ARCA—An Adaptive Routing Protocol for Converged Ad-Hoc and Cellular Networks

Yumin Wu, Kun Yang, and Hsiao-Hwa Chen

Abstract: **This paper proposes an adaptive routing protocol called ARCA for converged ad-hoc and cellular network (CACN). Due to the limitation of both bandwidth and transmission range in a cell, a mobile host (MH) may not be able to make a call during busy time. CACN offers a flexible traffic diversion mechanism that allows a MH to use the bandwidth in another cell to ease the congestion problem and increase the throughput in a cellular network. Based on the presentation of CACN's physical characteristics, the paper details the design issues and operation of the adaptive routing protocol for CACN (ARCA). Detailed numerical analysis is presented in terms of both route request rejection rate and routing overhead, which, along with the simulation results, have indicated the effectiveness and efficiency of the ARCA protocol.**

Index Terms: **Ad-hoc network, cellular mobile, heterogenous network, routing protocol.**

I. INTRODUCTION

With the rapid development of telecommunications and wireless communication technology, networks, such as general packet radio service (GPRS), asymmetric digital subscriber line (ADSL), IEEE 802.11, and 3G networks, have been widely applied. Diverse network structures and protocols cause problems on network integration and service integration. For a more efficient utilization of variable network resources, two or more network infrastructures can be integrated, such as integrated cellular and ad-hoc relaying system (iCAR) [1] and mobile-assisted data forwarding (MADF) [2]. For instance, the iCAR tries to introduce mobile ad-hoc networks into a cellular network to improve the latter's call blocking rate in hot spots such as sporting venues. The heterogeneity in network structures is a typical feature of 4G mobile networks. In recent years, small portable devices have been increasingly equipped with multiple communication interfaces. New multiple-interface servers are also under design and development. These facilities make the integration of different networks more feasible. Therefore, it is obvious that heterogeneous networks will become the main part of future mobile systems and offer people more flexible network services [3].

The work presented in this paper is largely inspired by iCAR, MADF, and ad-hoc global system for mobile communications (A-GSM) [4]. However, these converged network structures all have their own demerits. For example, mobile hosts (MH) in iCAR have only one air interface, i.e., ad-hoc interface or A-interface involved (at least other interface, if there is any, does not involve) in the relaying procedure, and the intermediate nodes of relaying routes are only composed of traffic diversion stations (TDS). This paper tries to design a novel converged ad-hoc and cellular network (CACN) structure by including another air interface into MHs–cellular network interface, or C-interface, and constructing relaying routes with both TDSs and MHs. This is driven by the trend of small mobile devices being increasingly equipped with multiple communication interfaces, such as cellular and IEEE 802.11. A CACN system presented in this paper can take advantage of this extra hardware feature and further increase cellular network's overall performance. The focus of this paper is mainly on the routing protocols of the CACN system. The introduction of an ad-hoc air interface into MHs and allowing such MHs involved in data routing brings flexibility and complexity to the routing protocols for heterogeneous wireless networks. Section VI presents a more detailed comparison of CACN and its routing algorithms against other routing algorithms in heterogeneous wireless networks.

The remainder of the paper is organized as follows. Based on an analysis of the main physical characteristics of a CACN system, Section II discusses about the special design issues the heterogeneity of CACN brings to the routing protocols. Section III details the operation of a dynamic and adaptive routing protocol for the CACN system, named as ARCA. Section IV provides a numerical analysis of ARCA in terms of both route request rejection rate and routing overhead, which are exemplified by simulation results in Section V. After a related work discussion in Section VI, this paper concludes in Section VII.

II. PRELIMINARY

This section discusses about the physical characteristics of a CACN system and its implication on design issues of routing protocols.

A. Physical Characteristics of an CACN System

The CACN system in this paper adopts similar system architectures such as iCAR, MADF, and A-GSM, but allowing both A-interface and C-interface on MHs involved in traffic diversion. Fig. 1 shows the physical nature of two cells in the CACN system, including the interfaces (type and number) for every individual node. Two air interfaces are utilized for the communication between nodes: C (cellular) interface that operates at a cellular network frequency (in-band), and A (ad-hoc) interface that operates at an ad-hoc network frequency (out-of-band, e.g., IEEE 802.11). The introduction of two air interfaces on MHs significantly increases the flexibility of routing.

Base station (BS) is the same as the present cellular network base stations with C-interface. A base station uses its C-interface to communicate with mobile handsets in a wireless

Manuscript received May 1, 2006.

Y. Wu and K. Yang are with the Department of Electronic Systems Engineering, University of Essex, U.K., email: {ywud, kunyang}@essex.ac.uk.

H.-H. Chen is with the Institute of Communications Engineering, National Sun Yat-Sen University, Taiwan, email: hshwchen@ieee.org.

1229-2370/06/$10.00

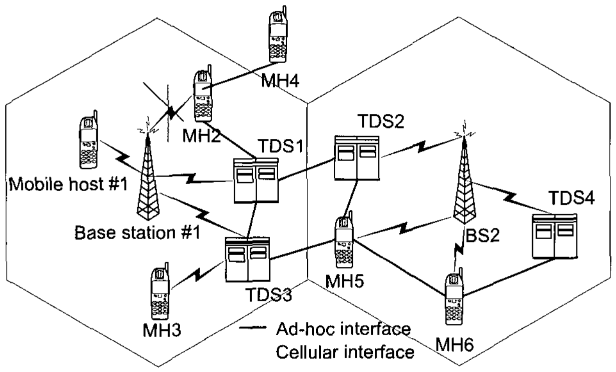


Fig. 1. CACN physical structure and route structure.

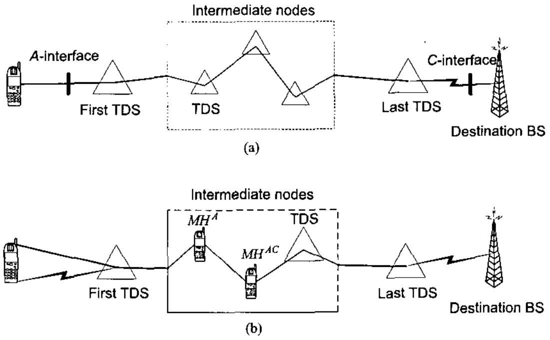


Fig. 2. Routing in (a) iCAR and (b) CACN.

mode. However, the communication between base stations is actually in wire mode, and controlled by a center control system (CC). Traffic diversion station (TDS) with managed mobility [5], which is similar to ad-hoc relaying station (ARS) in iCAR and GSM-MANET dual-mode terminal in A-GSM, uses both ad-hoc technologies and cellular network technologies with one A-interface and one C-interface. In a TDS, a C-interface is used for communicating with a BS or a MH with a C-interface. It consumes an in-band bandwidth in cellular networks, and operates at about 1900 MHz. A-interface, which operates at around 2.4 GHz in the unlicensed industry, science, and medical (ISM) band and consumes out-of-band bandwidth, is equipped for communicating between TDSs or MHs with an A-interface. Mobile host or mobile handset (MH) in the above mobile systems is designed with more flexibility. In summary, there are three types of MHs in CACN:

1. MHs with one C-interface and without A-interface (denoted as MH^C): They can be referred to most current mobile phones with simple voice service and limited data communication capability, only assigned with in-band frequencies.
2. MHs with one A-interface and without C-interface (denoted as MH^A): They can only communicate with TDSs in out-of-band frequencies, with examples being today's PDAs and laptops.
3. MHs with both A-interface and C-interface (denoted as MH^{AC}): Representing the emerging mobile handsets.

