Performance Analysis of Signaling Delay for Wireless Cellular Networks

Masashi Sugano[1], Doo Seop Eom[2], Masayuki Murata[3] and Hideo Miyahara[3]

[1]Faculty of Comprehensive Rehabilitation, Osaka Prefecture College of Nursing, Osaka, Japan
[2]Department of Electrical Engineering, Korea University, Seoul, Korea
[3]Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

In wireless cellular networks, signaling traffic such as location update, paging and handoff due to the user's mobility takes a considerable portion of the total traffic load. In addition, the maximum allowable delays may be different among the signaling packets. In this paper, we present an analytical model for evaluating a total processing delay of signaling packets of wireless cellular networks, which integrates the delays of the radio channel and the processing delay at the wired portion. Through numerical examples, we show that priority processing is effective for reducing the handoff processing delays. We also evaluate the delay difference among cells according to their position within the location area, and the influence of number of terminals upon the processing delays.

Keywords: wireless cellular network, signaling packet, location update, handoff, paging.

1. Introduction

In wireless cellular networks, signaling traffic such as location update, paging and handoff due
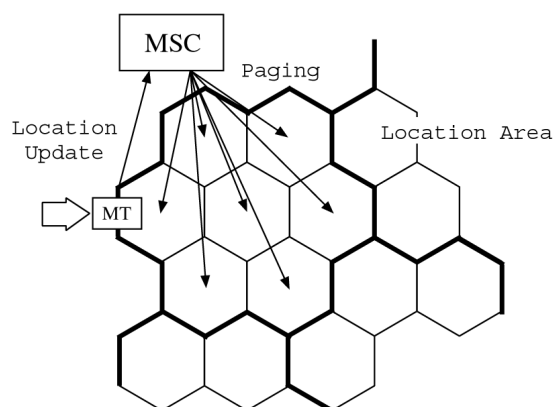

Fig. 1. Location management of mobile terminals.

to the user's mobility takes a considerable portion of the total traffic load. Such signaling traffic load is significantly increasing because of increase of mobile terminals and decrease of radio cell sizes [1]. To reduce the amount of location update signaling information, a notion of the location area has been introduced [2, 3]; the region covered by a wireless cellular network is first divided into location areas. Each of the location areas consists of several cells (Figure 1), and it is a basic unit for location management of the mobile terminals so that the larger location area could decrease the traffic load of location updates. However, uncertainty of the location of the mobile terminal increases as the size of the location area increases. The traffic load of paging will increase as a result. The load at the mobile switching center managing each location area is also increased. Consequently, the size of location area has a strong influence on signaling packet delays [4].

Past research was dedicated mainly to the signaling traffic of the wired portion of the cellular network. For example, the authors in [5] evaluated the signaling load of Signaling System 7 (SS7) in both the Pan–European GSM system and the North American IS–41. In [6, 7], the authors have proposed new paging schemes to reduce paging costs under delay bounds. Also, there was some research on evaluating the effect of signaling traffic in wireless ATM networks [8, 9, 10, 11, 12], but they only focused on the wired portion of ATM network, and did not consider the delays on wireless channels.

On the other hand, Rubin et al. have studied the signaling delay experienced on the radio channels in wireless cellular networks [13]. They analyzed the delay of the forward signaling channel which is a radio channel directed from a base station to mobile terminals. They also discussed the relation between the signaling delay and the size of location area [2]. In those works, however, they only evaluated the delay on radio channels, and did not consider the processing delay in the wired portion of the cellular networks, in spite of the fact that the location area size affects not only the traffic load of the radio channel, but also, the processing delay at the wired part of the network. Therefore, it becomes essential to consider the delays of the wireless channel and those in the wired network. Only such total modeling of the cellular network makes it possible to evaluate the processing delay of the wireless cellular network such as call setup delays.

In addition, the maximum allowable delays may be different among the signaling packets. For example, when a mobile terminal with communicating in progress crosses the boundary of a cell in order to avoid information loss, its handoff processing delay must be kept as small as possible. The processing of location updates, on the contrary, would not require such a stringent delay. The priority scheduling seems to be attractive to satisfy different kinds of QoS requirements. Actually, several priority assignment schemes on radio channels were presented in [2], but the effect of priority processing in the wired portion of the cellular network should also be clarified.

In this paper, we present an analytical model for evaluating the total processing delay of signaling packets of wireless cellular networks. By using this model, we can design a system satisfying the delay requirement of each type of processing. Through the analytical method, we derive the total processing delays for connection setup, location update and handoff. In addition, we introduce priority processing to reduce the handoff processing delay, and evaluate its effectiveness through numerical examples. The signaling delay should also be influenced by the position of the cells within the location area, because the arrival rate of location update packets is different. Therefore, we will compare the delays experienced in each cell according to the

position of its location area. We will also investigate the influence of number of terminals upon processing delays.

This paper is organized as follows. In the next section, we explain the model of the wireless cellular network and the processing delay under consideration. In Section 3 we present the analytical method to derive each processing delay. In Section 4, we discuss the performance improvement of the priority processing for handoff delay, the delay differences among cells in a location area, and the influence of number of terminals. Finally, we conclude our paper in Section 5.

2. Model Description

2.1. Model of Wireless Cellular Network

As illustrated in Figure 2, we assume that a wireless cellular network has the following components;

- Mobile terminal (MT)
- Base station (BS)
- Mobile switching center (MSC)

The region covered by the wireless cellular network is divided into location areas. Each location area consists of some adjacent cells, and has its dedicated MSC. Under this configuration, the following control should be performed;
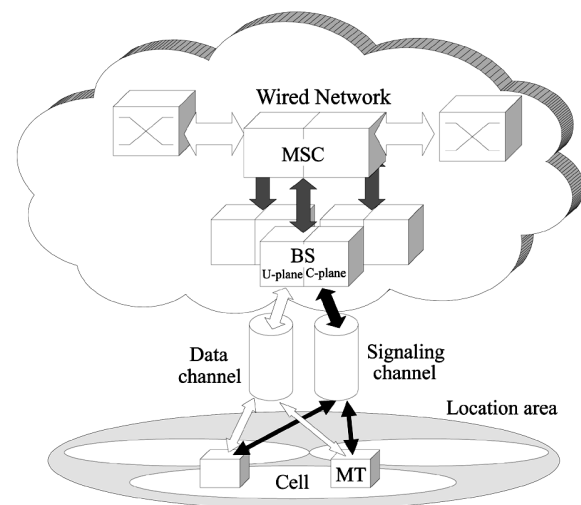

Fig. 2. Model of the wireless cellular network including signaling channels.

