Minimum Length Transmission Scheduling of Return Channels for Multicode MF-TDMA Satellite Interactive Terminals

Dongwoo Kim, Dong-Hyuk Park, Ki-Dong Lee, and Ho-Jin Lee

***Abstract*—In this paper, we propose a transmission scheduling method that deals with a problem of determining superframe length (SL) when allocating the return channel resources to the capacity request from satellite interactive terminals (SITs). The proposed method is shown to converge to getting an acceptable solution of the problem. A main purpose of the method is to minimize the SL in order to reduce scheduling-wait-time as well as improve resource utilization. The method provides great flexibility in scheduling by limiting the SL as short as possible and also achieves high resource utilization by smoothing the time-varying demands with an overload control. Extensive simulation results show that the method successfully converges to a desirable SL within a few iterations and attempts to reduce the SL to provide highest flexibility as well as diminish idle resources.**

***Index Terms*—Minimum length transmission scheduling, multicode MF-TDMA, radio resource allocation, return channel satellite terminals.**

I. Introduction

FOR MULTIMEDIA satellite distribution networks, an interactive return channel system has been standardized by ETSI [1]. The return link from a satellite interactive terminal (SIT) of an individual user to an interactive server at a hub station is relayed via a *return-path satellite* using Ka/Ku-band [2]. The return link air interface specifies both traffic and signaling data formats and is based on matched-filter time-division multiple access (MF-TDMA). MF-TDMA allows a group of SITs to communicate with a hub using a set of carrier frequencies, each of which is divided into time slots. Multicode MF-TDMA is a variant of MF-TDMA, where different users can further share the spectrum in the same time slot by spreading and scrambling their traffic data with specifically assigned orthogonal codes [3]. Multicode MF-TDMA enables SITs to be operated in more interference-limited environments and could make them portable and mobile with smaller antennas [4].

The hub in the return link will allocate to each active SIT a series of bursts, each of which is defined by an available combination of *radio resources*: a frequency, codes, a start time and a duration. Only after receiving the allocation table via the forward link, an SIT can transmit data on the return channel in a contentionless mode [3]. The allocation schedules and regulates data transmission from SITs, and needs to be adaptive in order to accommodate the constantly varying demand on the channel resources from SITs.

In scheduling the return channel, the hub generally segments the whole time resource into consecutive *superframes*. A superframe is a designated scheduling unit for which the *burst configuration* of each user, such as modulation/coding parameters and requested traffic rates, is unchanged. In each superframe, a fraction of the resource should be devoted to carrying signaling traffic from SITs for synchronization, acquisition and access control. The signaling overhead is usually proportional to the total number of log-on users and kept in a minimum level to obtain a maximum return channel capacity.

Recent works that deal with optimizing resource allocation in the satellite return channel assume a fixed superframe length (SL) [5]–[8]. Since SITs call for the return channel resources to send multimedia traffic, the demands may vary swiftly and constantly. An acceptable return channel scheduling thus should be flexible enough to admit the time-varying demands. Fixed-length superframes certainly fail to efficiently adapt to the fluctuating requests from SITs. Too long a superframe especially increases a mean waiting time that is required for the next resource allocation and might cause idle resources that are not assigned to a specific user. With short superframes, on the other hand, the allocation can be more adaptive. It however increases the ratio of the signaling overhead to the total carried data and often fails to accommodate all the requests (i.e., so-called *overload* occurs) as a result of reducing the capacity. A plausible scheduler is therefore considered to determine the SL as short as possible but, at the same time, to be capable of assigning maximal resources in order to carry for the demands.

In this paper, we propose a transmission scheduling method that adaptively determines the length of superframes, which eventually allocates the radio resource to active SITs based on requests. A main purpose of the proposed method is to reduce the mean waiting time as well as the occurrence of idling slots. The method thus pertains to a minimum SL. An optimality, with respect to a waiting cost and an overload cost, of the method will be discussed. Given burst configurations, the method computes an initial SL with resource assignment and further reduces it by rearranging the assignment. This procedure will be shown to converge to getting a desirable SL. The proposed method also achieves high efficiency in use of the limited resources. As a practical measure of the resource utilization, a *resource-idling ratio* (RIR) of a superframe, defined as a ratio of unassigned

Manuscript received May 12, 2003; revised May 11, 2004, September 23, 2004. The review of this paper was coordinated by Dr. Y.-P. Eric Wang.

D. Kim, D.-H. Park, and K.-D. Lee are with the Department of Electrical Engineering and Computer Science, Hanyang University, Ansan, 425-791 Korea (e-mail: dkim@hanyang.ac.kr).

H.-J. Lee is with Electronics and Telecommunications Research Institute, Daejon, Korea (e-mail: hjlee@etri.re.kr).

