

# Allocation and Scheduling Algorithms for IEEE 802.14 and MCNS in Hybrid Fiber Coaxial Networks

Ying-Dar Lin Chen-Yu Huang Wei-Ming Yin

**Abstract**— IEEE 802.14 and MCNS are two standards developed for the Hybrid Fiber Coaxial (HFC) CATV networks. Both standards model an upstream channel as a stream of minislots. But their philosophies on resolving collisions in the shared upstream channel are rather different, where IEEE 802.14 adopts the priority+FIFO first-transmission rule and the n-ary tree retransmission rule, and MCNS adopts the binary exponential backoff algorithm with adjustable window sizes. Both provide reservation access, while IEEE 802.14 and MCNS also support isochronous access and immediate access, respectively. In this paper, we try to prepare a suggestion list for vendors on how to allocate minislots for reservation access and immediate access and how to schedule the reserved bandwidth, which greatly affect the performance of a cable network and are left open by the standards.

**Keywords:** IEEE 802.14, MCNS, HFC, upstream, collision resolution, immediate access, bandwidth allocation

## I. INTRODUCTION

To facilitate interoperability between cable modems and headends designed by different vendors, standardization is required. Two major associations working on the hybrid fiber coax (HFC) networks are *IEEE 802.14 Working Group* [1], [2], [3] and *Multimedia Cable Network System Partners Ltd.* [4], [5], abbreviated as MCNS. The objective of these two standards is to provide data communication capabilities over the HFC network and to ensure the interoperability of conforming products built by different vendors. Physical and Medium Access Control (MAC) layers are specified in these standards. The MCNS standard has been approved as a standard by International Telecommunications Union (ITU) on March 19, 1998. IEEE 802.14 is still in its draft and expected to be finalized in 1998.

The HFC technology, followed by many cable companies, allows provision of upstream channels in a coaxial cable distribution network. Figure 1 represents an HFC system. A fiber node which can serve 500 to 2000 subscribers receives signals sent from the headend via a fiber. Then these signals are translated into cable signals by the fiber node and sent to amplified tree-and-branch feeder cables. Subscribers can retrieve or transmit signals by connecting their coaxial terminal units (CTU), i.e. set-top boxes or cable modems, to the taps on the network. With multiple access technologies, all subscribers within a branch can share the upstream bandwidth to send data back to the headend.

The HFC networks have the following important features. They certainly affect the protocol design.

- Point-to-multipoint and multipoint-to-point: It is a point-to-multipoint, tree-and-branch access network in the

downstream direction, but a multipoint-to-point, bus access network in the upstream direction.

- Inability to detect collisions by CTU: CTU can only listen to the downstream traffic, which is different from Ethernet where adaptors can tell by themselves whether collisions occur.

- Large propagation delay: According to [1], [5], the maximum round-trip-delay (RTD) is 0.4 ms in IEEE 802.14 and 0.8 ms in MCNS which is much longer than the RTD on Ethernet. Assuming the propagation speed of 0.005 ms/km, the IEEE 802.14 and MCNS networks may cover up to 40 and 80 km, respectively.

- Asymmetric upstream and downstream: The downstream bandwidth is much larger than upstream's.

- Non-uniform user distribution: Most of the subscribers are distributed over the last few miles of the network.

Given these two features, three important issues, including synchronization to compensate the large propagation delay, collision resolution to resolve collisions in the shared upstream channel, and upstream bandwidth management to improve throughput and access delay, in the MAC layer need to be resolved [6], [7]. In these two standards, the bandwidth allocation and transmission scheduling algorithms are left open to be designed by vendors instead of being explicitly specified. This is because allocation and scheduling do not affect interoperability. However, they are correlated with the performance of a cable network.

In this paper, the issues of how to properly allocate the upstream bandwidth and how to schedule requests from stations are studied. We discuss these issues after we present the network operation in section II. Section II also illustrates the similarities and differences between IEEE 802.14 and MCNS. We then simulate the systems by applying various bandwidth allocation strategies, collision resolution parameter settings, and transmission scheduling algorithms. Our simulation models are described in section III. Numerical results are presented in section IV and section V.

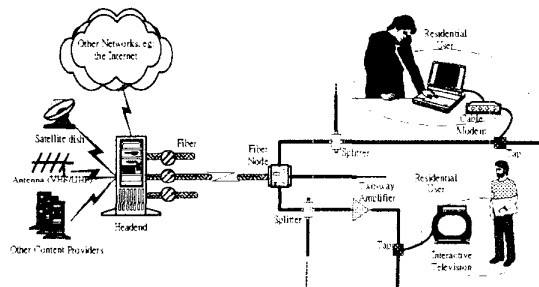


Fig. 1. An HFC network

The authors are with the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan.

Corresponding author: Ying-Dar Lin, email: ydlin@cis.nctu.edu.tw.

## II. IEEE 802.14 VERSUS MCNS

### A. Normal Operation

In IEEE 802.14 and MCNS, an upstream channel is modeled as a stream of minislots, the smallest transmission units in the upstream. The usage of upstream minislots is assigned by the headend, and basically there are two types of minislots—*request minislot* and *data minislot*. Data minislots are used to carry data. Request minislots are used to carry bandwidth requests made by stations for their *virtual queues* each of which is the elementary entity in the MAC protocol. From time to time, the headend broadcasts a *bandwidth allocation map*, which contains the usage assignments of upstream minislots. Stations learn the assignments from that map and work accordingly.