- When a MT crosses the boundary of the location area, it generates a location update packet, by which the involved MSCs are informed of the location change of the MT.

- When a MSC receives a call request for the MT within the location area, paging packets are broadcasted by BSs in the location area to alert the MT to the incoming call. Upon receipt of the response from the destined MT, the call setup becomes possible.

The wireless channel is divided into the data channel for user data and the signaling channel for the signaling packets. In the current study, we chose TDMA/TDD (Time Division Multiple Access/Time Division Duplex) [14] as a multiple access method, which implements an underlying physical structure. TDMA/TDD has been adopted by several wireless ATM prototype implementations, and we believe that the method is most suitable to realize the mechanism of QoS guarantees. As shown in Figure 3, TDMA/TDD protocol divides a fixed length frame into uplink (from MTs to BS) and downlink dynamically, and it is applicable to the environment where the demand of both directions changes, dependent on time. Both downlink and uplink consist of control slots and user data slots for each direction.
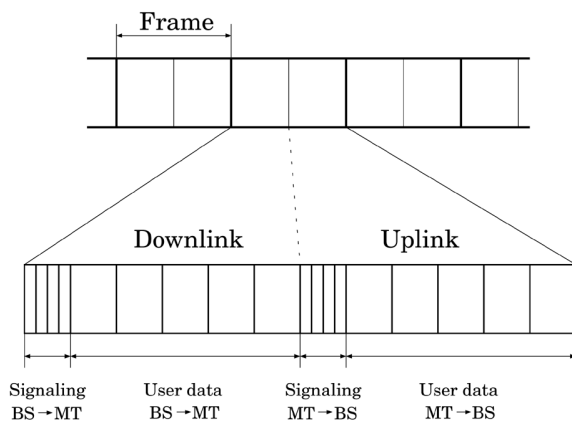

Fig. 3. Radio frame structure of TDMA/TDD.

When the MT has a new signaling packet to transmit, it first sends the packet on the control slot, according to a slotted ALOHA protocol. If the BS receives the signaling packet correctly, it returns the acknowledgment during the next downlink. On the other hand, if packet collision occurs, the BS sends no response, so that the MT will retry the transmission at the next frame. We will consider the delays caused by retransmissions to derive from the transmission delay on radio channel, as shown in Section 3.

2.2. Signaling Packet Flow and Processing Delays

The delay experienced by the signaling packet consists of the transmission delay on both uplink and downlink, the processing delays at BS and MSC, and the processing delay at the wired network. We evaluate the following signaling packet delays;

- Mobile originating call setup delay: T_{MW}
  Let T_{MW} represent the time duration from generation of CR (call request) at MT to arrival of CA (call acknowledgment) at MT as shown in Figure 4(a). For calls destined for the wired networks, the flow of the CR packet is; Uplink $\rightarrow$ BS $\rightarrow$ MSC. After being processed at wired network, CA is returned via MSC $\rightarrow$ BS $\rightarrow$ Downlink.

- Mobile terminating call setup delay: T_{WM}
  T_{WM} represents the time duration from the arrival of call request at MSC to the time when its response is sent to the wired network. For calls destined for some MT, PG (paging) packets are broadcasted from the MSC to BSs in the MSC's location area, and sent to MTs via downlink. Then, the corresponding MT generates CR which will be returned to the wired network (Figure 4(b)).

- Location update processing delay: T_{LU}
  The flow of LU (location update) requests is Uplink $\rightarrow$ BS $\rightarrow$ MSC, and corresponding LA (location update acknowledgment) is returned as BS $\rightarrow$ downlink (Figure 4(c)).

- Handoff processing delay: T_{HO}
  HO (handoff) requests are sent via Uplink $\rightarrow$ BS $\rightarrow$ MSC. After the handoff request is processed at the wired network, the corresponding HA (handoff acknowledgment) is returned as MSC $\rightarrow$ BS $\rightarrow$ downlink (Figure 4(d)).
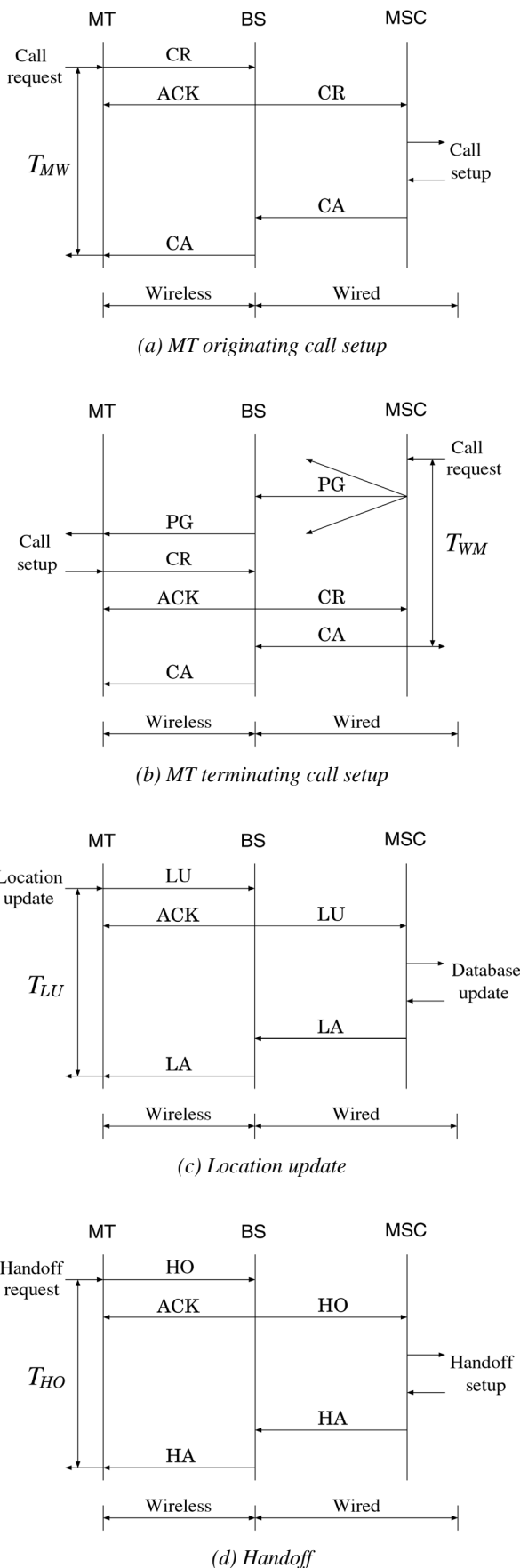
(a) MT originating call setup