B. Routing Design Issues in CACNs

The following three issues affects the routing strategy and process that try to find a route r from a requesting mobile host m to a destination TDS t locating next to the destination BS: 1) Using what kind of air interface for m to get access to its immediate next-hop node on route r; 2) how the intermediate nodes are composed of; 3) using what kind of node (i.e., TDS or MH) and what air interface at the last hop (i.e., connecting to the destination BS). In terms of Issue #1, two types of air interfaces can be utilized and as such two types of accesses respectively: Namely, cellular access (CA, like MH3-TDS3 in Fig. 1 where MH3 is the requesting node), and ad-hoc access (AA, like MH2-TDS1 in Fig. 1 where MH2 is the requesting node). While SOPRANO [6] employs CA only via cell splitting technique to reuse in-band frequencies and iCAR and MADF use AA only, CACN makes use of both. Regarding Issue #2, the intermediate nodes in iCAR are composed of only TDSs whereas CACN supports both TDSs and MHs as intermediate relaying nodes. Considering Issue #3, both iCAR and CACN use C-interface for t to connect to destination BS. Fig. 2 illustrates the main differences in terms of route composition between iCAR and CACN.

In mobile ad-hoc networks (MANETs), destinations are known to sources. However, in CACN mentioned above, a call or data can be diverted to any destination as long as the latter has free bandwidth to accommodate the call. As such a destination selection procedure is a necessity in CACN. The feature of multiple destinations also indicates the existence of multiple routes between a source and destination(s). While multi-path routing is a special requirement for MANET routing where multiple routes to one destination are required to be found by one single route request packet, it is an inherent part of routing in heterogeneous networks such as CACN. Therefore, a route selection procedure is needed in routing.

To provide the CACN system with a more efficient load balancing, the CACNs routing algorithm, as a design principle, establishes routes through both TDSs and MHs employing both C-interface and A-interface. However, for a more efficient use of out-of-band frequencies in ARCA, the C-interface of a TDS is only used to connect to the destination BS or the source MH. Intermediate nodes, including TDSs and MHs with A-interface, only use A-interface to communicate with each other. Hence, MHs which work as intermediate nodes in a route can not be the first hop or the last hop. In this way, a queuing delay on A-interface might get bigger. Wu *et al.* discussed this problem and proposed some solutions in [7].

If a source MH is unable to find a TDS of sufficient bandwidth within its transmission range (also called one-hop range), a source selection procedure is needed in CACNs routing algorithm to select a proper pseudo-source MH. This pseudo source MH should have a TDS within its one-hop range and be able to find a relaying route to divert its own traffic. Then it releases its occupied channel to the original source MH.

III. ARCA: AN ADAPTIVE ROUTING PROTOCOL FOR CACN

Based on the CACNs physical model and routing protocol design issues presented in Section II, this section details the design and operation of a routing protocol designed for CACN,

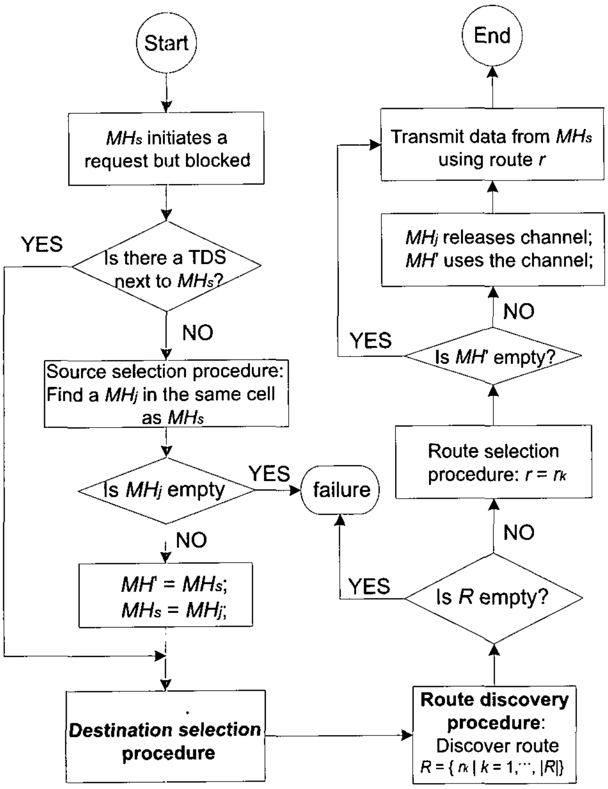


Fig. 3. ARCA operation flowchart.

named ARCA. In the CACN system, once a MH can not get enough bandwidth from its home base station for communication, it starts a route discovery process. First, it tries to find a route to another BS (destination) that has enough free bandwidth. Then, the MH utilizes the bandwidth in this destination BS through a discovered route. Finally, the MH works in normal cellular model through the destination base station, such as making a call or surfing the Internet.

Fig. 3 illustrates the operation procedure of ARCA. It starts from the point where a MH initiates a voice call or data transmission request whereas no sufficient bandwidth in its home cell is available to accommodate this request. We denote the request initiating MH (also called source MH) as MH_s. We name the cell where MH_s locates as home cell and denote it as C_h.

If there is a TDS of sufficient bandwidth within the transmission range of MH_s, then MH_s uses it directly as the next hop for relaying data. Otherwise, a source selection procedure (SSP) is triggered to find a proper pseudo-source MH (denoted as MH_j as illustrated in Fig. 3) within cell C_h so as to release this MH's channel to MH_s. In any case, a MH's data needs to be transmitted via relay as such a relaying route needs to be discovered. Unlike conventional ad-hoc routing algorithms where the destination of a route is known, MH_s, either it being the genuine one that is making the request or the pseudo one MH_j that is selected by SSP, does not know its route destination. All it needs is a route to a cell that can accommodate its data transmission regardless of the cell's location. In this light, a destination selection procedure (DSP) is a necessity. Once a destination or a list of destinations is selected, a route discovery procedure (RDP) is activated to find all routes satisfying the destination and bandwidth requirements. If multiple routes are found, then there needs a route selection procedure (RSP) to select the most proper route to relay MH_s's data. Obviously, if SSP gets involved in the beginning, the pseudo source MH (MH_j) will have to release the channel it uses to the genuine source MH (MH_s) before invoking the data transmission procedure to transmit data from MH_s using route r_k.

