Multihop Cellular Networks: Potential Gains, Research Challenges, and a Resource Allocation Framework

Long Le, University of Waterloo
Ekram Hossain, University of Manitoba

ABSTRACT

Recently, there has been increasing interest in integrating multihop relaying functionalities into cellular wireless networks. Multihop cellular networks can potentially enhance coverage, data rates, QoS performance in terms of call blocking probability, bit error rate, as well as QoS fairness for different users. However, in-depth investigations and careful system designs are required to exploit these potential advantages. Specifically, routing and resource allocation algorithms should be designed such that the maximum performance gain can be achieved. A number of different architectures, protocols, and analytical models for MCNs have been proposed in the literature where different system aspects were investigated. This article aims to present an overview of existing work in this area, pointing out key research issues and their possible solutions. Also, we present a resource allocation framework for out-of-band relaying. The throughput enhancement due to the proposed framework is demonstrated through numerical results.

INTRODUCTION

The next-generation cellular wireless networks will support high data rates and provide quality of service (QoS) for multimedia applications with increased network capacity. Under limited frequency resources, the conventional approach to increase network capacity is to install more base stations (BSs) to exploit spatial reuse. This solution is not very efficient because the cost of the BS transceiver is quite high. An alternative approach is to employ relay stations (RSs) as intermediate nodes to establish multihop communication paths between mobile hosts (MHs) and their corresponding BSs. This has spurred increasing interest in developing new architectures and corresponding protocols for future-generation multihop cellular networks (MCNs) [1].

Existing architectures and protocols proposed for MCNs are very diverse and different in several aspects. RSs can be preinstalled by network operators [2, 3] or simply be other idle MHs who are not transmitting their own data [4–8]. Also, depending on how radio resources are allocated for routing paths of active connections, different protocols at the medium access control and routing layers can be designed. Radio resources for MHs at different hops may be allocated in time-division duplex (TDD) or frequency-division duplex (FDD) mode. Frequency bands other than the cellular frequency band may be used for relaying [7, 8]. Finally, advanced techniques using cooperative diversity [9–12] can be employed to enhance network performance compared to simple relaying schemes. In this article we present an overview of recent advances in MCNs and discuss several key research issues.

The rest of this article is organized as follows. We describe different design alternatives for MCNs and highlight their pros and cons. The major research issues on relaying and cooperative implementations for MCNs are described. We present a resource allocation framework using out-of-band relaying for which the throughput performance is demonstrated through typical numerical results. Conclusions are then stated.

MULTIHOP CELLULAR NETWORK DESIGN

RELAYING FOR LOAD BALANCING AND QOS FAIRNESS ENHANCEMENTS

For cellular networks, relaying was proposed in [2] to balance traffic load among highly loaded (*hot*) cells and lightly loaded (*cool*) cells. The authors proposed *primary relaying* and *secondary relaying schemes* as illustrated in Fig. 1. Here, it was assumed that each cell is assigned a finite number of channels, and preinstalled RSs are available to regulate traffic from hot cells to cool cells using transmissions in unlicensed frequency bands. Each RS is equipped with two air interfaces, a C (cellular) interface for communications with a BS, and an R (relaying) interface for communications with MHs or other RSs. Mobile hosts also have a C interface to communicate with a BS and R interface to communicate with RSs.

In a conventional system, if an MH wishes to establish a new call and cannot find an available channel in its home BS, it is blocked. In an MCN using primary relaying, this MH switches to its R

