Improving TCP Performance in TDMA-based Satellite Access Networks

Jing Zhu, Sumit Roy
{zhuj, roy}@ee.washington.edu
Department of Electrical Engineering, University of Washington, Seattle, WA 98105

Abstract- **In TDMA-based MAC (media access control) protocols for satellite network access, slotted ALOHA which has been widely deployed in Very Small Aperture Terminal (VSAT) systems. More recently, Generalized Retransmission Announcement Protocol (GRAP) was proposed in [5], which regroups the immediate access by contention at low loads and the reservation access at higher loads to achieve a better efficiency. Inspired by GRAP, we propose a new MAC protocol (CA-GRAP) using *cumulative acknowledgment* (CA) technique to further improve the success probability of reservation request. Furthermore, the efficiency of random access in CA-GRAP is improved by introducing *contention index*. Our proposal does not require collision or error detection but is based only on correctly received packets, dramatically simplifying the procedure. The impact of all these protocols on TCP traffic is investigated and an upper bound of TCP throughput at high loads is derived for GRAP-based protocols. Simulation results show that our new method significantly outperforms GRAP.**

Key Words: satellite access networks, TCP/IP, MAC

I. Introduction

Satellite is increasingly used to complement terrestrial networks in providing a worldwide access to Internet. Very small aperture terminal (VSAT) system [2] is such an example, where a great number of wireless terminals share a common satellite channel such that a MAC protocol is needed to coordinate all active terminals to achieve the optimal utilization of limited channel resources.

One of the most widely implemented MAC protocols in satellite systems is slotted ALOHA for it's simplicity. Compared to reservation-based MAC protocols, it greatly reduces access delay and achieves good efficiency at low loads. However, the performance of slotted ALOHA dramatically degrades as traffic load increases because of the increased collision probability. Much research work has focused on performance evaluation of slotted ALOHA and innovations towards improving performance at high loads. The authors in [3] showed that slotted ALOHA is mono-stable if the number of retransmissions is limited to, at most, eight. In [4], a general analytical model was built for ALOHA-based satellite channel. However, like most other published research, these are based on the assumption that arrival traffic is a Poisson process independent of MAC procedure. While this may be reasonably true for open-loop traffic such as UDP, it is largely invalid for TCP that supports most of reliable data services (e.g. FTP, HTTP, email, telnet, etc.) on Internet. As is well known, the TCP traffic exhibits characteristic *burstiness* that is significantly different from Poisson arrivals. Therefore it is necessary to re-evaluate slotted ALOHA performance with respect to TCP traffic - a simulation study is conducted focusing on key parameters such as *maximum transmission number* and *maximum backoff time*. It is shown that for reliable end-to-end data traffic like TCP, both large maximum transmission number and long maximum backoff time are preferred to ensure high access reliability. Moreover, the theoretical maximum throughput 0.368 [6] of slotted ALOHA is still valid for TCP traffic.

The other protocol studied in the paper is generalized retransmission announcement protocol (GRAP) [5], derived from ARRA [1]. It regroups stations needing immediate access by contention at low loads and the reservation access at higher loads to achieve a better efficiency. In this paper, we incorporate the following techniques to further enhance GRAP.

a. Introduce cumulative acknowledgment (CA) in reservation request to improve the success probability of reservation.
b Allow each terminal to access multiple available slots to improve the slot utilization.
c. Associate reserved slots with a flow (i.e. flow-oriented), and not to a specified packet (i.e. packet-oriented) to simplify the procedure.
d. Employ Round Robin policy in slot allocation to achieve fairness.

Our proposal is called *cumulative acknowledgment* based GRAP. An upper bound of the total throughput is derived for TCP traffic at high loads for GRAP-based protocols. Simulation results show that our proposal has a performance much closer to the bound than GRAP. Furthermore, it is shown that CA-GRAP has enhanced resistance to decreasing the number of mini slots and increasing the number of data slots, both of which have negative impact on the total throughput.

The paper is organized as follows. The network scenario is shown in Section 2. We describe our new MAC protocol (CA-GRAP), and derive the upper bound in Section 3. In Section 4, simulation is performed with OPNET to evaluate the performance of slotted ALOHA, GRAP and CA-GRAP. Section 5 concludes the paper and outlines the future work direction.