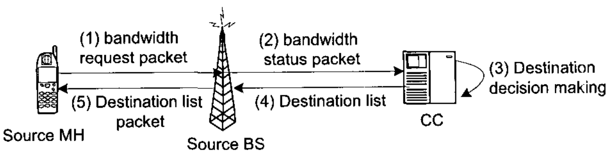


Fig. 4. Proactive method for destination decision.

As illustrated in Fig. 3, there are mainly four procedures involved in ARCA: Destination selection procedure or DSP, route discovery procedure or RSP, route selection procedure or RSP, and source selection procedure or SSP. The following subsections detail the design and operation of each. However, since SSP operates only in special case, it is to be discussed last.

A. Destination Selection Procedure

In ARCA, two destination decision mechanisms are designed for destination decision. The first one adopts a reactive method, i.e., destinations are chosen dynamically during route discovery process. The second one uses a proactive method, i.e., the destinations are decided before route discovery process.

Reactive method is applied to every TDS, and works after a TDS receives a route request (RREQ) packet. Because destinations are chosen by intermediate nodes, TDSs should judge whether the BS covering them has enough bandwidth to work as a destination. In order to avoid congesting another cell, the available bandwidth in the destination BS should be more than a threshold (α). The functioning of the reactive method requires that a TDS knows the bandwidth status of the BS covering it. To get the bandwidth status, BSs can broadcast bandwidth status packets to all TDSs periodically. However, this method causes lots of traffic overhead. Alternatively, after receiving a RREQ, a TDS which is not within the source cell can send bandwidth request packet to the current BS. After that, the current base station responds a bandwidth status packet back to the TDS. Then, the TDS decides the destination BS according to the bandwidth information contained in the bandwidth status packet.

On the other hand, because the central control system (CC) knows which cells are non-congested cells with available bandwidth more than a bandwidth threshold (α), the CC can inform the source MH a list of available destinations before the source broadcasts RREQs, as shown in Fig. 4. This is a proactive method for destination decision making and the decision is made by the CC. By pre-deciding destinations, TDSs can avoid frequently enquiring the bandwidth status of the BSs which they belong to. In this method, the "destination address" field in a RREQ is a list of available destinations. Following this, the route discovery process (detailed in Section III-B) is similar to that in ad-hoc model. Once a TDS finds that it stays in a destination cell according to the destination list in a RREQ, the TDS chooses this BS as a destination BS and responds a route reply

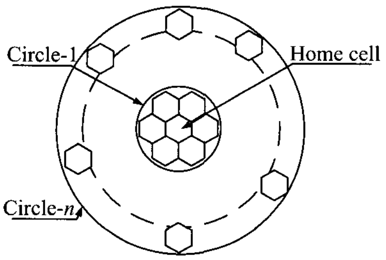


Fig. 5. Limited broadcast.

(RREP) packet including a single destination address following the reverse route back to the source MH. Because more than one route could be found during the route discovery, the source can choose a final route included in the first arrived RREP. Alternatively, the source can also choose a final route with minimum number of hops so as to reduce the transmission delay.

Furthermore, as the CC can decide destinations for the source MH, a limited broadcasting can be applied to route discovery process, as shown in Fig. 5. Cells surrounding the source cell are divided into several circles according to the distance to the source cell. Circle-1 includes six cells of the first ring with the closest distance to the source cell. The n-th ring includes $6n$ cells. Circle-n includes all cells from the 1st ring to the n-th ring. RREQs are only broadcasted within one of these circles (namely the limited broadcast area). The limited broadcast area can be chosen according to the number of destinations required for a route discovery. More destinations included in RREQs lead to more destination BSs discovered, but also bring more routing overhead. However, in order to broadcast RREQs within the pre-decided circle, the CC has to give a full list of cells, in which RREQs can be broadcasted. Hence, with limited broadcast, once a TDS or MH receives a RREQ, it first checks whether it is within the cells listed in the RREQ. All TDSs and MHs that locate out of the circle simply drop this RREQ.

B. Route Discovery Procedure

After choosing a destination decision method according to the design requirements, the source MH starts its route discovery process similar to DSR [8]. This route discovery process involves four steps:

- Step 1: The source node broadcasts a RREQ. At the same time, the source sets the amount of time spent on route discovery, also for receiving multiple routes. When a RREQ crosses through a hop, the address of this very intermediate node is recorded in the RREQ. RREQs can also be broadcasted within limited area as depicted in Fig. 5. A RREQ contains the following fields: {Source address, source cell, destination address, intermediate node list, MH node count, hop count, maximum hop count, bandwidth requirement, sequence number, limited broadcast cells}.
- Step 2: As a destination receives a RREQ, it will respond a RREP where the route set is copied from the received RREQ, and the packet will be sent back to the source following the reverse route. Here, the destination is the TDS next hop to the destination BS on the route. However, the destination BS is only a theoretical destination. The RREP is actually broadcasted by a TDS which finds the destination BS. In reactive destination making method, the destination is checked by a TDS once it receives a bandwidth status packet from the current BS. In proactive destination making, a TDS only checks whether the current BS is within the destination list in the receiving RREQ. No matter what destination decision module a TDS chooses, there is only one destination in a RREP. A RREP packet contains the following fields: {Source address, destination address, intermediate node list, MH node count, hop count, sequence number}.

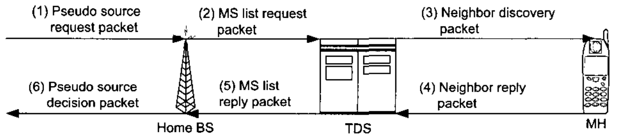


Fig. 6. Source selection procedure.

- Step 3: Route selection procedure is carried out in this step if an original source receives more than one RREP as a result of the route discovery process. The basic principle utilized in ARCA is shortest-path, i.e., minimum MH node count. The motivations are twofold: 1) To minimize the end-to-end delay in traffic diversion; 2) to increase the robustness.
- Step 4: As the source that has initiated the route request process selects the final "optimal" route, the routing process succeeds. The source records the route set in the table and divert traffic to the destination along the routing path by putting the routing path in the header of the data packet.

C. Source Selection Procedure

Because in the home cell, there are many MHs in service within the transmission range of TDSs, the BS in the congested cell should choose one proper MH as a pseudo source (MH2 in Fig. 1) to release its bandwidth for the use of the original source (MH1 in Fig. 1). As mentioned above, SSPs are designed to run in BSs. It aims at choosing pseudo sources to further reduce the call block rate in hot (or congested) cells, and is triggered after receiving a pseudo source request packet (PSREQ) from the original source (see (1) in Fig. 6). To get the addresses of all available pseudo sources, the home BS broadcasts a MH list request packet to all TDSs within the home cell (see (2) in Fig. 6). Following this, by broadcasting a neighbor discovery packet (see (3) in Fig. 6), TDSs within the home cell can get a list of available pseudo sources after receiving neighbor reply packets from all neighbor MHs (see (4) in Fig. 6). Then, TDSs respond MH list reply packets (MLREP) to inform the home BS which MHs can be pseudo sources (see (5) in Fig. 6). Thus, the home BS chooses a final pseudo source from the list included in MLREP's according to a source selection algorithm. This final pseudo source starts a route discovery process to find a relaying route to neighbor cells. Then, it releases its occupied channel to the original source just after diverting its call (or traffic) to another BS through relaying routes.