(b) MT terminating call setup

(c) Location update

(d) Handoff

Fig. 4. Signaling packet flow and processing delay.

3. Analysis

3.1. Queuing Network Model

To analytically derive processing delays, we build a queuing network model to evaluate the delays of signaling packets in the wireless cellular networks (see Figure 5). In what follows, we define each queue in turn.
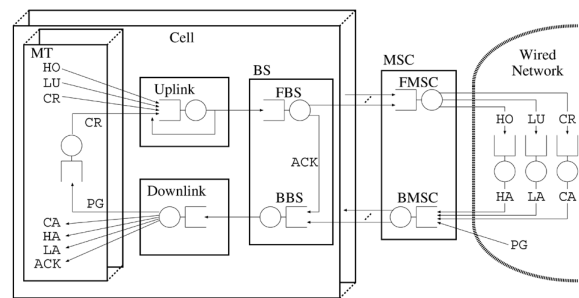

Fig. 5. Queuing model of the signaling traffic of the wireless cellular networks.

1. Uplink queue of the wireless channel
   The uplink is contended by CR, LU, and HO packets. CR packets to respond to PG packets should also be considered. The delays of those packets are determined by the approximate analysis which takes into account collision of the packets due to the multiple access protocol of the TDMA/TDD system.

2. FBS (forward base station) queue
   When BS processes the packet, it returns acknowledgment (ACK) to the sending MT via the corresponding BS. At the same time, BS forwards the packet to MSC. The packets are processed in an FIFO manner.

3. FMSC (forward mobile switching center) queue
   MSC handles all incoming packets from BSs within the location area. Those packets are processed in an FIFO manner and forwarded to the wired network,

4. CR, LU and HO queues
   CR, LU and HO queues represent the delays of the corresponding packets experienced at the wired networks. HO, LU and CR packets are assumed to be processed separately. The exact representation cannot be known, and we use IS (Infinite Server) queues for the scheduling discipline and the service demands are treated as given parameters.

5. BMSC (backward mobile switching center) queue
CA, LA and HA signaling packets are returned to BSs by the MSC. PG packets coming from the wired network are also forwarded to the appropriate MT via BS. Those are processed in an FIFO manner.

6. BBS (backward base station) queue
Signaling packets coming from MSC are processed in an FIFO manner at the BS.

7. Downlink queue of the wireless channel
The packet delay at the downlink queue can be modeled by FIFO queue since the BS station can actively schedule the packet transmission on the wireless channel.

To formulate the problem, we introduce some notations and assumptions as follows;

1. Each location area consists of N_c cells.

2. MTs are uniformly spread within the network, so that each cell has N_{MT} mobile terminals.

3. Arrivals of MT originating and terminating call requests follow Poisson processes with rate λ_{MW} and λ_{WM}, respectively.

4. Call duration has an exponential distribution with mean $1/\mu$. Furthermore, the duration of MT stay in the wireless cell before moving into other cells follows an exponential distribution with mean $1/h$. Generally, these approximation is introduced by the performance evaluation of cellular networks (see e.g., [15, 16, 17]). By introducing these approximation, due to the PASTA (Poisson arrivals see time-averages) property, the handoff call and the new call can be treated identically.

5. The processing time at MT, BS, MSC, and wired portion follow exponential distributions with means.

   - $1/\mu_{MT}$: Mean processing time at MT.

   - $1/\mu_{BS}$: Mean processing time at BS.

   - $1/\mu_{MSC}$: Mean processing time at MSC.

   - $1/\mu_{CR}$: Mean call setup time.

   - $1/\mu_{HO}$: Mean handoff processing time.

   - $1/\mu_{LU}$: Mean location update processing time.

6. The MT engaged in the active call does not generate CR packets. Therefore, when we consider the arrivals of CR packets at MSC, such a possibility should be considered.

7. The transmission delay at wired network is assumed to be negligible.

Furthermore, we use the following notations to describe the delay at each component. In these notations, the symbol '$*$' represents the types of signaling packets (e.g., CR). In addition, we introduce the index k ($k = 1, \ldots, N_c$) to distinguish the delays at each cell, since the arrival rates of location update packets are different at each cell, according to its position in the location area.

- $D_{Ru}^k(*)$: Transmission delay on uplink.

- $D_{Rd}^k(*)$: Transmission delay on downlink.

- $D_{Bu}^k(*)$: Processing delay at FBS.

- $D_{Bd}^k(*)$: Processing delay at BBS.

- $D_{Mu}(*)$: Processing delay at FMSC.

- $D_{Md}(*)$: Processing delay at BMSC.

- $D_{Mp}(*)$: Processing delay for call setup, location update, and handoff at wired network.

- $D_{MT}(*)$: Processing delay of call setup at a MT.

If we assume the exponential distributions of processing delay at all queues, we can have a product–form network solution, for which we have well–known computational methods such as the convolution method and the MVA algorithm [18]. However, we have the Uplink queue not satisfying the condition of the product–form solution. Nevertheless, we adopt the widely used approximation approach that the uplink queue is modeled as an IS (infinite server) queue, where work demands of the signaling packets are determined by separately performing delay analysis of the multiple access system. Then, our model can be numerically analyzed. To consider priority scheduling for several signaling packets, we need to introduce another approximate analysis [19]. As shown in the numerical examples, queues for possibly introducing priority scheduling are FBS, FMSC, BMSC and BBS.

