and are referred to as satisfied.

For definiteness, assume that, of the *n* requesting connections, *m* are satisfied and the total bandwidth requested by the *m* satisfied connections is *S* units. The residual bandwidth, *R*, is partitioned among the remaining applications, each receiving $R/(n - m)$ units. Notice that, in the current example, the addition of the new connection E brings the total number of connections to five and, consequently, each is guaranteed four units of bandwidth. As illustrated in Fig. 4, connections B, C, and D are satisfied, while connections A and E are not. Since connections C and D are requesting less than their equal share, there are three units of residual bandwidth.

As shown in Fig. 5, the residual bandwidth is now partitioned between the unsatisfied connections A and E, each receiving 1.5 units of additional bandwidth. With this new allo-

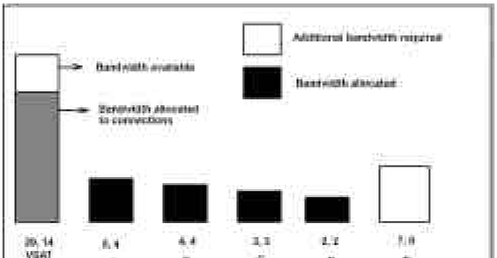

FIG. 3. Insufficient bandwidth to satisfy new connection.

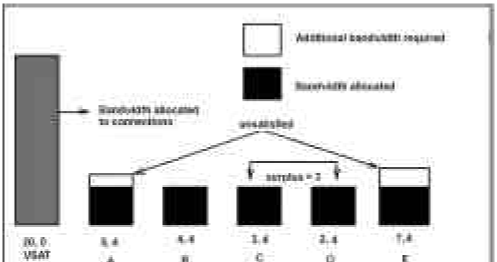

FIG. 4. Allocation of the equal share to all connections.

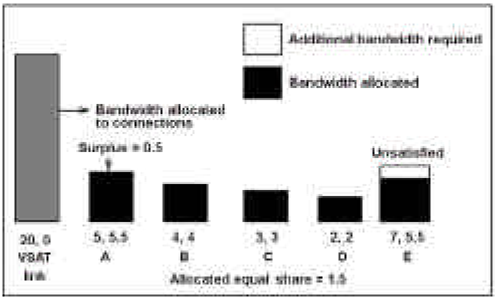

FIG. 5. Redistribution of the residual bandwidth.

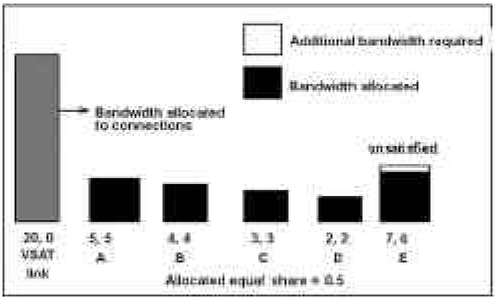

FIG. 6. Final redistribution of the residual bandwidth.

cation, connection A gets satisfied, leaving E as the only unsatisfied connection. Finally, the residual 0.5 unit of bandwidth is now given to connection E. The final max-min fair bandwidth allocation is shown in Fig. 6. Notice that connections A, B, C, and D are satisfied and that the only unsatisfied connection is the relatively bandwidth-intensive connection E.

As illustrated by the example above, max-min fairness attempts to maximize the bandwidth allocation to the connections requesting the least amount of bandwidth. The max-min scheme considers that each connection is entitled to an equal share of the limited bandwidth. Some connections request less bandwidth than others [7]. In the end, connections with a low bandwidth demand receive their desired amount of bandwidth, thus getting satisfied, while the remainder of the bandwidth is equally apportioned among the unsatisfied connections.

One may want to associate the weights $w1$, $w2$, $w3, \ldots wn$ with sources 1, 2, …, n, respectively, which reflect their relative sharing of resources. The concept of max-min fairness including weights is known as max-min weighted fair share allocation as given by Keshav [21]. Here resources are allocated in order of increasing demand normalized by the weight, no source gets a resource share larger than its demand and sources with unsatisfied demands get resource shares in proportion to their weights.