A simplified SSP may only take the location information of MHs into consideration to select a MH, which has the furthest distance to the home BS. This is because the MH furthest from the home BS has the most possibility of moving into a neighbor

Table 1. Notations of the system.

R	The center-to-vertex distance of a cell
r	The transmission range of an TDS or an MH with an A-interface (assuming both of them have the same transmission range)
N_T	The average density of TDSs in the whole network (number per km^2)
N_{MA}	The average density of MHs with an A-interface in the whole network (number per km^2)
N_M	The average density of MHs in the whole network (number per km^2)

Table 2. Notations for RRRR analysis.

P_{RR}	Route request rejection rate in a cell
R_1	The average maximum transmission range of a route request packet without limited broadcast (how far a route request packet can get to)
P_b	The average coverage rate of route request packets within the maximum transmission range
N_C	The average number of cells with enough free bandwidth (number per km^2)
N	The maximum circle number within the transmission range of a route request packet in Protocols 1 and 2 (Fig. 5)
N_D	The number of destinations included in the "destination address" field of a route request packet in Protocols 2 and 3
$N_{\min}$	The minimum number of available destinations within the limited broadcast area in Protocol 3
T_0	The current traffic in the home cell (Erlangs)
$T_{\max}$	The maximum traffic that can be burdened by every cell (Erlangs)
T_C	The current traffic in a certain cell (Erlangs)
T_T	The maximum traffic that a cell can have in order to work as a destination cell (Erlangs) (if the current traffic in a cell is more than T_T, it is unable to be a destination cell)

cell, and subsequently utilizes the bandwidth in the neighbor cell according to the operation principles of cellular networks. To design a more practical SSP, pseudo sources chosen by the home BS should have the most possibility of discovering a relaying route, and a large number of pseudo sources do not partially block TDSs. This is realized by making use of the information contained in MLREP. The following shows a simple example of MLREP: {TDS ID, MH ID list, bandwidth info}, where "TDS ID" refers to the address of a TDS in the home BS. "MH ID list" includes a list of the addresses of MHs within the transmission range of the TDS. "Bandwidth info" is the available free bandwidth (or channel) in the TDS. Only TDS's with enough free bandwidth for the original calling traffic can offer traffic diversion service. The MH's covered by TDS's without enough bandwidth are not able to divert its ongoing call through relaying routes. Then, they cannot release the bandwidth to original sources. As a result, these MH's cannot be chosen as pseudo sources.

IV. NUMERICAL ANALYSIS

In this section, two metrics are analyzed for the evaluation of the proposed routing protocols, namely the route request rejection rate (RRRR) of a congested cell, and routing overhead. In a general condition, a cell X and its neighbor cells are controlled by a center control system. We assume that TDSs are randomly placed in each cell, and mainly analyze the DSP part of ARCA. Then, the routing protocol using reactive method is Protocol 1. The routing protocol using proactive method without limited broadcast is Protocol 2. The routing protocol using proactive method with limited broadcast is Protocol 3. Table 1 gives the basic parameters of this integrated system.

A. Route Request Rejection Rate (RRRR) Analysis

The RRRR is the possibility that a source MH fails to find a route to any destination BS after broadcasting a RREQ packet. The RRRR is affected by several factors, such as the density of TDSs and MHs with an A-interface, the number of cells with enough free bandwidth, and the performance of routing algorithms. Table 2 shows the parameters used for the RRRR analysis.

According to the symbols shown above, the area of a cell is $\frac{3\sqrt{3}}{2}R^2$. P_{RR} is one minus route request successful rate (P_{RS}). P_{RS} is the area of cells with enough free bandwidth within the transmission range of a route discovery process (A_A), divided by the difference between transmission area of a route discovery process (πR_1^2) and the area covered by RREQs within the transmission range of a route discovery process ($\pi R_1^2 P_b$), as shown in Fig. 7. Also, A_A is the number of cells with enough free bandwidth within the transmission range of a route discovery process, multiplied by the area of a cell, namely $(\pi R_1^2 N_C)(\frac{3\sqrt{3}}{2}R^2)$. Thus, the route request rejection rate (P_{RR}) in a congested cell is

$$P_{RR} = 1 - \frac{\pi R_1^2 N_C \frac{3\sqrt{3}}{2} R^2}{\pi R_1^2 - \pi R_1^2 P_b} = 1 - \frac{N_C \frac{3\sqrt{3}}{2} R^2}{1 - P_b} \qquad (1)$$

where $N_C\frac{3\sqrt{3}}{2}R^2 < 1 - P_b$ and $P_{RR} = 1 - \frac{N_C\frac{3\sqrt{3}}{2}R^2}{1-P_b}$. If $N_C\frac{3\sqrt{3}}{2}R^2 \geq 1 - P_b$, namely the area covered by cells with enough free bandwidth is more than the area uncovered by RREQ packets, which means that RREQ packets can get to at least one of these cold cells, then $P_{RR} = 0$.

To simplify the analysis, it is assumed that the traffic in each cell surrounding the home cell is all at the average traffic T_1, and every cell within the same circle has the same possibility to be selected as a destination cell. Hence, cells within the same circle have the same possibility of becoming unavailable when $T_c > T_T$. Also, we assume that the closest cold (or non-congested cell) cell has the highest possibility to be chosen as a destination. Ignoring the distance between the home BS and the source MH, both normal broadcast in Protocols 1 and 2 and limited broadcast in Protocol 3 have the same center (the home BS). Because the number of cells in each circle during limited broadcast is $6i$, the overall number of cold cells inside the transmission range of route request packets is $\sum_{i=1}^{N} 6i$. According to Table 2, the traffic that a destination cell can bur-

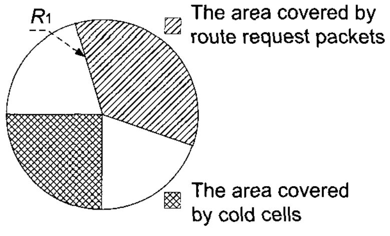


Fig. 7. RREQ broadcast analysis.

den is $T_T - T_1 (T_T < T_{\max})$. Then, the traffic that can be diverted to the i-th ring (Ring-i) is $6i(T_T - T_1)$. Thus, the overall traffic that can be diverted by cold cells within the transmission range of route request packets is $\sum_{i=1}^{N} 6i(T_T - T_1)$. According to the assumptions given above, because the current overall diversion traffic from the home cell is $T_0 - T_{\max}$, then $\sum_{i=1}^{a-1} 6i(T_T - T_1) \leq T_0 - T_{\max} < \sum_{i=1}^{a} 6i(T_T - T_1)$, and cells from Circle-1 to Circle-$(\alpha - 1)$ all become unavailable $(\alpha \geq 1)$. Thus, the current number of available cold cells is $\sum_{i=\alpha}^{N} 6i$, and the number of cells within the limited broadcast area in Protocol 3 is $\sum_{i=1}^{\alpha} 6i$.

