이동 호스트를 위한 가상 셀 시스템의 성능 분석

임 경 식[†]

요 약

본 논문은 이동 컴퓨터 통신 환경에서 IP(Internet Protocol) 데이터그램을 전송하기 위한 가상 셀 시스템(virtual cell system)의 성능 분석을 다룬다. 하나의 가상 셀은 이웃한 다수의 물리적 셀(physical cell)들의 집합으로서, 원격 브리지(remote bridge)로 구현된 기지국(base station)들을 멀티캐스트 기능을 갖는 고속 데이터그램 패킷 데이터 망으로 연결하여 구성된다. 가상 셀 시스템에서의 호스트 이동성은 기지국들 사이에 분산되어 있는 계층적 위치정보를 기반으로 동작하는 데이터 링크 계층의 가상 셀 프로토콜(Virtual Cell Protocol)을 통하여 지원된다[1]. 이러한 가상 셀 시스템은 물리적 셀들 사이에 임의의 호스트 이동성 패턴과 데이터 전송 패턴이 주어진 경우에 전체 시스템의 통신 비용을 최소화할 수 있도록 논리적으로 유연한 가상 셀 시스템의 구축을 가능하게 한다[2]. 본 논문에서는 가상 셀 시스템의 성능 모델로서 BCMP 개방 복합 클래스 대기 행렬 네트워크(BCMP open multiple class queueing network)를 채택하고, 물리적 셀들 사이의 호스트 이동성 패턴과 데이터 전송 패턴에 대하여 임의의 토폴로지와 최적화된 토폴로지로 구축된 가상 셀 시스템의 성능을 비교 분석한다. 특히 이동 호스트 수, 이동 속도, 그리고 데이터 전송 양과 같은 다양한 시스템 파라메타를 변화시키면서 이에 따라 생성되는 데이터 메시지, 핸드오프(handoff) 메시지, 그리고 주소 용해(address resolution) 메시지 각각에 대하여 망 구성요소의 이용도(utilization)와 시스템 처리 시간(system response time)을 비교 분석한다.

A Performance Analysis of the Virtual Cell System for Mobile Hosts

Kyung-Shik Lim[†]

ABSTRACT

In this paper, we analyze the performance of the virtual cell system[1] for the transmission of IP datagrams in mobile computer communications. A virtual cell consists of a group of physical cells whose base stations are implemented by remote bridges and interconnected via high-speed datagram packet-switched networks. Host mobility is supported at the data link layer using the distributed hierarchical location information of mobile hosts. Given mobility and communication patterns among physical cells, the problem of deploying virtual cells is equivalent to the optimization problem of finding a cover of disjoint clusters of physical cells. The objective is to minimize the total communication cost for the entire system where intercluster communication is more expensive than intracluster communication[2]. Once an optimal partition of disjoint clusters is obtained, we deploy the virtual cell system according to the topology of the optimal partition such that each virtual cell corresponds to a cluster. To analyze the performance of the virtual cell system, we adopt a BCMP open multiple class queueing network model. In addition to mobility and communication patterns among physical cells, the topology of the virtual cell system is used to determine service transition probabilities of the queueing network model. With various system parameters, we conduct interesting sensitivity analyses to determine network design tradeoffs. The first application of the proposed model is to determine an adequate network

---

† 정 회 원 : 경북대학교 컴퓨터과학과 교수
논문접수 : 1997년 8월 27일, 심사완료 : 1998년 1월 19일

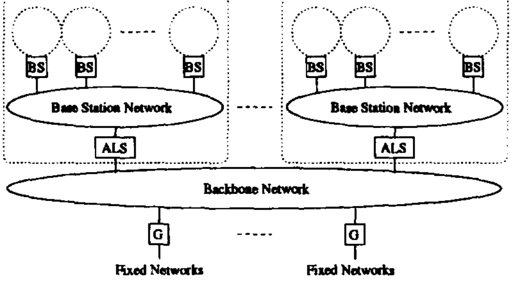
bandwidth for base station networking such that the networks would not become a bottleneck. We also evaluate the network utilization and system response time due to various types of messages. For instance, when the mobile hosts begin moving fast, the migration rate will be increased. This implies more handoff messages and more forwarding operations. The network traffic should be increased accordingly. The results of the performance analysis provide a good evidence to demonstrate the system efficiency under different assumptions of mobility and communication patterns.

1. Introduction

As computers become more powerful and portable with the appearance of high-speed wireless interfaces, there has been an increasing demand on the provision of mobile computer communications in TCP/IP environments. As an effort to meet this demand, we have proposed a virtual cell approach to the transmission of IP datagrams for mobile computer communications[1]. As depicted in Figure 1, a virtual cell consists of a group of physical cells whose base stations are implemented by remote bridges and interconnected via high-speed datagram packet-switched networks. Examples are Asynchronous Transfer Mode (ATM) networks and Switched Multi-megabit Data Services(SMDS) networks. Host mobility is supported at the data link layer using the distributed hierarchical location information of mobile hosts. It eliminates the necessity of IP-level mobile host protocols and achieves a logically flexible coverage area according to mobility and communication patterns among physical cells.

Mobility and data traffic patterns among physical cells can be represented by the *move* and *find* frequencies among base stations, respectively, because base stations are served as the interfaces between mobile hosts and base station networks. In other words, base stations are regarded as traffic sources and destinations from the prospective of the virtual cell system. There are three types of messages entering or leaving the virtual cell system via base stations: the *handoff*, *data*, and *address resolution* messages. The handoff messages are generated due to move operations and the data and address resolution messages are due to find operations.

Given the move and find frequencies among n base stations, the problem of deploying m virtual cells is equivalent to the optimization problem of finding a cover of m disjoint clusters of contiguous base stations, so as to minimize the total communication cost for the entire system where intercluster communication is more expensive than intracluster communication for each type of operation. Note that the optimization problem differs from general graph partitioning problems in that it additionally considers underlying topology constraints, such as the linear arrangement of n base stations in highway cellular systems and the hexagonal mesh arrangement of n base stations in hexagonal cellular systems. In [2], we have presented an optimal partitioning algorithm of $O(mn^2)$ by dynamic programming for a linear array of n base stations and several heuristics for multiway partitioning of a hexagonal mesh of n base stations.