2.2. *Proposed bandwidth allocation scheme*

The proposed bandwidth allocation scheme assigns priority to applications based on application metrics (metrics are bandwidth, delay, jitter, reliability, etc.) and its weights. Later, it applies max-min fairness technique for bandwidth allocation among the applications in the same priority class starting from the higher priority classes. In the scheme, if a new node enters the network, there remains a provision of its inclusion in network at the cost of agreeable degradation of service (i.e. all the other nodes agree to it) till a user-defined minimum threshold bandwidth below which no new node will be included.

To illustrate the computation of priority values for an application, consider the applications of three nodes A, B, and C, each having four application metric values (a1, a2, a3, a4), (b1, b2, b3, b4) and (c1, c2, c3, c4), respectively. The four metrics of application of each of the nodes are attached with weights W, X, Y, Z, respectively. The weights are a measure of priorities assigned to the various application metrics. Table I depicts weights of metrics for the nodes A, B and C.

Table I
Weights and metric assigned to nodes A, B and C

Weight attached	Metric	Nodes		
		A	B	C
W	1st	a1	b1	c1
X	2nd	a2	b2	c2
Y	3rd	a3	b3	c3
Z	4th	a4	b4	c4

The priority for each of the nodes' application is computed as follows.

$$PA = (a1)(W) + (a2)(X) + (a3)(Y) + (a4)(Z)$$
$$PB = (b1)(W) + (b2)(X) + (b3)(Y) + (b4)(Z)$$
$$PC = (c1)(W) + (c2)(X) + (c3)(Y) + (c4)(Z)$$

and the value of 'W + X + Y + Z' must be equal to 1.

After computing priority, applications are grouped into different priority classes. Then, total available bandwidth will be allocated to the nodes of same priority using the max-min fairness technique. Pseudocode in Algorithm 1 presents the functioning of the proposed scheme.

Algorithm 1: Dynamic bandwidth allocation

```
BEGIN
1. For $i$ = 1 to $n$ do        /*$n$ = number of applications*/
   ■ Define the weights of QoS parameters and their values (bandwidth: BW, delay: D, jitter: J, reliability: Rl, queue length: QL) to assign priority of the QoS parameter for application of node $i$;
   ■ Define various flags for each application, which are required for defining the priorities and for other operations like request granting, etc.;
   ■ Define the degradation level ($Deg$) of the application of node $i$;
   ■ Compute acceptable bandwidth ($ABW_i$) for node $i$, $ABW_i = BW - BW*Deg$;
2. Calculate the total bandwidth requirement by all applications (considering requesting applications at each node);
3. If (total requested bandwidth <= total available bandwidth), Then grant the applications with the required bandwidth and go to step 9, Else go to step 4;
4. Find all the old requested applications (of different nodes) for bandwidth allocation and their successful allocation or rejection (it is assumed that history of allocation/rejection of applications and the nodes is maintained at the VSAT within a given time window as configured by the network administrator);
5. If there are old requested applications that are rejected (for 50% of the time from a node), and are currently requesting, Then apply max-min fairness to the applications of such nodes and allocate the bandwidth (giving priority to the nodes which are rejected earlier for allocation);
6. If bandwidth is available after allocation, go to step 7 Else go to step 9;
```

7. Classify the rest of the requesting applications from the nodes under each priority category (*nc* categories are considered). Categories are formed by using *z* = *[(Higher_priority value among the applications) – (Lower_priority value among the applications)]/nc*.
   - Category 1: Lower_value to Lower_value+z,
   - Category 2: Lower_value+z to Lower_value+2z,
   - Category 3: Lower_value+2z to Lower_value+3z, so on up to
   - Category nc: Lower_value+3z to Lower_value+*nc**z;

8 **For** *j* = 1 to *nc* /**nc* = number of categories*/

   - Apply the max-min fairness to the applications in category *j* and allocate the bandwidth;
   - If the allocated bandwidth is less than the acceptable bandwidth of an application from a node, then application is rejected;
   - The allocated bandwidth for rejected applications is reutilized for allocating other unsatisfied applications;