Figure 2 shows a simple state diagram for the stations of MCNS and IEEE 802.14. Initially, the station enters the

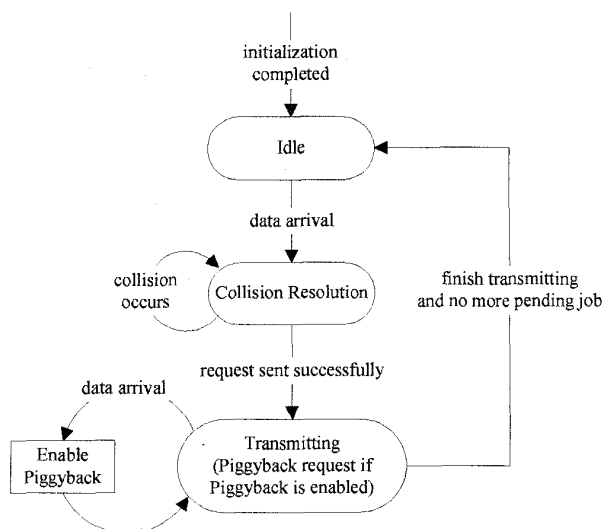


Fig. 2. A simple state diagram for stations of MCNS and IEEE 802.14

*Idle* state. When a data message arrives, the station enters the *Collision Resolution* state and repeatedly tries to send a bandwidth request to the headend until the request is successfully received and acknowledged by the headend. Then the station enters the *Transmitting* state. If extra data messages arrive during the station's transmission, the station may enable the Piggyback mechanism and piggyback the extra bandwidth request. Once the station finishes sending data messages and there are no more pending jobs, the station switches back to the *Idle* state.

### B. The Similarities and Differences

There are a few differences between IEEE 802.14 and MCNS. Key features of their physical layer specifications are summarized in table 1. Similarities and differences between their MAC layer specifications are illustrated in table 2 and table 3, respectively. More detailed comparisons can be found in [6]. We only discuss the immediate access

		IEEE 802.14	MCNS
Downstream	RF range	88–860 MHz	50/54–860 MHz
	Modulation	64 and 256 QAM	64 and 256 QAM
	Channel width	6 or 8 MHz	6 MHz
Upstream	RF range	5–42 MHz	5–30 or 5–42 MHz
	Modulation	QPSK and 16 QAM	QPSK and 16 QAM
	Symbol rate*	160 × M Kbaud M=1,2,4,8,16,32	160 × M Kbaud M=1,2,4,8,16

\*: The minimum channel spacing is  $(1+\alpha) \cdot R_s$  where  $\alpha$  is the spectral roll-off factor, and  $R_s$  is the symbol rate.

Table 1. Key features in the physical layer of IEEE 802.14 and MCNS

	IEEE 802.14	MCNS
Data-link-layer security	Encryption algorithm: DES algorithm.	
	Key exchange protocol: Main Key Exchange or Quick Key Exchange protocol.	Key exchange protocol: base on RSA and DES algorithms.
Upstream as a stream of minislots	Minislot size = time of transmitting 8 bytes of data and other overhead.	Minislot size = time of transmitting 16 bytes of data and other overhead.
Downstream MPEG-2 format	MPEG-2 PID = 0x1FFF	MPEG-2 PID = 0x01FFE
Virtual queue	Each SID of a station maps to a virtual queue.	Each (LID,LQ) of a station maps to a virtual queue.
Upstream bandwidth management	The minislot usage assignment is described in the Bandwidth Management Cell, composed of several kinds of Information Elements	The minislot usage assignment is described in the Allocation Map, composed of several kinds of Information Elements.
	The Information Element is used to describe the usage of some contiguous upstream minislots.	
Piggyback	Stations can piggyback extra bandwidth requests when they transmit data in data minislots.	

Table 2. Similarities between the MAC layers of IEEE 802.14 and MCNS

mode and collision resolution algorithms here because they are among the issues studied here.

### Immediate Access Mode in MCNS

In addition to the normal reservation and piggyback reservation modes supported in both standards, MCNS also provides immediate access mode. When the network traffic load is low, the headend may assign a region of minislots as immediate access minislots. This assignment, including 1) where the immediate access region starts, 2) how many minislots are in the region, 3) the maximum message length permitted to use the immediate access, and 4) the allowed starting points within the immediate access region, is encapsulated in the *bandwidth allocation map* multicast to stations. That is, any immediate access transmission must start at some specific minislots in the region and must fit within a limited number of minislots. Stations can transmit not only data but also their bandwidth requests in these immediate access minislots. Depending on the transmission result, a positive or negative feedback will be issued to the station. Even if the transmission is collided, the station can retransmit its request or data in the next region of immediate access mode, i.e. no further collision resolution protocol is enforced, or give up the immediate access mode and try the reservation access mode.

### Collision Resolution Algorithms

For the collision resolution algorithms, the latest draft of IEEE 802.14 combines four techniques, which are prioritized admission control mechanism, FIFO mechanism, n-ary tree algorithm, and the idea of multiple collision resolution engines, into its collision resolution algorithm. The

	IEEE 802.14	MCNS
Ranging	P-persistent collision resolution	Binary exponential backoff collision resolution
Access Modes	Normal reservation Piggyback reservation CBR access	Normal reservation Piggyback reservation Immediate access
Collision Resolution Algorithms	Latest draft [2]: First transmission rule : Priority + FIFO Retransmission rule : n-ary tree + p-persistent Multiple collision resolution engines Old draft [1]: P-persistent + ternary tree algorithm + multiple collision resolution engines	Binary exponential backoff + backoff window

Table 3. Differences between the MAC layers of IEEE 802.14 and MCNS

prioritized admission control mechanism is used to discriminate requests of different priorities. The FIFO mechanism is used to prevent excessive collisions. These two mechanisms are used as the first transmission rule, i.e. the rule that applies to the newly arriving requests. The n-ary tree algorithm is used to resolve the collided requests.