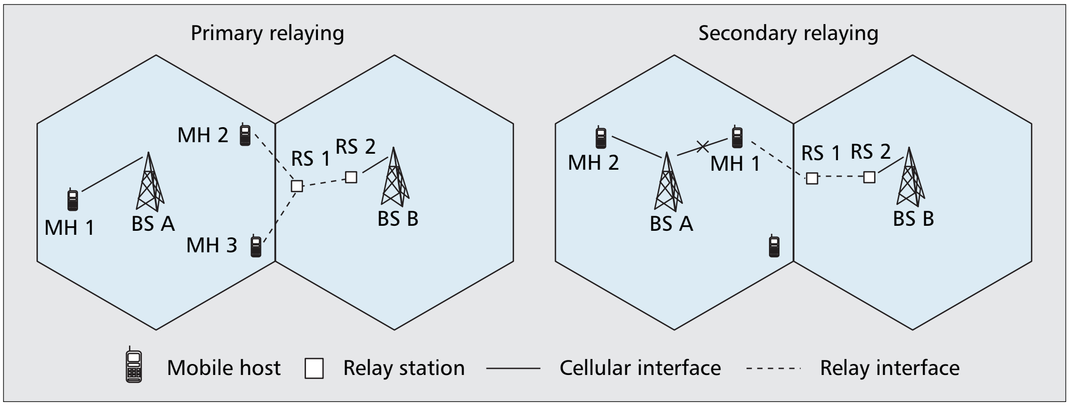

Figure 1. *Examples of primary and secondary relaying schemes.*

interface and establishes multihop communication with a neighboring BS through multiple RSs. In Fig. 1, if MH 2 cannot find an available channel in its congested BS A, it will try to communicate with the noncongested BS B through RSs 1 and 2. Here, RS 2 communicates with BS B by using its C interface on a channel allocated by BS B.

If primary relaying is not possible because a new call cannot be diverted from a congested cell to neighboring cells, the secondary relay scheme will be activated, as shown in Fig. 1. Here, a new call initiated by MH 2 could not be accommodated by either BS A or its neighboring cells using the primary relaying scheme. In this case the secondary relaying scheme is initiated as follows. An ongoing call from MH 1 may be diverted to BS B by using multihop connection through RSs 1 and 2. The channel allocated for MH 1 is released and re-allocated to MH 2.

The implementation of these relaying schemes was shown to reduce call blocking probability significantly from the case where no relaying scheme is employed, as in conventional cellular networks. These proposed relaying schemes also improve QoS fairness in terms of call blocking probability by balancing traffic among congested and noncongested cells. These schemes are most suitable for time-division multiple access (TDMA)-based cellular systems.

IN-BAND RELAYING VS. OUT-OF-BAND RELAYING

The major motivation for integrating multihop transmission in cellular networks is to enhance coverage and network capacity. Relaying can be used to assist communications to and from MHs at the cell edge or MHs experiencing deep fading in their home BS. An illustrative example is shown in Fig. 2 where four fixed RSs are installed at four street corners to provide radio coverage around the street corners due to the effect of shadowing on radio propagation through the buildings [1].

The capacity advantage of multihop relaying comes from the reduction of path loss due to the employment of multiple hops to transmit data to/from the corresponding BS compared to the one-hop transmission option in conventional cellular networks. However, it requires more radio resources to transmit data in different hops. Also, more interference is created due to a larger number of simultaneous transmissions in the network. The ultimate gain of multihop relaying, therefore, becomes unclear. In fact, it was shown that relaying is not always beneficial, especially if the target MH is close to the BS, and all RSs share common cellular bandwidth [4, 5]. Therefore, a smart resource allocation scheme and an adaptive implementation, where relaying is only employed if necessary, is important to achieve maximum capacity.