3.2. Delay Analysis at Each Component

3.2.1. Analysis of Uplink Transmission Delay

We first consider λ_{CR}, the CR packet arrival rate at uplink. CR packets are generated by the call request in both directions. Furthermore, from our assumption that active MT does not generate CR packets, only idle MTs generate CR packets. Let p_{busy} be the probability that the MT has the active call. We can derive the average number of idle MTs in the cell as $(1 - p_{busy})N_{MT}$. Then, we have the CR arrival rate as

$$\lambda_{CR} = (\lambda_{MW} + \lambda_{WM})\left(1 - p_{busy}\right)N_{MT} \quad (1)$$

where p_{busy} is derived by approximating each MT as $M/M/1/1$ queuing system;

$$p_{busy} = \frac{(\lambda_{MW} + \lambda_{WM})/\mu}{1 + (\lambda_{MW} + \lambda_{WM})/\mu}. \quad (2)$$

When the MT having the active call moves into another cell, the MT generates the HO packet. Therefore, the arrival rate of HO packets is given by

$$\lambda_{HO} = h \cdot p_{busy} \cdot N_{MT}. \quad (3)$$

On the contrary, when the idle MT moves into another location area, it generates the LU packet. However, the arrival rates of LU packets at cells are different according to the cell's position within the location area. Let us introduce c_k to represent the fraction of the length of the boundary of the k–th cell when k–th cell is located at the boundary of the location area. If k–th cell is not at the boundary of the location area, the MT in that cell does not generate the LU packets. The coefficient c_k can be determined by the position of the k–th cell in the location area. For example, in the case of Figure 1, $c_k = 0$ for the centered cell, and $c_k = 1/2$ for other cells.

Thus, we can determine the arrival rate of LU packets at k–th ($k = 1, \ldots, N_c$) cell in the location area. We have

$$\lambda_{LU}^k = h \cdot (1 - p_{busy}) \cdot N_{MT} \cdot c_k. \quad (4)$$

Total packet arrival rate on the uplink is finally given as $\lambda_{CR} + \lambda_{HO} + \lambda_{LU}^k$. Since the slotted ALOHA is considered for the multiple access method, collisions of the packets should be taken into consideration. Let p_{col}^k be the probability of the collision at k–th cell (its determination is given later). By assuming that the number of retransmissions until the packet is successfully admitted by FBS follows the geometric distribution, the average number of transmissions is given by $1/(1 - p_{col}^k)$. We then obtain the packet arrival rate including retransmissions as

$$\lambda_{up}^k = \frac{1}{1 - p_{col}^k}\left(\lambda_{CR} + \lambda_{HO} + \lambda_{LU}^k\right). \quad (5)$$

Let T_f be the frame length, and n_u be the number of control slots of uplink in the frame. By considering that the packets having arrived in the previous frame compete for the control slots in the current frame, the packet collision probability, p_{col}^k, can be determined as

$$p_{col}^k = 1 - e^{-T_f \lambda_{up}^k / n_u}. \quad (6)$$

Since p_{col}^k depends on λ_{up}^k, we should derive λ_{up}^k by an iterative calculation using Eqs.(5) and (6).

After BS receives the control packet, it returns ACK in the downlink part of the next frame. Thus, the MT not receiving the ACK immediately retransmits the packet in the next control slots. Then we finally have the uplink delay $D_{Ru}^k(*)$ as follows:

$$D_{Ru}^k(*) = \frac{p_{col}^k}{1 - p_{col}^k}T_f + \frac{T_f}{2}. \quad (7)$$

The second term in the above equation represents the average delay from the packet arrival to the control slots at the next frame.

3.2.2. Analysis of Downlink Transmission Delay

When the MT terminating call arrives at MSC, it broadcasts PG packets to all BSs within the location area. The PG arrival rate is thus obtained as

$$\lambda_{PG} = \lambda_{WM}(1 - p_{busy})N_{MT}N_c. \quad (8)$$

Since MSC receiving CR, HO, and LU packets generates corresponding ACK packets, those arrival rates are obtained as

$$\lambda_{CA} = \lambda_{CR}, \quad \lambda_{HA} = \lambda_{HO}, \quad \lambda_{LA}^k = \lambda_{LU}^k. \quad (9)$$

Furthermore, BS returns the ACK packet when it receives CR, LU, and HO correctly. The arrival rate of ACK packets is then given as

$$\lambda_{ACK}^k = \lambda_{CR} + \lambda_{HO} + \lambda_{LU}^k. \qquad (10)$$

We have the arrival rate of downlink at k–th cell as follows:

$$\lambda_{down}^k = \lambda_{PG} + \lambda_{CA} + \lambda_{HA} + \lambda_{LA}^k + \lambda_{ACK}^k. \qquad (11)$$

Let n_d be the number of control slots in downlink. We approximately model the downlink queue as an M/D/1 queue where transmission time of signaling packets is T_f/n_d. Then we obtain the downlink delay at k–th cell as follows:

$$D_{Rd}^k(*) = \frac{T_f \lambda_{down}^k}{2n_d(n_d - T_f \lambda_{down}^k)} + \frac{T_f}{2}. \qquad (12)$$

3.2.3. Analysis of FBS Queue

The arrival rates of CR, HO, and LU packets at the FBS queue are given by λ_{CR}, λ_{HO}, and λ_{LU}^k, respectively. The priority processing may be introduced to satisfy the different delay requirements according to the kind of signaling packet. In that case, we can derive the processing delay at FBS by modeling it as a non-preemptive priority scheduling queuing system.