II. MAC Protocols in TDMA-based Satellite Access Networks

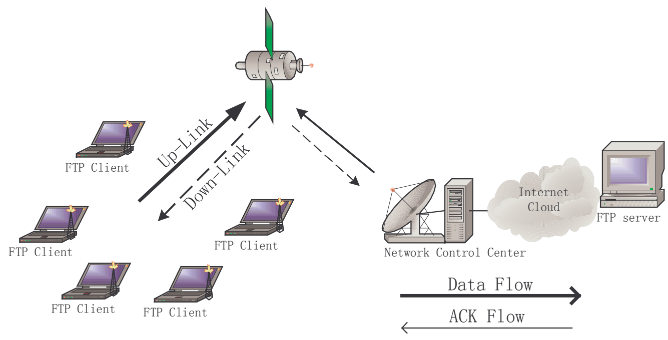

Fig. 1. Network Scenario of Multiple TCP Connection Sharing A Satellite Up-Link

Fig.1 shows the network scenario we study in this paper, where a number of terminals access Internet through a satellite access network. Here, we consider FTP service, a widely used file transfer application based on TCP. Since all FTP clients share a common satellite up-link channel, a MAC protocol works at each client to make them cooperate in a proper way. The network control center (NCC) takes responsibility for all MAC control aspects, such as collision detection and broadcast of system status. Collided packets are re-sent till successful reception or reaching a predefined maximum transmission number. A buffer is needed at the MAC layer of each terminal to store data packets. The key system parameters are:

- B: Buffer capacity at the MAC layer of each terminal running a FTP client;
- N: Number of terminals sharing one satellite up-link channel;
- R_d: Down-link bandwidth;
- R_u: Up-link bandwidth;
- RTT: Round trip time between a terminal and the NCC.

We consider a slotted TDMA system, where the whole transfer period is divided into successive time slots. Data packets from an upper layer are encapsulated into MAC packets with fixed length (equal to the length of a data slot field), and each MAC packet is transmitted only at the beginning of a slot, namely slot *synchronization* is assumed.

III. Cumulative Acknowledgment based GRAP (CA-GRAP)

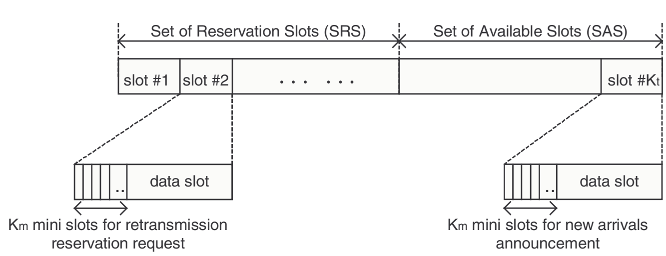

Fig. 2. GRAP Frame Structure

CA-GRAP uses the same frame structure as GRAP (see Fig.2). A frame is divided into K_t slots with the same length. Each slot consists of a data slot (DS) field and K_m minislots (MS) field, leading to total K_t data slots and K_mK_t mini slots in a frame. Hence, the payload ratio is

$$\frac{K_tL_d}{K_tL_d + K_mK_tL_m} = \frac{L_d}{L_d + K_mL_m}, \quad (1)$$

where L_d and L_m are the length of a data slot and a mini slot, respectively. Generally, K_m, L_m and L_d are chosen such that the consequent payload ratio is as high as possible. There are two sets of slots in a MAC frame: SAS (set of available slots) for random access, and SRS (set of reservation slots) consisting of reserved slots. For each upcoming frame, the NCC performs slot allocation according to correctly received reservation requests; the remaining slots of the frame then constitute SAS for random access.

The key parameters needed for CA-GRAP operation are

- **pkt_status**: State of a packet: "new","waiting", and "reserved".
- **rtx_timer**: Retransmission timer of the last reservation request.
- **rqt_num[addr]**: The number of correctly received but unacknowledged (i.e. not allocated) reservation requests for the terminal with address "addr" (the initial value is 0).
- **last_sn[addr]**: The largest sequence number of acknowledged (i.e. allocated) reservation request for the terminal with address "addr" (the initial value is -1).

The first two are used by terminals and the last two by the NCC.

A. Description of CA-GRAP

Random Access: New packets from uplayer are encapsulated into MAC packets with fixed length (equal to the length of a DS field). A new MAC packet is assigned an increasing sequence number (from zero), and marked as "new". For each upcoming frame, a terminal monitors the down-link stream to read the set of available slots (SAS) for contention.