(Fig. 1) The Virtual Cell System Architecture

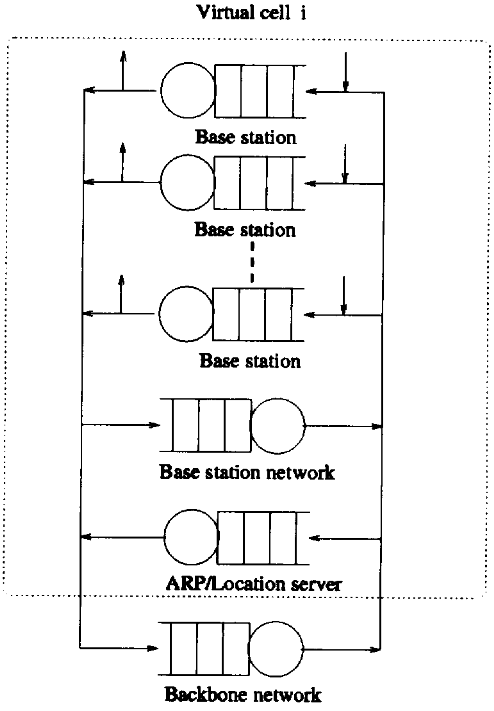
Once an optimal partition of m disjoint clusters is obtained by the algorithms in [2], we can deploy a virtual cell system as shown in Figure 1. Each cluster corresponds to a virtual cell. Virtual cell i, where $1 \leq i \leq m$, is implemented by interconnecting the base stations of the ith cluster and an ARP/Location server by a base station network. These m virtual cells are interconnected by the backbone network.

In this paper, we deal with the performance analysis of the virtual cell system deployed according to an optimal partition. The virtual cell system is modeled as an open multiple class queueing network in Section 2. The move and find frequencies in conjunction with the topology of the virtual cell system are used to determine service transition probabilities in the queueing network model as well as the arrival rate for each type of messages. The performance measures and evaluations are described in Section 3 and 4, and we conclude in Section 5.

2. Performance Model

We adopt a BCMP open multiple class queueing network to model the virtual cell system, as depicted in Figure 2. A virtual cell is modeled as a number of base station nodes, an ARP/Location server node, and a base station network node which captures traffic characteristics among physical cells in the same virtual cell. But, in order to capture traffic characteristics between virtual cells, a separate service node is used to model the backbone network. Messages from mobile hosts enter and leave the network model, only going through base station nodes in the virtual cell system.

The base station nodes of virtual cell i are sequentially indexed as $1, 2, \ldots, n_i$ in an arbitrary order, where n_i is the number of base stations in virtual cell i such that $\sum_{i=1}^{m} n_i = n$. Denote ij as



(Fig. 2) The Performance Model of the Virtual Cell System

service node j in virtual cell i. The network model with $N = n + 2m + 1$ nodes is defined as follows:

1. A base station node is labeled ij, where $1 \leq i \leq m$ and $1 \leq j \leq n_i$, and modeled as an FCFS type of service station with a fixed service rate μ_{ijr} for message class r. The messages in the queue are transmitted to its base station network node or leave the network model.

2. A base station network node is labeled iB, where $1 \leq i \leq m$, and modeled as a PS type of service station with a fixed service rate μ_{iBr} for message class r. The messages in the queue are transmitted to its base station nodes or ARP/Location server node.

3. An ARP/Location server node is labeled iS, where $1 \leq i \leq m$, and modeled as an FCFS type of

service station with a fixed service rate μ_{iSr} for message class *r*. The messages in the queue are transmitted to its base station network nodes or the backbone network node.

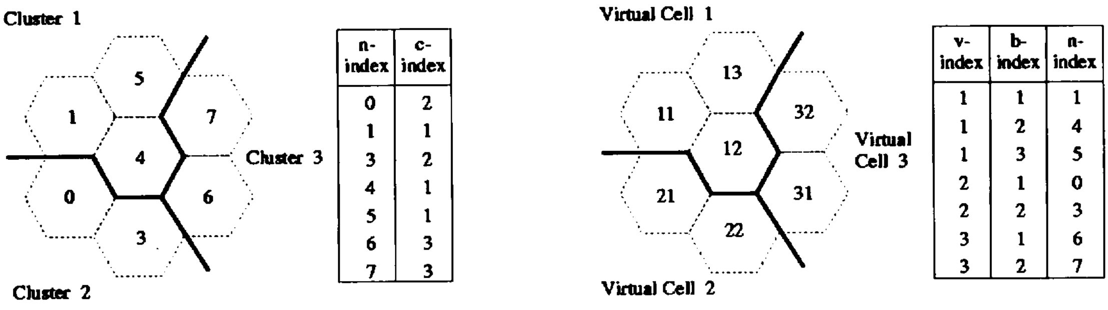
4. The backbone network node is labeled *B1* and modeled as a PS type of service station with a fixed service rate μ_{B1r} for message class *r*. The messages in the queue are transmitted to ARP/ Location server nodes of virtual cells.

The analysis is based on the following assumptions:

1. Given a hexagonal mesh of base stations of size *s*, it is known that the number of base stations are $n=3s^2-3s+1$ and the number of columns in each three directions is $d=2s-1$. From left to right, each column is indexed as $0,1,2,\dots,d-1$. Then the nodes of every column *i* are sequentially labeled $i\times d, i\times d+1, i\times d+2,\dots,$ from bottom to top. Denote *nindex* and *cindex* as the node and cluster indices, respectively. An optimal partition produced by the algorithms[2] can be represented by an index pair (*nindex, cindex*) for every base station. Figure 3(a) depicts an optimal partition for a hexagonal mesh of size 2. Note that the node index reflects the underlying topology of the hexagonal mesh of base stations. Thus, given the node index of a base station, its neighboring base stations can be directly identified by the labeling scheme of the algorithms.