In (1), N_C equals to the number of overall available destinations (N_A) divided by the overall area of cells within the transmission range of a route discovery process (A_C). A_C is the area of one cell multiplied by the overall number of cells within the transmission range of a route discovery process (N_N). Hence,

$$N_C = \frac{N_A}{N_N \frac{3\sqrt{3}}{2} R^2}. \tag{2}$$

Additionally, the average maximum transmission range (R_1) of a RREQ packet in Protocol 1 is related to the maximum hop count of a route discovery process and the density of TDSs and MHs with an A-interface. In terms of P_b, it drops with the increase of the diversion traffic from the home cell $(T_0 - T_{\max})$. To simplify the analysis of P_b, assuming

$$P_b = P_{Default} \frac{b}{T_0 - T_{\max}} \tag{3}$$

where b is a variable related to the distribution density and bandwidth of TDSs and MHs with an A-interface, and $P_{Default}$ is the default value of P_b without the effect of diversion traffic.

In Protocol 1, N_A is the number of current available cold cells within the transmission range of RREQs, namely $\sum_{i=\alpha}^{N} 6i$. Moreover, N_N is $1 + \sum_{i=1}^{N} 6i$. Hence,

$$N_C = \frac{\sum_{i=a}^{N} 6i}{(1 + \sum_{i=1}^{N} 6i) \frac{3\sqrt{3}}{2} R^2}. \tag{4}$$

In Protocol 2, because destinations pre-chosen by the center control system (CC) are those closest to the home cell, the broadcast range of RREQs can be considered as an area from the home cell to the furthest circle, including most destinations from the destination list pre-decided by the CC. However, these destinations can not be all in the transmission range of a route discovery process. Then, if $\sum_{i=\alpha}^{N} 6i > N_D$, then $N_A = N_D$, $N_N = 1 + \sum_{i=1}^{\alpha} 6i$, thus

$$N_C = \frac{N_D}{(1 + \sum_{i=1}^{a} 6i) \frac{3\sqrt{3}}{2} R^2}. \tag{5}$$

If $\sum_{i=a}^{N} 6i \leq N_D$, then $N_A = \sum_{i=\alpha}^{N} 6i$ and $N_N = 1 + \sum_{i=1}^{N} 6i$, thus

$$N_C = \frac{\sum_{i=a}^{N} 6i}{(1 + \sum_{i=1}^{N} 6i) \frac{3\sqrt{3}}{2} R^2}. \tag{6}$$

In Protocol 3, the whole routing algorithm is similar to that in Protocol 2, but RREQs are only broadcasted in a limited area pre-decided by the CC. To avoid greatly increasing the route request rejection rate, the number of available destination cells should be more than $N_{\min}$. According to the assumptions, the overall number of available cold cells in this limited broadcast area is $6a$.

If $6a > N_D$, then $N_A = N_D$ and $N_N = 1 + \sum_{i=1}^{\alpha} 6i$, thus

$$N_C = \frac{N_D}{(1 + \sum_{i=1}^{a} 6i) \frac{3\sqrt{3}}{2} R^2}. \tag{7}$$

If $N_{\min} \leq 6a \leq N_D$, then $N_A = 6a$ and $N_N = 1 + \sum_{i=1}^{\alpha} 6i$, thus,

$$N_C = \frac{6a}{(1 + \sum_{i=1}^{a} 6i) \frac{3\sqrt{3}}{2} R^2}. \tag{8}$$

Comparing these 3 protocols, according to (1), (2), (3), and N_C computed in each protocol, the RRRR in Protocol 2 is general more than that in Protocol 1 with the increase of traffic in the home cell, because the limited number of destinations pre-decided by the CC in Protocol 2 causes that a source MH has higher possibility of not finding a route to a destination cell. In terms of Protocol 3, the RRRR should be more than that in Protocol 2, because the number of available destination cells decreases in the limited broadcast area as the diversion traffic from the home cell increases.

B. Routing Overhead Analysis

Routing overheads are packets or messages used for route discovery and establishment. Too much routing overhead can dramatically deteriorate the performance of a network, such as causing longer queue time and bigger drop rate. This part gives a general routing overhead analysis on Protocols 1–3. Table 3 shows the parameters used for the routing overhead analysis.

The routing overhead of a routing protocol mainly includes route request (RREQs) packets, route reply (RREPs) packets, and other control packets. N_{RREQ} is the overall number of nodes broadcasting route request packets (N_{NRREQ}) multiplied by the average number of a node's neighbors (N_{NB}). N_{NRREQ} is the difference between overall number of TDSs and MHs with an A-interface within the route discovery area (N_{ALL}) and the

Table 3. Notations for routing overhead analysis.

N_{RREQ}	The average number of route request packets during a route discovery process
N_{RREP}	The average number of route reply packets during a route discovery process
N_{BREQ}	The average number of bandwidth request packets in Protocol 1
N_{BST}	The average number of bandwidth status packets in Protocol 1
N_{MREQ}	The average number of MH list request packets during the source selection procedure
N_{MREP}	The average number of MH list reply packets during the source selection procedure
N_{NREQ}	The average number of neighbor request packets during the source selection procedure
N_{NREP}	The average number of neighbor reply packets during the source selection procedure
N_H	The average number of hops of a relaying path
N_R	The average number of routes found during a route discovery process
N_{DF}	The average number of destination TDSs found

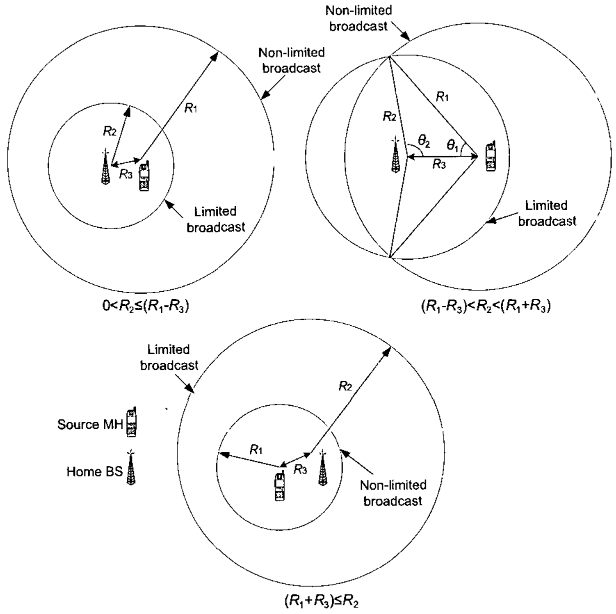


Fig. 8. Routing overhead analysis in Protocol 3.

number of destinations discovered during a route discovery process (N_{DF}), or

$$N_{RREQ} = (N_{ALL} - N_{DF})N_{NB}. \tag{9}$$

According to the symbols shown in the above tables, we have

$$N_{RREP} = N_R N_H. \tag{10}$$

The transmission area of a TDS or MH with A-interface is πr^2. Then, the average number of neighbors (N_{NB}) of a TDS or MH with A-interface is $\pi r^2(N_T + N_{MA})$. To simplify the analysis, it is assumed that the source MH is a MH with A-interface, and all TDSs and MHs with A-interface receiving RREQs simply rebroadcast them or reply RREPs after judging destinations and maximum hops.