If SAS is not empty, m new MAC packets are allowed to be transmitted by each terminal in data slot field of m randomly chosen SAS slots of the following frame. Each packet is also accompanied by an anticipating retransmission reservation request placed in a mini slot randomly chosen in the same slot. The value of m is given by Eq. 2, where N is the number of terminals, K_a is the number of SAS slots and $\mathcal{C}$ is a new parameter called *contention index*, which is a ratio of the maximum number of packets transmitted in a SAS to the SAS capacity. All these information are broadcast by the NCC.

$$m = max(1, \lfloor\frac{\mathcal{C}K_a}{N}\rfloor). \quad (2)$$

Obviously, a larger contention index leads to higher utilization of SAS, but higher collision probability as well.

If SAS is empty, no random access will happen. Instead, a reservation request with the sequence number of the first "new" MAC packet in the buffer will be transmitted by a randomly chosen mini slot in a reserved slot.

After the above procedure (random access or sending reservation request), the processed MAC packet will be marked as "waiting". Whenever a reservation request is sent out,

rtx_timer must be updated. Usually, the value of **rtx_timer** is set by an estimate of RTT. If **rtx_timer** expires, the reservation request of the last "waiting" packet in the buffer will be resent.

Frame Analysis and Feedback Formulation: The uplink frame is completely received by the NCC before starting any analysis or process. First of all, acknowledgment (ACK) are sent back to notify terminals towards removing those correctly received data packets. Then both "rqt_num[addr]" and "last_sn[addr]" are updated by scanning each slot in the frame with the following algorithm, where i is the terminal address of a MS or DS slot and sn is the sequence number of packet or reservation request in the slot .

```
If (sn > last_sn[i])
   rqt_num[i] = rqt_num[i]+(sn -last_sn[i]);
   last_sn[i] = sn;
   If (DS)
      rqt_num[i] − −;
   end
end
```

The quantity rqt_num[i] indicates the total number of reservation requests at terminal with address "i". Using Round Robin policy, the NCC fairly distributes reserved slots to all terminals and rqt_num[addr] will decrease accordingly until all waiting requests are served (i.e. rqt_num[addr] falls to zero).

A correctly received reservation request includes not only the packet with the sequence number of the request but also all previously corrupted requests so that there is no need to resend a corrupted request except for the last one resulting in *cumulative acknowledgments* that greatly improve the reliability while reducing overhead and simplifying the protocol.

Reservation Access: Each terminal monitors the down-link stream to read the status of the upcoming frame and the set of reserved slots (SRS) broadcast by the NCC. If x slots have been assigned to a terminal, the first x "waiting" MAC packets of the terminal will be transmitted in the reserved slots of the upcoming frame with the packets marked as "reserved". Since no contention occurs in any reserved slot, all "reserved" MAC packets will be successfully received. Please note that ACK information arrives earlier than SRS information so that successfully transmitted "waiting" packets have no impact on reservation access.

In summary, the main innovations of CA-GRAP compared to GRAP, is the introduction of the *contention index* and using *cumulative acknowledgement* in reservation request.

B. Asymptotic Performance Analysis of GRAP-based Protocols at High Loads

First we introduce some necessary notations.

- T: The persistence time of a packet, measured from the first transmission to the arrival of ACK for that packet.
- W: The TCP window size.
- L: MAC frame length (given by $L_d K_t + L_m K_t K_m$).
- η: Normalized total throughput (throughput normalized by the channel bandwidth).
- α: Payload ratio which is the maximum value of total throughput (given by $\frac{L_d}{L_d + K_m L_m}$).
- μ: Channel bandwidth (we assume down-link bandwidth equals up-link bandwidth).

By simulation, we find that GRAP-based protocols (GRAP and CA-GRAP) have good fairness performance, which allows us to evaluate the whole system by studying a single connection.

At high loads, most packets will be transmitted by reservation access, which requires at least two round trips to finish a successful transmission, one for reservation request and the other for data transmission. On the other hand, a MAC packet can be transmitted in any slot of a frame. The transmission delay of the first slot is 0 and that of the last slot is L/μ. For simplicity, we assume the transmission delay of a packet to be uniformly distributed on range [0, L/μ], leading to the average value $\frac{L}{2\mu}$. Thus, the minimum value of T can be approximated by

$$min(T) = 2(\frac{L}{2\mu} + RTT). \quad (3)$$

MAC packets can not be removed from the buffer until the arrival of ACK so that the total number of in-flight packets are limited by the buffer size. Consequently, TCP window size should be equal to the buffer size at time overflow occurs.

$$W = B. \quad (4)$$

Then TCP window will be reduced by half to $B/2$. Fig.3 shows the cumulative distribution function of TCP window size in a simulation, implying that it approximately follows uniform distribution on range [$W/2$, W].

