Electronics and Communications in Japan, Part 1, Vol. 86, No. 2, 2003
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J84-B, No. 3, March 2001, pp. 344–353

Mobile IP Route Optimization Method for Next-Generation Mobile Networks

Yasushi Takagi, Takeshi Ihara, and Hiroyuki Ohnishi

NTT Network Service Systems Laboratories, Musashino, 180-8585 Japan

SUMMARY

Mobile IP has gained attention as a technology that can provide mobility to universal users independently of the access network. The introduction of Mobile IP in mobile core networks is currently under investigation. This paper discusses route optimization methods when Mobile IP technology is applied to mobile core networks. More specifically, it discusses means of resolving the "Triangle Routing" problem, which occurs when communications from a correspondent node to a mobile node must be routed through a Home Agent (HA). This paper first describes the conventional route optimization method being investigated by the Internet Engineering Task Force (IETF) and indicates problems with that method. It then proposes a new Mobile IP Border Gateway (MBG) method for solving the problems of the conventional method. The MBG method enables packets that arrive from another network to be directly received at the mobile node by having a Border Gateway router maintain Binding Information consisting of the care-of address and home address that were originally being maintained by the Home Agent for the communicating node. This paper also shows that introducing the MBG method enables the mean route length within the network to be shortened and the traffic volume routed through the Home Agent to be reduced. Finally, it shows methods of reducing the burden of registering Binding Information at the Mobile IP Border Gateway and indicates the applicability of the MBG method to large-scale mobile communications networks. © 2002 Wiley Periodicals, Inc. Electron Comm Jpn Pt 1, 86(2): 31–41, 2003; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ecja.1151

Key words: mobile communications network; IP mobility; Mobile IP; route optimization.

1. Introduction

Recently, mobile communications services have experienced remarkable growth, and among these, services providing Internet access from mobile terminals are steadily increasing by tens of thousands of subscribers per day [1]. In Japan, the third-generation mobile communications system, IMT-2000 [2], was launched in 2001. The main goals of IMT-2000 services are to provide high-quality multimedia services at speeds up to 2 Mbit/s, to provide a global roaming service spanning mobile communications carriers worldwide, and to enable technologies to be examined by a coalition of telecommunications standards-setting bodies [3, 4]. Other wireless access technologies that are also under investigation include MMAC [5] and HIPERLAN/2 [6], which can provide even faster services at speeds on the order of 10 Mbit/s. Mobile communications seem to be entering an era of genuine high-speed, wide-area communications.

To provide these wide-area and multimedia communications services and to deal with the dramatic increase in Internet users, the construction of IP-based networks instead of conventional circuit-switching networks is being considered even for mobile core networks [7, 8]. The creation of IP-based core networks can be divided into (1) the creation of IP-based packet transfer mechanisms and (2) the creation of IP-based mobility management mechanisms. To create IP-based packet transfer mechanisms, which will provide multimedia services that include voice by creating and transferring IP packets, different types of servers are under investigation for providing a variety of speeds and QoS classes and offering connection-type services for voice and images on IP networks [7]. To create IP-based mobility management mechanisms, efforts are being made to implement mobility management at the IP level (hereafter referred to as IP Mobility) by using IP addresses, for example, instead of managing the movement of terminals within a mobile network according to the conventional mobile terminal number or subscriber number [9]. Advantages of implementing IP Mobility include the ability to implement universal mobility management mechanisms independent of specific access networks and the ability to provide seamless mobile services spanning wireless access network generations and even heterogeneous access networks including both mobile and fixed networks. Since IP Mobility supports the implementation of network integration and services between mobile and fixed networks, it is becoming an important technology not only for mobile networks but also for fixed networks.

Currently, the most promising technology for providing IP Mobility is Mobile IP [10]. Mobile IP had first been investigated as a so-called roaming technology for enabling IP users to move among multiple private LANs. However, recently, it is also being investigated as a technology for implementing IP-based mobile core networks [9]. Although Mobile IP enables wide-area mobility to be implemented at the IP level, it does not have functions characteristic of wireless access networks such as high-speed handover or paging functions. Therefore, to apply Mobile IP in mobile networks, investigations are being carried out with the goal of implementing high-speed mobility and wide-area (including other networks) mobility according to mobility management techniques that combine the advantages of both, namely, by applying Mobile IP to the core network portion and following the policies of conventional wireless access networks for the access network part [11].

Problems encountered when applying Mobile IP to mobile core networks include (i) the increase in route length and concentration of traffic at the Home Agent (HA) due to the fact that all communications from a correspondent node to a mobile terminal are necessarily routed through the Home Agent, (ii) the implementation of high-speed, high-quality handover mechanisms capable of supporting multimedia [12–14], and (iii) the implementation of security mechanisms such as user authentication at the IP level. Among these, the increase in route length and concentration of traffic at the Home Agent in (i), which is a problem that depends on the fact that IP is a connectionless protocol, had been resolved by measures [15] related to methods of establishing connections using conventional mobile management mechanisms, which assume that a connection to the mobile terminal is established in advance. The solution of this "Triangle Routing" problem is to use network resources efficiently and eliminate Home Agent processing bottlenecks. The IETF has also been discussing a solution method known as "route optimization" [12]. However, the IETF method requires not only mobile terminals, but also correspondent nodes to have route optimization mechanisms, and considering the global nature of IP communications, requiring all terminals to have this function is not realistic.