Consider the case where HO packets are given higher priority than other signaling packets. By assuming the exponential distribution for the processing time with identical means $1/\mu_{BS}$ for all signaling packets, we can obtain the packet delay of HO packets as (see, e.g., [19])

$$D_{Bu}^k(\text{HO}) = \frac{1}{\mu_{BS}} + \frac{R^k}{1 - \rho_1}. \qquad (13)$$

CR and LU packet delays with lower priority are identically given by

$$D_{Bu}^k(\text{CR}) = \qquad D_{Bu}^k(\text{LU})$$
$$= \frac{1}{\mu_{BS}} + \frac{R^k}{(1-\rho_1)(1-\rho_1-\rho_2^k)} \qquad (14)$$

where

$$\rho_1 = \frac{\lambda_{HO}}{\mu_{BS}}, \quad \rho_2^k = \frac{\lambda_{CR} + \lambda_{LU}^k}{\mu_{BS}},$$

and

$$R^k = \frac{\lambda_{CR} + \lambda_{HO} + \lambda_{LU}^k}{\mu_{BS}^2}.$$

3.2.4. Analysis of BBS Queue

In the BBS queue, ACK packets should be considered in addition to CA, LA, HA, and PG packets. However, the delay analysis can be performed in a way similar to the analysis of the FBS queue. It is true, even in the case of introducing the priority processing.

3.2.5. Analysis of FMSC Queue

The total packet arrival rate at the FMSC queue from all BSs belonging to the same location area is given by $N_c \lambda_{CR}$, $N_c \lambda_{HO}$, and $\sum_{k=1}^{N_c} \lambda_{LU}^k$. We can also adopt the priority processing.

3.2.6. Analysis of BMSC Queue

CA, LA, HA, and PG packets arrive at BMSC. Priority processing is also allowed.

3.2.7. Analysis of Processing Delay at the Wired Network

According to our assumption, the average processing delay for each type of signaling packets at wired network is derived by using the analysis results of the M/M/1 queue. For example, the call setup time at the wired network is given by

$$D_{Mp}(\text{CR}) = \frac{1/\mu_{CR}}{1 - N_c \cdot \lambda_{CR}/\mu_{CR}}. \qquad (15)$$

3.2.8. Analysis of Processing Delay at MT

The arrival rate of PG packets at the MT is λ_{WM}, and therefore we obtain the average processing delay at MT as

$$D_{MT}(\text{PG}) = \frac{1/\mu_{MT}}{1 - \lambda_{WM}/\mu_{MT}}. \qquad (16)$$

3.3. Derivation of Total Processing Delay

Based on the analysis of each component, we finally derive the total average processing delay for each type of signaling packets under consideration.

- MT originating call setup delay
  By simply summing up the delay at each component, we can obtain the total average processing delay for MT originating call setup at the k–th cell in the location area as follows:

$$
\begin{aligned}
T_{MW}^k &= D_{Mu}(\text{CR}) + D_{Mp}(\text{CR}) + D_{Md}(\text{CA}) \\
&\quad + D_{Ru}^k(\text{CR}) + D_{Bu}^k(\text{CR}) \\
&\quad + D_{Bd}^k(\text{CA}) + D_{Rd}^k(\text{CA}) \qquad (17)
\end{aligned}
$$

- MT terminating call setup delay
  Similarly, we have the total average processing delay for the MT terminating call setup at the k–th cell in the location area as follows:

$$
\begin{aligned}
T_{WM}^k &= D_{Md}(\text{PG}) + D_{MT}(\text{PG}) + D_{Mu}(\text{CR}) \\
&\quad + D_{Bd}^k(\text{PG}) + D_{Rd}^k(\text{PG}) \\
&\quad + D_{Ru}^k(\text{CR}) + D_{Bu}^k(\text{CR}) \qquad (18)
\end{aligned}
$$

- Handoff processing delay
  The handoff signaling packets are transferred via the same route as MT originating call setup packets. Then we have

$$
\begin{aligned}
T_{HO}^k &= D_{Mu}(\text{HO}) + D_{Mp}(\text{HO}) + D_{Md}(\text{HA}) \\
&\quad + D_{Ru}^k(\text{HO}) + D_{Bu}^k(\text{HO}) \\
&\quad + D_{Bd}^k(\text{HA}) + D_{Rd}^k(\text{HA}) \qquad (19)
\end{aligned}
$$

- Location update processing delay
  Similarly, we can obtain the average delay by applying CR $\rightarrow$ LU and CA $\rightarrow$ LA to Eq.(17).

4. Numerical Examples and Discussions

In the previous section, we have presented analytical techniques for deriving the total processing delay for each type of signaling packets. Based on those results, we now present numerical examples. Parameters used for the numerical examples are summarized in Table 1. Most parameters are drawn from [2, 8].

|	
Frame length	20 ms
Control slots in uplink	2
Control slots in downlink	2
Ave. number of MTs per cell	100
MT terminating call rate	0.0013 calls/sec
MT originating call rate	0.002 calls/sec
Ave. call duration	180 sec
Ave. cell sojourn time	20 sec
Ave. processing time in MT	10 msec
Ave. processing time in BS	5 msec
Ave. processing time in MSC	10 msec
Ave. connection setup time	10 msec
Ave. location update time	10 msec
Ave. handoff processing time	10 msec

Table 1. Parameter settings.

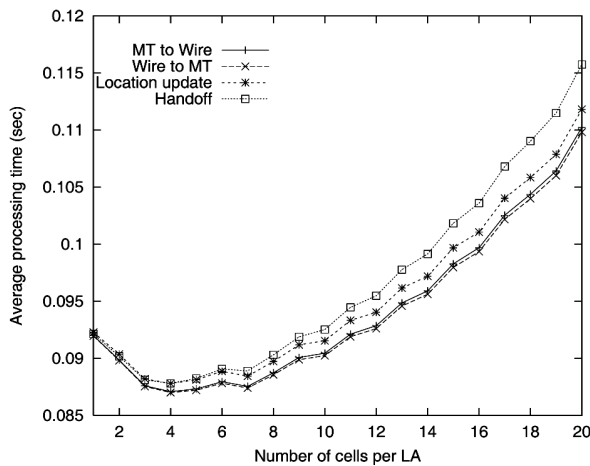## 4.1. Effect of Priority Processing

First, we show the case without priority processing. In this case, all kinds of signaling messages are processed by the FIFO discipline at every component. In what follows, we present the overall average delay in the location area. For example, the overall average delay of MT originating call setup delay is given by