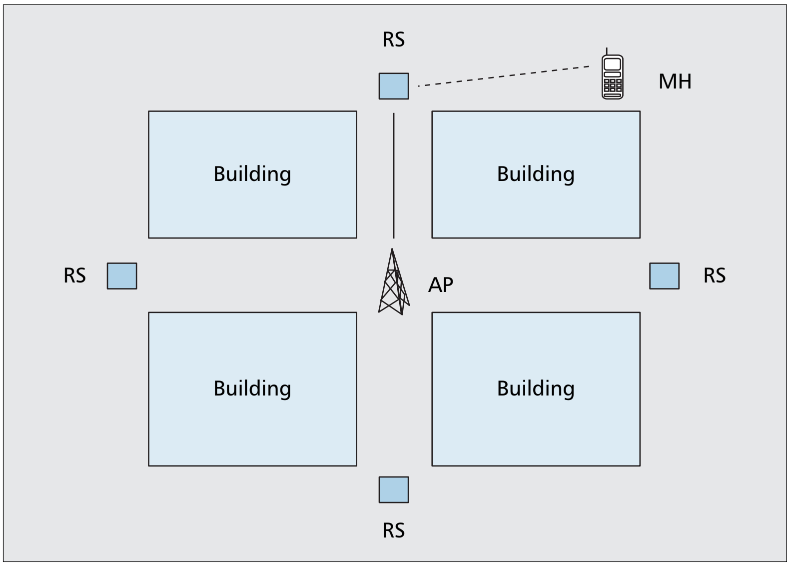

Figure 2. *Relaying for extending coverage around street corners.*

In fact, the multihop relaying method for cellular wireless networks was considered by the Third Generation Partnership Project (3GPP) under the name *opportunity-driven multiple access* (ODMA) [13]. In ODMA and similar relaying methods proposed in [4, 6], different hops on a routing path share the wireless channel (e.g., in code domain) in TDD mode (e.g., TDD code division-multiple access [TDD-CDMA]). We refer to this method as *in-band relaying*. The advantage of in-band relaying is that no modification of MHs is required, and MHs can serve as RSs if they are not transmitting their own data. Here, spatial reuse should be exploited so that the performance gain due to path loss reduction outweighs the capacity reduction due to multiple simultaneous transmissions on different hops.

In order to provide ubiquitous wireless ser-

Recently, cooperative diversity has emerged as an efficient way to achieve diversity gain through forming a virtual antenna array. The advantage of this type of cooperation is that each node needs only one antenna and a virtual antenna array is formed through multiple nodes in the network.

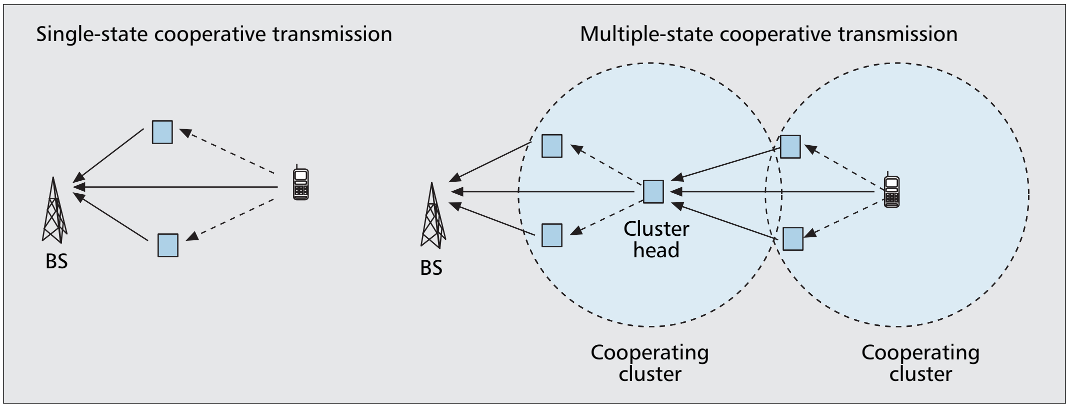

Figure 3. *Implementation of cooperative diversity in multihop cellular networks.*

vices, future MHs are likely to be equipped with multiple radios/interfaces to communicate with different wireless systems on different frequency bands. For example, an MH may have two interfaces, one for 3G cellular and the other for IEEE 802.11 (WiFi) networks. This implementation is called *out-of-band relaying* in the rest of this article. With this multiple-radio/interface capability, MHs can enjoy high data rates using the WiFi interface when they are in the coverage areas of WiFi hot spots. If an MH experiences bad channel conditions, its data can be relayed via other MHs by using the high-rate WiFi interface. It was shown that a significant performance gain in terms of outage probability can be achieved from this implementation even with one-hop relaying [7]. Similarly, the authors in [8] showed that ad hoc relaying can be employed to greatly enhance multicast throughput in cellular networks.