This paper proposes a solution to the above route optimization problem based on the assumption that Mobile IP is implemented in the mobile core network, and describes the results. Section 2 describes the Mobile IP-based mobile core network configuration that is assumed and the Triangle Routing problem. Section 3 outlines the conventional Triangle Routing solution discussed by the IETF and indicates problems with this solution. Section 4 proposes a new Mobile IP Border Gateway (MBG) method, which solves the problems of the conventional method. Section 5 uses numerical calculations to evaluate the effect of introducing the MBG method.

2. Mobile IP-Based Mobile Core Network Configuration

Figure 1 shows the configuration of the mobile communications network using Mobile IP in the core network, which is assumed in this paper. Since a Mobile IP core network provides universal mobile management functions at the IP level, it accommodates both mobile and fixed access networks, enabling IP users to move (roam) between different access networks. The Mobile IP core network consists of routers for performing mobility management of users called Home Agents (HA) and Foreign Agents (FA) [10] and routers for providing Border Gateway (BG) functions with other networks. With Mobile IP, individual For-

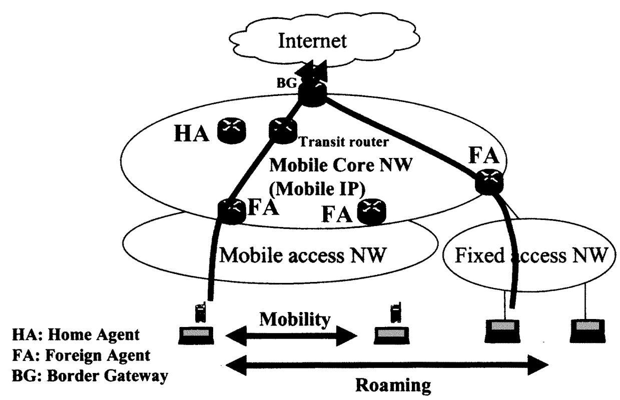


Fig. 1. Mobile IP-based mobile communications network configuration.

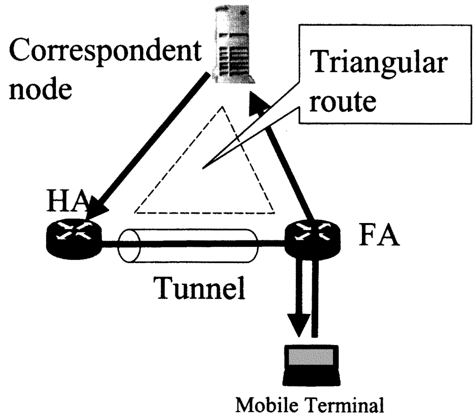


Fig. 2. Triangle routing problem in Mobile IP.

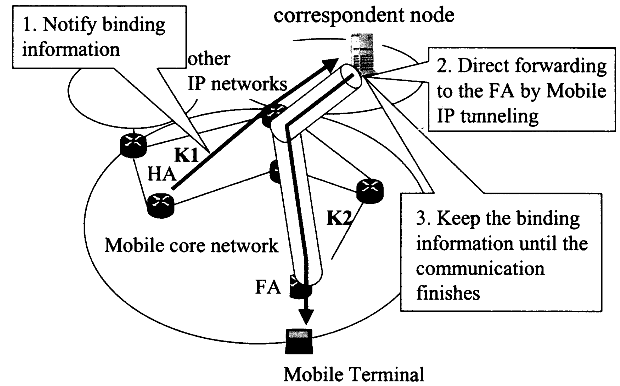


Fig. 3. Conventional route optimization method.

eign Agents, which are located corresponding to individual subnets that mobile nodes will visit, manage position information and maintain the locations of visiting nodes within the access networks. Also, Home Agents maintain combinations of the home addresses of all mobile nodes under their jurisdiction and the addresses (hereafter called care-of addresses) of the visited Foreign Agents. A characteristic of packet transfers in Mobile IP is that downstream packets are always routed through the Home Agent as shown in Fig. 2. In other words, in a Mobile IP network, upstream packets from a mobile node to a correspondent node are transferred directly to the correspondent node in a similar manner as in a normal IP network. However, for downstream packets, after the correspondent node sets the mobile node's home address as the IP packet's destination address and transfers the packet to the Home Agent, the Home Agent derives the address on the visited Foreign Agent (care-of address) and tunnels the packet to the care-of address so that the IP packet is delivered to the mobile terminal. In this way, Mobile IP causes Triangle Routing, in which the upstream and downstream routes differ. This Triangle Routing is a phenomenon caused by the connectionless nature of Mobile IP. While the need to establish a connection prior to communication can be eliminated, transfer quality and network cost are adversely affected as the increase in route length increases transfer delay and wastes network resources. A means of dealing with these problems is desirable.

3. Route Optimization and Problems with Conventional Method

The IETF is investigating the route optimization method shown in Fig. 3 for dealing with this problem [12]. The operation of the IETF method is outlined below.

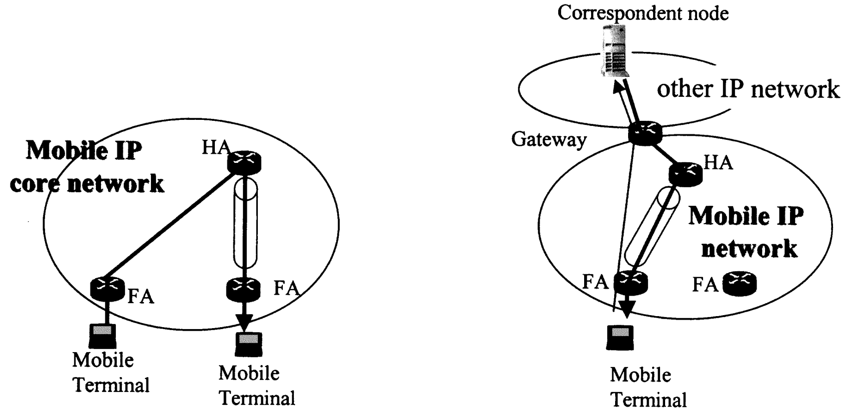
(E1) When the Home Agent receives a packet destined to the mobile terminal from a correspondent node, it reports (Binding Update) a combination of the relevant mobile terminal's home address and care-of address (Binding Information), which the Home Agent itself is maintaining, to the correspondent node.