In addition to the node index, the network model of the virtual cell system also requires base stations to be logically indexed because a base station in a virtual cell can change into another virtual cell according to traffic patterns. Denote *vindex* and *bindex* as the virtual cell and base station indices, respectively. The virtual cell index is directly mapped to the cluster index and the base station index is logically assigned so that the base stations in a virtual cell is sequentially labeled $1,2,\dots,$ in an arbitrary order. Figure 3(b) depicts the labeling of the network model which corresponds to an optimal partition in Figure 3(a). Thus, given base station *xy* in the network model, where *x* and *y* are the virtual cell and base station indices, respectively, its neighboring base stations are directly identified by the node index. Let *Adj*(*xy*) be a set of base stations adjacent to *xy*. From *Adj*(*xy*) we can identify which neighboring base station belongs to which virtual cell.

2. We use a flow-based mobility model[5] which assumes that mobile hosts are uniformly distributed in the area of a physical cell and the travel direction



n-index	c-index
0	2
1	1
3	2
4	1
5	1
6	3
7	3

v-index	b-index	n-index
1	1	1
1	2	4
1	3	5
2	1	0
2	2	3
3	1	6
3	2	7

(a) The labeling scheme in the optimization algorithms

(b) The corresponding labeling in the network model

(Fig. 3) Labeling the Topology of the Virtual Cell system of each mobile host with respect to the border is uniformly distributed. If we define ρ_{xy} to be the density of the mobile hosts per km^2 at the area of base station *xy*, v_{xy} to be the average speed in *km*/sec of a mobile host at the area of base station *xy*, and L to be the length of the perimeter of the area of a physical cell, then the average number of mobile hosts per *sec* leaving base station *xy* is given by $M_{xy}=\rho_{xy}v_{xy}L/\pi$. Denote $f_m(xy, x'y')$ as the move frequency from base station *xy* to one of its neighboring base station $x'y' \in Adj(xy)$. Then $f_m(xy, x'y') = M_{xy}/|Adj(xy)|$, where $|Adj(xy)|$ is the number of base stations adjacent to *xy* and set to 6 in the hexagonal arrangement of *n* base stations.

3. If we define b to be the data rate in *bits/sec* of a wireless channel, l to be the average message length in bits, and E to be the in-call probability in Erlangs for a mobile host, then the non-blocking data arrival rate in *messages/sec* at base station *xy* is given by $F_{xy}=\pi R^2\rho_{xy}Eb/l$, where R is the radius in *km* of a physical cell. Denote $f_f(xy, x'y')$ as the find frequency from base station *xy* to base station $x'y'$. Then $f_f(xy, x'y')$ is a fraction of F_{xy}, such that $\sum_{x'y'} f_f(xy, x'y') = F_{xy}$.

4. The address resolution frequency is derived from the find frequency. If we define F_{x^*y} to be the arrival rate in *messages/sec* of data messages at base station *xy*, which are destined to base stations in the same virtual cell *x*, *c* to be the average number of messages over a conversation from the source mobile host to the destination mobile host, and h_m to be the miss ratio of the address resolution table of a mobile host, then the arrival rate in *messages/sec* of the address resolution messages at base station *xy* is given by $A_{xy}=F_{x^*y}h_m/c$, where $F_{x^*y}=\sum_{xy} f_f(xy, x'y')$. Denote $f_a(xy, xy')$ as the address resolution frequency from base station *xy* to base station *xy′*. Then $f_a(xy, xy')$ is a fraction of A_{xy}, such that $\sum_{xy} f_a(xy, xy') = A_{xy}$.

It should be noted that the move frequency not only affects the arrival rate of the handoff messages but also the forwarding rate of the data and address resolution reply messages. For example, consider the move frequency $f_m(xy, xy')$. When a mobile host at base station *xy* moves into base station *xy′*, it initiates a handoff request message to *xy′*. But a data message destined to the mobile host may be forwarded to base station *xy′* at base station *xy* during the handoff process. In the same way, the address resolution reply message for the mobile host from the ARP/Location server should be also forwarded to base station *xy′* at base station *xy*.

2.1 Multiple Class Traffic Model

There are three classes of messages entering the network: the handoff request message, the data message, and the address resolution request message. As each class of message traverses through the network, it not only requires different service requirements and different routing behavior, but changes its class. For example, the base station received a handoff request message from a mobile host sends the message to the previous base station of the mobile host. Then the previous base station multicasts the message in its virtual cell to update the distributed location information of the mobile host, and at the same time it sends a handoff response message to the new base station from which the mobile host receives a handoff confirmation message. Thus, as a message of the *handoff request class* progresses through the network, it is changed into a message of the *multicast class* and next into a message of the *handoff response class*.

After the handoff completes, a data message destined to the mobile host will be directly delivered to the new base station. However, during the handoff the data message will be first delivered to the previous base station, which in turn forwards it to the new base station. Thus, as a message of the *data class* progresses through the network, it might be changed into a forwarding message which may be involved in a virtual cell or between virtual cells. The forwarding message involved in a virtual cell is referred to as the *intra-forwarding class* message, while that involved between virtual cells is referred to as the *inter-forwarding class* message. In the same way, as a message of the *address resolution request class* progresses through the network, it is changed into a message of the *address resolution reply class* at ARP/Location servers and possibly the *address resolution reply forwarding class* at base stations.

Hence, the set of message classes in the network can be partitioned into three subsets which contain the message classes related to handoff, data, and address resolution, respectively. Denote z = 1, 2, and 3 as the subsets of the message classes related to handoff, data, and address resolution, respectively. Even though any class r in a subset z can visit the entire set of nodes N, we cannot define a routing chain for each subset z because a class r message in a subset z may require different routing behavior due to the topology of the virtual cell system.

For example, consider a handoff request message from a mobile host which moves from an adjacent base station into base station 12 in the example of Figure 3(b). If the mobile host is from base station 11, the message will be directly delivered to its previous base station 11 via base station network 1*B*. However, if the mobile host is from base station 32, the message will be delivered to its previous base station 32, going through base station network 1*B*, ARP/Location server 1*S*, backbone network *B*1, ARP/Location server 3*S*, and finally base station network 3*B*. Thus, in addition to the message class, the topology of the virtual cell system should be considered to determine the routing behavior of messages in the network.