In the old version of IEEE 802.14 specification (draft 2), the first transmission rule and the algorithm used to resolve collided requests are p-persistent and ternary tree algorithms, respectively. Both versions allow multiple collision resolution engines to work in parallel to resolve different sets of contending stations.

MCNS adopts a binary exponential backoff algorithm to resolve collisions in the request minislot contention process. *Data Backoff Start* (DBS) and *Data Backoff End* (DBE) in the bandwidth allocation map are used to indicate the initial and maximum backoff window sizes, respectively. The station that wants to send its request does not contend for any request minislots until it has deferred  $T$  contention transmission opportunities, where  $T$  is an integer randomly selected between 0 and its backoff window size.

### C. Studied Issues

In these two specifications, the headend controls three factors that affect the performance of a cable network. They are 1) the usage of each upstream minislot, 2) the parameters of collision resolution procedures, and 3) the scheduling of received bandwidth requests. The upstream channel is a limited resource, therefore the more upstream minislots are assigned as request minislots, the less upstream minislots can be assigned as data minislots. With proper settings of the collision-resolution parameters, not only less upstream minislots would be wasted on resolving collided requests but also shorter request access delay, to be defined in section III, can be obtained.

In this paper, we examine the following issues in IEEE 802.14 (draft 2) and MCNS:

1. Minislot allocation in the collision resolution process: How much of the upstream bandwidth should be allocated as request minislots? How to set the parameters of collision resolution algorithms?
2. Minislot allocation for immediate access in MCNS: What is the proper time for the headend to conduct immediate access allocation? How much of the upstream band-

width should be allocated as immediate access? How to set the parameters of the immediate access?

3. The effect of the piggyback mechanism on the transmission scheduling algorithms: The probability that a request can be piggybacked is equal to the probability that the request arrives to a non-empty queue of a station. Thus, the scheduling algorithm needs to reduce the data transfer delay but may also want to reduce the request access delay, where the latter might be achieved by increasing the chance of piggyback.

## III. MODELS FOR SIMULATION STUDIES

The common simulation parameters are listed in table 4. In the simulation, a MiniPDUTime, i.e. the duration of one minislot, is the elementary time unit, and all simulation results are measured in the unit of MiniPDUTime.

Parameter	Description	Default value
N	The number of stations	20
MiniPDUTime	The time length of a minislot	8 micro second
RTD	Round-trip delay	400 micro second
Traffic model	Poisson process	Mean packet size=10 minislots Mean inter-arrival time = (N*MeanPacketSize)/Load

Table 4. Common parameters in the simulation study

### A. Model for Collision Resolution and Request Scheduling

We measure the *throughput*, *request access delay* (RAD), and *data transfer delay* (DTD) of the simulated systems which are applied various minislot allocation strategies and transmission scheduling algorithms. The throughput of the system is defined as what percentage of the upstream bandwidth is used for successful data transmission. The RAD of a request is defined as how much time a station takes to successfully transmit the request. Once a request has been successfully transmitted to the headend, it becomes a granted request. The DTD of a granted request is defined as how much extra time the headend takes to serve the request. RAD is used to measure the efficiency of collision resolution algorithms, while DTD is used to measure the efficiency of transmission scheduling algorithms. For example, in figure 3, a data message arrives at a station at  $T_1$ . After one or several request transmissions by the station, the headend receives the request and puts it into the scheduling queue at  $T_2$  and then sends back a feedback message which is received by the station at  $T_3$ . The request is then scheduled and the station is notified about when and how much it can transmit. Finally, the station finishes the transmission for the request at  $T_4$ . Therefore, the RAD is  $T_3 - T_1$ , and the DTD is  $T_4 - T_2$ .

### B. Model for Immediate Access

In this model, an upstream channel is modeled as cyclically repeated and fixed-length frames, as shown in figure 4. The first  $P\%$  of each frame is reserved for immediate access. Any station with a pending bandwidth request may 1) try immediate access if the data message length is small

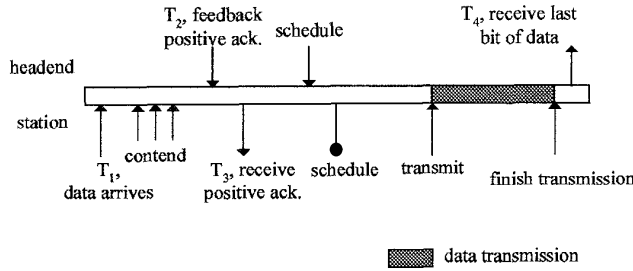


Fig. 3. A sequence of events for request, schedule and transmission

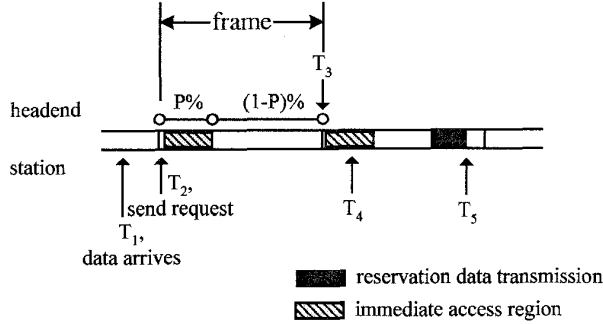


Fig. 4. Model for immediate access

enough, and the number of retransmissions does not exceed the upper limit or 2) conduct reservation access by sending its request at the beginning of each frame, as shown in figure 5. For the headend, it adopts first come first serve