FIXED VERSUS MOBILE RELAY IMPLEMENTATIONS

Relay stations can be fixed pre-installed ones or simply normal MHs that are not transmitting their own data. Fixed RSs can be much cheaper than normal BSs because their function is just to decode received packets, then re-encode and forward them to the next station along the routing path. Fixed RSs can be installed in each edge between each pair of cells as proposed in [2] or in multiple rings in each cell centered by the corresponding BS. For mobile RSs, a significant performance gain can be achieved free of charge in low traffic load and high node density because many idle MHs are available to relay data from active MHs. The average power consumption of each MH increases due to this extra relaying functionality. However, it is expected that the increase in power consumption is not significant because the transmission range of each hop is now decreased. In high traffic load the performance gain may reduce because idle MHs are less likely to be available.

RELAYING VS. COOPERATIVE TRANSMISSION

Most existing work on MCNs focuses on developing relaying schemes where packets from the first station are forwarded along a routing path to reach the last station (i.e., the BS if in the uplink direction). The possible capacity gain comes from the decrease of path loss, and there-fore the increase of transmission rate on each hop over direct transmission.

Recently, cooperative diversity has emerged as an efficient way to achieve diversity gain through forming a virtual antenna array [9, 10]. The advantage of this type of cooperation is that each node (i.e., MH or RS) needs only one antenna, and a virtual antenna array is formed through multiple nodes in the network. Compared to a conventional multiple-input multiple-output (MIMO) setting, where each mobile node is equipped with multiple antennas, implementation of cooperative diversity may therefore be easier since it is difficult to install multiple antennas in a small mobile unit.

In a simple single-state cooperative transmission, a transmitting node employs several cooperating RSs to assist its transmission to the destination (e.g., BS if in the uplink direction as in Fig. 3). The initial work on cooperative diversity was done in [9] where the authors proposed a two-user cooperation strategy for CDMA cellular networks [9]. It was shown that significant performance gains in terms of capacity and/or coverage extension can be achieved. So far, the two most popular cooperative strategies are amplify-and-forward (AF) and decode-and-forward (DF) [11]. In [11] the authors proved that the same diversity order can be achieved by these schemes as in a conventional MIMO setup.

In the AF strategy the transmitting node broadcasts its signal in the first time slot. Relay nodes will amplify the signals they have received in the first time slot and forward them to the receiving node. The receiving node will combine the signals received in the first and second time slots to decode the signal (e.g., a simple method is to employ a maximum ratio combiner to form a decision variable at the receiving node). For the DF strategy, relay nodes first try to decode the signal they have received in the first time slot. Relay nodes that have successfully decoded the signal in the first time slot will re-encode the signal and forward it to the receiving node. The receiving node processes the signals it has received in both time slots to decode the message. In the DF scheme relay nodes could employ a distributed space-time code to transmit to the receiving node [10].

The relay nodes used in these transmission strategies can be active or inactive MHs. In low traffic conditions, inactive MHs can serve as relay nodes for their neighboring MHs. If the network is highly loaded, MHs can take turns serving as relay nodes for each other. As shown in Fig. 3, the cooperative transmission concept can be employed in single- or multiple-stage cooperative transmissions. For multiple-stage cooperative transmissions, data packets are forwarded through multiple clusters to reach the receiving node (i.e., BS in this figure). In each cluster/stage, a cluster head chooses several slaves (i.e., other MHs in the cluster) to perform cooperative transmission to another cluster head in the forward direction. The formation of cooperating clusters and routing paths are coordinated by the BS.

OPEN RESEARCH ISSUES

In this section we point out some key research issues related to designing and engineering MCNs. In particular, we describe research problems involved in both relaying and cooperative transmission strategies. The research problems center around challenges in developing routing and resource allocation schemes for MCNs.