In order to incorporate the topology of the virtual cell system into the queueing network model, it is necessary to identify which base station generates the message classes for each subset z. Thus, we define a routing chain for the message classes of each subset z generated by each base station. Then there are 3n routing chains in the network, denoted as $E_{x,y,z}$, where $1 \le x \le m$, $1 \le y \le n_x$, and z = 1, 2, and 3. For each routing chain $E_{x,y,z}$, the service transition probabilities are defined by the set $\{p_{ijr,kls}\}$, which describes the probability that a class r message at node ij goes next to node kl as a class s message, where $r, s \in z$ and $ij, kl \in N$.

2.2 Arrival Process

We assume a Poisson state-independent arrival stream for each routing chain. Denote $\lambda_{x,y,z}$ as the arrival rate corresponding to a routing chain $E_{x,y,z}$. Then the arrival rate $\lambda_{x,y,z}$ is determined by the move, find, and address frequencies among n base stations.

1. A mobile host migrating from base station $x'y' \in Adj(xy)$ to base station xy initiates a handoff request message to base station xy. Thus, the arrival rate for the handoff request class of messages at base station xy can be represented by $\lambda_{x,y,1} = \sum_{x'y' \in Adj(xy)} f_m(x'y', xy)$.

2. A data message transmitted by a mobile host at base station xy needs a find operation to locate base station $x'y'$ to which the destination mobile host belongs. Thus, the arrival rate for the data class of messages at base station xy can be represented by $\lambda_{x,y,2} = \sum_{x'y'} f_f(xy, x'y')$.

3. The source mobile host performs address resolution operations only when it resides in its

native virtual cell and the network address of the destination mobile host indicates the same native virtual cell. Thus, the arrival rate for the address resolution request class of messages at base station *xy* can be represented by $\lambda_{x,y,3} = \sum_{x'y', x'=x} f_a(xy, x'y')$.

2.3 Service Transition Matrix

A service transition matrix $P_{x,y,z} = [p_{ijr,kls}]$ is to be defined for each routing chain $E_{x,y,z}$. To determine service transition probabilities, it is necessary to consider not only the move, find, and address resolution frequencies, but also the topology of the virtual cell system. For the handoff message classes, base station *xy* may send and receive a handoff request class message and a handoff response class message to and from an adjacent base stations $st \in Adj(xy)$ on a different route in the network, depending on whether base station *st* is located in the same virtual cell or a different virtual cell. For the data message classes, in the same way, base station *xy* may deliver a data class message to base station $st \in Adj(xy)$ on a different route, depending on whether base station *st* is located in the same virtual cell or a different virtual cell. If the data class message arriving at base station *st* is to be forwarded to an adjacent base station $s't' \in Adj(st)$, the forwarding message may also take a different route in the network model depending on whether *s't'* and *st* are in the same virtual cell or different virtual cells.

(1) Handoff Message Classes

Consider the handoff request, multicast, and handoff response message classes originated from base station *xy*, i.e, a routing chain $E_{x,y,1}$. Define $\delta_{st} = f_m(st, xy)/\lambda_{x,y,1}$ to be a probability that a mobile host at base station $st \in Adj(xy)$ moves into base station *xy*. Denote *xy'* as a base station adjacent to *xy* in the same virtual cell *x* and *x'y'* as a base station adjacent to *xy* in a different virtual cell *x'*. Then we can compute the following probabilities from the move frequency:

- The probability that a mobile host at base station *xy'* moves into base station *xy* is given by $\delta_{xy'} = f_m(xy', xy)/\lambda_{x,y,1}$.

- The probability that a mobile host in virtual cell *x* moves into base station *xy* is given by $\delta_x = \sum_{xy' \in Adj(xy)} \delta_{xy'}$.

- The probability that a mobile host at base station *x'y'* moves into base station *xy* is given by $\delta_{x'y'} = f_m(x'y', xy)/\lambda_{x,y,1}$.

- The probability that a mobile host in a different virtual cell *x'* moves into base station *xy* is given by $\delta_{x'} = \sum_{x'y' \in Adj(xy)} \delta_{x'y'}$.

The service transition matrix $P_{x,y,1}$ for the handoff message classes originated from base station *xy* is given as follows:

- For the handoff message classes when a mobile host moves from base station $xy' \in Adj(xy)$ into base station *xy*, service transition probabilities are as follows:

  - $p_{xy1;xB1} = 1.0$ and $p_{xB1;xy'1} = \delta_{xy'}$ for the handoff request class,
  - $p_{xy'1;xB2} = 1.0$ and $p_{xB2;xy'2} = \delta_{xy'}/\delta_x$ for the multicast class, and
  - $p_{xy'2;xB3} = 1.0$ and $p_{xB3;xy8} = 1.0$ for the handoff response class.

- For the handoff message classes when a mobile host moves from base station $x'y' \in Adj(xy)$ into base station *xy*, service transition probabilities are as follows:

  - $p_{xy1;xB1} = 1.0$, $p_{xB1;xS1} = 1 - \delta_x$, $p_{xS1;B11} = 1.0$,

$p_{B1;x'S1} = \delta_{x'} / \sum_{x' \in Adj(xy)} \delta_{x'}$, $p_{x'S1;x'B1} = 1.0$, and $p_{x'B1;x'y'1} = \delta_{x'y'} / \delta_{x'}$ for the handoff request class,

- $p_{x'y'1;x'B2} = 1.0$ and $p_{x'B2;x'y'2} = \delta_{x'y'} / \delta_{x'}$ for the multicast class, and
- $p_{x'y'2;x'B3} = 1.0$, $p_{x'B3;x'S3} = 1.0$, $p_{x'S3;B13} = 1.0$, $p_{B13;xS3} = 1.0$, $p_{xS3;xB3} = 1.0$, and $p_{xB3;xy3} = 1.0$ for the handoff response class.