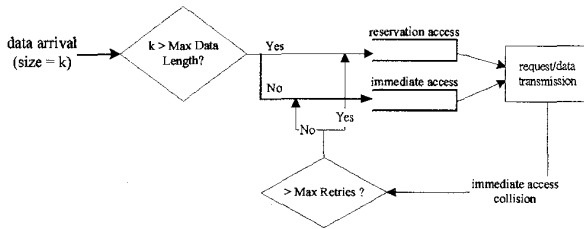


Fig. 5. Operations of a station

(FCFS) algorithm to schedule the requests at the beginning of each frame. In figure 4, a data message (size= $K$ ) arrives at a station at  $T_1$ . If the station conducts immediate access and successfully transmits the data message at  $T_4$ , the *access delay* (AD) is calculated as  $T_4 - T_1$ . If reservation access is used, and the station finishes transmission at  $T_5$ , the AD is calculated as  $T_5 - T_1 - K$ . The *throughput* of the system is defined as what percentage of the upstream bandwidth is used for successful data transmission. AD and throughput are used to measure the performance of the allocation strategies for immediate access.

#### IV. MINISLOT ALLOCATION IN IEEE 802.14 AND MCNS

##### A. Collision Resolution in IEEE 802.14

The collision resolution algorithm adopted in IEEE 802.14 (draft 2) consists of the p-persistent algorithm which

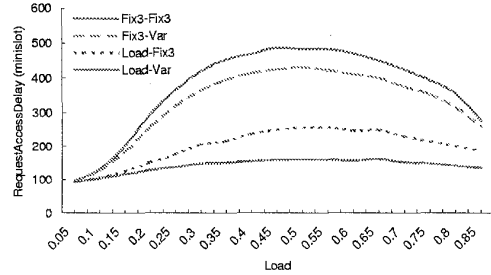


Fig. 6. RAD of different allocation strategies in IEEE 802.14

is the first transmission rule, and the n-ary tree algorithm which is the retransmission rule. Thus, the minislot allocation in the collision resolution process of IEEE 802.14 (draft 2) can be further divided into two parts:

1. The initial allocation for the newly arriving requests, and
2. The collision-resolution allocation for the collided requests.

The applied allocation schemes in the simulation are defined in table 5. Besides, we assume that the headend has the talent to guess the proper  $p$  value in the p-persistent first transmission rule, i.e.  $p = \min(\text{potential contention entry range} / \text{number of allocated minislots}, 1)$ .

In figure 6, four combinations of initial and collision-resolution minislot allocations are simulated. For example, the line labeled as “Fix3-Var” represents the schemes of initial and collision-resolution allocations are Fix3 and Var, respectively. We can see that the RAD is high when the load is medium. This is because when the traffic load is high, most stations can piggyback their bandwidth requests instead of contending request minislots. Thus, the headend has to allocate more request minislots when the traffic load is medium. The Load-Var has better performance than the other allocation strategies. Also, comparing the improvement amounts of Fix3-Var over Fix3-Fix3 and Load-Fix3 over Fix3-Fix3, we can find that the initial allocation has a greater impact on the RAD than the collision-resolution allocation.

Allocation schemes	How to allocate?
FixM	The headend allocates M request minislots in each minislot cluster.
Var*	The headend allocates $3 \times C$ request minislots, where C is the number of collisions in last allocated minislot cluster.
Load**	The headend allocates N request minislots, where N depends on the current traffic load. The value of N is calculated as $3+38 \times \text{Load} - 56 \times \text{Load}^2 + 21 \times \text{Load}^3$ whose value at medium load is larger than the value at light/heavy load.

\*: This only applies to the collision-resolution allocation of IEEE 802.14.

\*\* : This only applies to the initial allocation of IEEE 802.14.

\*\*\*: This polynomial expression is the approximation of the curve “Fix3-Fix3” in figure 6 and is obtained through mathematical package Origin.

Table 5. Minislot allocation strategies in collision resolution of IEEE 802.14 (draft 2)

##### B. Collision Resolution in MCNS

MCNS runs binary exponential backoff to resolve collisions in request minislots. However, it allows flexibility in

specifying the backoff start and end window sizes. Here we investigate how to allocate request minislots and how to adjust the backoff window under Ethernet-like and random-select types of collision resolution, as defined in table 6.

#### Request minislot allocation strategies

For random-select type of collision resolution, the simulation results of throughput and RAD are shown in figure 7 and figure 8, respectively. When the load is high, most requests are piggybacked through the scheduled data transmission instead of contending request minislots. Hence, some request minislots are wasted and the throughput is not proportionally increased. The performance of Fix8, in terms of RAD, is the best among these allocation schemes. For random-select type of collision resolution with window size fixed at 8, we can just allocate 8 minislots in every cluster. Note that when the load is above 0.75, the RAD of Fix12 is not accurate because almost all requests are sent to the headend through piggyback instead of contending request minislots.