9. Stop

END

Number of priority categories (*nc*) to be chosen may depend on the administrator of the network. The proposed scheme is a stateful one. Maintaining the states of allocation for applications facilitates the allocation scheme to give more preference to the suffered applications. Hence, states are maintained. To reduce the states, we can maintain a time window of shorter duration, may be 2 to 5 min, during which states are recorded, i.e. maximum state information is of recent 2 to 5 min. Other option is, states can be maintained by the clients and can pass on the information to the router whenever it requires a network application to be executed.

The proposed scheme can be implemented in the router/switch connected to the VSAT indoor unit. Router can be configured to include various parameters for bandwidth allocation, number of priority categories, predefined priority to several nodes, etc.

3. Simulation

The network model assumes a corporate network consisting of *n* nodes that are connected to VSAT hub through a switch/router. The nodes combine star and bus topologies, where *n/x* are connected using bus topology and one of the nodes in each bus is connected to router connected to VSAT. VSAT bandwidth is considered for three cases: 64 (1-voice channel), 192 (3-voice channels) and 512 kbps (8-voice channels). Network model uses four application metrics (bandwidth, delay, jitter, loss) attached to an application generated by a node. Weights for the metrics are generated randomly between 0 and 1 such that summation of all the weights is equal to 1. The number of priority categories considered is *nc*.

Maximum number of applications generated randomly for *n* nodes of the network is *napp*. Application arrivals are Poisson distributed with mean interarrival time *mit* units. Minimum bandwidth available to data applications is 64 kbps in the case of 192 and

512 kbps channels. Application requests for a bandwidth are randomly distributed between *b1* and *b2* kbps; other parameter (delay, jitter, loss) values are generated randomly between 0 and 1. Application acceptable degradation is considered to be in the range of 0 to deg. History of information about each node is randomly generated which includes the number of times applications is rejected, bandwidth allocated to accepted applications, and quality of service status given to a node.

The performance parameters considered for evaluation of the proposed scheme are as follows.

- *Application accepted* (%): It is defined as the ratio of number of applications accepted for bandwidth allocation to the ratio of number of applications requesting for bandwidth allocation.
- *Application rejection* (%): It is defined as the ratio of number of applications rejected for bandwidth allocation to the ratio of number of applications requesting for bandwidth allocation.
- *Bandwidth allocated:* It is called the sum of the allocated bandwidth to all the applications.
- *Bandwidth utilization* (%): It is defined as the ratio of the total bandwidth allocated to the maximum bandwidth of the network.

The following inputs are considered in the simulation model to test the proposed scheme: $n = 100$, $x = 10$; napp = 100; $b1 = 6$, $b2 = 10$; $nc = 4$; mit = 10 s; deg = 0, 0.2 and 0.4; for instance, degradation = 0.4 means the application is ready to accept 40% less than the required.

4. Results

The acceptance of applications decreases with rise in available bandwidth and the increase in the number of requesting applications (Fig. 7). Figures 8–10 present the application rejection for available bandwidths (64, 192 and 512 kbps), respectively, with different acceptable degradation levels (0, 0.2 and 0.4). It is observed that as the acceptable degradation level increases, application rejection reduces.

We observe from Fig. 11 that sum of bandwidth allocated to all applications is limited to maximum of 64, 192, 512 kbps for available bandwidths of 64, 192 and 512 kbps, respectively. Bandwidth utilization increases with increase in the number of applications and the increase in acceptable degradation levels for different available bandwidths (Fig. 12).

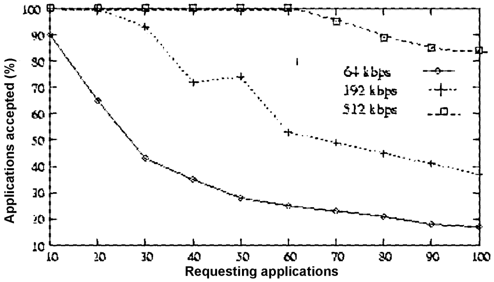

FIG. 7. Applications accepted versus requesting applications for different available bandwidths (64, 192 and 512 kbps).