(E2) The correspondent node maintains the Binding Information of (E1) and tunnels subsequent packet transmissions to the mobile terminal's care-of address.

(E3) Although the correspondent node maintains the Binding Information for a fixed interval, it will update the timer before a timeout occurs when it receives another Binding Update from the Home Agent.

In the above description, the Home Agent decides to send out a Binding Update. However, there is also an option in which the correspondent node issues a request (Binding Request) to the Home Agent to send out a Binding Update. These functions described in (E1) to (E3) enable the correspondent node to directly send packets to the mobile terminal without routing them through the Home Agent.

When the application of this IETF method to a mobile core network is considered, communication between mobile terminals as shown in Fig. 4(a) can be implemented by incorporating the functions described in (E2) and (E3) in mobile terminals belonging to the relevant mobile core network. However, for communication with a terminal belonging to an external network as shown in Fig. 4(b), the functions described in (E2) and (E3) must also be added to these external terminals. Currently, the majority of IP communications from mobile terminals in the global Internet environment are believed to be these kinds of communications with terminals in external networks. However, adding the functions described in (E2) and (E3) to various terminals (such as WWW servers) worldwide in order to make this method operate efficiently is currently unrealistic. This is a problem with the conventional method. In addition, since this method reports care-of addresses, which indicate user position information within the mobile network, a security-related problem also exists.



(a) Between two mobile terminals

(b) Between mobile terminal and correspondent node in other IP network

Fig. 4. Types of communication in mobile networks.

4. Proposed Route Optimization Technique

This section proposes a new method, which solves the problems of the conventional route optimization (IETF) method described in Section 3. The objective of the proposed method is to implement route optimization within a mobile network without adding functions to terminals in external networks by locating the functions that must be added to correspondent nodes in the conventional method in Mobile IP Border Gateways (MBG), which are devices within the mobile network [16, 17].

Outline of proposed method

A feature of the proposed MBG method is that incoming packets to the mobile network from another IP network are routed through a border gateway router located at the interconnection point between the networks (Fig. 5). Since this border gateway router, which is called a Mobile IP Border Gateway (MBG), maintains the association of the communicating mobile terminal's home address and current care-of address (Binding Information), incoming packets from the external network are tunneled from the Mobile IP Border Gateway and delivered directly to the mobile terminal. As a result, even incoming packets from an external network need not be routed through the Home Agent, thereby enabling the Triangle Routing problem to be solved. The operation of the MBG method is outlined below.

(M1) When the Home Agent receives a packet destined to the mobile terminal from a correspondent node, it reports (Binding Update) a combination of the relevant mobile terminal's home address and care-of address (Binding Information), which the Home Agent itself is maintaining, to the Mobile IP Border Gateway.

(M2) The Mobile IP Border Gateway maintains the Binding Information of (M1) and tunnels subsequent packet transmissions to the mobile terminal's care-of address.

(M3) Although the Mobile IP Border Gateway maintains the Binding Information for a fixed interval, it will update the timer before a timeout occurs when it receives another Binding Update from the Home Agent.

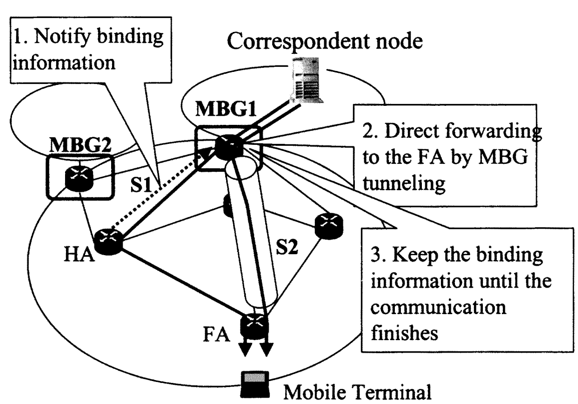


Fig. 5. Proposed route optimization method using a Mobile IP Border Gateway.

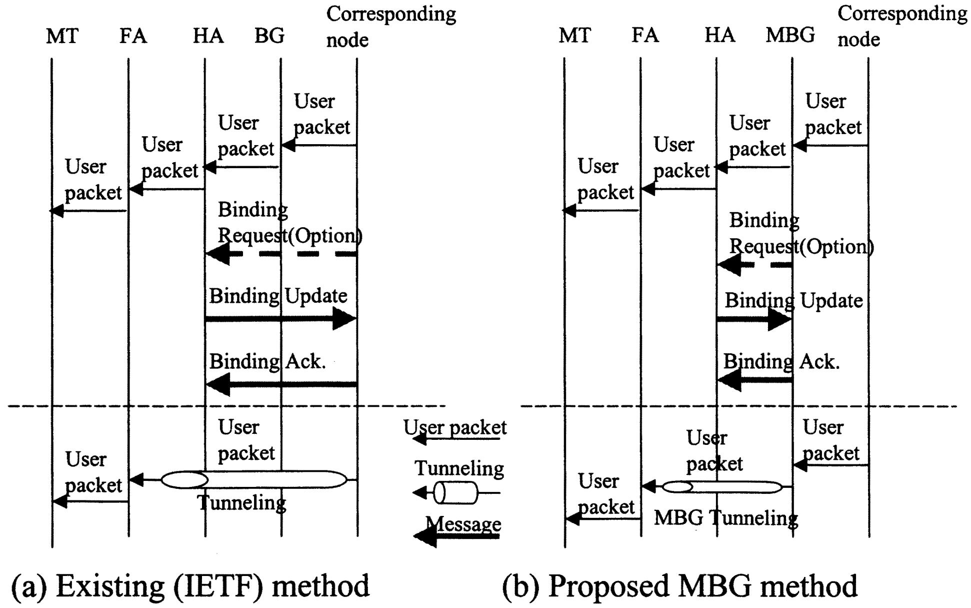


Fig. 6. Route optimization processing sequences (existing method and MBG method).