RESEARCH ISSUES FOR RELAYING SCHEMES

The fundamental question in any relaying strategy in an MCN is how to perform joint resource allocation and routing such that maximum performance gains in terms of network capacity, coverage, and QoS performance can be achieved. Resource allocation depends on the physical layer design where either TDD or FDD is employed for transmissions on different hops of each routing path between an MH and its corresponding BS. It also depends on whether out-of-band relaying is employed or not, and how many radios (interfaces) each MH carries. Since all 3G cellular networks employ CDMA technologies, the network capacity is interference limited [14]. In general, resource allocation should be done such that the best tradeoff between spatial reuse gain and capacity reduction due to interference effects can be achieved.

Several existing routing algorithms proposed in the literature aim to minimize total transmission power or maximize the transmission rate on each routing path while ignoring interference due to concurrent transmissions on different hops and among different routing paths [15]. When the effects of interference are not considered, the optimum routing path and/or optimum number of hops can usually be found given high node density. These achievable capacity gains are, however, very optimistic and much higher than what could be achieved in real networks. When both intra- and intercell interference as well as self-interference on each routing path are taken into account, there is a tight coupling between the aggressiveness of spatial reuse for radio resource and the congestion level in the network [7, 14]. In fact, the congestion level of the network can be quantified through a Perron-Frobenius eigenvalue of the system path gain matrix [14]. Therefore, the design of a joint resource allocation and routing scheme should be done such that the congestion level is low enough and the desired QoS performance in terms of bit error rate (BER) or signal-to-interference-and-noise ratio (SINR) can be achieved.

There are two popular approaches to modeling interference in an MCN. In the first approach interference is explicitly captured by SINR, and the feasibility of a QoS constraint can be checked through the Perron-Frobenius eigenvalue of the channel gain matrix [7, 14]. This approach was employed to develop an interference-aware routing algorithm in [4]. In that paper the authors first obtained the minimum path loss routing solution. Then this initial routing solution was renavigated to find a routing path that improves the congestion level (i.e., interference level) in the network based on the Perron-Frobenius eigenvalue. Because two-hop relaying schemes could achieve a major portion of possible performance gains [5], limiting the number of hops to two may be a good design choice. In this case the routing problem degenerates into a relay selection one [7], which can simplify the protocol design and minimize the communication overhead significantly. We discuss this relay selection scheme further in the resource allocation framework in the next section.

For the second approach, the joint resource allocation and routing problem is solved by using a graph-theoretic approach [8]. In this approach transmission links that interfere with each other are assumed to be known (e.g., based on interference range). Given this information, only links that do not interfere with one another are allowed to be active (i.e., transmitting data) at the same time. Given a routing path for end-to-end data delivery (i.e., from the source node to the destination node), there is an optimal transmission schedule of minimum length where in each time slot of the schedule only noninterfering links are allowed to transmit. Thus, the joint resource allocation and routing problem is equivalent to finding routing paths for all active MHs and a transmission schedule such that the total number of time slots required to activate each link once on these routing paths is minimized. If all links in the network transmit at the same rate (i.e., single-rate transmission), the end-to-end throughput for each active MH is equal to the ratio between this transmission rate and the length of the schedule (i.e., the minimum number of time slots used in the schedule). If we map each time slot in the schedule to one color, the underlying problem is equivalent to a graph-coloring problem which is usually NP-hard [8]. Therefore, good polynomial-time heuristic algorithms with provable performance bounds are usually developed to solve the problem. The penalty of suboptimality is, however, quite high in many cases, which may ultimately result in very poor performance. For example, the algorithm proposed in [8] for the multicast problem achieves only a quarter of maximum throughput in the worst case, which may be unacceptable considering the potential gain due to multihop implementation.

RESEARCH ISSUES FOR COOPERATIVE TRANSMISSION SCHEMES

When cooperative diversity is employed, several research issues arise in different layers of the protocol stack. In particular, an efficient algorithm to find a routing path through multiple clusters should be constructed for end-to-end data transmission. In each cluster, a cluster head should choose several slaves to serve as gateways for cooperative transmission in the forward direction. Here, the resource allocation, clustering, and routing problems should be tackled jointly. As in the relaying schemes, interference should be carefully considered in solving this joint problem.