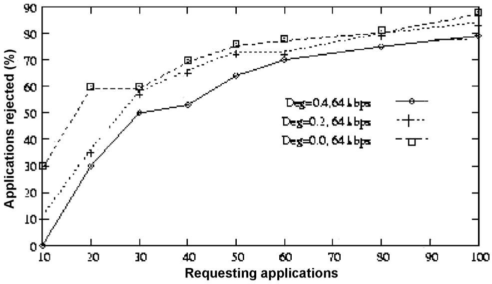

FIG. 8. Application rejection versus requesting applications for different acceptable degradation levels (0, 0.2, 0.4) with available bandwidth = 64 kbps.

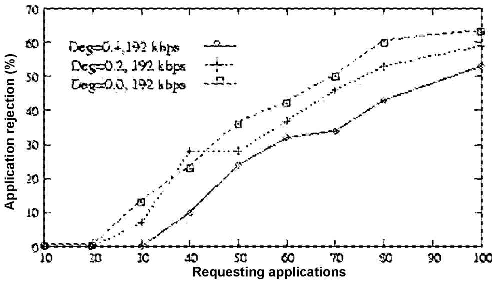

FIG. 9. Application rejection versus requesting applications for different acceptable degradation levels (0, 0.2, 0.4) with available bandwidth = 192 kbps.

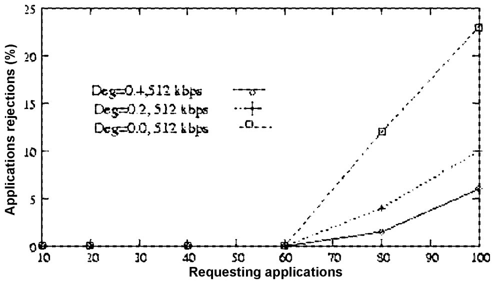

FIG. 10. Application rejection versus requesting applications for different acceptable degradation levels (0, 0.2, 0.4) with available bandwidth = 512 kbps.

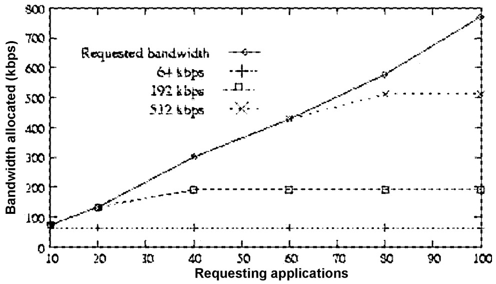

FIG. 11. Bandwidth allocated (sum of allocation to all applications) versus requesting applications with different available bandwidths (64, 192 and 512 kbps) with degradation level = 0.2.

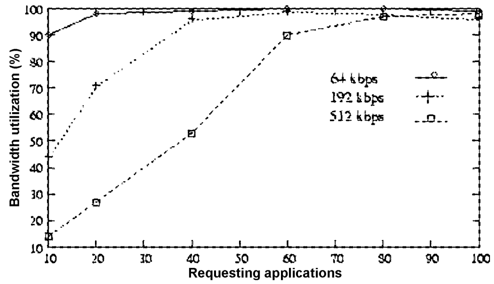

FIG. 12. Bandwidth utilization versus requesting applications for different available bandwidths (64, 192 and 512 kbps) with degradation level = 0.2.

5. Conclusions

The paper presents the VSAT-based corporate networks and bandwidth allocation problem. The proposed bandwidth allocation scheme prioritizes the applications based on the application parameters according to the organization's needs and applies max-min fairness scheme in each priority class to allocate the bandwidth to the applications. Parameters like *bandwidth*, delay, *error rate*, *queue length* and *jitter* can be considered. It fairly allocates the bandwidth among the same priority applications and allows using any level of defining priorities to the applications while granting the requested bandwidth.