In Protocol 1, because destinations are decided during route discovery process, every TDS receiving a RREQ needs to enquire the bandwidth status of the current BS. This brings some extra overheads including bandwidth request (BREQs) packets and bandwidth status (BSTs) packets. Additionally, N_{ALL} is the area covered by RREQs, multiplied by the density of both TDSs and MHs with an A-interface, namely $(\pi R_1^2 P_b)(N_T + N_{MA})$. Because the average number of destination TDSs discovered in Protocol 1 is N_{DF1}, then N_{NRREQ} is $[\pi R_1^2 P_b(N_T + N_{MA})] - N_{D1}$. In (9), the RREQ overhead in Protocol 1 is

$$\begin{aligned} N_{RREQ1} &= (N_{ALL} - N_{DF1})N_{NB} \\ &= \{[\pi R_1^2 P_b(N_T + N_{MA})] - N_{DF1}\}[\pi r^2(N_T + N_{MA})]. \end{aligned} \tag{11}$$

In (10), the RREP overhead in Protocol 1 is $N_{RREP1} = N_{R1}N_{H1}$. The extra overhead used for obtaining the bandwidth status of current BSs is the number of TDSs within the area covered by RREQs (N_{TR}) multiplied by two, because each TDS needs to send a BREQ and receive a BST. N_{TR} is the area covered by RREQs multiplied by N_T, namely $(\pi R_1^2 P_b)N_T$. Thus, overheads for getting the bandwidth status of current BSs is $N_{BREQ} + N_{BST} = 2[\pi R_1^2 P_b N_T]$.

In Protocol 2, instead of discovering destinations during a route discovery process, the source MH has to get a destination list from the home BS. Because only two extra packets (one is bandwidth request packet, another is destination list packet in Fig. 4) are brought to the system, these extra control overheads in Protocol 2 can be ignored. Hence, an intermediate TDS only checks whether it is inside a destination cell. Similar to Protocol 1, RREQs in Protocol 2 is also broadcasted over the whole network, and only limited by the maximum hop count. Hence, the RREQ overhead in Protocol 2 is

$$\begin{aligned} N_{RREQ2} &= (N_{ALL} - N_D)N_{NB} \\ &= \{[\pi R_1^2 P_b(N_T + N_{MA})] - N_{DF2}\}\{\pi r^2(N_T + N_{MA})\}. \end{aligned} \tag{12}$$

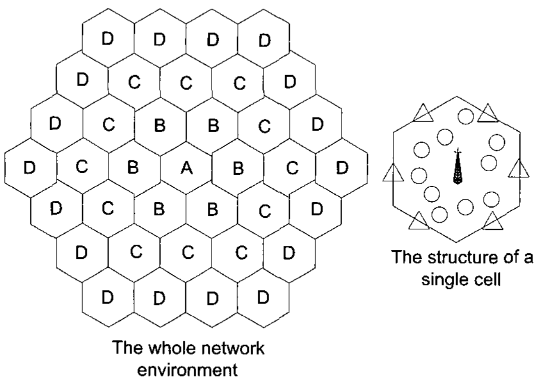
The RREP overhead in Protocol 2 is $N_{RREP2} = N_{R2}N_{H2}$.

In Protocol 3, RREQs are broadcasted within a circle at a radius R_2, and the whole routing process is similar to that in Protocol 2. However, the center of the circle is the home base station, not the source MH. As shown in Fig. 8, the real transmission range of RREQs is the cross area between the non-limited broadcast circle and the limited broadcast circle. In Protocol 3, the average number of destinations discovered is N_{DF3}, and the average number of routes found is N_{R3}. Hence, the RREP overhead in Protocol 3 is $N_{RREP3} = N_{R3}N_{H3}$.

In the situation of $0 < R_2 \leq R_1 - R_3$ (shown in Fig. 8), the cross area is the limited broadcast circle, namely πR_2^2. Thus, according to (9), the RREQ overhead in Protocol 3 is

$$\begin{aligned} N_{RREQ3} &= (N_{ALL} - N_{DF3})N_{NB} \\ &= \{[\pi R_2^2 P_b(N_T + N_{MA})] - N_{DF3}\}\{\pi r^2(N_T + N_{MA})\}. \end{aligned} \tag{13}$$

In the situation of $R_1 - R_3 < R_2 < R_1 + R_3$, the cross area is the area (A) between the limited broadcast circle and the



The whole network environment

The structure of a single cell

Fig. 9. Simulated network environment.

non-limited broadcast circle. In Fig. 8, we have

$$
\begin{aligned}
A &= [(\frac{2\theta_1}{2\pi}\pi R_1^2) - \frac{R_1 \sin\theta_1 R_1 \cos\theta_1}{2} 2] \\
&+[(\frac{2\theta_2}{2\pi}\pi R_2^2) - \frac{R_2 \sin\theta_2 R_2 \cos\theta_2}{2} 2] \\
&= \theta_1 R_1^2 + \theta_2 R_2^2 - (R_1^2 \sin\theta_1 \cos\theta_1 + R_2^2 \sin\theta_2 \cos\theta_2)
\end{aligned} \tag{14}
$$

where $\theta_1 = \arccos\left(\frac{R_1^2+R_3^2-R_3^3}{2R_1R_3}\right)$, $\theta_2 = \arccos\left(\frac{R_2^2+R_3^2-R_3^3}{2R_2R_3}\right)$.

Then, the RREQ overhead in Protocol 3 is

$$
\begin{aligned}
N_{RREQ3} &= (N_{ALL} - N_{DF3})N_{NB} \\
&= \{[AP_b(N_T + N_{MA})] - N_{DF3}\}\{\pi r^2(N_T + N_{MA})\}.
\end{aligned} \tag{15}
$$

In the situation of $R_2 \geq R_1 + R_3$, the cross area is the area of the non-limited broadcast circle. Thus, the RREQ overhead in Protocol 3 is similar to that in Protocols 1 and 2:

$$
\begin{aligned}
N_{RREQ3} &= (N_{ALL} - N_{DF3})N_{NB} \\
&= \{[\pi R_1^2 P_b(N_T + N_{MA})] - N_{DF3}\}\{\pi r^2(N_T + N_{MA})\}.
\end{aligned} \tag{16}
$$

Considering the SSP, it is launched because the original source MH is not within the transmission range of any TDS. The following gives the traffic overheads used for the SSP procedure. According to Fig. 6, the number of packets sent and received by the home BS (N_{BS}) is $N_{BS} = 2 + 2\frac{3\sqrt{3}}{2}R^2N_T = 2 + 3\sqrt{3}R^2N_T$. The number of packets sent and received by TDSs in the home cell (N_{TDS}) is $N_{TDS} = 2\frac{3\sqrt{3}}{2}R^2N_T + 2\pi r^2N_M$.