The fundamental question in any relaying strategy in an MCN is how to perform joint resource allocation and routing such that maximum performance gains in terms of network capacity, coverage, and QoS performances can be achieved.

When cooperative diversity is employed, several research issues arise in different layers of the protocol stack. In particular, an efficient algorithm to find a routing path through multiple clusters should be constructed for end-to-end data transmission.

At the physical layer, several design implementations can be considered to achieve the potential diversity gain from cooperative diversity. Specifically, a distributed space-time code or distributed phased array (beamforming) technique can be employed to realize the diversity gain [10]. Space-time code implementation, which is a specific implementation of the decode-and-forward scheme, is challenging because of the distributed nature of relay nodes. In addition, development of an optimal space-time code even in the traditional MIMO context is still an active research issue. For beamforming implementation, synchronization of simultaneous transmissions from multiple relay nodes for coherent summation of their signals at the receiving side is a very challenging task. Note that beamforming can be used to implement DF or AF schemes. If relay nodes transmit asynchronously, a sophisticated decoding technique should be employed. An example of a decoding scheme for asynchronous cooperative diversity can be found in [12] where a novel minimum mean squared error (MMSE) receiver was proposed for an ad hoc network setting.

RESOURCE ALLOCATION FRAMEWORK FOR OUT-OF-BAND RELAYING IN CDMA NETWORKS

In this section we present a resource allocation framework for out-of-band relaying in CDMA cellular networks. We restrict the framework to the scenario where each MH either transmits directly to the BS or relays its traffic through only one neighboring RS (i.e., two-hop relaying). In [5] it was shown that allowing routing paths longer than two hops results in only marginal improvement in system throughput, but may lead to significant implementation complexity as well as communication overhead. Under this assumption, a routing problem degenerates into a relay selection problem, which is easier to implement for an existing cellular network.

In fact, several relay selection schemes were proposed in [7] for this out-of-band relaying implementation. However, a single transmission rate was employed for all MHs. The efficacy of the relay selection schemes was demonstrated in terms of outage probability. This assumption has several limitations. First, most 3G cellular systems such as HDR and high-speed downlink packet access (HSDPA) employ link adaptation techniques where users adapt their transmission rates with channel conditions. Second, under the assumption of single-rate transmission, the throughput enhancement of the relaying schemes, which is one of the main motivations of multihop implementation for cellular networks, remains unknown. The extension to multirate transmission is, however, nontrivial and is the focus of this section.

SYSTEM MODEL

We assume that each MH has two radios and two interfaces. One interface is used to communicate with the BS using the cellular frequency band, and another is used for relaying purposes using the ad hoc frequency band. These two radios/interfaces are used simultaneously if a particular MH serves as an RS for another MH (i.e., one radio is used to receive data from the BS using the cellular interface, and the other is used to forward data to the MH using the ad hoc interface). We further assume that MHs can relay traffic for others when they are not transmitting their own data. In addition, data rates achieved by the ad hoc frequency band are assumed to be much higher than those achieved by the cellular frequency band. This is usually true because, for example, transmission rates up to 11 Mb/s and 54 Mb/s can be achieved, respectively, by the 802.11b and 802.11g interfaces, while the maximum data rate of the 3G HDR interface is only 2.4 Mb/s.

We consider downlink transmission in a cellular network with B BSs where the maximum transmission power for each BS is P_M. Analysis for the uplink case can be conducted in a similar manner with the power constraint at each MH [7]. Now, let P_c and P_{ci} denote the total transmission power used by the BS in cell c and transmission power used to transmit data for MH i in cell c, respectively. Also, let g_{ci} denote the channel gain from BS c to MH i and W be the total cellular bandwidth. We use the terms *link gain* and *channel gain* interchangeably. We assume that there are K transmission modes, and transmission mode k corresponds to transmission rate R_k ($k = 1, 2, \ldots, K$). The SINR at MH i in cell c can be written as