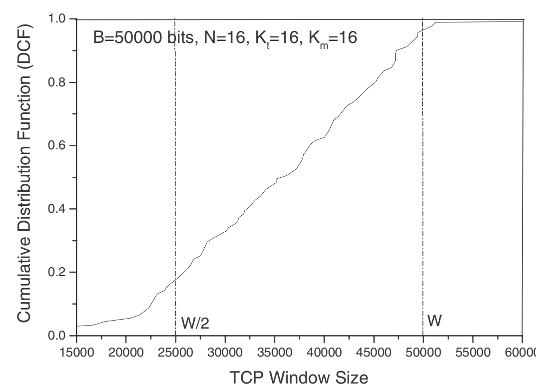

Fig. 3. Cumulative Distribution Function of TCP Window Size

Using w_i to denote the TCP window size of the ith terminal, we have the total throughput computed as

$$\eta = min(1, \frac{\sum_{i=1}^{N} w_i}{\mu T})\alpha. \quad (5)$$

Given a specified T, the mean value of total throughput can be calculated as

$$E[\eta|T] = min(1, \frac{E(\sum_{i=1}^{N} w_i)}{\mu T})\alpha. \quad (6)$$

Assuming $\{w_i, 1 \le i \le N\}$ to be independent random variables with the same distribution, Eq.6 turns into

$$E[\eta|T] = min(1, \frac{NE(w_i)}{\mu T})\alpha = min(1, \frac{N\int_{B/2}^{B} \frac{w}{B/2} dw}{\mu T})\alpha. \quad (7)$$

Thus we have

$$\begin{aligned} E[\eta] &\leq E[\eta|min(T)]. \\ &= min(1, \frac{3}{4}\frac{NB}{(L+2\mu RTT)})\alpha \end{aligned} \quad (8)$$

Eq.8 gives an upper bound of total throughput for GRAP-based protocols in consideration of TCP traffic.

IV. Simulation Results and Discussion

The round trip time from terminal to the NCC is fixed with 0.2s. The delay in terrestrial portion is ignored. The bandwidth of satellite downlink/uplink is 2Mbps. We also assume that all terminals have the same buffer size. TCP packet length is fixed with 576 bytes (i.e. 4608 bits) including TCP header and IP header, and the length of a DS field is set equal to the length of a TCP packet to avoid segmentation. In our simulation, we find that slotted ALOHA, GRAP and CA-GRAP all have good fairness performance for TCP traffic. Therefore, the only metric studied here is *total throughput*.

A. Slotted Aloha

In our simulation, ACK (acknowledgement) /NACK (negative acknowledgement) packets are employed to notify the transmitter whether the data packet is received correctly or collided with other data packet. If receiving a NACK, the transmitter will retransmit the corrupted packet but with a delay, given by an uniformly distributed random variable on the range $[1, \tau_b]$ where τ_b is the *maximum backoff time* in the unit of time slot. If the total number of retransmission reaches the *maximum transmission number*, the packet will be discarded.

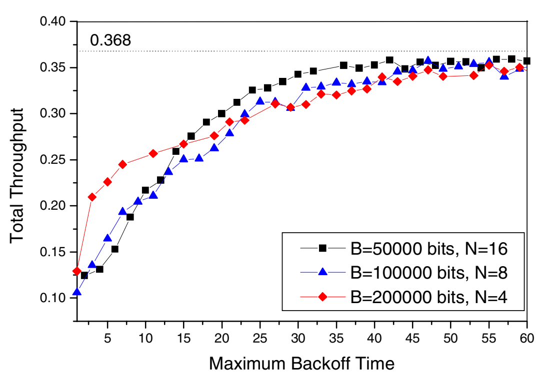

Fig. 4. Effect of Maximum Backoff Time

The effect of maximum backoff time on total throughput is illustrated in Fig.4 with the assumption of unlimited retransmission (each packet is transmitted till success) . Three values of N are investigated (16, 8, 4). It is clearly observed that the longer maximum backoff time helps improve the total throughput-the main reason being that the collision probability is reduced by extending backoff time. In our following simulation, we use 60 as the maximum backoff time.