(2) Data Message Classes

Consider the data, intra-forwarding, and inter-forwarding message classes originated from base station *xy*, i.e, a routing chain $E_{x,y,2}$. Define $a_{st} = f_f(xy, st) / \lambda_{x,y,2}$ to be a probability that a mobile host at base station *xy* sends a data message to base station *st*. Denote *xy'* as a base station in the same virtual cell *x* and *x'y'* as a base station in a different virtual cell *x'*. Then we can calculate the following probabilities for the data message class from the find frequency:

- The probability that a data message from base station *xy* is destined to base station *xy'* is given by $a_{xy'} = f_f(xy, xy') / \lambda_{x,y,2}$.

- The probability that a data message from base station *xy* is destined to the base stations in virtual cell *x* is given by $a_x = \sum_{xy'} a_{xy'}$.

- The probability that a data message from base station *xy* is destined to a base station *x'y'* is given by $a_{x'y'} = f_f(xy, x'y') / \lambda_{x,y,2}$.

- The probability that a data message from base station *xy* is destined to the base stations in a different virtual cell *x'* is given by $a_{x'} = \sum_{x'y'} a_{x'y'}$.

Let us now consider forwarding message classes in the network. Consider the case that the source mobile host at base station *xy* tries to send a data message to the destination mobile host which is moving from base station *st* into base station $s't' \in Adj(st)$. If the message transmission occurs after the handoff completes, the message should be directly sent to base station *s't'*. If the message transmission occurs before the handoff completes, however, the message will be sent to the previous base station *st* where it will be forwarded to the new base station *s't'*. Define τ_{st} to be the average time in second for a mobile host to stay at base station *st* before it leaves. Then $\tau_{st} = H_{st} / M_{st}$, where H_{st} is the total number of mobile hosts at base station *st* and M_{st} is the average rate of mobile hosts per second moving out of base station *st*. Given the average handoff time τ_h, the probability that a message destined to base station *st* is forwarded to base station $s't' \in Adj(st)$ is given by $\sigma_{st} = \tau_h / \tau_{st} |Adj(st)|$. This should be a valid estimate if an MH migrates independent of when it receives data messages.

Define $q^{out}_{st,s't'}$ to be the forwarding frequency going out of base station *st* to base station $s't' \in Adj(st)$, q^{out}_{st} to be the forwarding frequency going out of base station *st* to all neighboring base stations in $Adj(st)$, and q^{out}_{s} to be the forwarding frequency going out of all base stations in virtual cell *s* to different virtual cells. Then $q^{out}_{st,s't'} = f_f(xy, st)\sigma_{st}$, $q^{out}_{st} = \sum_{s't' \in Adj(st)} q^{out}_{st,s't'}$, and $q^{out}_{s} = \sum_{st} \sum_{s't' \in Adj(st), s' \neq s} q^{out}_{st,s't'}$.

Define $q^{in}_{st,s't'}$ to be the forwarding frequency going out of base station $s't' \in Adj(st)$ into base station *st*, q^{in}_{st} to be the forwarding frequency going out of all neighboring base stations in the same virtual cell *s* into base station *st*, q^{in}_{st*} to be the forwarding frequency going out of all neighboring base stations in different virtual cells into base station *st*, and q^{in}_{s} to be the forwarding frequency going out of all neighboring base stations in different virtual cells into all base stations in virtual cell *s*. Then $q^{in}_{st,s't'} = f_f(xy,$

$s't')\sigma_{s't'}$, $q^{\text{in}}_{st} = \sum_{s't' \in Adj(st), s'=s} q^{\text{in}}_{st,s't'}$, $q^{\text{in}}_{st*} = \sum_{s't' \in Adj(st), s' \neq s} q^{\text{in}}_{st,s't'}$, and $q^{\text{in}}_{s} = \sum_{st} q^{\text{in}}_{st*}$.

From the above forwarding frequencies, we can compute the following probabilities used to determine service transition probabilities for the forwarding message classes:

- The probability that a forwarded message at base station *st* is routed to a different virtual cell is given by $\beta^{\text{inter}}_{s} = q^{out}_{s} / \sum_{st} q^{out}_{st}$.

- The probability that a forwarded message at base station *st* is routed to base station *st′* in the same virtual cell *s* is given by $\beta^{intra}_{st'} = q^{\text{in}}_{st'} / \sum_{st'} q^{\text{in}}_{st'}$.

- The probability that a forwarded message at base station *st* is routed to the base stations in a different virtual cell *s′* is given by $\gamma^{\text{inter}}_{s'} = q^{\text{in}}_{s'} / \sum_{s'} q^{\text{in}}_{s'}$.

- The probability that a forwarded message at base station *st* is routed to base station *s′t′* is given by $\gamma_{s't'} = q^{\text{in}}_{s't'*} / \sum_{s't'} q^{\text{in}}_{s't'*}$.

The service transition matrix $P_{x,y,2}$ for the data message class generated at base station *xy* is given as follows:

- When a data class message is destined to base station *xy′* in the same virtual cell *x*, service transition probabilities are as follows:

  - $p_{xy1;xB1} = 1.0$ and $p_{xB1;xy'1} = \alpha_{xy'} / \alpha_{x}$ for the data class,
  - $p_{xy'1;xB2} = \sigma_{xy'} \mid Adj(xy') \mid$, $p_{xB2;xy'2} = (1 - \beta^{\text{inter}}_{x}) \beta^{intra}_{xy'}$, and $p_{xB2;xS2} = \beta^{\text{inter}}_{x}$ for the intra-forwarding class, and
  - $p_{xS2;B13} = 1.0$, $p_{B13;x'S3} = \gamma^{\text{inter}}_{x'}$, $p_{x'S3;x'B3} = 1.0$, and $p_{x'B3;x'y'3} = \gamma_{x'y'}$ for the inter-forwarding class.