As with the conventional method, there is also an option in which the Mobile IP Border Gateway issues a request (Binding Request) to the Home Agent to send out a Binding Update.

Figure 6 shows the processing sequences for the conventional and MBG methods. As seen, the introduction of the MBG method does not necessitate the addition of functions to either the Foreign Agent or the mobile terminal, and messages that have already been defined by the existing IETF method can be used for all messages required for route optimization processing. In addition, with the proposed method, user position information remains within the mobile network, thereby solving the problem with the conventional method in which this information was sent out to the external network.

5. Evaluation of Effect and Numerical Examples

This section evaluates the effect of introducing the MBG method proposed in Section 4 and discusses the applicability of the MBG method in a mobile core network. As indices indicating the effect of route optimization, this section evaluates the reduction of the mean route length within the mobile network and the reduction of traffic routed through the Home Agent.

5.1. Route length reduction effect

The route length reduction effect due to the MBG method was evaluated from the standpoints of (1) network scale and (2) Home Agent and Foreign Agent locations.

The three-layer network model shown in Fig. 7 was used for the hypothetical large-scale mobile communications network. The conditions for this model are described below.

- The first layer, which consists of n nodes connected in a mesh topology, is connected to a border gateway (with or without MBG functions). The mean distance between nodes in the first layer is assumed to be a, and the distance from the border gateway is assumed to be zero.

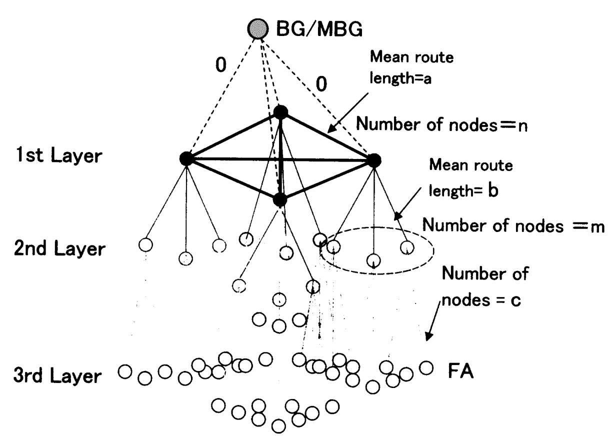


Fig. 7. Network model used for evaluations.

• The second layer consists of n nodes connected in a star topology to each node of the first layer for a total of $n \times m$ nodes. The mean distance from the first-layer node is assumed to be b, and this is assumed to represent the network scale of the second layer.

• The third layer consists of nodes connected in a star topology to each node of the second layer. The mean distance from the second-layer node is assumed to be c, and this is assumed to represent the network scale of the third layer.

Foreign Agents are assumed to constitute all nodes of the third layer, and two cases are assumed for Home Agents. In case 1, Home Agents constitute all nodes of the first layer, and in case 2, they constitute all nodes of the second layer. In each case, an equal number of mobile terminals is assumed to belong to each Home Agent, and the mobile terminals are assumed to be uniformly distributed under all Foreign Agents regardless of the Home Agents to which they belong. The evaluations that use this model assume that communications are only between mobile terminals and correspondent nodes in external networks. In addition, the distribution of traffic exchanged between correspondent nodes and all terminals is assumed to be uniform.

In a relatively small-scale network or area, the Home Agents and Foreign Agents may also be considered to be connected in a mesh topology. This can be represented by setting $b = c = 0$ and $m = 1$ for case 1 of the model in Fig. 7.

The mean value (mean route length) of the route lengths from the border gateway to the visited Foreign Agents for all mobile terminals is used as the index indicating route length. Evaluating this mean route length enables the usage efficiency of network resources (such as link cost) or the effect on transfer delay to be estimated. In addition, the effect of the layer in which the Home Agents are located or of the network scales of each layer (a, b, c) is evaluated at the same time.

The mean route length for each case is as follows.

(a) Without MBG functions

For case 1:

The mean route length is $[(n - 1)/n] \times a + b + c$.

For case 2:

The mean route length is

$$[(n - 1)/n] \times a + b + c + 2 \times [1 - 1/(n \times m)] \times b.$$

Appendix 1 describes the specific calculations.

(b) With MBG functions

For both cases 1 and 2, the mean route length is $b + c$.

The route length reduction effect due to the MBG method will be the difference of the mean route length for a border gateway without MBG functions and the mean route length for a border gateway with MBG functions, namely, $[(n - 1)/n] \times a$ (for case 1) and $[(n - 1)/n] \times a + 2 \times [1 - 1/(n \times m)] \times b$ (for case 2).

Figure 8 shows the relationship between the ratio ($r = b/a$) of the network scales of the first and second layers when $c = 0.1a$ and the mean route length normalized by a, which represents the network scale of the first layer. The relationship $0.1 < r = b/a < 1$ is assumed to hold, and the values of n and m are set as $(n, m) = (3, 4)$ and, for sufficiently large values, $(n = m = \infty)$.