Digital Object Identifier 10.1109/TVT.2005.853463

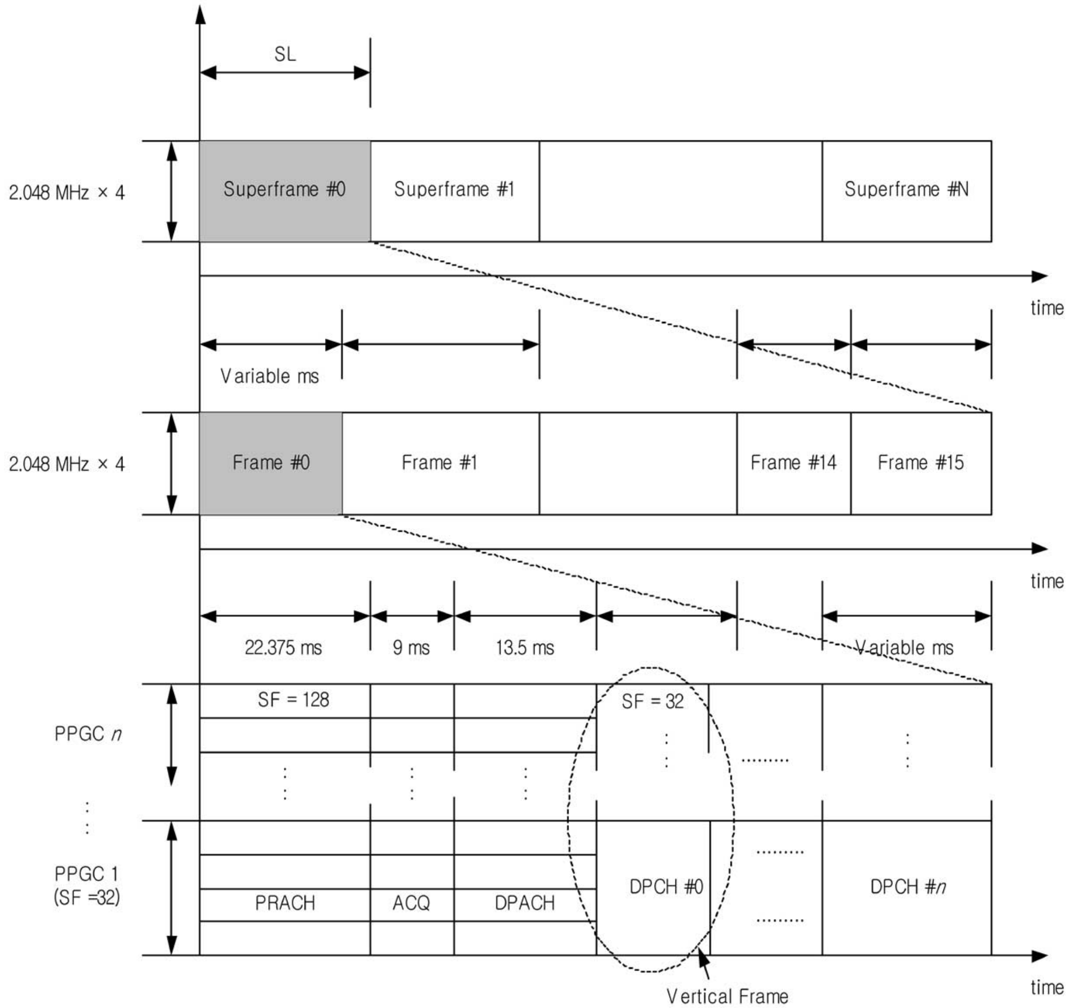


Fig. 1. Illustrative structure of multicode MF-TDMA superframes.

bandwidth to the total bandwidth available in a superframe, is also numerically obtained to evaluate the proposed method in various traffic environments.

Since the maximum capacity of the return link is normally limited, large requests from SITs cannot always be allocated the necessary radio resources. To prevent such an overload, call admission control can be used for screening the bandwidth request [9]. Temporal overload, however, is inevitable since multimedia traffic typically varies in offered rate and volume. The proposed method is also capable of dealing with the overload. It can detect the overload when calculating SL and autonomously reduces the portion of radio resources allocated to unurgent users while relieving the overload. The overload control helps to provide rather a stable input for the scheduler, which contributes to reducing RIR.

Many scheduling algorithms developed for different communication systems also have focused on quality of service (QoS) and fairness among contending terminals [10], [11]. Though those issues are crucial in designing a scheduler dealing with multimedia traffic, we restrict, in this paper, our attention to the "responsiveness" of scheduling, that is, how to wisely allocate the limited resource to contending terminals in order to achieve the efficient utilization. To include thorough discussions on QoS and fairness, we need to define more functions on higher-layer protocol stacks and further modify the network architecture given by [1] and [3], which will be the future study.

The rest of this paper is organized as follows. In the next section, we briefly introduce the multicode MF-TDMA system, provide the concept of minimum length transmission scheduling (MLTS) and derive an optimality condition. We describe the proposed method in detail in Section III starting with an overall sketch, provide a convergence proof the algorithm and discuss some implementational issues. Numerical evaluation of the method is presented in Section IV and we give a summary of our results and offer some concluding remarks in Section V.

II. Multicode MF-TDMA System and Minimum Length Transmission Scheduling

A. *Multicode MF-TDMA Return Channel System*

In multicode MF-TDMA [3], a frame is characterized by a set of frequency assignments (FA), spreading codes and time slots. Fig. 1 gives an example of a valid frame. In a frame, there are

TABLE I
AVAILABLE NUMBERS OF PPGCs

SINR (dB)	SF=32	SF=64	SF = 128
4	13	26	52
5	11	21	41
6	8	17	33
7	7	13	26

TABLE II
THE DATA SLOT LENGTH (IN ms) FOR TYPICAL CODE RATES

Code rate	SF=32	SF=64	SF = 128
1/3	14.00	28.00	57.88
2/5	12.34	24.69	51.25
1/2	10.69	21.38	44.63
2/3	9.03	18.06	38.00
3/4	8.48	16.96	35.79
4/5	8.20	16.41	34.69
5/6	8.04	16.08	34.03
6/7	7.93	15.85	33.58
7/8	7.85	15.70	33.27

four types of burst channels, similar to MF-TDMA [1]: physical random access channel (PRACH) for common signaling, acquisition channel (ACQ) for coarse synchronization, dedicated physical acquisition channel (DPACH) for fine synchronization, and dedicated physical channel (DPCH) for delivering user data. The overhead slots, PRACH, ACQ, and DPACH are located consecutively at the start of each frame and should occur more than once for a log-on SIT in a superframe. The minimum volume of the overhead in a superframe is usually determined by the number of log-on SITs.

All the burst channels are spread by preferentially phased gold code (PPGC) in 2.048 MHz per FA [3]. The three signaling channels use the spreading factor (SF) 128. The SF of DPCH is normally 32 but adaptively changes to the return link condition. When the link condition goes worse (e.g., in case of a rainfall), it is increased stepwise into 64 and 128. Given SF, the number of PPGCs that can be simultaneously assigned with keeping the received signal-to-noise-plus-interference ratio (SINR) at a required value is obtained under a specific link budget. Table I, illustrates the target SINR and the available number of PPGCs for the respective SF.

Both the signaling and the traffic channels are encoded for error protection and modulated using QPSK like in [1]. The synchronization scheme in multicode MF-TDMA is also the same as in MF-TDMA based on a special 27 MHz program clock reference (PCR) signaled in DVB/MPEG2-TS private sections [12]. Different coding methods as well as different SFs result in different slot lengths. The signaling slot length is depicted in Fig. 1, and the data slot length that is needed to carry a 53-byte ATM cell is illustrated in Table II, for different burst configurations.

The return channel resources in Fig. 1 can be viewed in two dimensions: time is the horizontal resource, but FAs and codes are the vertical resource. Since the height of vertical dimension is fixed and given in a superframe, time is an important decision variable in scheduling. A *vertical frame* (VF) is defined as a set of time slots that can be aligned to have the same start time. A VF is called *full* if its vertical resources are wholly occupied at least on the start time by the respective time slots. Since signaling slots are usually positioned at the start of each frame, they may constitute three full VFs. When slots with the same length build a VF, it is called a *complete* VF.

An SIT that has data to send on a return channel initially accesses the satellite system in a slotted ALOHA mode using PRACH slot. In case of a successful access, the SIT is instructed to move to ACQ for coarse synchronization and finally gets to use repeatable DPACHs. After synchronization, SITs could make requests on the return traffic channel resources through satellite access control (SAC) messages on the signaling slots. The scheduling algorithm in the hub then allocates an adequate amount of resource on each request. The scheduling period may be an integer multiple of the SL. In this paper, we assume the scheduling occurs at the start of every superframe for achieving desirable flexibility.