- When a data class message is destined to base station *x′y′* in a different virtual cell *x′*, service transition probabilities are as follows:

  - $p_{xy1;xB1} = 1.0$, $p_{xB1;xS1} = 1 - \alpha_{x}$, $p_{xS1;B11} = 1.0$, $p_{B11;x'S1} = \alpha_{x'} / (1 - \alpha_{x})$, $p_{x'S1;x'B1} = 1.0$, and $p_{x'B1;x'y'1} = \alpha_{x'y'} / \alpha_{x'}$ for the data class,
  - $p_{x'y'1;x'B2} = \sigma_{x'y'} | Adj(x'y') |$, $p_{x'B2;x'y'2} = (1 - \beta^{\text{inter}}_{x'}) \beta^{intra}_{x'y'}$, and $p_{x'B2;x'S2} = \beta^{\text{inter}}_{x'}$ for the intra-forwarding class, and
  - $p_{x'S2;B13} = 1.0$, $p_{B13;x'S3} = \gamma^{\text{inter}}_{x'}$, $p_{x'S3;x'B3} = 1.0$, and $p_{x'B3;x'y'3} = \gamma_{x'y'}$ for the inter-forwarding class.

(3) Address Resolution Message Classes

Consider the address resolution request, address resolution reply, and address resolution reply forwarding message classes originated from base station *xy*, i.e., a routing chain $E_{x,y,3}$. Denote h_{BS} as the hit ratio of the cache table at a base station for a mobile host. The service transition matrix $P_{x,y,3}$ for address resolution message classes originated from base station *xy* is given as follows:

- $p_{xy1;xB1} = 1 - h_{BS}$ and $p_{xB1;xS1} = 1.0$ for the address resolution request class,
- $p_{xS1;xB2} = 1.0$ and $p_{xB2;xy2} = 1.0$ for the address resolution reply class, and
- $p_{xy2;xB3} = \sigma_{xy} | Adj(xy) |$ and $p_{xB3;xy'3} = 1 / | Adj(xy) |$ for the address resolution reply forwarding class.

2.4 Traffic Equations

Suppose that e_{ijr} is the average throughput of class *r* messages through node *ij* in a routing chain $E_{x,y,z}$. Then for node *ij* and class *r*, $(ij, r) \in E_{x,y,z}$, we can write the following traffic equations:

$$\sum_{(kl,s) \in E_{x,y,z}} e_{kls} p_{kls;ijr} + p_{0;ijr} = e_{ijr},$$

where $p_{0;ijr}$ is the probability that an external arrival is for node *ij* and class *r* and determined by the rate of external arrivals of class r messages to node ij. Since $p_{0,ijr}>0$ for some $(ij,r)\in E_{x,y,z}$, $E_{x,y,z}$ is open and so the traffic equations have a unique solution for $\{e_{ijr}\}$.

3. Performance Measures

The BCMP theorem states that multiple class queueing networks with the FCFS, PS, IS, and LCFS- PR types of nodes have a product form solution for the steady state joint probability distribution of the node states[3]. Since in an open network a message sees the network in the steady state when it leaves or enters the network, Little's result can be applied to any given class of messages at a node.

Node ij has a fixed service rate of μ_{ijr} and relative throughput of e_{ijr} for class r messages. The service time of FCFS nodes is independent of the class and so $1/\mu_{ijr}=1/\mu_{ij}$ for all classes r. The relative throughput e_{ijr} is the average number of times that a class r message visits node ij before leaving the network. Then the total service demand of a class r message at node ij is $D_{ijr}=e_{ijr}/\mu_{ijr}$.

Let η_r be the external arrival rate of each class r. The performance measures we can obtain from the described queueing network model include:

- The throughput of class r messages in the steady state is $T_r=\eta_r$.
- The utilization of node ij by class r messages is $U_{ijr}=\eta_r D_{ijr}$ by Little's result. Thus, the utilization of node ij is given by $U_{ij}=\sum_{r\in R}\eta_r D_{ijr}$, where R is the set of all message classes.
- The mean waiting time of a class r message at node ij for each visit is given by $W_{ijr}=(1/\mu_{ijr})/(1-U_{ij})$.
- The mean time that a class r message spends at node ij during its stay in the network, i.e., the mean residence time at node ij for class r is given by $Q_{ijr}=e_{ijr}W_{ijr}$.
- The mean time that a class r message spends in the network, i.e., the system response time for class r is given by $Q_r=\sum_{ij\in N}Q_{ijr}$.
- The mean number of class r messages at node ij is $L_{ijr}=\eta_r Q_{ijr}$ by Little's result.

4. Performance Evaluations

This section gives the results of computations from the performance measures of the queueing network model. By changing the density, velocity and in-call probability of mobile hosts among mobility and traffic parameters, we evaluate the utilization of various network components and the system response time due to various messages. With the same model parameters we also compare two different virtual cell systems: one deployed according to an initial partition and the other deployed according to an optimal partition with respect to the initial partition. The initial model parameters are listed in Table 1 and Table2.

The analysis assumes that the initial density of mobile hosts in each of n physical cells is randomly varied between 325 mh/km^2 and 650 mh/km^2 such that the average density becomes 500 mh/km^2. The initial velocity of mobile hosts in each of n physical cells is also randomly varied between 5 km/hr and 25 km/hr such that the average velocity becomes 15 km/hr. The initial density of each physical cell is increased with the same ratio at each step such that the average density is raised from 500 mh/km^2 to 2000 mh/km^2. After the average density increased up to 2000 mh/km^2, the average velocity is then increased from 15 km/hr up to 30 km/hr in the same way and then the in-call probability is raised from 0.5 up to 1.0. The base station network service rate μ_{iB} and the backbone network service rate μ_{Bl}

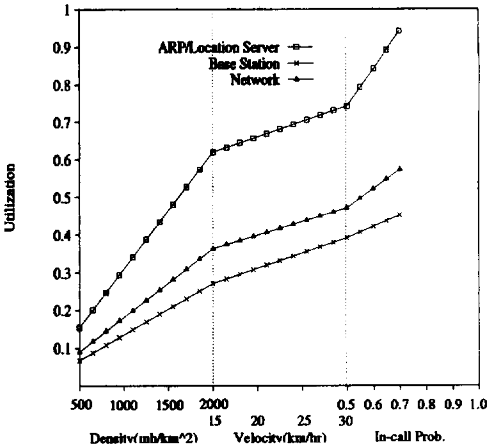
represent the service rate of the same physical transport network.