Name	Description
Ethernet-like	Backoff start window size is $S$ and backoff end window size is $E$ . ( $S \neq E$ )
Random-select	Both backoff start and end window sizes are equal to some $W$ .

Table 6. Two types of backoff window setting in collision resolution of MCNS

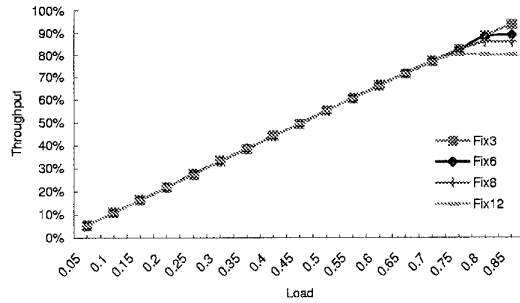


Fig. 7. Throughput for different numbers of allocated request minislots in MCNS (adopting random-select type of collision resolution)

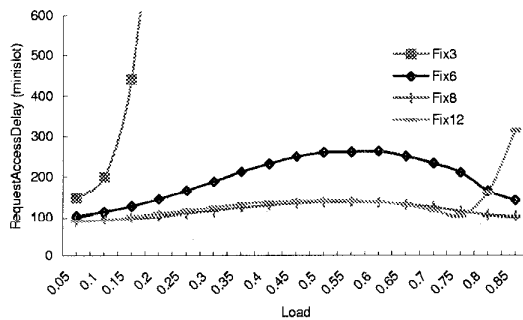


Fig. 8. RAD of different numbers of allocated request minislots in MCNS (adopting random-select type of collision resolution)

For the Ethernet-like type of collision resolution with backoff start,  $S$ , and backoff end,  $E$ , the simulation results of throughput and RAD under 6 minislot allocation strategies, defined in table 7, are shown in figure 9 and figure 10, respectively. The SE scheme which resolves collisions efficiently performs the best among these allocation strategies. On the other hand, the E and MeanSE schemes that allocate too many request minislots lead to lower throughput. Because there is no idea about how many stations collided within one minislot, allocating as many request minislots as possible, i.e. the SE scheme, is an efficient strategy to quickly resolve collisions. Therefore, the SE scheme performs better than Dbl and Exp schemes.

Allocation schemes	How to allocate?
S	The headend allocates $S$ request minislots in each minislot cluster.
E	The headend allocates $E$ request minislots in each minislot cluster.
MeanSE	The headend allocates $(S+E)/2$ request minislots in each minislot cluster.
Dbl	Normally, the headend allocates $S$ request minislots in each minislot cluster. If there were $C$ collisions in last minislot cluster, the headend allocates $\max(2^C, E)$ minislots in this minislot cluster.
Exp	Normally, the headend allocates $S$ request minislots in each minislot cluster. If there were $C$ collisions in last minislot cluster, the headend allocates $\max(2^C, E)$ minislots in this minislot cluster.
SE	Normally, the headend allocates $S$ request minislots at each minislot cluster. If there were collisions in last minislot cluster, the headend allocates $E$ minislots in this minislot cluster.

Table 7. Minislot allocation strategies in collision resolution of MCNS

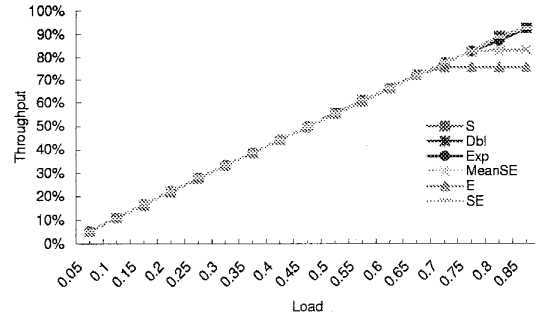


Fig. 9. Throughput of various minislot allocation strategies in MCNS (adopting Ethernet-like type of collision resolution)

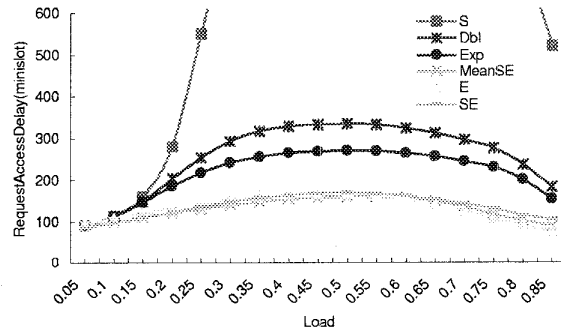


Fig. 10. RAD of various minislot allocation strategies in MCNS (adopting Ethernet-like type of collision resolution)

### Backoff window settings

In this simulation, the headend adopts the Fix6 initial minislot allocation strategy and applies random-select and Ethernet-like window settings. From figure 11, we find that the window size should be enlarged when the load is medium and shrunk when the load is light or very heavy under the random-select type of collision resolution. Note that over-sized window, such as the W32, only results in lower utilization of request minislots; consequently, the request opportunities are deferred and the RAD is therefore larger.