From Fig. 8, it is clear that route optimization due to the MBG method can reduce the mean route length regardless of the positions where the Home Agents are located as indicated by cases 1 and 2. Also, for case 1 in which the Home Agents are in the first layer, the route length reduction effect accompanying an increase in r remains a fixed value that depends only on the scale of the first layer. However, for case 2 in which the Home Agents are in the second layer, the effect of introducing the MBG method increases as r increases because the effect increases according to the scale of the second layer as well as the route length for the case without MBG functions. When the values of n and m increase, the number of Home Agents increases, and the probability that the mobile terminal will be visiting a Foreign Agent under its own Home Agent decreases. Therefore, since the mean route length for the case without MBG functions increases, the MBG method is more effective.

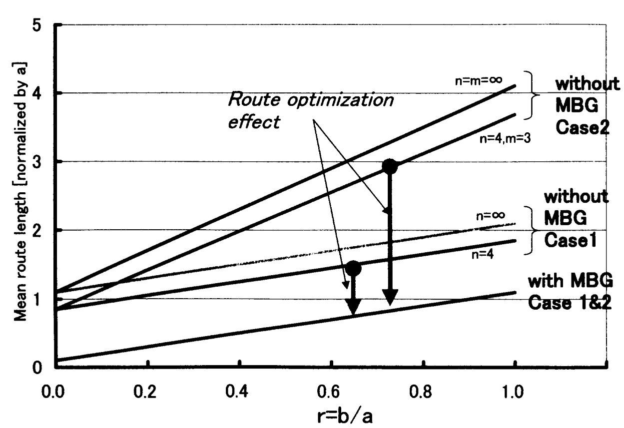


Fig. 8. Mean route length reduction effect due to MBG method.

5.2. Effect on reduction in traffic routed through home agents

This section uses the same network model as in Section 5.1 to examine the effect on the reduction in traffic routed through Home Agents due to the application of the MBG method. The aggregate number of Home Agents transited by the communication routes for all traffic is used as an index indicating the amount of Home Agent traffic.

First, when MBG functions are not used, there are two cases for determining the number of Home Agents transited by the communication routes, namely, (i) when the mobile terminal is visiting a Foreign Agent under its own Home Agent, only the relevant Home Agent will be transited, and (ii) when the mobile terminal is visiting a Foreign Agent under a Home Agent other than its own Home Agent, after its own Home Agent is transited, the Home Agent accommodating the visited Foreign Agent will be transited. The number of transited Home Agents per communication will be 1 for case (i) and 2 for case (ii), and the probability of each case will be 1/(number of Home Agents) for case (i) and (number of Home Agents – 1)/(number of Home Agents) for case (ii). Therefore, if the total number of communications in the network is M, the aggregate number of Home Agent transits can be calculated as follows for both cases 1 and 2.

$$\text{Aggregate number of Home Agent transits} = M \times [1 \times (1/\text{number of Home Agents}) + 2 \times (\text{number of Home Agents} - 1)/(\text{number of Home Agents})] = M \times [2 - (1/\text{number of Home Agents})]$$

Here, the number of Home Agents is n in case 1 and $n \times m$ in case 2.

On the other hand, when MBG functions are used, since the only Home Agent transited by a communication route will be the single Home Agent under which the visited Foreign Agent is located regardless of the mobile terminal's visiting position, the aggregate number of Home Agent transits will be M for both cases 1 and 2.

Figure 9 shows the relationship between the number of Home Agents located in the network and the aggregate number of Home Agent transits for cases when MBG functions are used and are not used. However, the number of Home Agent transits has been normalized by the total number of communications M. When the MBG method is applied, the aggregate number of Home Agent transits is a fixed value regardless of the number of Home Agents. However, when the MBG method is not applied, the aggregate number of Home Agent transits monotonically increases as the number of Home Agents increases and

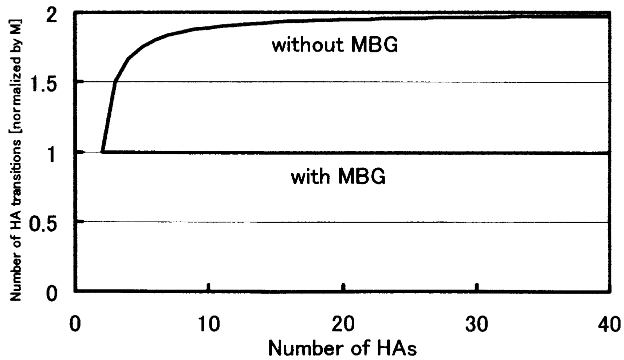


Fig. 9. Reduction in traffic routed through Home Agents due to MBG Method.

asymptotically approaches a value that is twice the value obtained when the MBG method is applied. From this, it is clear that when the number of Home Agents is sufficiently large, the amount of traffic transiting Home Agents can be reduced by approximately 1/2 by applying the MBG method.

5.3. Method of reducing MBG position registration load

The previous sections showed the route optimization effect due to the introduction of the MBG method. However, since the Mobile IP Border Gateway executes an operation for registering the previously described binding information for all communications between the mobile network and the external network for which the Mobile IP Border Gateway itself is the connection point, the Mobile IP Border Gateway's processing load can become quite large in a large-scale network. This section describes a procedure for solving this problem and evaluates its effect.

Each time a mobile terminal moves through the area (FA area) covered by a single Foreign Agent, which is the unit used for mobility management, it registers its position with its Home Agent. When any of these moving mobile terminals is communicating, the Home Agent executes an operation (Binding Update) for registering the position with the Mobile IP Border Gateway. As a means of reducing the processing load, the FA area can be enlarged to reduce the transit frequency, which also reduces the frequency with which the Home Agent executes operations for registering the position with the Mobile IP Border Gateway.