As a future scope for our proposed scheme we would like to extend it to different kinds of networks depending on the parameters and metrics of that network, such as wireless networks. Futuristic methods of allocating bandwidth like swarm- or mobile agent-based techniques can be used to allocate the bandwidth to facilitate flexible and adaptable intelligent services.

References

1. Gerard Maral, *Satellite communication systems: Systems, techniques and technology*, Wiley, 3rd edition (1998).

2. T. T. Ha, Personal computer communications via VSAT networks, *IEEE J. Selected Areas Commun.*, **7**, 235–245 (1989).
3. C. J. Foo, H. Lim, S. Sharp, K. M. Lye, B. T. Tan and K. C. Chua, Two way calling public CT2 telepoint system, *IEEE J. Selected Areas Commun.*, **13**, 923–931 (1995).
4. A. H. Rana, J. S. Mcoskey and W. A. Check, VSAT technology trends and applications, *Proc. IEEE*, **78**, 1987–1995 (1990).
5. Yi Qian, Rose Hu, Hosame Abu-Amara, Simulation study of broadband multimedia VSAT network. *Proc. OPNETWORK Conf.* (1999).
6. S. S. Manvi and P. Venkataram, A review of bandwidth allocation schemes in multimedia communications, *IETE Golden Jubilee Compendium* (2003).
7. D. P. Bertsekas and R. Gallager, *Data networks*, Prentice Hall (1992).
8. S. Chandramathi and S. Shanmugavel, Fuzzy based dynamic bandwidth allocation for heterogeneous sources in ATM networks, *Applied Soft Computing*, **3**, 53–70 (2003).
9. Mingyu Zhai, Guanqun Gu and Yuan Yuan, Queue length-based fair queuing in core-stateless networks, *Proc. IFIP Con. Management of Multimedia Networks and Services (MMNS)*, Chicago, USA (2001).
10. Sanjeev Arora and Bo Brinkman, A randomized online algorithm for bandwidth utilization, *Proc. ACM SIAM Symp. on Discrete Algorithms*, San Francisco, USA (2002).
11. Mesut Gunes, Udo Sorges and Imed Bouazizi, ARA–The ant-colony-based routing algorithm for MANETS, *IEEE Int. Conf. on Parallel Processing Workshops* (ICPPW’02), Canada (2002).
12. Payman Arabshahi, Swarm-based routing algorithms for sensor and satellite networks. http://esto.gsfe.nasa.gov/conferences/estc-2001.
13. Peter Marbach, Priority service and max-min fairness, *Proc. IEEE INFOCOM*, New York, USA (2002).
14. F. P. Kelly, A. Maulo and D. Tan, Rate control for communication networks, *J. Op. Res. Soc.*, 237–252 (1998).
15. Anjlica Malla, Mona El-Kadi, Stephan Olariu and Petia Todorova, A fair resource allocation protocol for multimedia wireless networks, *IEEE Trans. Parallel Distributed Systems*, **14**, 63–71 (2003).
16. G. J. Miller and M. Paterakis, A dynamic bandwidth-allocation-based priority mechanism for the P_i-persistent protocol for MAN’s, *IEEE J. Selected Areas Commun.*, **11**, 1229–1239 (1993).
17. Chun-Ting Chou and Kang G. Shin, Analysis of combined adaptive bandwidth allocation and admission control in wireless networks, *Proc. IEEE Globecom* (2002).
18. Jon Klienberg and Eva Tardos, Fairness in routing and load balancing, *Proc. 41st IEEE Symp. Foundations of Computer Science* (2000).
19. Marios Lestas, Petros Iannous and Andres Pittsilides, Distributed congestion control algorithm for maxmin allocation of resources, *Proc. Am. Control Conf.*, Boston, USA (2004).
20. L. Tassiulas and S. Sarkar, Maxmin fair scheduling in wireless networks, *Proc. IEEE INFOCOM* (2002).
21. S. Keshav, *An engineering approach to computer networking*, Pearson Education (1997).