Comparing the performance of Protocols 1 and 2, the routing overhead in Protocol 1 is generally more than that in Protocol 2, because every intermediate TDS in Protocol 1 should enquire the bandwidth status of the current BS by sending BREQs and receiving BSTs. Considering Protocol 3, because RREQs are broadcasted within a limited area, the routing overhead in Protocol 3 depends on how large the limited broadcast area is. If the limited broadcast area is smaller than the non-limited broadcast area (Fig. 8), the routing overhead in Protocol 3 is less than that in Protocol 2. If the limited broadcast area is bigger than the non-limited broadcast area, the routing overhead in Protocol 3 should be the same as that in Protocol 2.

V. SIMULATION RESULTS

To evaluate the performance of routing protocols in a CACN system, a simulated network environment is built up as shown in Fig. 9. Cell-A is the home cell with traffic density more than the maximum traffic density that a cell can bear. Hence, Cell-A starts to divert traffic to its neighbor cells through relaying routes. For the convenience of the system analysis, neighbor cells are divided into subsystems (as shown in Fig. 9). In this environment, there are overall 37 cells, and each cell has the transmission radius of two km. TDSs with the transmission radius of 500 m are deployed at each shared border of two adjacent cells. MHs with an A-interface are randomly distributed over the whole networks, and the average number of these MHs with random speed and random moving direction is ten in a cell. The maximum traffic that a cell can support is 100 Erlangs. The average traffic in each neighbor cell is 60 Erlangs. The threshold which a cell can work as a destination cell is 80 Erlangs.

The results of route request rejection rate are shown in Fig. 10. It is observed that the RRRR in Protocol 1 is less than that in Protocols 2 and 3. This is because the number of available destinations discovered in Protocol 1 is more than the number of pre-decided destinations in Protocol 2. Additionally, because RREQs in Protocol 1 are broadcasted over the whole network, this leads to more routes discovered. However, with the increase of the traffic in the home cell, the number of available destinations in Protocol 1 drops due to more and more traffic diverted from the home cell to neighbor cells. Finally, the route request rejection rate in Protocol 1 can be same as that in Protocols 2 and 3. Furthermore, the performance of Protocol 3 is affected by the minimum number of available destinations within the limited broadcast area, which is decided by the CC. In Protocol 3, the increase of the minimum number of available destinations within the limited broadcast area can reduce the route request rejection rate, but may increase the transmission range of the limited broadcast area. As more overloaded traffic is diverted from home cell to neighbor cells, some of neighbor cells within a limited broadcast region could become unavailable. When the number of available destination BS's is lower than the minimum number of available destinations in the limited broadcast area, the range of the limited broadcast area should be extended. Thus, the RRRR in Protocol 3 could firstly rise in a certain limited broadcast area, and then drops when the range of the limited broadcast area is extended, as shown in Fig. 10.

It can be observed from Fig. 11 that the routing overhead in Protocol 1 is much more than that in Protocols 2 and 3, because most intermediate TDSs in Protocol 1 has to obtain the bandwidth status of the current BS by sending BREQs to the BS. Considering Protocols 2 and 3, due to the limited broadcast, the routing overhead in Protocol 3 is less than that in Protocol 2. However, the transmission range of the limited broadcast increases with the amount of traffic diverted to neighbor cells. Hence, if the transmission ranges of the limited broadcast is equal to (or more than) the transmission range of RREQs in Protocol 2, the routing overhead in Protocol 3 is the same as that in Protocol 2 according to the numerical analysis.

VI. RELATED WORK

Much work has been carried out in the integration of cellular networks and ad-hoc networks but of various objectives. For example, iCAR [1] positions itself on the extra throughput

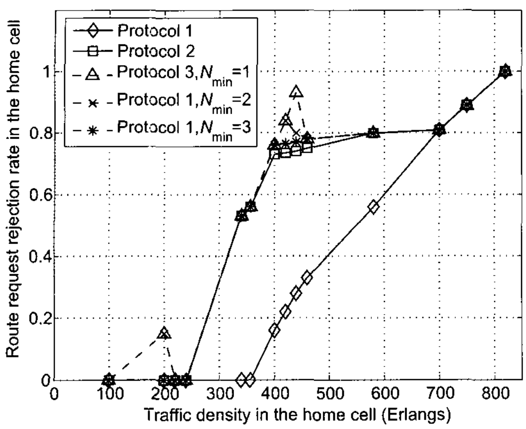


Fig. 10. Results of route request rejection rate.

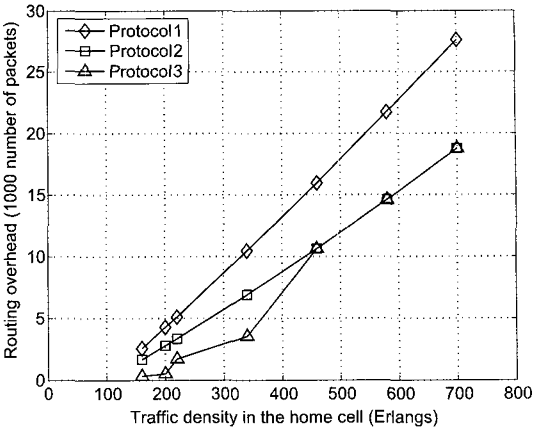


Fig. 11. Results of routing overhead.

gain and the improvement of call block rate in a busy cell. The same applies to MADF [2], which utilizes a different mechanism called "forwarding channel" to divert traffic. Another salient feature of MADF is the fact that it realizes load balancing without a need of specific diversion equipment such as the ARS in iCAR or the TDS in CACN. However this comes at the cost of pre-allocated forwarding channels in each cell. *A*-GSM [4] tackles how to expand the transmission range of each cell in cellular networks via ad-hoc relay. Different motivations render these heterogeneous networks employ different techniques when carrying out relaying or routing.

With a different objective from that of iCAR and MADF, a unified cellular and ad-hoc network UCAN [9] architecture focuses on higher data transmission. The relaying routes in UCAN are built only inside a cell rather than across cells as in iCAR and CACN. And the triggering event for a route discovery in the UCAN is deterioration of downlink channel of a MH rather than a call being blocked. Ioannisdis *et al.* [10] presented some better routing schemes so as to gain higher throughput based on a UCAN-alike system. The routing issues considered in [10] are still within a cell. Moreover, two heterogeneous network structures can also work together to gain extra benefits. Yanmaz *et al.* [11] presented a "location dependent dynamic load balancing" scheme, which makes iCAR and CBWL [12] work together to further reduce the call block rate in the current cellular networks.