There are in general five categories of capacity request [1]: continuous rate assignment (CRA), rate based dynamic capacity (RBDC), volume based dynamic capacity (VBDC), absolute volume based dynamic capacity (AVBDC) and free capacity assignment (FCA). CRA and RBDC constantly require the rates, though RBDC request expires after a time-out period (e.g., 2 superframes). VBDC and AVBDC are volume capacities. VBDC is cumulative. AVBDC is used instead of VBDC when the SIT senses that its previous VBDC request might be lost. In order to decide how much volume among VBDC and AVBDC requests is carried on the considered superframe, we assume that a target rate for each volume request exists. The target rates are autonomously adjusted by the proposed algorithm during scheduling and, as a result, VBDC and AVBDC requests are assigned dynamically changing volume of resources. Capacity assigned in FCA category is intended as a bonus which can be used to improve QoS on any traffic. In this paper, FCA is treated as an idle assignment.

B. Definition and Optimality of Minimum Length Transmission Scheduling

The purpose of this subsection is to provide a basic idea that is essential in constructing the proposed method. Especially, MLTS is defined as a reasonable rule of determining SL and its optimality is examined with respect to minimizing costs of waiting time for the next scheduling as well as overload that causes uncarried traffic.

When the burst configuration to be employed in a superframe is given according to the requests, the mean waiting time and the resource utilization are mainly determined by the SL (in seconds), l. Let R (ksps) be the total available capacity and B (ksps) the sum of valid capacity requests. Given the burst configuration, R is accurately estimated since the available FAs and PPGCs are normally fixed. Let S (kilo-symbols) be needed to locate the necessary signaling slots in a superframe, which is not dependent on the SL but usually determined by the number of log-on users.

When a value for l is chosen, assuming the capacity request from SITs is uniformly inserted into a superframe, the mean waiting time for the next scheduling is $l/2$. Then a waiting cost $W(l)$ can be modeled as

$$W(l) = \begin{cases} c_w \frac{l}{2}, & \text{if } l \le l_{\max} \\ \infty, & \text{if } l > l_{\max} \end{cases} \tag{1}$$

where $c_w (> 0)$ is a unit cost for the mean waiting time. $W(l)$ linearly increases for $l \le l_{\max}$. $l_{\max}$ is an engineering parameter that represents the maximum allowable waiting time before the scheduling starts. Mean waiting time larger than $l_{\max}/2$ is prohibited with the cost function $W(l)$.

Small l evidently reduces the waiting cost. It however incurs *overload* when

$$S + Bl > Rl. \tag{2}$$

If $B \ge R$, the system cannot accommodate the whole request B in the return channel whatever l is selected. Otherwise, the overload can be avoided by selecting l such that

$$l \ge \frac{S}{R-B}. \tag{3}$$

Assuming the overload cost is proportional to the overload ratio, $S + Bl/Rl$, we have an overload cost function

$$V(l) = \begin{cases} c_v \frac{S+Bl}{Rl}, & \text{if } 0 < l < \frac{S}{R-B} \\ 0, & \text{if } l \ge \frac{S}{R-B} \end{cases} \tag{4}$$

where $c_v (> 0)$ is a unit cost for the overload ratio.

The total cost $C(l) = W(l) + V(l)$ is then

$$C(l) = \begin{cases} c_w \frac{l}{2} + c_v \frac{S+Bl}{Rl}, & \text{if } l < \min\{l_{\max}, \frac{S}{R-B}\} \\ c_w \frac{l}{2}, & \text{if } \frac{S}{R-B} \le l \le l_{\max} \\ \infty, & \text{elsewhere.} \end{cases} \tag{5}$$

An optimal l that minimizes $C(l)$ can be found as

$$l^* = \begin{cases} \frac{S}{R-B}, & \text{if } \frac{S}{R-B} \le l_{\max} \\ \min\{\sqrt{\frac{c_v}{c_w}\frac{2S}{R}}, l_{\max}\}, & \text{if } \frac{S}{R-B} > l_{\max}. \end{cases} \tag{6}$$

If c_v is sufficiently greater than c_w

$$l^* = l_{\min} \overset{\text{def}}{=} \min\{\frac{S}{R-B}, l_{\max}\}. \tag{7}$$

An MLTS algorithm in this paper is defined as a scheduling algorithm that selects $l_{\min}$ in (7) as an SL. MLTS is named since $S/R - B$, a minimal length that can avoid the overload from (3), is selected as $l_{\min}$ unless it violates the limitation on the waiting time. MLTS is primarily intended to prevent the overload as taking great c_v in (5). And it also gives a short superframe as possible in order to provide maximal flexibility for the resource allocation in the return channel. When

$$c_v \ge c_w l_{\max}^2 \frac{R}{2S} \tag{8}$$

MLTS is an optimal strategy in selecting SL, which minimizes the total cost in (5). It is noted that MLTS provided in the following does not assume specific values for c_v and c_w. Performance of MLTS will be evaluated primarily in terms of RIR without the assumed costs.

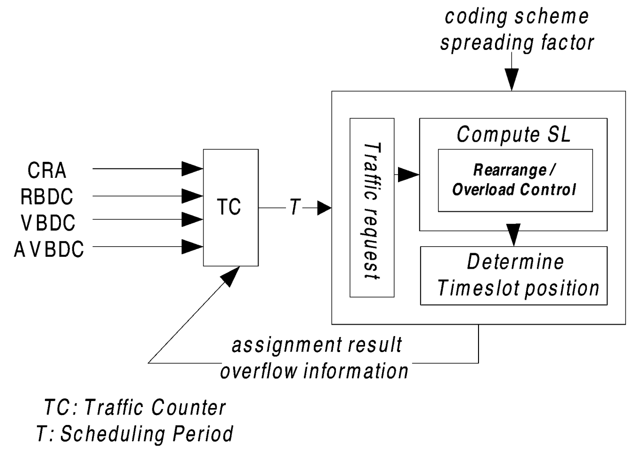


Fig. 2. Scheduler structure.

III. Proposed MLTS Method

A. Overall Structure

Fig. 2 shows a scheduler structure considered in this paper. The traffic counter (TC) gathers the capacity request information on SAC from SITs and transfers it to the scheduler. TC converts the volume capacity requests of VBDC and AVBDC into rate requests that will be provided for the scheduler. TC operates timers for RBDC and updates rate targets for VBDC and AVBDC. A short list of variables that TC maintains is provided in Table III.

Schedulers are evoked periodically. The proposed algorithm assumes it works at every superframe. The scheduler is also informed of the coding scheme and SF currently used by each SIT from the upper protocol layer via a management function. It determines an SL with the rule provided in (7) and tries to further reduce the SL by rearranging individual DPCH slots.