To quantitatively evaluate the position registration processing load on the Mobile IP Border Gateway, Home Agent, and Foreign Agent, the large city model (total area = 600 km^2, number of mobile terminals = 3,000,000, and mean velocity of mobile terminals = 10 km/h) shown in Fig. 10 was used.

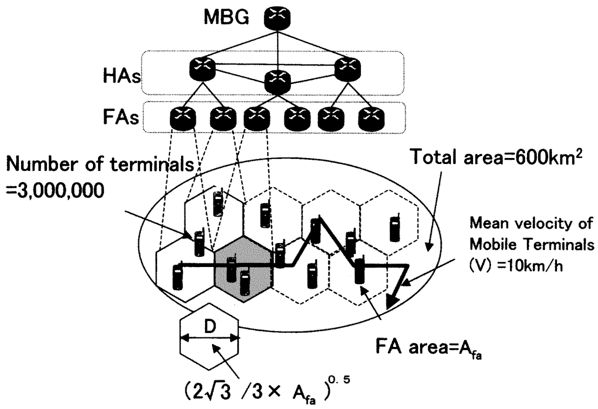


Fig. 10. Large city model.

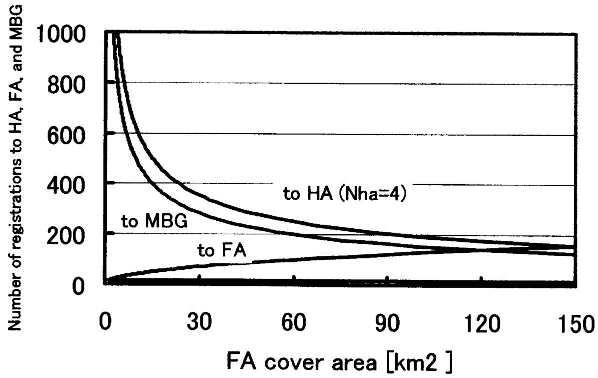


Fig. 11. Relationship between FA area size and numbers of position registrations with a Home Agent, Foreign Agent, and the Mobile IP Border Gateway.

If the FA area is approximated by the regular hexagon model with area A_{fa} shown in Fig. 10, the frequencies (R_{ha}), (R_{fa}), and (R_{mbg}) with which a position is registered (Binding Update) with a Home Agent, Foreign Agent, and the Mobile IP Border Gateway, respectively, per unit time are obtained by the following equations. Appendix 2 describes the specific calculations.

$$R_{ha} = M_{ha} \times V/(2\sqrt{3}/3)^{0.5}/\sqrt{A_{fa}}/N_{ha}$$

(per Home Agent)

$$R_{fa} = M_{ha} \times V/(2\sqrt{3}/3)^{0.5} * \sqrt{A_{fa}}/A$$

(per Foreign Agent)

$$R_{mbg} = N_{ha} \times R_{ha} \times p$$
$$= p \times M_{ha} \times V/(2\sqrt{3}/3)^{0.5}/\sqrt{A_{fa}}$$

Here, M_{ha} is the number of mobile terminals belonging to the Home Agent, V is the mean velocity of the terminals, p is the probability that a mobile terminal is communicating ($0 \leq p \leq 1$), N_{ha} is the number of Home Agents in the network, and A is the total area. When M_{ha}, V, p, and N_{ha} are constants in the above equations, the frequencies with which a position is registered (Binding Update) with a Home Agent and the Mobile IP Border Gateway increase in inverse proportion to the square root of A_{fa}, and the frequency with which a position is registered (Binding Update) with a Foreign Agent increases in direct proportion to the square root of A_{fa}.

Figure 11 shows the relationship between the frequencies with which a position is registered (Binding Update) with a Home Agent, Foreign Agent, and the Mobile IP Border Gateway in the large city model and the size of an FA area. In the figure, the constant values $N_{ha} = 4$ and $p = 0.2$ are assumed. The reason that the frequencies of registrations with a single Home Agent and the Mobile IP Border Gateway are almost identical is that the Home Agent registers the positions of all mobile terminals while the Mobile IP Border Gateway processes only terminals that are communicating. Figure 11 clearly shows that increasing the size of the FA area reduces the frequencies with which a position is registered with a Home Agent and the Mobile IP Border Gateway. Generally, although the number of users per Foreign Agent is thought to grow in direct proportion to the increase in size of the FA area, since the number of positions registered with the Foreign Agent is directly proportional to the square root of this size, an enlargement of the FA area seems to have a relatively small effect on the number of positions registered with the Foreign Agent.

When Foreign Agents must be located in terms of relatively small areas due to restrictions on the number of users accommodated by a Foreign Agent or restrictions due to the configuration of the access network, a Foreign Agent hierarchy as shown in Fig. 12 can be used effectively. Creating a Foreign Agent hierarchy is a means of increasing the FA area size viewed from the Home Agent or Mobile IP Border Gateway by establishing Gateway Foreign Agents (GFA) that accommodated multiple Foreign Agents [18]. Each time a mobile terminal crosses into a new FA area, its position is registered with the Gateway Foreign Agent. However, its position is only registered with the Home Agent or Mobile IP Border Gateway when the mobile terminal crosses into a new GFA area covered by the group of Foreign Agents that were grouped together by a Gateway Foreign Agent. This is equivalent to enlarging the FA area size. The effect on reducing the number of positions registered with the Home Agent or Mobile IP Border Gateway