VII. CONCLUSION AND FUTURE WORKS

This paper proposes an adaptive routing protocol (called ARCA) based on a CACN system. The CACN system offers MHs dynamic load balancing to increase the throughput and to decrease the call block rate of a cellular network. ARCA enables fast and efficient routing and avoids bringing lots of routing and signaling overheads on a CACN. Actually, ARCA is a cluster of three main types of routing protocols (as mentioned in Section IV) according to the method applied to RREQ broadcasting and the destination decision method adopted in the routing protocol. For reactive method and proactive method, the reactive method needs extra packets to get the bandwidth status of the current BS but offers faster routing. With limited broadcast, routing overheads are dramatically decreased but at the cost of the further increase of route request rejection rate. As such, routing protocols in a CACN should be chosen according to network environments.

The near future work is to look into how the in-band frequencies are used by TDSs and how the placement of TDSs affects the performance of the overall system. Further improvement on routing protocols based on the outcomes of above near-future research is also our planned next step work.

REFERENCES

[1] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated cellular and ad-hoc relaying systems: iCAR," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 2105–2113, Oct. 2001.

[2] X. Wu, S.-H. G. Chan, B. Mukherjee, and B. Bhargave, "MADF: Moblie-assisted data forwarding for wireless data networks," *J. Commun. Networks*, vol 6, no. 3, pp. 216–225, Jan. 2002.

[3] D. Cavalcanti, D. Agrawal, C. Cordeiro, B. Xie, and A. Kumar, "Issues in integrating cellular networks, WLANs, and MANETs: A futuristic heterogeneous wireless network," *IEEE Wireless Commun.*, pp. 30–41, June 2005.

[4] G. N. Aggelou and R. Tafazolli, "On the relaying capability of next-generation GSM cellular networks," *IEEE Pers. Commun.*, Feb. 2001.

[5] H. Wu, S. De, C. Qiao, E. Yanmaz, and O. Tonguz, "Managed mobility: A novel concept in integrated wireless systems," *IEEE J. Select. Areas Commun.*, pp. 537–539, 2004.

[6] A. N. Zadeh, B. Jabbari, R. Pickholtz, and B. Bojcic, "Self-organizing packet radio ad-hoc networks with overlay (SOPRANO)," *IEEE Commun. Mag.*, vol. 40, no. 6, pp. 149–157, 2002.

[7] H. Wu, S. De, C. Qiao, E. Yanmaz, and O. Tonguz, "Queuing delay performance of the integrated cellular and ad-hoc relaying system," *IEEE J. Select. Areas Commun.*, pp. 923–927, 2003.

[8] D. B. Johnson, D. A. Maltz, and Y.-C. Hu, "The dynamic source routing protocol for mobile ad-hoc networks (DSR)," IETF MANET Working Group, Internet-Draft, 15 Apr. 2003.

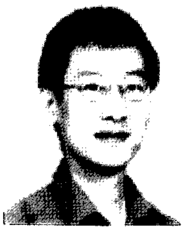
[9] H. Luo, R. Ramjee, P. Sinha, L. (E.) Li, and S. Lu, "UCAN: A unified cellular and ad-hoc network architechture," in *Proc. MobiCom 2003*, Sept. 14–19, 2003.

[10] I. Ioannisdis, B. Carbunar, and C. Nita-Rotaru, "High throughput routing in hybrid cellular and ad-hoc networks," in *Proc. WoWMoM 2005*, June 15, 2005.

[11] E. Yanmaz and O. K. Tonguz, "Location dependent dynamic load balancing," in *Proc. GLOBECOM 2005*, vol. 1, Dec. 2005, pp. 587–591.

[12] H. Jiang and S.S. Rappaport, "CBWL: A new channel assignment and sharing method for cellular communication systems," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 313–322, May 1994.

Yumin Wu was born in Hunan, China, on November 15, 1979. He received the B.Sc. degree in Telecommunication System in 2002 from Huazhong University of Science and Technology and the M.Sc. in Electronic Systems Engineering in 2005 from the University of Essex. Currently, he is working toward the Ph.D. degree in the University of Essex. His research interests are physical structure design, routing protocols, and system modelling in heterogeneous networks.



Kun Yang has been working as an academic staff in the Department of Electronic Systems Engineering, University of Essex, UK since 2003. Before that, he worked for 3 years at the Department of Electronic and Electrical Engineering, University College London (UCL)—UK on several European Union IST projects, such as MANTRIP, FAIN, CONTEXT in a wide area of networking, network and service management. He received his Ph.D. from the Electronic and Electrical Engineering Department of University College London (UCL), UK in the networks and services engineering area and an earlier Ph.D. from the Computer Science Department of Jilin University, China in the area of distributed artificial intelligence. His M.Sc. and B.Sc. degrees were both from Jilin University, China in the field of distributed systems and computer networks in 1994 and 1991, respectively. His main research interests include telecommunications and computer networks, wireless networks, mobile networks and mobile computing, mobile ad-hoc and sensor networks, multimedia over wireless networks, and pervasive service engineering. He has published journal papers and co-authored book chapters, amongst other conference papers. He serves on the editorial board of IEEE Communications Surveys and Tutorials, John Wiley Journal of Wireless Communications and Mobile Computing (WCMC), and serves as a special issue guest editor in several international journals. He is a chair (e.g., ICPS 2006) or a TPC member of many international conferences (e.g., ICC 2007 and GLOBECOM 2006). He is a member of IEEE, IEE, and ACM.

Hsiao-Hwa Chen is currently a full professor in Institute of Communications Engineering, National Sun Yat-Sen University, Taiwan. He received his B.Sc. and M.Sc. degrees from Zhejiang University, China, and Ph.D. degree from the University of Oulu, Finland, in 1982, 1985, and 1990, respectively, all in Electrical Engineering. He worked with Academy of Finland for the research on spread spectrum communications as a research associate during 1991–1993 and the National University of Singapore as a lecturer and then a senior lecturer from 1992 to 1997. He joined Department of Electrical Engineering, National Chung Hsing University, Taiwan, as an associate professor in 1997 and was promoted to a full-professor in 2000. In 2001, he moved to National Sun Yat-Sen University, Taiwan, as the founding director of the Institute of Communications Engineering of the University. Under his leadership, the institute was ranked the 2nd place in the country in terms of SCI journal publications and National Science Council funding per faculty in 2004. He has been a visiting professor to Department of Electrical Engineering, University of Kaiserslautern, Germany, in 1999, the Institute of Applied Physics, Tsukuba University, Japan, in 2000, and Institute of Experimental Mathematics, University of Essen, Germany in 2002. He is a recipient of numerous research and teaching awards from the National Science Council and Ministry of Education, Taiwan. He has authored or co-authored over 150 technical papers in major international journals and conferences, and four books and several book chapters in the areas of communications. He served as a symposium chair of major international conferences, including IEEE VTC, IEEE ICC, IEEE Globecom, IEEE WCNC, etc. He served or is serving as the editor or guest editor for IEEE Communications Magazine, IEEE Journal of Selected Areas in Communications, IEEE Networks Magazine, IEEE Wireless Communications Magazine, IEEE Transactions on Wireless Communications, IEEE Vehicular Technology Magazine, Wireless Communications and Mobile Computing Journal and International Journal of Communication Systems, etc. He is an honorary guest professor of Zhejiang University, China, and Shanghai Jiao-Tong University, China.