In MLTS, the maximum return channel capacity is achieved by $l_{\max}$. When all the capacity requests from SITs cannot be supported even with $l_{\max}$ (i.e., $S/(R-B) > l_{\max}$), a certain portion of the request is not allocated to the present superframe. The scheduler detects this overload during the computation of SL and then reduces the target rates of VBDC and AVBDC sequentially until

$$\frac{S}{R-B} = l_{\max}. \tag{9}$$

The scheduler assumes traffic priority in the order of CRA, RBDC, AVBDC, and VBDC. When the overload is detected, the lower order traffic request is reduced earlier.

The proposed method consists of four modules: computing a size of overhead vertical frames, computing a coarse SL (CSL), reducing the CSL by slot rearrangement, and checking convergence as well as overload. The method is iterative in nature: it starts with any SL, l_{in}, and produces an SL, l_{RSL}, after the rearrangement; l_{RSL} is checked for convergence in module 3 where the method is possibly restarted with l_{RSL} as a new input SL. When convergence is achieved, a resulting SL, l_{out}, is obtained. Convergence of the method will be discussed later.

TABLE III
TYPICAL VARIABLES THAT TC MAINTAINS

Variable	Description
N_{SIT}	the number of log-on SITs
$r_{i,\cdot}$	rate request from SIT i (ksps)
$v_{i,\cdot}$	remaining volume request from SIT i (kilo-symbols)
$h_{i,\cdot}$	the number of superframes elapsed after a capacity request from SIT i is initiated
$d.$	a counting threshold for rejecting an aged capacity request
$t_{i,\cdot}$	a target rate for volume request from SIT i (ksps)

B. Module 0: Computing A Size of Overhead Vertical Frames

Let define a *vertical signaling frame* (VSF) as a complete VF that consists of the three sequenced signaling slots in each vertical layer. Since every log-on SIT should be assigned at least one of the signaling sequences

$$N_{\mathrm{VSF}} \geq \frac{N_{\mathrm{SIT}}}{N_{\mathrm{FA}} \cdot N_{\mathrm{PPGC}}} \tag{10}$$

where N_{A} denotes the respective number of A. When N_{VSF} is determined, a signaling overhead ratio (SOR) is achieved as follows:

$$\mathrm{SOR} = \frac{N_{\mathrm{VSF}} \cdot D_{\mathrm{VSF}}}{l_{\mathrm{in}}} \tag{11}$$

where D_{VSF} (in seconds) denotes the duration of VSF (e.g., 44.875 ms in Fig. 1). Since a large SOR is a burden to obtain efficiency in resource utilization, SOR is kept below a target SOR_r, and then N_{VSF} is upper bounded by

$$N_{\mathrm{VSF}} \leq \frac{\mathrm{SOR}_r \cdot l_{\mathrm{in}}}{D_{\mathrm{VSF}}}. \tag{12}$$

Thus, an acceptable N_{VSF} is obtained with (10) and (12). Note that if (10) and (12) make a conflict, (10) may have a priority since it is a system requirement.

C. Module 1: Computing CSL

In this module, a coarse SL is computed with the following inputs: l_{in}, rate requests given by TC, and N_{VSF} obtained from module 0. Let I_k be a set of SITs using coding scheme k. Then, for each coding scheme k, the sum of the symbol rate capacity (in kilo-symbols) requested for l_{in} is

$$b_k = l_{\mathrm{in}} \sum_{i \in I_k} (r_{i,\mathrm{CRA}} + r_{i,\mathrm{RBDC}} + r_{i,\mathrm{VBDC}} + r_{i,\mathrm{AVBDC}}) \tag{13}$$

where $r_{i,\mathrm{A}}$ (ksps) denotes the rate request of type A capacity category from SIT i and is given by TC. Given a coding scheme and an SF, the number of symbols required in a data slot to send a 53-byte ATM cell, e_k, is determined by adding preambles and guard symbols. Then the number of data slots needed to send b_k is

$$q_k = \lceil b_k / e_k \rceil \tag{14}$$

where $\lceil X \rceil$ gives the smallest integer greater than or equal to X.

If $q_k \geq N_{\mathrm{FA}} \cdot N_{\mathrm{PPGC}}$, a complete *vertical data frame* (VDF) can be obtained, where all the data slots have the same duration D_k (in second). Let $N_{k,\mathrm{VDF}}$ be the minimal number of VDF that is needed to send q_k data slots, then

$$N_{k,\mathrm{VDF}} = \lceil \frac{q_k}{N_{\mathrm{FA}} \cdot N_{\mathrm{PPGC}}} \rceil. \tag{15}$$

If $N_{k,\mathrm{VDF}}$ VDFs are not full, the number of empty data slots in these VDFs is

$$N_{k,\mathrm{EMT}} = N_{k,\mathrm{VDF}} \cdot N_{\mathrm{FA}} \cdot N_{\mathrm{PPGC}} - q_k. \tag{16}$$

Given $N_{k,\mathrm{VDF}}$ for each k, CSL (l_{CSL}) is obtained by

$$l_{\mathrm{CSL}} = D_{\mathrm{VSF}} \cdot N_{\mathrm{VSF}} + \sum_{k=1}^{K} (D_k \cdot N_{k,\mathrm{VDF}}) \tag{17}$$

where K is the number of different coding schemes. Note that the l_{CSL} certainly overestimates the necessary SL and may produce idle resources as $\sum_{k=1}^{K}(N_{k,\mathrm{EMT}} \cdot D_k)$ in a time unit. To reduce this inefficiency, the following module 2 shuffles data slots in the above VDFs to achieve full VDFs as many as possible.

D. Module 2: Reducing CSL by Rearrangement

We assume, without loss of generality, the index k denoting the coding scheme is given such an order as $D_1 \geq D_2 \geq \cdots \geq D_K$. Let an intermediate variable $N_{k,\mathrm{MRG}}$ denote the number of marginal slots that need rearrangement

$$N_{k,\mathrm{MRG}} = N_{\mathrm{FA}} \cdot N_{\mathrm{PPGC}} - N_{k,\mathrm{EMT}} \tag{18}$$

and repeat the following until there is no k such that $k > k^*$ and $N_{k,\mathrm{MRG}} > 0$ as well as $N_{k,\mathrm{EMT}} > 0$:

- $k^* \leftarrow \min\{k : N_{k,\mathrm{MRG}} > 0 \text{ and } N_{k,\mathrm{EMT}} > 0, 1 \leq k \leq K\}$;
- for the smallest k such that $k > k^*$ and $N_{k,\mathrm{MRG}} > 0$ as well as $N_{k,\mathrm{EMT}} > 0$, move $v_k = \min\{N_{k^*,\mathrm{EMT}}, N_{k,\mathrm{MRG}}\}$ data slots of coding type k into not a complete VDF of coding type k^*;
- $N_{k^*,\mathrm{EMT}} \leftarrow N_{k^*,\mathrm{EMT}} - v_k, N_{k,\mathrm{MRG}} \leftarrow N_{k,\mathrm{MRG}} - v_k, N_{k,\mathrm{VDF}} \leftarrow N_{k,\mathrm{VDF}} - 1$ if $N_{k,\mathrm{MRG}} = 0$, and check the repetition condition;

where $\leftarrow$ denotes an assignment operation.