Fig.5 shows that total throughput is monotonically improved by increasing maximum transmission number, implying that no limit should be put on retransmission when supporting TCP

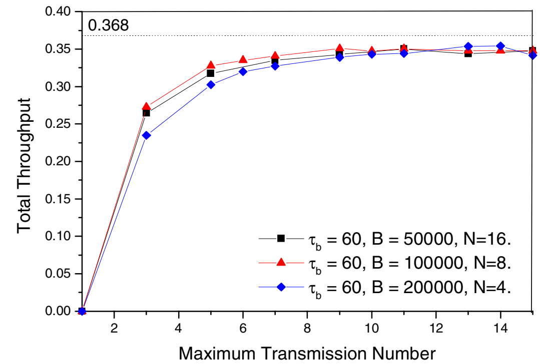

Fig. 5. Effect of Changing Maximum Transmission Number

traffic. It seems to conflict with the conclusion in [3] that slotted ALOHA is mono-stable if the number of retransmissions is limited to, at most, eight. The intuitive explanation is that unlike a Poisson process used in [3], the closed-loop control of TCP traffic has guaranteed the system stability so that the access reliability will be the only factor to dominate end-to-end performance. Therefore, it is necessary to set the maximum transmission number large enough for desired performance with TCP traffic. In addition, it is clearly demonstrated in Fiq.4 and 5 that the theoretical maximum throughput of slotted ALOHA 0.368 is still valid for TCP traffic.

In conclusion, the total throughput of TCP over slotted ALOHA is very low. Next, we will focus on the performance of GRAP-based protocols, which uses reservation access at high loads.

B. CA-GRAP vs GRAP

First of all, we study the effect of contention index on total throughput. The number of terminals is 16, and K_m is fixed with 16. Fig.6 shows total throughput as a function of the contention index for different K_t (say 16, 32, and 64). We can see the lower total throughput for both too large and too small value of contention index. And the contention index of 4 acquires the best performance for all three K_t, which will be used in our following simulation.

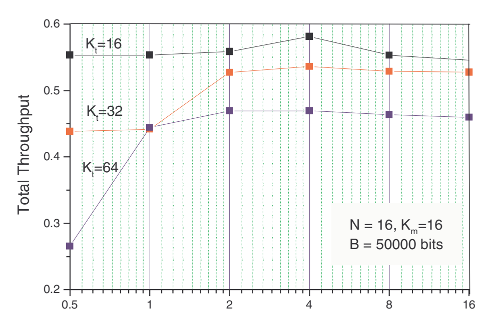

Fig. 6. Effect of contention index $\mathcal{C}$

In order to demonstrate the major benefits of our proposal (CA-GRAP), a total throughput comparison between GRAP and CA-GRAP is performed in Fig.7, and the total throughput

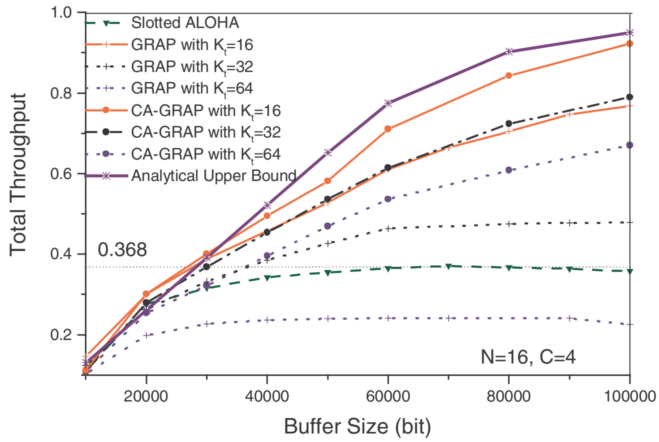

Fig. 7. Total Throughput Comparison

of slotted ALOHA is given as well. We also include the upper bound derived in Section3B; it is clearly demonstrated that CA-GRAP acquires a performance much closer to the upper bound than GRAP, showing the performance improvement with CA-GRAP. Furthermore, larger the value of K_t, the greater the improvement. Additionally, the total throughput of slotted ALOHA cannot exceed 0.368.

Fig.8 and 9 show the effect of K_m and K_t on total throughput. Here we set the length of a mini slot small enough to ignore the adverse effect of increased overhead as the number of mini slots increases. Generally, fewer the mini slots, the less the success probability of reservation request. Therefore, generally the total throughput degrades with K_m decreasing. However CA-GRAP has strong ability to resist such degradation. If we take the overhead introduced by mini slots into consideration, fewer mini slots will be preferred so that CA-GRAP is preferred over GRAP.