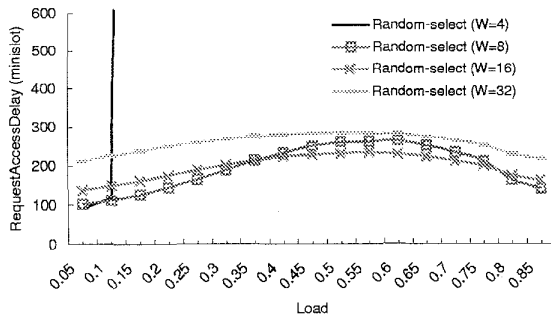


Fig. 11. RAD of different window sizes for the random-select type of collision resolution in MCNS (adopting Fix6 allocation)

For the Ethernet-like type of collision resolution, the headend should not assign a backoff start window size that is too small to avoid the collision at first transmissions and a backoff end window size that is too large to avoid deferring the transmission unnecessarily. As shown in figure 12, the window size range of 4 to 32 is the best.

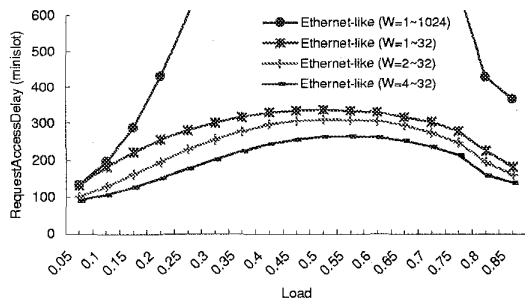


Fig. 12. The RAD of different Ethernet-like type of collision resolution in MCNS (adopting Fix6 allocation)

Comparing these two backoff window setting strategies, from figure 13, the fine-tuned random-select setting, which sets a smaller window at light and heavy load, and a large window at medium load, performs better than the Ethernet-like setting. Basically speaking, the random-select setting tries to avoid collisions at the first transmission, but the Ethernet-like setting aims at efficiently resolving collisions.

### C. Immediate Access in MCNS

MCNS offers immediate access mode to access upstream bandwidth. The headend can determine the immediate

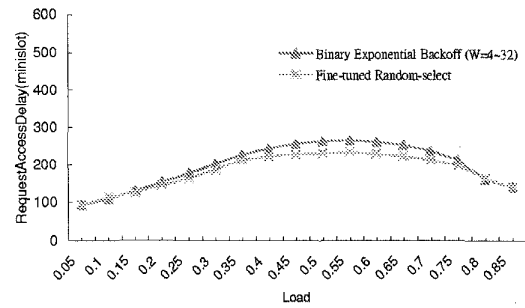


Fig. 13. RAD of fine-tuned random-select and Ethernet-like in MCNS

access region and the maximum message length permitted to use this access region. We investigate the performance of cable networks under the following three configurations:

1. Different sizes of immediate access regions
2. Different maximum message lengths
3. Different numbers of retries for immediate access

### Size of immediate access regions

From figure 14, we can find that the headend should turn off the immediate access mode when the load is heavy. In other words, the bandwidth should be allocated to the reservation-based transmission when the load is heavy. From figure 15, we can find that the size of immediate access regions should be dynamically adjusted. Designating too much bandwidth for immediate access is not efficient because there might not be so many messages qualified to use the immediate access regions. On the other hand, under-sized immediate access regions lead to severe collisions and, in turn, wasted the bandwidth. Observed from the simulation results, the headend should designate 28% and 0% of upstream bandwidth for immediate access when the load is under 0.55 and above 0.55, respectively.

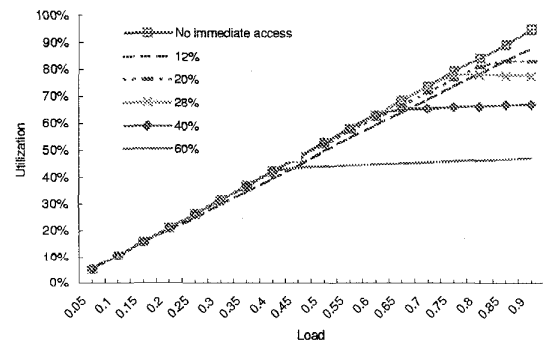


Fig. 14. Throughput for different sizes of immediate access regions in MCNS

### Maximum message length

From figure 16, we can find that setting a smaller value of maximum message length, when the load is heavy, has an effect on preventing messages from contending immediate access regions. Hence, severe collisions can be avoided.

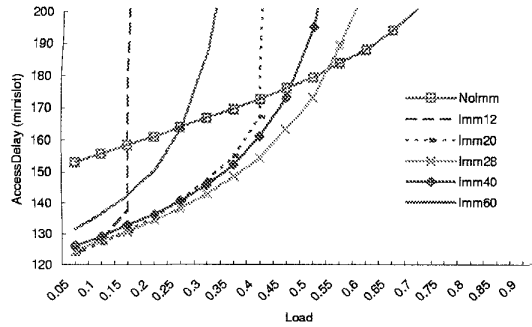


Fig. 15. AD for different sizes of immediate access regions in MCNS

On the other hand, when the load is light, setting a larger value of maximum message length leads to higher utilization of immediate access bandwidth and the AD is therefore lower. In this simulation, the headend should assign the maximum message length as 5, 4, 2 when the load is under 0.15, between 0.15 and 0.35, between 0.35 and 0.6, respectively. For the load over 0.6 or 0.7, immediate access regions should be closed.