The above rearrangement activity makes VDFs as full as possible. As a result, at most one VDF remains not full. After rearrangement, the SL is re-computed as (17), which is denoted by l_{RSL}. Evidently, $l_{\mathrm{RSL}} \leq l_{\mathrm{CSL}}$. With l_{RSL}, idle resources are produced due to two factors: the unequal length of data slots in newly filled VDFs during the above rearrangement and the final $N_{k^*,\mathrm{EMT}} > 0$. To measure the idleness, RIR is defined as a ratio of the total of the mismatched time length plus $D_{k^*} \cdot N_{k^*,\mathrm{EMT}}$ to total resource $l_{\mathrm{RSL}} \cdot N_{\mathrm{FA}} \cdot N_{\mathrm{PPGC}}$.

E. Module 3: Termination Condition and Overload Control

If $l_{\rm RSL} = l_{\rm in}$, a converged SL is obtained. Otherwise, let $l_{\rm in} = l_{\rm RSL}$ and repeat module 0. When a convergence occurs, let $l_{\rm out} = \min\{l_{\rm in}, l_{\max}\}$ as MLTS. In case of $l_{\rm in} > l_{\max}$ at convergence, overload is detected and some capacity request should not be assigned the channel resources.

The proposed overload control gradually reduces the capacity request to achieve $l_{\rm in} \leq l_{\max}$ but to keep the throughput in a maximum level as the follows:

- $l_{\rm RDC} \leftarrow l_{\rm in} - l_{\max}$, where $l_{\rm RDC}$ is interpreted as a target time in reducing $l_{\rm in}$;
- $$r_{\rm RDC} \leftarrow \frac{l_{\rm RDC} \cdot N_{\rm FA} \cdot N_{\rm PPGC} \cdot e_{\rm AVG}}{N_{\rm SIT} \cdot l_{\max} \cdot D_{\rm AVG}} \tag{19}$$

  where $r_{\rm RDC}$ is a rate reduction target, and $e_{\rm AVG}$ and $D_{\rm AVG}$ are average values for e_k and D_k, respectively

  $$e_{\rm AVG} = \frac{1}{K}\sum_{k=1}^{K} e_k; \quad D_{\rm AVG} = \frac{1}{K}\sum_{k=1}^{K} D_k; \tag{20}$$
- reduce the capacity request $r_{i,\cdot}$ by the amount of $r_{\rm RDC}$ in the traffic priority order.

In (19), time reduction target $l_{\rm RDC}$ is converted into rate reduction target $r_{\rm RDC}$ with considering both bandwidth and an average coding parameter. After reducing the rate, the method resumes at module 0 in order to achieve an un-overloaded convergent result. Convergence discussion for the overall procedure will be provided in the next subsection.

F. Convergence and Other Implementatinal Issues

Proposition: The iterative method consisting of the above modules converges to $l^*_{\rm RSL} \leq l_{\max}$ if the sum of capacity request is finite.

Proof: Let iterations of the method be indexed by $(t), t = 0, 1, \ldots$, and $N_{\rm VSF}$ be fixed without loss of generality. Once we have $l^{(t)}_{\rm RSL} < l^{(t)}_{\rm in}$ from any starting $l^{(0)}_{\rm in}$, the method generates a monotone decreasing real sequence of $l^{(\tau)}_{\rm RSL}$ for $\tau \geq t$ since the request $b^{(\tau)}_k$ in (13) also decreases hereafter. As the $l^{(\tau)}_{\rm RSL}$ is certainly bounded below, it converges, say, $l^*_{\rm RSL}$. If $l^*_{\rm RSL} \leq l_{\max}$, we have the desired result. Otherwise, the overload control is applied.

On the contrary, if the method generates a monotone increasing sequence of $l^{(t)}_{\rm RSL}$. Then the sequence either converges to $l^*_{\rm RSL} \leq l_{\max}$ or gets to $l^{(t)}_{\rm RSL} > l_{\max}$ after some t. In the former case, we obtain the desired result. But in the latter case the overload control is applied again.

Suppose that the overload control is evoked infinitely. Since $l^{(t)}_{\rm RDC}$ is always positive at the overload control, so is $r^{(t)}_{\rm RDC}$, which means the sum of rate reduction made by (19) also goes to infinity, which is impossible if the sum of capacity request is finite.

At a light load, the method converges very fast. Although the computational burden of the method slightly increases when the overload control starts, it also can be speeded up by setting rather a longer time reduction target than $l_{\rm RDC}$. Moreover, with adopting a conservative recovery process from the reduced request, the method would avoid frequent overload computation. The recovery process, not discussed in detail in the paper, is exactly the reverse operation of the overload control, which recovers the reduced requests when a convergent $l_{\rm in} < (1-\alpha)l_{\max}$ where $0 < \alpha < 1$. α may control the speed of recovery. Let a target time for recovery $l_{\rm INC} \leftarrow (1-\alpha)l_{\max} - l_{\rm in}$ and restore the capacity request with the rate increase target obtained from (19) where $l_{\rm RDC}$ is substituted for $l_{\rm INC}$.

After the convergence, resulting VDFs can be easily distributed to each request if the changed positions of data slots in module 2 are appropriately recorded. This procedure finalizes the resource allocation in scheduling. When locating VSFs in a superframe, a certain optimized strategy that considers the distribution of VSFs or the scheduling interval can be applied. Though the proposed method is targeting the multicode MF-TDMA system in [3], it can be directly used for an MF-TDMA system based on [1] by setting $N_{\rm PPGC} = 1$, replacing Table II, and neglecting the discussion related to SF. In case of a rainfall, the satellite link may suffer from more attenuation and the SITs then need to increase SF or the transmitter power. Allowing a higher SF for the SITs in bad link condition, the method can be easily modified to support different SFs by letting the index k used for a coding scheme indicate a combination of an SF and a coding scheme. This may increase RIR, which will be numerically examined later.

IV. Numerical Results

To evaluate performances of the proposed method, we use, in the simulation, the data presented in Fig. 1 as well as Tables I and II, which are employed in the developing mobile satellite internet access (MSIA) systems [3].