$$
T_{MW} = \frac{1}{N_c} \sum_{k=1}^{N_c} T_{MW}^k. \qquad (20)
$$

Figure 6(a) shows how the location area size affects the average processing delays. In obtaining the figure, we set the originating call rate at each cell to be fixed, and the number of cells in each location area is increased. Thus, the total originating rate at each location area is increased as the location area size becomes larger. Nevertheless, the processing delays are decreased for small location area sizes, as shown in the figure. It is because the load of location update packets can be decreased at radio channel and BS by the larger location area. However, too large a location area introduces the increase of PG load, which introduces the bottleneck at MSC. It leads to the increase of the packet delays.
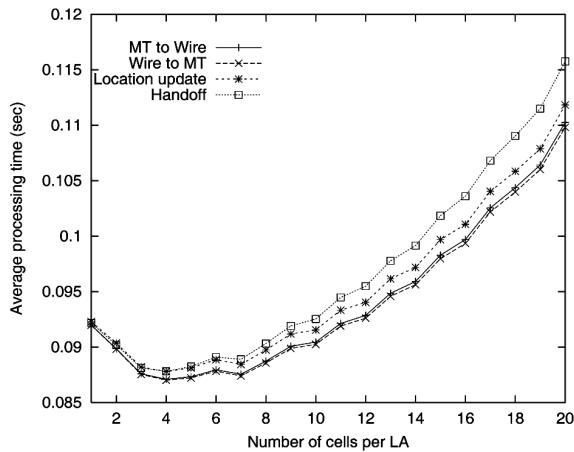
(a) Without priority processing

(b) Higher priority is given to handoff packets at BS
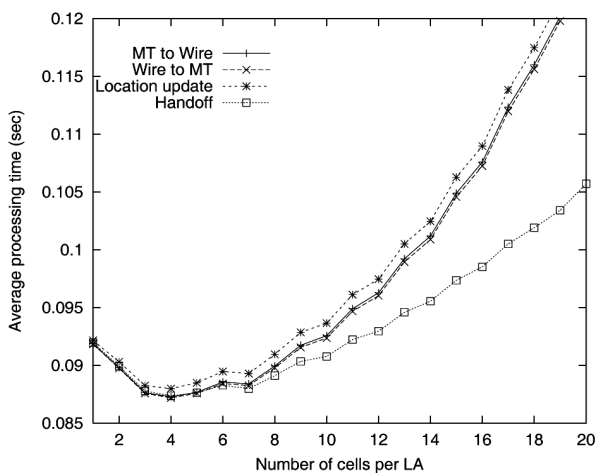
(c) Higher priority is given to handoff packets at MSC

Fig. 6. Effect of priority processing on average processing delay.

Next, we investigated the effect of priority processing. First, we performed priority processing to the handoff and the ACK packet at base stations. That is, we applied the priority higher than CR and LU packets to HO packets in the FBS queue, and applied the priority higher than CA and LA packets to HA and an ACK packets in the BBS queue. However, as shown in Figure 6(b), the priority processing at BS was almost ineffective. The next example shown in Figure 6(c) clearly demonstrates the effect of priority processing at MSC. In this case, HO packet is given higher priority than CR and LU at FMSC queue, and HA packet is given higher priority than CA and LA packets. The handoff delays can be much improved at the expense of increased processing delays of other packets. If the location area is small (N_c=1 to 7 in Figure 6(c)), however, we can see that the priority processing is not so effective. This is because the delays of uplink and downlink are dominant in the total processing delays in this region. From these results, it became clear that the priority processing in MSC is effective, because all the packets in a location area concentrated on the MSC and the load of MSC increased.

4.2. Delay Difference Among Cells

In the previous subsection, we have evaluated the overall average delay in the location area. However, the processing delays are actually different, according to the position of the cell within the location area. It is because the location update signaling packets only arrive at the edge of the location area, which affects the processing delays of other signaling packets. We examine the differences of the processing delays at cells. Figure 8 shows the maximum, minimum, and average of handoff processing delays within the location area. It can be obtained by using Eq.(19). For given number of cells per the location area, the structure illustrated in Figure 7 was used.

The minimum delay is observed in the cell located in the center of the location area. In the figure, the minimum delay is almost linearly decreased as N_c becomes larger. Then, when N_c exceeds seven, the delay is increased. It can be explained as follows. When N_c is smaller than
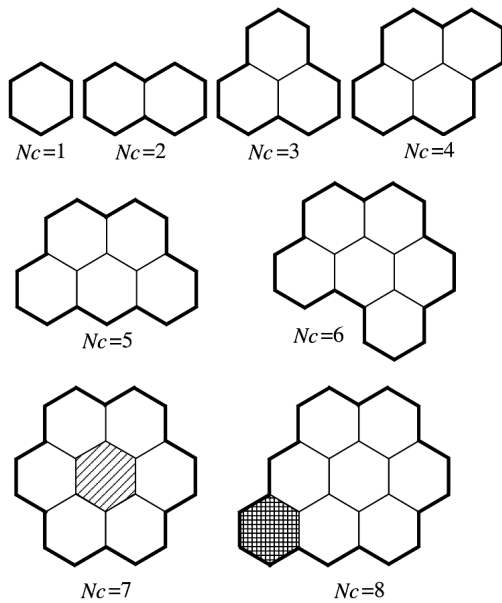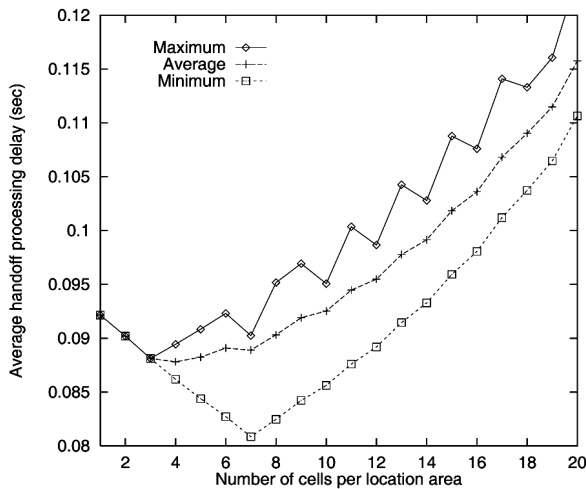
Fig. 7. Structure of location areas.