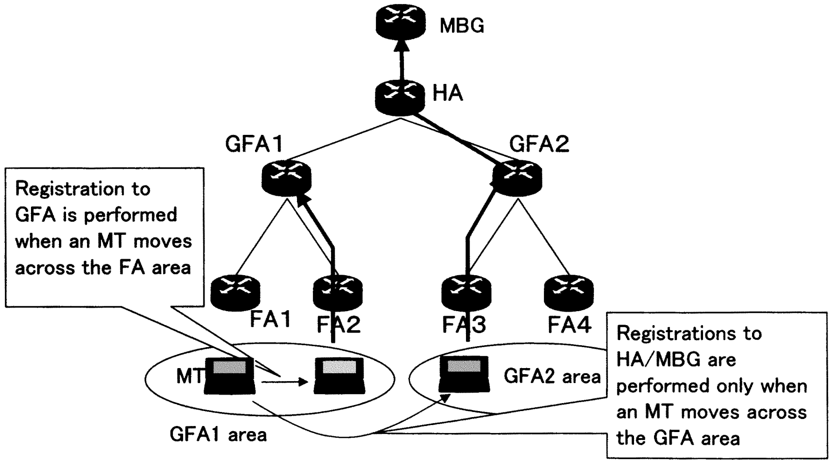


Fig. 12. Reduction in position registration frequency by creating a Foreign Agent hierarchy (GFA).

is that when N Foreign Agents, for example, are accommodated by a Gateway Foreign Agent in Fig. 11, the frequency with which positions are registered with the Home Agent or Mobile IP Border Gateway is reduced by $1/\sqrt{N}$ because the apparent FA area size is N times larger.

Therefore, since a processing load concentrated on the Mobile IP Border Gateway or Home Agent can be avoided by enlarging the FA area size or creating a Foreign Agent hierarchy, these are effective techniques to use when applying Mobile IP to a large-scale mobile network.

6. Conclusions

In this paper, a route optimization method using a Mobile IP Border Gateway (MBG) was proposed as a means of solving the Triangle Routing problem that occurs when Mobile IP is applied to a mobile core network. In addition, the route length reduction effect, Home Agent traffic load reduction effect, and method of reducing the processing load for registering positions with the Mobile IP Border Gateway were evaluated using numerical examples based on a model. The evaluation results indicated the effectiveness and applicability of the proposed method on a large-scale network.

Future subjects of study include investigation of methods of distributing the processing load by locating multiple Mobile IP Border Gateways in the network, clarification of specific conditions (such as the node for determining the start of route optimization or the number of packets required to start route optimization) for starting route optimization using a Mobile IP Border Gateway after communications had started, opportunities for sending Binding Updates out to the Mobile IP Border Gateway, and the establishment of specific methods of combining the IETF method and MBG method for communications between mobile terminals. Investigations of these subjects are continuing.

Acknowledgments. The authors are sincerely grateful to Assistant Professor Hiroshi Ezaki of the Tokyo University Mainframe Computer Center for valuable discussions during the course of this research. Useful guidance was also provided for a long time by Project Manager Kou Miyake and Group Leader Katsuyuki Okada at NTT Network Service Systems Laboratories. The authors also thank Mr. Minoru Matsumoto of NTT Network Systems Laboratories for helpful cooperation while this paper was being written.

REFERENCES

1. NTT DoCoMo Gateway Business Department. DoCoMo i-mode. 3GPP SA2 S2-99E94.
2. ITU-T Recommendation Q.1701, Version 4.2, Framework for IMT-2000 Network, 1998.5
3. 3GPP home page (http://www.3gpp.org/).
4. 3GPP2 home page (http://www.3gpp2.org/).
5. Umehira M, Aikawa S, Matsumoto Y, Nakura M, Kobayashi T. Multimedia mobile access trends. 2000 IEICE General Conference, TB-2-2, p 698–699, 1999.
6. ETSI TS 101.475 (v.1.1.1) Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer, 2000.

7. 3GPP TR 23.821 (v.1.0.0) Architecture Principles for Release 2000.
8. 3GPP2 Completes cdma2000 specification. 3GPP2 news release, April 18, 2000 (http://208.45.131.70/docs/news/2000-0.4% 20release.PDF).
9. Combined GSM and Mobile IP Mobility Handling in UMTS IP CN (3G TR 23.923 version 1.2.0).
10. Perkins C. IETF RFC 2002 IP Mobility Support, 1996.
11. Matsumoto M, Ihara T, Ohnishi H, Takagi Y. Investigations of mobile data network architectures from a mobility viewpoint. 2000 IEICE General Conference, B-7-166, p 261, 2000.
12. Perkins C. Route optimization in mobile IP. IETF draft · ietf · mobileip · optim-09.txt.
13. Dommety G. Local and indirect registration for anchoring handoffs. IETF draft · dommety · mobileip · anchor · handoff-0.1txt.
14. El Malki K. Fast handoffs in Mobile IPv4. IETF draft elmalki · mobileip · fast · handoffs-03.txt.
15. Sawada H, Hirata S, Miki M, Tani N. Network control methods and signaling methods in IMT-2000. NTT DoCoMo Tech J 1998;6:19–23.
16. Ohnishi H, Ihara T, Takagi Y. Proposal for route optimization methods in mobile IP networks. Trans IEICE Commun SSE99-123, 7-12, 1999.
17. Ohnishi H, Ihara T, Matsumoto M, Takagi Y. A proposal for a route optimization method in mobile IP networks. 2000 IEICE General Conference, B-7-166, p 259, 2000.
18. Tan WL, Chuah HT, Chuah MC. Mobile-IP: Hierarchical mobility management. APSITT '99.