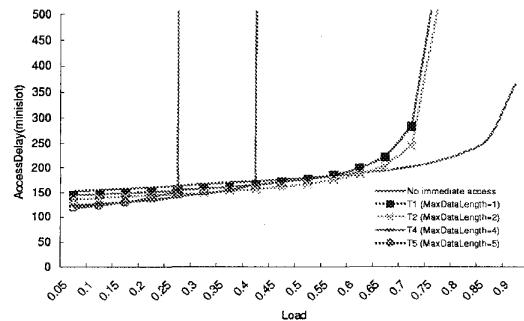


Fig. 16. AD for different values of maximum message length

#### Number of retries

Investigating our previous simulation results, we find that with the explosion of the AD for immediate access at heavy traffic load, the overall AD gets extremely high. We expect that setting the retry limit on the immediate access might alleviate the problem. From figure 17, it is observed that the difference of AD between various retry limits is not notably when the load is low. However, when the load starts to increase, setting a smaller retry limit can prevent AD from increasing dramatically and the system should eventually turn off the immediate access mode when the load is high enough ( $>0.7$ ). No retry for immediate access seems to be a good policy. Note that, from figure 18, the throughput is mainly constrained by the size of immediate access regions instead of retry limit.

#### D. Comparisons

We now examine the RAD and throughput of IEEE 802.14 and MCNS under their fine-tuned parameter settings and the same minislot allocation strategy—Fix6. Figure 19 shows that they have the same throughput. As

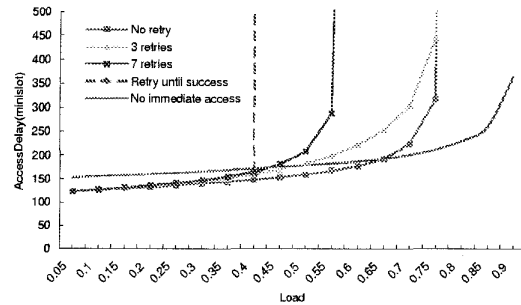


Fig. 17. AD for different limits on the number of retries

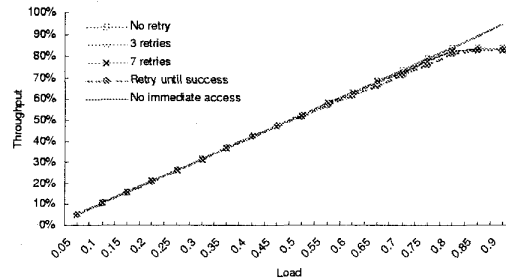


Fig. 18. Throughput for different limits on the number of retries

what we have discussed earlier, there are more potential contenders for request minislots when the load is medium. We can find that when the load is medium, IEEE 802.14's RAD is lower than MCNS's, as shown in figure 20. That is because the first transmission rule can avoid too many requests contending the same minislot cluster. The difference of the RAD between IEEE 802.14 and MCNS is small when the load is light or heavy.

#### V. SCHEDULING IN IEEE 802.14 AND MCNS

In IEEE 802.14 and MCNS cable networks, the headend not only allocates request minislots but also schedules upstream data transmissions. Here we investigate how different scheduling disciplines, as defined in table 8, affects the performance of data transmission. From figure 21 and figure 22, we find that the system adopting shortest job first (SJF) has the lowest DTD, but its RAD is the worst. And vice versa for longest job first (LJF). This phenomenon is due to that when the system has small DTD, the queue at a station has a higher probability of being empty, hence, the probability that a station can piggyback its request is also small. Hence, we introduce a modified SJF discipline which distributes scheduled data minislots to increase the probability for a station to piggyback its request. For example, the headend may grant 30 minislots for some station. These granted minislots might be distributed over minislots numbered from 20 to 29, 40 to 49, and 60 to 69, instead of from 20 to 49 contiguously. As what we expected, the RAD does get reduced, while the DTD gets increased. The modified SJF seems to be a good policy to balance RAD and DTD.

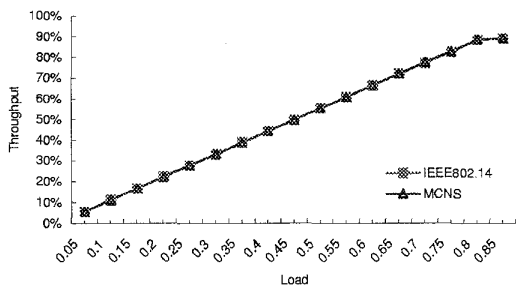


Fig. 19. Throughput of IEEE 802.14 and MCNS. (IEEE 802.14 adopts Fix6-Var. MCNS adopts Fix6.)

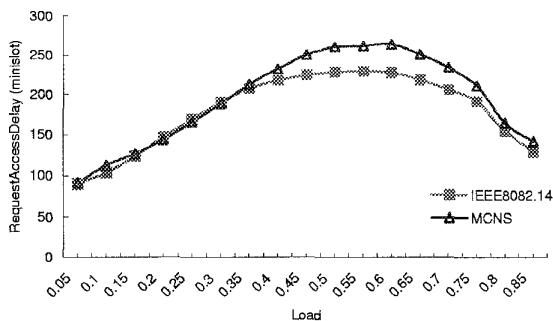


Fig. 20. RAD of IEEE 802.14 and MCNS. (IEEE 802.14 adopts Fix6-Var. MCNS adopts Fix6.)

VI. CONCLUSION

The effects induced by applying various strategies for minislot allocation, collision resolution parameter setting, and transmission scheduling in IEEE 802.14 and MCNS cable networks are studied in this paper. From the simulation results, we conclude that many allocation algorithms should be load-dependent and the scheduling algorithms are influenced by the piggyback mechanism where there is a tradeoff between request access delay and data transfer delay. We now summarize the major results and suggestions for the vendors of cable headends and modems.