During the simulation, we assume the number of log-on SITs is constantly 20 and each capacity request is made independently. The duration between two consecutive rate requests is assumed to have a geometric distribution with means 10 and 5 superframes for CRA and RBDC, respectively. For VBDC and AVBDC, the requests are assumed not overlapped to avoid unrecoverable overload. The amount of the request is generated by a uniform distribution with different parameters. To indicate different distributions, we use a notation CPA, where "CPA" denotes one of capacity categories, and a and b are respectively a lower and an upper end of a uniform distribution. Both a and b have a unit of either ksps or kilo-symbols.

Fig. 3 illustrates the convergence property of the proposed system. We capture some arbitrarily loaded cases that activate the overload control procedure with a starting SL $l^{(0)}_{\rm in} = 2.0$. In most of the cases, the method converges to $l_{\rm RSL} \leq l_{\max} = 3.0$ within seven iterations. With 10 iterations, it successfully converges in all the cases. Moreover, a sample mean of $l_{\rm out}$ over the 1000 overload cases is 2.985, which shows that the method allocates the capacity as large as possible even when overload is detected.

In Table IV, the resource utilization performance between the proposed MSL method and a fixed-SL method is compared. In the fixed SL method, the SL of 3 is constantly used. The data

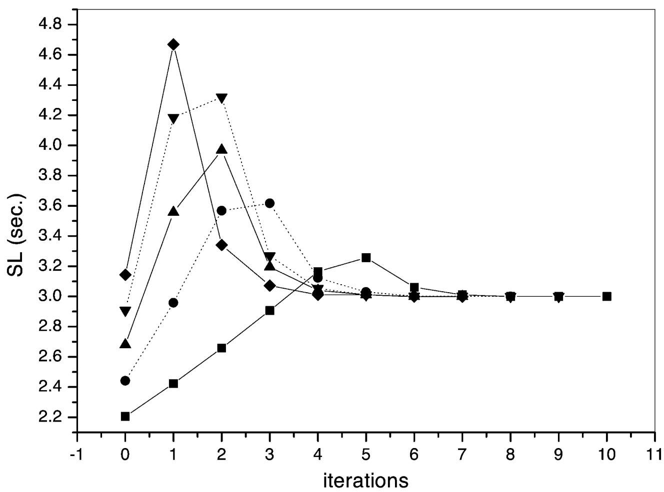


Fig. 3. Convergence illustration ($\mathrm{SF} = 32, l_{\mathrm{in}}^{(0)} = 2.0, l_{\max} = 3.0$)

TABLE IV
RIR (%) PERFORMANCE COMPARISON

Load		Fixed	Proposed
CRA(20, 60)	RBDC(50, 160)	23.96	0.25
CRA(20, 80)	RBDC(50, 180)	18.62	0.24
CRA(20, 100)	RBDC(50, 200)	11.84	0.23
CRA(20, 120)	RBDC(50, 220)	6.75	0.23
CRA(20, 140)	RBDC(50, 240)	5.40	0.24
VBDC(1, 400)	AVBDC(1, 400)	22.54	0.45
VBDC(1, 4K)	AVBDC(1, 4K)	8.80	0.23
VBDC(1, 40K)	AVBDC(1, 40K)	2.93	0.26

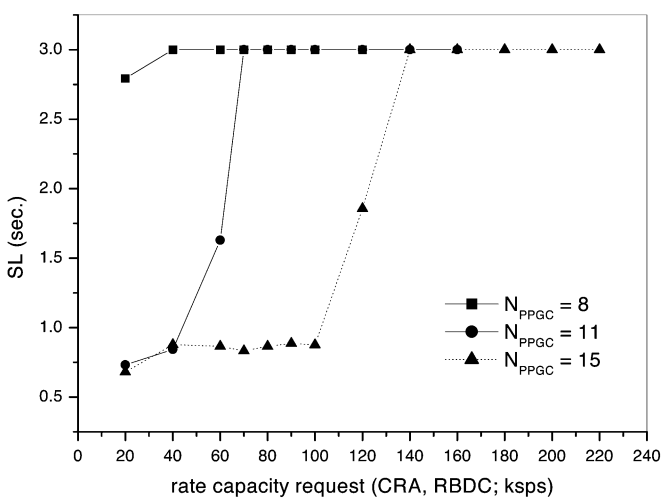


Fig. 4. SL behavior when offered load increases ($\mathrm{SF} = 32, l_{\mathrm{in}}^{(0)} = 2.0, l_{\max} = 3.0$).

traffics offered for simulating various load conditions are shown at the first column. Upper 5 rows are designated for increasing rate capacity requests: CRA and RBDC. Lower 3 rows are for varying volume capacity requests: VBDC and AVBDC. RIR generally decreases as the capacity demands increase in both methods. The proposed one outperforms the fixed SL method.

In Fig. 4, the SL generated by the proposed method is illustrated for increasing demands on CRA and RBDC. The numbers on x-axis are means of the uniform distribution for CRA

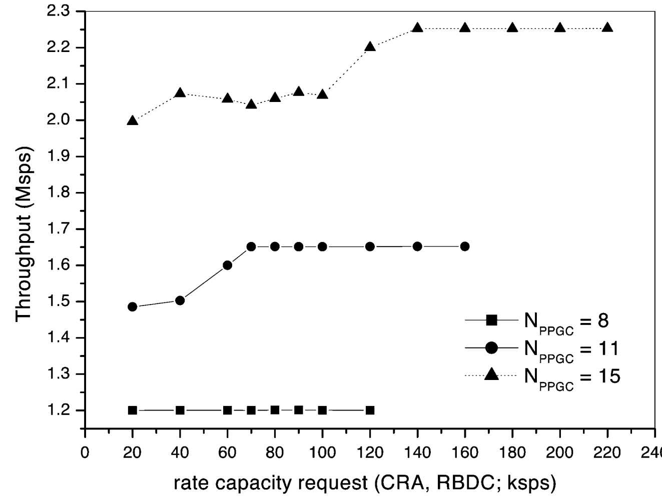


Fig. 5. Throughput behavior when offered load increases ($\mathrm{SF} = 32, l_{\mathrm{in}}^{(0)} = 2.0, l_{\max} = 3.0$).

and RBDC. When load increases, the method with the smaller N_{PPGC} reaches a neighbor of $l_{\max}$ (2.985 in the figure) at the earlier step. In Fig. 4, the method using $N_{\mathrm{PPGC}} = 15$ produces the limited SL when the load becomes 140, which is double of the load where the one with $N_{\mathrm{PPGC}} = 11$ reaches the limit. Fig. 4, also shows the SL that supports the capacity request is very sensitive to the load variation. The SL suddenly jumps from 0.8 to $l_{\max}$ for small increase in the load. Since the multimedia traffic volumes are very fluctuating, without the delay bound $l_{\max}$ defined in cost function $W(l)$, any method only to maximize the throughput frequently results in unfavorably long SL.