APPENDIX

1. Mean Route Length for Each Case without MBG Functions

Case 1: The mobile terminal can be at either of the following two positions.

(i) Visiting a Foreign Agent under its own Home Agent

(ii) Visiting a Foreign Agent under a Home Agent other than its own Home Agent

The route length from the border gateway to the mobile terminal's visited Foreign Agent will be $b + c$ for (i) and $a + b + c$ for (ii). Also, the probability of each situation is $1/n$ for (i) and $(n - 1)n$ for (ii). Therefore, the mean route length will be as follows:

$$[1/n] \times (b + c) + [(n - 1)/n] \times (a + b + c)$$
$$= [(n - 1)/n] \times a + b + c$$

Case 2: Like case 1, the calculations are divided according to the mobile terminal's visiting position.

(i) Visiting a Foreign Agent under its own Home Agent

(ii-1) Visiting a Foreign Agent under a Home Agent other than its own Home Agent, and the two Home Agents are connected to the same first-layer node

(ii-2) Visiting a Foreign Agent under a Home Agent other than its own Home Agent, and the two Home Agents are connected to different first-layer nodes

The route length from the border gateway to the mobile terminal's visited Foreign Agent will be $b + c$ for (i), $3 \times b + c$ for (ii-1), and $a + 3 \times b + c$ for (ii-2). Also, the probability of each situation is $1/(n \times m)$ for (i), $(m - 1)/(n \times m)$ for (ii-1), and $(n - 1)/n$ for (ii-2). Therefore, the mean route length will be as follows:

$$[1/(n \times m)] \times (b + c) + [(m - 1)/(n \times m)] \times (3 \times b + c) + [(n - 1)/n] \times (a + 3 \times b + c)$$
$$= [(n - 1)/n] \times a + b + c + 2 \times [1 - 1/(n \times m)] \times b$$

2. Frequencies with which a Position Is Registered with a Home Agent, Mobile IP Border Gateway, and Foreign Agent

If the diameter (distance between opposite sides) of the FA cover area approximated by the regular hexagon model with area A_{fa} is represented by D, then D will be $(2\sqrt{3}/3 * A_{fa})^{0.5}$. Also, a mobile terminal is assumed to be moving with mean velocity V along the solid line joining the centers of various areas, regardless of whether or not it is communicating.

The frequency (R_{ha}) with which a position is registered with a single Home Agent per unit time can be obtained by using the following equation in which the quotient of the distance moved V divided by the area diameter D multiplied by the number of mobile terminals M_{ha} is divided by the number of Home Agents in the network N_{ha}:

$$R_{ha} = M_{ha} \times V/D$$
$$= M_{ha} \times V/(2\sqrt{3}/3)^{0.5}/\sqrt{} A_{fa}/N_{ha}$$

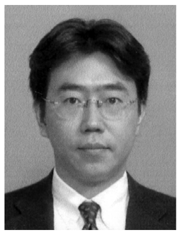
Next, the frequency (R_{fa}) with which a position is registered with a single Foreign Agent is obtained by dividing the frequency with which a position is registered with a single Home Agent (R_{ha}) by the number of Foreign Agents under a Home Agent (N_{fa}). Since N_{fa} here is obtained according to $N_{fa} = A/N_{ha}/A_{fa}$ from the total area A, R_{fa} will be as follows:

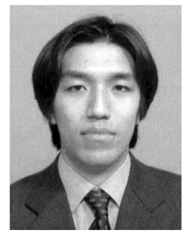
$$R_{fa} = R_{ha}/N_{fa}$$
$$= M_{ha} \times V/(2\sqrt{3}/3)^{0.5} * \sqrt{\ } A_{fa}/A$$

Finally, since a position is only registered with the Mobile IP Border Gateway when a mobile terminal is communicating, the frequency (R_{mbg}) with which a position is registered with the Mobile IP Border Gateway per unit time is obtained according to the following equation by using the probability that a mobile terminal is communicating p and the number of Home Agents in the network N_{ha}:

$$R_{mbg} = N_{ha} \times R_{ha} \times p$$
$$= M_{ha} \times V/D$$
$$= M_{ha} \times V/(2\sqrt{3}/3)^{0.5}/\sqrt{\ } A_{fa} \times p$$

AUTHORS (from left to right)





Yasushi Takagi (member) graduated from the Department of Industrial and Mechanical Engineering of Osaka University in 1985 and completed the M.S. program in 1987. He then joined Nippon Telegraph and Telephone Corporation, where he has been engaged in research and development of ATM networks and system configuration methods and research in IP Network architecture and Next-Generation Mobile Networks architecture. He is now a senior research engineer at NTT Network Service Systems Laboratories. He received an Academic Encouragement Award from IEICE in 1993, and a Switching System Research Award from IEICE in 1999.

Takeshi Ihara (member) graduated from the Department of Mechanical Engineering of Tokyo University in 1990 and completed the M.S. program in 1992. He then joined Nippon Telegraph and Telephone Corporation, where he has been engaged in research and standardization of ATM signaling methods, mobile communication signaling methods, and mobile management methods in Next-Generation Mobile Communications. He is now a senior research engineer at NTT DoCoMo Network Laboratories. He received a Switching System Research Award from IEICE in 1999.

Hiroyuki Ohnishi (member) graduated from the Department of Mechanical Engineering of Waseda University in 1996 and completed the M.S. program in 1998. He then joined Nippon Telegraph and Telephone Corporation, where he has been engaged in research on Next-Generation Mobile IP Network architecture. He is now a research engineer at NTT Network Service Systems Laboratories. He received a Switching System Research Award from IEICE in 1999.