1. Avoid collision in the first place : It is observed that the initial allocation is more important than the collision-resolution allocation for both IEEE 802.14 and MCNS. The initial allocation, however, should depend on traffic load. Due to the piggyback effect, the demand for request min-

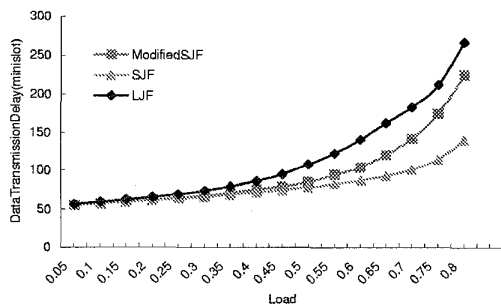


Fig. 21. DTD for different scheduling algorithms

Name	Scheduling procedure description
SJF	The headend always schedules the station with the minimum bandwidth request.
LJF	The headend always schedules the station with the maximum bandwidth request.
MSJF	The headend schedules the station with the minimum bandwidth request but distributes the scheduled data minislots into several bursts.

Table 8. Scheduling algorithms upstream data transmission

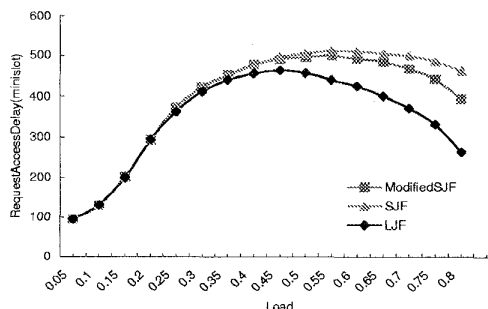


Fig. 22. RAD for different scheduling algorithms

islots under medium load is higher than under light and heavy load. The first transmission rule can be combined with the initial allocation to effectively reduce the chance of collisions. For IEEE 802.14 and MCNS, the first transmission rules are priority FIFO and backoff window, respectively. The backoff window in MCNS can be fixed or variable in a binary range.

2. Resolve collisions quickly : Contrast to the initial allocation, the collision-resolution allocation should not depend on traffic load. Instead, it should depend on the number of observed collisions. Variable collision-resolution allocation combined with load-dependent initial allocation achieves the best result. Interleaved multiple collision resolution engines can be executed under medium traffic load, i.e. high demand for request minislots.

3. Allocate the right size of MCNS immediate access regions at the right time : The percentage of immediate access regions within an upstream channel should also depend on traffic load. It should be lowered as load increases and closed under heavy load. The maximum size of messages permitted to use immediate access regions can be lowered too as load increases. It has the effect to decrease the chance of collisions in immediate access regions. Finally, no retry is suggested for messages collided in the first try in immediate access regions. They can redirect themselves to use the reservation access.

4. Distribute the schedule for a single request into several pieces : Contrast to collision resolution, it is not necessary to schedule data transmission for a request as quickly as possible. For a station, the sooner the data transmission is finished, the less likely the next request can be piggybacked. Thus there is a tradeoff between data transfer delay and request access delay. Fortunately, distributing the schedule for a single request into several pieces such that the requested data transmission is broken into several bursts can increase the chance for a new request to be piggybacked.

REFERENCES

[1] IEEE, *IEEE Project 802.14/a Draft 2 Revision 2*, July 1997.



- [2] IEEE, *IEEE Project 802.14/a Draft 3 Revision 1*, Apr. 1998.
- [3] "IEEE 802.14 cable TV working group," <http://www.walkingdog.com/>.
- [4] "Data over cable service interface specifications," <http://www.cablemodem.com/>.
- [5] Cable Television Laboratories, Inc., *Data-Over-Cable Service Interface Specifications—Radio Frequency Interface Specification*, Oct. 1997.
- [6] Chen-Yu Huang and Ying-Dar Lin, "Comparing IEEE 802.14 and MCNS standards for hybrid fiber coaxial networks," submitted for publication.
- [7] Ying-Dar Lin, "On IEEE 802.14 medium access control protocol," to appear in *IEEE Communications Surveys*, Sep. 1998.
- [8] John O. Limb and Dolores Sala, "A protocol for efficient transfer of data over hybrid fiber/coax systems," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 872-881, Dec. 1997.
- [9] Ying-Dar Lin, Chia-Jen Wu, and Wei-Ming Yin, "PCUP: Pipelined cyclic upstream protocol over hybrid fiber coax," *IEEE Network Magazine*, vol. 11, no. 1, pp. 24-34, January/February 1997.



**Ying-Dar Lin** received the Bachelor's degree in Computer Science and Information Engineering from National Taiwan University in 1988, and the M.S. and Ph.D. degrees in Computer Science from the University of California, Los Angeles in 1990 and 1993, respectively. He joined the faculty of the Department of Computer and Information Science at National Chiao Tung University in August 1993 and is now Associate Professor. His research interests include traffic characterization, service and

network management, distributed multimedia applications, and high-speed LANs/MANs/WANs. Dr. Lin is a member of ACM and IEEE.



**Chen-Yu Huang** received B.A. and M.S. in Computer and Information Science from National Chiao Tung University in 1996 and 1998, respectively. His research interests include high-speed LANs and interactive CATV protocols design. He is now on a two-year duty in the army of Taiwan.



**Wei-Ming Yin** received B.A. and M.S. in Computer and Information Science from National Chiao Tung University in 1995 and 1997, respectively. He is now a Ph.D. student in the same department. His current research interest is protocol design and analysis of residential networks.