The throughput achieved by the proposed method is shown in Fig. 5. At the same points where the method reaches the SL limit, maximum throughputs are provided. Since N_{PPGC} mainly determines the amount of vertical resources, the maximum throughputs are obtained proportionally to N_{PPGC}. Since the throughput is measured in QPSK symbol rate, the assumed multicode MF-TDMA return channel with $N_{\mathrm{PPGC}} = 11$ achieves about 38% of bandwidth efficiency in Fig. 5. When the throughput and the SL arrive at the respective limited values, the overload control is probably being activated. For the same simulation data in Figs. 4 and 5, Fig. 6, shows the overload control behavior.

In Fig. 6, reducing average target rates for VBDC and AVBDC is plotted as the load increases. When the system reaches its maximum performance, overload control starts to be evoked. The rate target for VBDC is reduced first and then that of AVBDC is reduced only after the former gets to 0. This kind of overload control gives higher priority to AVBDC than VBDC since SITs could request in AVBDC if VBDC request is not responded. We let the simulation terminate if $\mathrm{AVBDC} = 0$, which needs to reduce RBDC rates. RBDC and CRA may require high QoS, which may not allow rate reduction. To maintain adequate service quality, capacity requests should be screened before $\mathrm{AVBDC} = 0$.

The overload control gradually adapts to fluctuating high load by adjusting the amount of volume capacity request. Thus, the scheduler actually treats rather a stable level of input traffic

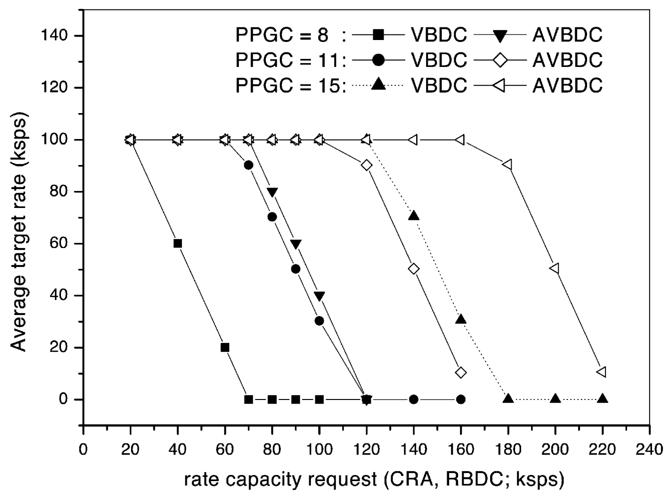


Fig. 6. Overload control behavior when offered load increases ($SF = 32, l_{\rm in}^{(0)} = 2.0, l_{\max} = 3.0$).

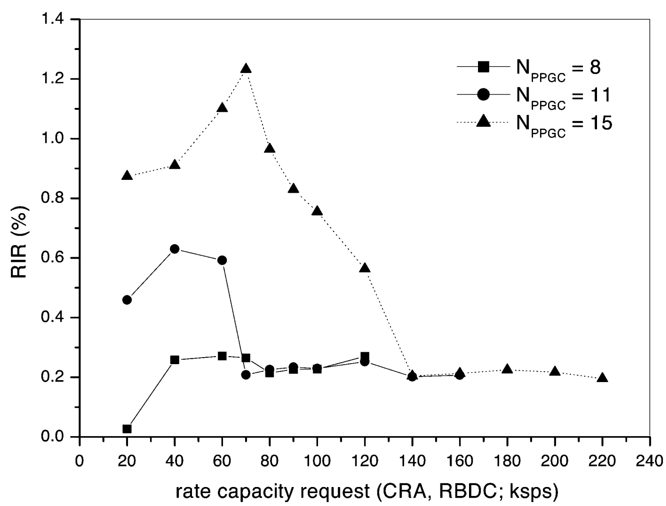


Fig. 7. RIR performance when offered load increases ($SF = 32, l_{\rm in}^{(0)} = 2.0, l_{\max} = 3.0$).

when high load is offered, which contributes to reducing RIR as depicted in Fig. 7. By comparing Fig. 6, with Fig. 7, we know RIR is minimized after the overload control starts.

In Fig. 8, the throughputs achieved when using different SFs are illustrated. When using PPGCs with SF 64 and 128, greater SINRs are assumed such as 6 and 7 dB, respectively. With this assumption, the throughput decreases about 30% for each increase of SF. If assuming the same SINR (e.g., 5 dB), $N_{\rm PPGC} = 21$ and 41 for SF 64 and 128, respectively, in our model. Thus, just a slight reduction is expected when SF increases with the same SINR requirement.

Using PPGCs of different SFs in a superframe is not usual since those PPGCs used in the same time slot on a certain FA cause severe cross interference. Thus, capacity requests from SITs using different SFs should be allocated to different time slots or different FAs. If adopting this separate allocation rule, RIR is expected to greatly increase as a function of employed

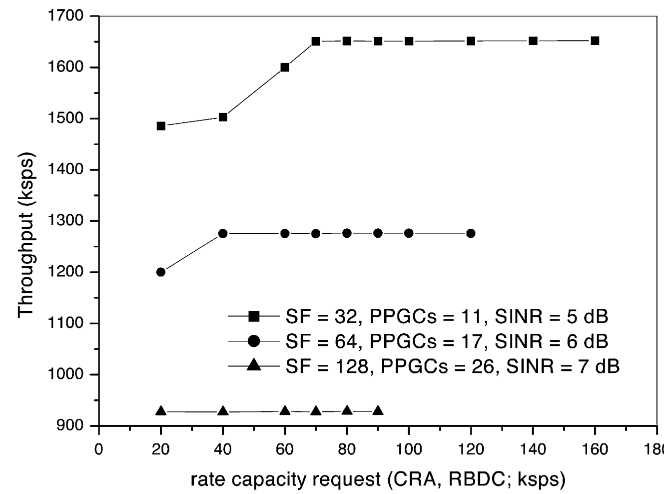


Fig. 8. Comparison of throughput when different SF is employed ($l_{\rm in}^{(0)} = 2.0, l_{\max} = 3.0$).

TABLE V
RIR OF A SYSTEM WITH MIXED SFs

Load		RIR (%)
CRA(10, 30)	RBDC(20, 80)	4.77
CRA(10, 40)	RBDC(20, 90)	4.37
CRA(10, 50)	RBDC(20, 100)	3.67
CRA(10, 60)	RBDC(20, 110)	3.35
CRA(10, 70)	RBDC(20, 120)	2.60
CRA(10, 80)	RBDC(20, 130)	2.54
CRA(20,60)	RBDC(50,160)	1.70
CRA(20, 80)	RBDC(50, 180)	1.70

number of SF types. But, as discussed above, RIR also goes to a certain minimum value if overload is detected and input request to the scheduler remains stable at a high level for all the SF types. Table V supports this conjecture. In Table V, an equal number of SITs use SF 32, 64, and 128, respectively. At a light load, RIR is much higher than the result shown in Table IV, and Fig. 7. But RIR diminishes when the load increases.