*Fig. 8.* Delay differences due to the position of each cell in a location area.

seven, the larger size of the location area leads to decrease of the location update traffic load since all cells partly have the boundary of the location area. As the location area size gets larger than seven, the location update packet does not arrive at the cell around the center of the area (the hatched cell in Figure 7). Therefore, larger location area leads to increase of the processing delays due to the increase of PG load and the bottleneck of MSC.

On the contrary, the maximum delay varies irregularly. This is because the structure of the location area leads to the various lengths of the boundary of the location area (see, e.g., the meshed cell in Figure 7). Since the location update load becomes heavy in such a cell, the average processing delays become large. In a regular structure (such as the case of $N_c = 7$), on the other hand, the average processing delay is small. From the figure, we can observe that in designing the system, it is important to take account of the structure of the location area since there exists an appropriate structure to minimize the location update traffic load.
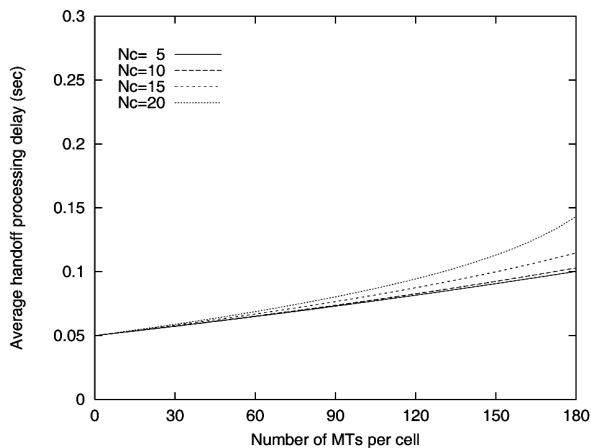
4.3. Influence of the Number of Terminals Upon Processing Delay

We last investigate the influence of the average number of terminals upon processing delay. Figure 9(a) shows how the number of terminals affects the average handoff delays for several location area sizes (N_c=5, 10, 15, and 20). In this example, we do not apply priority processing at any component, and assume that the other parameters are same as in the previous case (Table 1). As the number of terminals increases, the load of signaling packets also increases and then the average processing delay becomes larger. The degradation of performance is more serious at larger location areas. This is because all signaling packets are concentrated at MSC, and processing load of MSC becomes heavy as the location area size increases.

To keep the robustness against the increase of the number of terminals, we finally examine the effect of processing capability of MSC upon the total processing delay. Figure 9(b) shows the average handoff processing delays where the processing delay of MSC, $1/\mu_{MSC}$, is reduced to 5 msec. By comparing Figure 9(a) and (b), it can be observed that the total processing delays become small, especially at larger location areas. It also indicates that the delay of MSC is dominant in the total processing delays when the location area size is large. Therefore, the processing capability of MSC must be determined carefully for designing a robust signaling system.
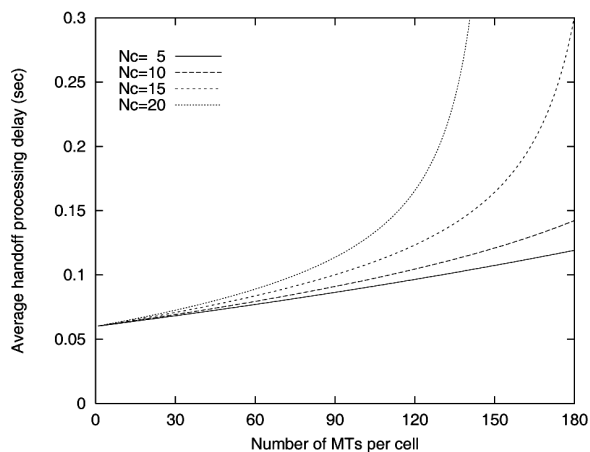
(a) Average handoff processing delay vs. number of terminals (Mean processing time at MSC: $1/\mu_{MSC}$=10 msec)

(b) Effects of processing capability of MSC ($1/\mu_{MSC}$=5 msec)

Fig. 9. Influence of the average number of terminals upon handoff processing delay.

5. Conclusion

In this paper, we have presented an analytical model for evaluating a total processing delay of signaling packets of the wireless cellular networks. By using our proposed model, we have derived the delays of call setup, location update, and handoff packets. Furthermore, we have introduced priority processing in the model to reduce the handoff processing delay, and evaluated its effectiveness through numerical examples. We have also evaluated the delay difference among cells within the location area to show how the location area structure affects the

processing delays. Finally, we have shown the influence of the number of terminals upon processing delays.

For future research topics, we need to analyze the bottleneck under various system parameters, to find out the component which dominates the total processing delays. While in the current paper we have simply assigned higher priority to handoff signaling packets than to the others, it is important to investigate more appropriate priority assignment based on the practical delay requirements of signaling packets.

References

[1] K.S. MEIER-HELLSTERN, E. ALONSO, AND D.R. O'NEIL, The use of SS7 and GSM to support high density personal communications, *Proceedins of the IEEE International Conference on Communications*, (1992) Chicago. pp. 1698–1702.

[2] I. RUBIN AND C.W. CHOI, Impact of the location area structure on the performance of signaling channels in wireless cellular networks, *IEEE Communications Magazine*, **35** (1997), no. 2, pp. 108–115.

[3] L.P. ARAUJO AND J.R.B. DE MARCA, Paging and location update algorithms for cellular systems, *IEEE Transactions on Vehicular Technology*, **49** (2000), pp. 1606–1614.

[4] A. MUKHERJEE, S. BANDYOPADHYAY, AND D. SAHA, *Location Management and Routing in Mobile Wireless Networks*, Artech House, 2003.

[5] G.P. POLLINI, K.S. MEIER-HELLSTERN, AND D.J. GOODMAN, Signaling traffic volume generated by mobile and personal communications, *IEEE Communications Magazine*, **33** (1995), no. 6, pp. 60–65.

[6] W. WANG AND I.F. AKYILDIZ, Reducing the paging costs under delay bound for PCS networks, *Proceedings of the IEEE Wireless Communications and Networking Conference*, (2000) Chicago, IL.