〈Table 1〉 Initial System Parameters

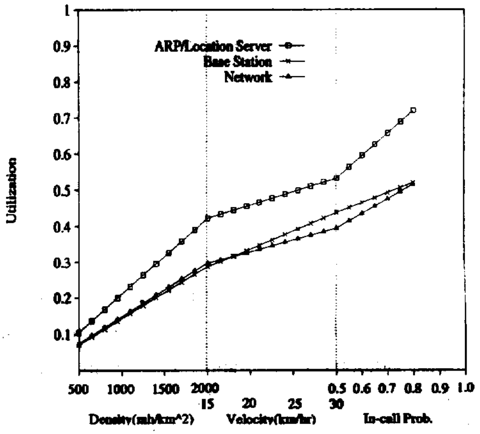
system parameters	initial value
the number of base stations(n)	19
the number of virtual cells(m)	3
base station network service rate(μ_{iB})	45Mbps
backbone network service rate(μ_{B1})	45Mbps
base station service rate(μ_{ij})	4.5Mbps
ARP/Location server service rate(μ_{iS})	4.5Mbps

〈Table 2〉 Initial Mobility and Traffic Parameters

mobility and traffic parameters	initial value
average density of mobile hosts at physical cell xy(ρ_{xy})	500 mh/ km^2
average velocity of mobile hosts at physical cell xy(v_{xy})	15 km/hr
cell radius(R)	1 km
data rate of a wireless channel(b)	2 Kbps
average message length(l)	4 Kbits
Erlangs per mobile host(E)	0.04
average number of messages over a conversation(c)	240
address resolution miss ratio at a mobile host(h_m)	0.5
address resolution hit ratio at a base station(h_{BS})	0.5
probability that a mobile host is powered on	0.5
average handoff time(τ_h)	1 sec



(Fig. 4) The Utilization of the Network Components in the Virtual Cell System When an Initial Partition Is Used



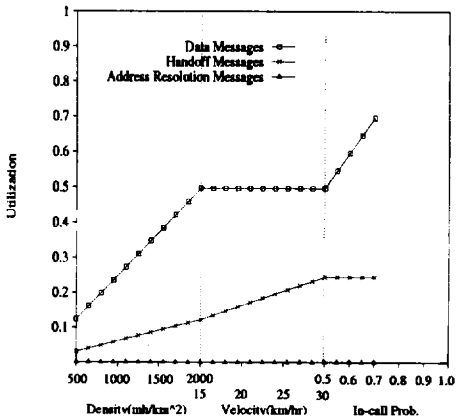
(Fig. 5) The Utilization of the Network Components in the Virtual Cell System When an Optimal Partition Is Used

Figure 4 and Figure 5 depict the utilization of the network components in the virtual cell system which is deployed according to an initial partition and an optimal partition, respectively. Since the same physical transport network is used for both base station networks and the backbone network to construct a virtual cell system, the network utilization of Figure 4 and Figure 5 represents the sum of all utilizations for three base station networks and one backbone network. The utilization of the ARP/Location server and base stations represents the average utilization of an ARP/Location server or a base station.

As depicted in Figure 4, the high utilization of the ARP/Location server for the initial partition implies that there is a large volume of mobility and data traffic between clusters before the optimization process. It is shown in Figure 5 that the inter-cluster traffic is significantly reduced after the optimization process. The less inter-cluster traffic means that the utilizations of the backbone network and base station networks are also reduced. This is due to the fact that for handoff or data transfer operations, intra-cluster traffic involves operations at only one base station network while inter-cluster traffic involves operations at two base station networks and one backbone network.

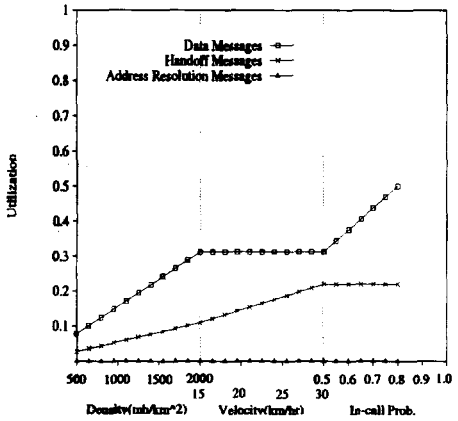
It is interesting to observe that the average utilization of a base station for the optimal partition is slightly higher than that for the initial partition and it is especially sensitive to the velocity of mobile hosts. This is due to the fact that the optimization process produces as large clusters as possible within the cluster size constraint in order to reduce the total communication cost for the entire system. Consider the impact of handoff multicast operations in terms of cluster size. The number of mobile hosts leaving a base station is determined by the density and velocity of mobile hosts in its physical cell, not the cluster size. For each leaving mobile host, the base station invokes a multicast operation to all other base stations in the same cluster in order to maintain the consistency of the distributed location information of the mobile host. Thus, a handoff multicast operation within a large cluster involves more base stations than that within a small cluster. This increases the total utilization of base stations. When considering only the handoff multicast operation, the best partition would consist of equal-weighted clusters where the weight of a cluster is the product of the number of base stations in the cluster and the number of mobile hosts moving out of physical cells in the cluster. As the velocity of mobile hosts increases, more handoff multicast operations will be needed, which in turn results in more utilization of base stations.

The utilization of the ARP/Location server for the optimal partition is less likely saturated compared to that for the initial partition. However, the average utilization of the ARP/Location server for the optimal partition approaches to approximately 0.72 at the saturation point of the in-call probability 0.8. This means that the ARP/Location server for the largest cluster of the optimal partition is saturated at the in-call probability 0.8. The bottleneck in the ARP/Location server comes from the fact that the destination of data messages are randomly distributed over the entire system and the locality of data traffic patterns is not considered. Thus the ARP/Location server may use a faster processor to resolve it. Adjusting the cluster size constraint in the optimization process could also alleviate the effect of the largest cluster on the utilization of the ARP/Location server. In this analysis, the cluster sizes of the initial partition are $n_1 = 6$, $n_2 = 6$, and $n_3 = 7$, and the cluster sizes of the optimal partition obtained by using the cluster size constraint, $4 < |P_i| < 12$, are $n_1 = 4$, $n_2 = 4$, and $n_3 = 11$.