V. CONCLUSION

The problem of determining SL in allocating the return channel resources to the capacity request from SITs has been addressed. The proposed MLTS method has been shown to converge to a reasonable solution of the problem. The method provides great flexibility in scheduling by limiting the SL at a desired level and achieves high resource utilization by smoothing the time-varying demands with the overload control.

The performance of the proposed method has been studied by computer simulation. Our result reveals that the method successfully converges to an SL within a few iterations and attempts to reduce the SL to provide highest flexibility. It also reduces RIR and achieves a maximal throughput as possible. Extensive testing shows that the MLTS is robust to fluctuating capacity requests and, moreover, the overload control facilitates rather a stable input to the scheduler, which contributes to reducing RIR. The proposed method can also serve as a useful tool for achieving a desirable tradeoff among throughput and waiting time by selecting an adequate $l_{\max}$.

The proposed method opens an issue of resource-scheduling for return channels in broadband satellite systems. Further study should be devoted to how to adapt the scheduling method in order to incorporate specific applications with explicit QoS requirements, support user mobility and cooperate with terrestrial mobile systems.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments that improve the clarity of this paper.

REFERENCES

[1] DVB, "Interaction Channel for Satellite Distribution Systems," ETSI EN 301 790 (v1.2.2), Dec., 2000.

[2] P. Maveddat, G. Boray, and F. Homayoun, "Advanced return channel system architecture for broadband multimedia communications via satellite," in *Proc. IEEE Globecom'99*, 1997, pp. 1103–1108.

[3] ETRI, "Wireless Interface Specifications of Interactive Satellite Channels for Mobile Satellite Internet Access," Radio & Broadcasting Lab., May, 2002.

[4] J. Neale, R. Green, and J. Landovskis, "Interactive channel for multimedia satellite networks," *IEEE Commun. Mag.*, pp. 192–198, Mar. 2001.

[5] K.-D. Lee and K.-N. Chang, "A real-time algorithm for timeslot assignment in multirate return channels of interactive satellite multimedia networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 3, pp. 518–529, Apr. 2004.

[6] K.-D. Lee, "An efficient real-time method for improving intrinsic delay of capacity allocation in interactive GEO satellite networks," *IEEE Trans. Veh. Technol.*, vol. 53, no. 2, pp. 538–546, Mar. 2004.

[7] K.-D. Lee, H.-J. Lee, Y.-H. Cho, and D. G. Oh, "Throughput-maximizing timeslot scheduling for interactive satellite multiclass services," *IEEE Commun. Lett.*, vol. 7, no. 6, pp. 263–265, Jun. 2003.

[8] K.-D. Lee, H.-J. Lee, Y.-H. Cho, and D. G. Oh, "Improving efficiency of timeslot assignment for non-realtime data in a DVB- RCS return link: Modeling and algorithm," *ETRI J.*, vol. 25, no. 4, pp. 211–218, Aug. 2003.

[9] D. Kim, "Efficient interactive call admission control in power-controlled mobile systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 3, pp. 1017–1028, May 2000.

[10] D. Kim, D. S. Kwon, and S. K. Kim, "A simple minimum rate supporting scheduler for UTRA/TDD high speed downlink packet access," in *Proc. 57th IEEE Veh. Technol. Conf.*, Apr. 2003.

[11] Y. Cao and V. O. K. Li, "Scheduling algorithms in broad-band wireless networks," *Proc. IEEE*, vol. 89, no. 1, pp. 76–87, Jan. 2001.

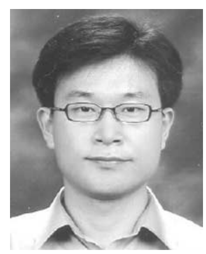
[12] J. Neale, C. Black, and P. Siebert, "Satellite interactive access by frame marker or program clock reference," *Proc., Sat. Commun. Architectures and Networks, ICC'99*, Vancouver, BC, Canada, Jun. 1999.

Dongwoo Kim received the B.A. degree from Seoul National University, Seoul, Korea, in 1987, and the M.S. and Ph.D. degrees in telecommunications engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejon, Korea, in 1989 and 1994, respectively.

From 1994 to 2000, he was with Shinsegi Telecomm, Inc., Seoul, Korea and involved in the design, development, test, and optimization of commercial CDMA cellular systems in Korea. Since March 2000, he has been with Department of Electrical Engineering, Hanyang University, Ansan, Korea. His present research interest is in the area of radio resource management, ad-hoc mobile networking and diversity techniques for future advanced mobile systems. He has served as an Editor for the *Journal of Communications and Networks.*

Dong-Hyuk Park received the B.S. and M.S. degrees in electrical engineering from Hanyang University, Ansan, Korea, in 2003 and 2005, respectively.

In March 2005, he joined with Samsung Electronics, Ltd. His research interest is in the area of system level simulation for evaluating cdma2000 1x, HSDPA, WiBro, and HMm.



Ki-Dong Lee received the B.S., M.S., and Ph.D. degrees in operation research (OR) and management science and the Ph.D. degree in industrial engineering (with applications to broadband wireless communications) from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1995, 1997, and 2001, respectively.

He has been with the Digital Broadcasting Research Division (formerly the Radio and Broadcasting Research Laboratory), Electronics and Telecommunications Research Institute (ETRI), Daejeon, as a Senior Member of Engineering Staff since his graduation from KAIST. At ETRI, he has been involved with research and development of interactive broad-band satellite access systems, where he develops various real-time algorithms for radio resource management and control.

Dr. Lee is a recipient of the IEEE ComSoc AP Outstanding Young Researcher Award. His research interests include stochastic modeling, performance analysis, scheduling algorithms, and optimization techniques for radio resource management in broad-band wireless and/or satellite networks.

Ho-Jin Lee received the B.S., M.S., and Ph.D. degrees in electronics engineering from Seoul National University (SNU), Korea, in 1981, 1983, and 1990.

He joined ETRI in 1983 and has been involved with TDX, a full electronic digital switching system development project, satellite ground mission/network control system development, KOMPSAT ground control system development, and satellite communication earth stations/service development. He has been with TRW, as a Visiting Engineer for 2 years. He works as the Director of the Satellite Communications Research Group of ETRI and is now the Project Manager of a mobile DVB-RCS system development project applicable to satellite mobile services.