[7] W. WANG, I.F. AKYILDIZ, G.L. STÜBER, AND B.-Y. CHUNG, Effective paging schemes with delay bounds as QoS constraints in wireless systems, *Wireless Networks*, **7** (2001), pp. 455–466.

[8] M. VEERARAGHAVAN, T.F. LA PORTA, AND R. RAMJEE, A distributed control strategy for wireless ATM Networks, *Proceedings of the IEEE International Conference on Communications*, (1995) Seattle, WA.

[9] M. VEERARAGHAVAN, M.J. KAROL, AND K.Y. ENG, Mobility and connection management in a wireless ATM LAN, *IEEE Journal on Selected Areas in Communications*, **15** (1997), pp. 50–68.

[10] R. YUAN, S.K. BISWAS, AND D. RAYCHAUDHURI, A signaling and control architecture for mobility support in wireless ATM networks, *Proceedings of the IEEE International Conference on Communications*, (1996) Dallas, TX.

[11] B.A. AKYOL, AND D.C. COX, Signaling alternatives in a wireless ATM network, *IEEE Journal on Selected Areas in Communications*, **15** (1997), pp. 35–49.

[12] N.H. LOUKAS, N.I. PASSAS, L.F. MERAKOS, AND I.S. VENIERIS, Design of call control signaling in wireless ATM networks, *Proceedings of the IEEE International Conference on Communications*, (1997) Quebec, pp. 1554–1559.

[13] I. RUBIN AND C.W. CHOI, Delay analysis for forward signaling channels in wireless cellular network, *Proceedings of the IEEE INFOCOM*, (1996), San Francisco, CA.

[14] D. RAYCHAUDHURI, Wireless ATM networks: architecture, system, design and prototyping, *IEEE Personal Communications*, **3** (1996), no. 4, pp. 42–49.

[15] D. HONG AND S.S. RAPPAPORT, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures, *IEEE Transactions on Vehicular Technology*, **35** (1986), no. 3, pp. 77–92.

[16] G. FOSCHINI, B. GOPINATH, AND Z. MILJANIC, Channel cost of mobility, *IEEE Transactions on Vehicular Technology*, **42** (1993), no. 4, pp. 414–424.

[17] M. NAGHSHINEH AND A.S. ACAMPORA, Design and control of micro-cellular networks with QOS provisioning for real-time traffic, *Jornal of High Speed Networks*, **5** (1996), Issue 1, pp. 53–71.

[18] S.C. AGRAWAL, *Metamodeling*, MIT Press, 1985.

[19] D. BERTSEKAS AND R. GALLAGER, *Data Networks Second Edition*, Prentice-Hall, 1992.

Contact address:
Masashi Sugano
Faculty of Comprehensive Rehabilitation
Osaka Prefecture College of Nursing
3-7-30, Habikino, Habikino-shi
Osaka, 583-8555
Japan
e-mail: sugano@osaka-hsu.ac.jp
Web: http://www.osaka-hsu.ac.jp/~sugano

MASASHI SUGANO received the B.E., M.E., and D.E. degrees in Information and Computer Sciences from Osaka University, Japan, in 1986, 1988, and 1993, respectively. In 1998, he joined Mita Industrial Co., Ltd. (currently, Kyocera Mita Corporation) as a Researcher. From 1996 to 2003, he was an Associate Professor in Osaka Prefecture College of Health Sciences. He moved to Faculty of Comprehensive Rehabilitation, Osaka Prefecture College of Nursing in April 2003. His research interests include design an performance evaluation of computer communication networks, network reliability, and wireless network systems. He is a member of IEEE, ACM, IEICE and IPSJ.

DOO SEOP EOM received the B.S. and M.S. degrees in Electronics Engineering from Korea University, Seoul, Korea, in 1987 and 1989, respectively. In 1999, he received the Dr.E. degree in Information and Computer Science from Osaka University, Osaka, Japan. He joined the Communication Systems Division, Electronics and Telecommunications Research Institute (ETRI), Korea, in 1989. From September 1999 to August 2000, he was an Associate Professor of Wonkwang University, Korea. From September 2000 to August 2002, he was an Assistant Professor in the School of Electrical Engineering, Korea University, Seoul, Korea. Since September 2002, he has been an Associate Professor in the School of Electronics Engineering, Korea University, Seoul, Korea. His research interests include communication network design, wireless ATM, and Internet QoS.

MASAYUKI MURATA received the M.E. and D.E. degrees in Information and Computer Sciences from Osaka University, Japan, in 1984 and 1988, respectively. In April 1984, he joined Tokyo Research Laboratory, IBM Japan, as a Researcher. From September 1987 to January 1989, he was an Assistant Professor with Computation Center, Osaka University. In February 1989, he moved to the Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University. From 1992 to 1999, he was an Associate Professor in the Graduate School of Engineering Science, Osaka University, and from April 1999, he has been a Professor at Osaka University. He moved to Graduate School of Information Science and Technology, Osaka University in April 2004. He has presented more than three hundred papers in international and domestic journals and conferences. His research interests include computer communication networks, performance modeling and evaluation. He is a member of IEEE, ACM, The Internet Society, IEICE and IPSJ.

HIDEO MIYAHARA received the M.E. and D.E. degrees from Osaka University, Osaka, Japan in 1969 and 1973, respectively. From 1973 to 1980, he was an Assistant Professor in the Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto, Japan. From 1980 to 1986, he was an Associate Professor in the Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University, Osaka, Japan. From 1986 to 1989, he was a Professor at the Computation Center, Osaka University. Since 1989, he has been a Professor in the Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University. From 1995 to 1998, he was a director of Computation Center of Osaka University. From 1998 to 2000, he was the dean of the Faculty of Engineering Science, Osaka University. From 2002 to 2003, he is the dean of Graduate School of Information Science and Technology, Osaka University. From 2003, he is the president of Osaka University. From 1983 to 1984, he was a Visiting Scientist at IBM Thomas J. Watson Research Center. His research interests include performance evaluation of computer communication networks, broadband ISDN, and multimedia systems. Prof. Miyahara is a Fellow of IPSJ and an IEEE Fellow.