(Fig. 6) The Utilization of the ARP/Location Server for Each Type of Message When an Initial Partition Is Used

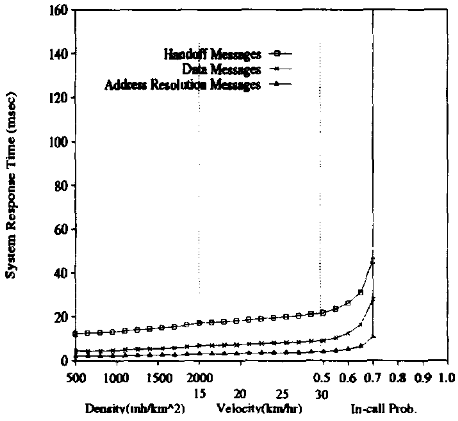
From Figure 6 and Figure 7, it should be noted that the significant difference in the utilization of the



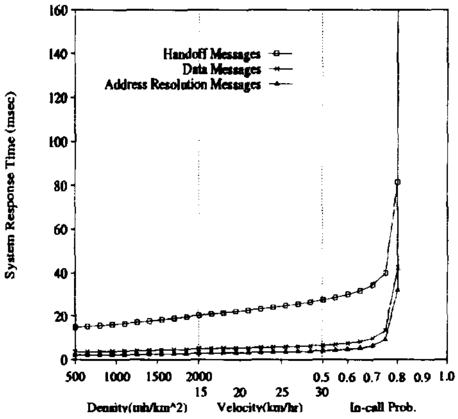
(Fig. 7) The Utilization of the ARP/Location Server for Each Type of Message When an Optimal Partition Is Used

ARP/Location server between the initial partition and the optimal partition comes from much less inter-cluster data traffic in the optimal partition. Thus, the sensitivity of data messages to the density, velocity, and in-call probability variations dominates the utilization of a network component in the virtual cell system.

For data messages, the increases in the density and in-call probability of mobile hosts directly affect the arrival rate of data messages. However, the increase in the velocity of mobile hosts does not raise the arrival rate but the forwarding rate of data messages. Thus, as the velocity of mobile hosts increases, the utilization of data messages at the ARP/Location server is slightly increased by the inter-cluster forwarding data messages, but it could be negligible compared to that of data transfer class of messages. The utilization of the ARP/Location by handoff messages is more sensitive to the velocity variation than the density variation. The in-call probability does not affect handoff messages because it only raises the arrival rate of data messages. Compared to data and handoff messages, the utilization of the ARP/Location server by address resolution messages could be negligible.



(Fig. 8) The System Response Time for Each Type of Message When an Initial Partition Is Used



(Fig. 9) The System Response Time for Each Type of Message When an Optimal Partition Is Used

The system response time Q_r for each type of messages are shown in Figure 8 and Figure 9. In the optimal partition, the system response time for data and address resolution messages is reduced at the cost of slightly higher response time for handoff messages. This is due to the impact of handoff multicast operations on the relatively larger cluster size with respect to the cluster size of the initial partition, as described previously. When considering high mobility and data parameters, i.e., 2000 mh/km^2, 30 km/hr, and 0.8 in-call probability of mobile hosts, the mean handoff response time 80 *msec* is quite acceptable. Given some mobility and data traffic parameters, the analysis shows a trade-off between the data and handoff response times. The much larger volume of data traffic over handoff traffic in mobile data communications reveals that even the small difference in the data response time could improve the overall system performance significantly. To reduce the handoff response time, more strict constraint on the cluster size is needed in the optimization process.

5. Conclusion

To analyze the performance of the virtual cell system we adopt a BCMP open multiple class queueing network. Since the same type of message may have a different routing behavior depending on which base station belongs to which cluster, a routing chain is defined for each type of message generated by each base station. Both mobility and data traffic patterns among base stations and the topology of the virtual cell system are used to determine service transition probabilities in the queueing network model. With various performance measures such as the utilization of network components in the virtual cell system and the system response time for various types of messages, we have conducted sensitivity analyses of those performance measures as mobility and data traffic parameters vary. We also compared the performance measures of two different virtual cell systems which are one deployed according to an initial partition used for the optimization process and the other deployed an optimal partition with respect to the initial partition, respectively.

References

[1] Kyungshik Lim, Young-Hwan Lim, and Yann-Hang Lee, "Virtual Cell in Mobile Computer Communications," Computer Communications, Vol. 20, No.7, pp.586-598, Aug. 1997.

[2] Kyungshik Lim and Yann-Hang Lee, "Optimal Partitioning of Heterogeneous Traffic Sources in Mobile Communications Networks," *IEEE Trans. on Computers*, Vol.46, No.3, pp.312-325, Mar. 1997.

[3] Forest Baskett, K. Mani Chandy, Richard R. Muntz, and Fernando G. Palacios, "Open, Closed, and Mixed Networks of Queues with Different Classes of Customers," J. ACM, Vol.22, No.2, pp. 248-260, Apr. 1975.

[4] Peter G. Harrison and Naresh M. Patel, Performance Modeling of Communication Networks and Computer Architectures, Addison-Wesley, 1993.

[5] R. Thomas, H. Gilbert, and G. Mazziotto, "Influence of Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network," Proc. 3rd Nordic Seminar, Paper 9.4, Copenhagen, Denmark, Sep. 1988.

임경식

1982년 경북대학교 전자공학과 졸업(공학사)
1985년 한국과학기술원 전산학과 졸업(공학석사)
1994년 University of Florida at Gainesville, 전산학과 졸업(공학박사)
1985년~1998년 한국전자통신연구원(ETRI) 책임연구원
1998년~현재 경북대학교 컴퓨터과학과 전임강사
관심분야 : mobile wireless computing and communications, high-speed communication networks, parallel and distributed systems