Fig.9 shows that the total throughput degrades as K_t increases; because the length of MAC frame increases with K_t, this leads to the extended persistence time of a MAC layer packet (see Eq.3). Again, CA-GRAP works better than GRAP.

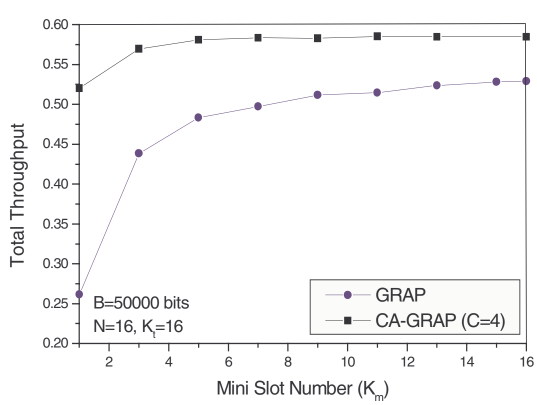

Fig. 8. Effect of Mini Slot Number (K_m)

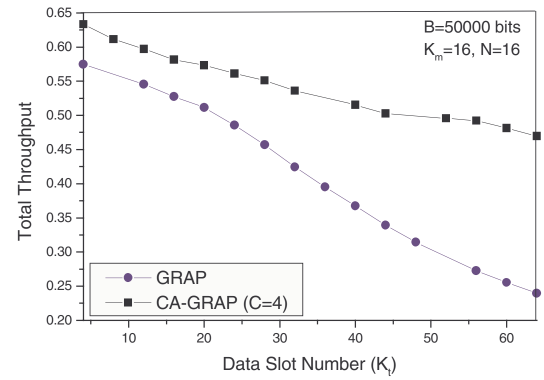

Fig. 9. Effect of Data Slot Number (K_t)

V. Conclusion

In this paper, two TDMA-based MAC protocols, slotted ALOHA and GRAP were evaluated in a satellite access network. Simulation results showed that the theoretical maximum throughput of slotted ALOHA is still valid for TCP traffic. Moreover, a large enough maximum transmission number and long backoff time are preferred in supporting TCP traffic.

Compared with slotted ALOHA, GRAP combining random access and reservation access acquires higher performance for TCP traffic. Inspired by GRAP, we proposed a new MAC protcol, CA-GRAP. By introducing *contention index* and *cumulative acknowledgment* technique, the efficiency and reliability are further improved. Moreover, an upper bound of total throughput is derived for TCP over GRAP-based protocols at high loads. Simulation results showed that CA-GRAP achieves a performance much closer to the upper bound than GRAP. Furthermore, CA-GRAP demonstrates strong ability to resist mini-slot decrease and data-slot increase, both of which have adverse impact on the total throughput. In addition, CA-GRAP only works on correctly received packets without any collision or error detection, greatly simplifying the procedure.

In conclusion, CA-GRAP is a promising MAC protocol for support of Internet traffic (e.g. TCP) in satellite access networks. In the future, we will investigate the issue of optimal *contention index*.

References

[1] D. Raychaudhuri, "Announced retransmission random access protocol", IEEE Trans. on Comm., vol. 33, no. 11, pp 1183-1190, 1985.

[2] N. Abramson, "Internet Access Using VSAT", IEEE Commu.Mag. pp. 60-68, July 2000.

[3] K. Sakakibara, H. Muta, and Y. Yuba, "The Effect of Limiting the Number of Retransmission Trials on the Stability of Slotted ALOHA Systems", IEEE Trans. on Vehi. Tech., vol. 49, no. 4, pp. 1449-1454, July 2000,

[4] S. Assavaruk, S. Sittichevapak, R. Varakulsiripunth, "Some general approaches for performance analysis of the ALOHA-based satellite channel", IEEE Asia-Pacific Conference on Circuits and Systems, 1998.

[5] H. Koraitim, S. Tohme, "Performane Analysis of Multiple Access Protocols for Multimedia Satellite Networks", IEEE JSAC, vol. 18, no.9, pp 1751-1763, Sep. 2000.

[6] A. Jamalipour, "Low Earth Orbital Satellites for Personal Communication Networks", Artech House Publishers, pp 121-122, 1998.

[7] S. Lepaja, K. Bengi, "A Random-Reservation Medium Access Protocol for Satellite Networks To Accomodate Real-Time Traffic", VTC01.