$$\gamma_i = \frac{W / R_{m_i} \cdot g_{ci} P_{ci}}{\eta (P_c - P_{ci}) + \sum_{b=1, b \neq c}^{B} g_{bi} P_b + N_0}, \tag{1}$$

where R_{m_i} is the transmission rate of user i who is using transmission mode m_i, B is the total number of cells in the network, N_0 is the additive white Gaussian noise (AWGN) power, and η is a factor capturing the imperfect orthogonality of spreading codes. In Eq. 1, the first term in the denominator denotes the intracell interference and the second term denotes the intercell interference. In order to achieve a desired BER level for transmission mode k, the SINR should be maintained higher than some particular SINR threshold Γ_k. Hence, if MH i uses transmission mode m_i at a particular time instant, we should have $\gamma_i \geq \Gamma m_i$. Using Eq. 1 and after some manipulations, we can write this constraint for active MHs in a matrix form as follows:

$$(\mathbf{I} - \mathbf{H})\mathbf{P} \succeq \mathbf{N}, \tag{2}$$

where $\succeq$ denotes the element-wise inequality, $\mathbf{I}$ is a $B \times B$ identity matrix, $\mathbf{P}$ and $\mathbf{N}$ are B-dimensional column vectors, and $\mathbf{H}$ is a $B \times B$ matrix where elements of $\mathbf{N}$ and $\mathbf{H}$ can be written, respectively, as follows:

$$\mathbf{H}_{cb} = \begin{cases} 0, & \text{if } c = b \\ \dfrac{\dfrac{1}{W / (R_{m_i} \Gamma_{m_i}) + \eta} \sum_{i \in B_C} \dfrac{g_{ib}}{g_{ic}}}{1 - \eta \sum_{i \in B_C} \dfrac{1}{W / (R_{m_i} \Gamma_{m_i}) + \eta}}, & \text{if } c \neq b \end{cases}$$

$$\mathbf{N}_c = \frac{\dfrac{N_0}{W / (R_{m_i} \Gamma_{m_i}) + \eta} \sum_{i \in B_C} \dfrac{g_{ib}}{g_{ic}}}{1 - \eta \sum_{i \in B_C} \dfrac{1}{W / (R_{m_i} \Gamma_{m_i}) + \eta}},$$

where B_c is the set of active MHs in cell c and $P = [P_1, P_2, \ldots, P_B]^T$. It is known that Eq. 2 has a non-negative solution if the Perron-Frobenius eigenvalue λ of matrix $\mathbf{H}$ satisfies the following condition [14]: $\lambda < 1$.

In this case the minimum power vector achieves the equality and $\mathbf{P} = (\mathbf{I} - \mathbf{H})^{-1}\mathbf{N}$ [7, 14]. In general, the Perron-Frobenius eigenvalue λ increases when the network becomes more congested (i.e., more active connections, and/or more connections use high transmission modes, and/or high transmission powers are used in many faded links). If $\lambda > 1$, the desired BER levels cannot be achieved regardless of the power levels in use. Now we are ready to present a resource allocation framework.

Resource Allocation Algorithm

The presented resource allocation framework in this subsection captures relay selection and rate/power control tasks. Relay selection aims at improving the data rate transmitted by the BS to an MH by relaying traffic via an idle MH that experiences a more favorable channel condition. In [7] the authors proposed several relay selection metrics where the scheme that chooses the MH with best link gain (ARLG) or low relative interference (ARRI) results in best performance in terms of outage probability. When these relay selection criteria are used in our framework, we have found that these two metrics achieve roughly the same throughput performance. Therefore, we only present relay selection based on the ARLG metric and the proposed rate/power control scheme.

The relay selection protocol works as follows [7]. An MH that wishes to establish a new connection with the corresponding BS periodically broadcasts an RS SOLICIT message which contains its own link gain from the BS. Other idle MHs in the network estimate their link gains from the BS. Upon receiving the RS SOLICIT message, an idle MH responds with an RS RESPONSE message to serve as the RS for the requesting MH if its link gain from the BS is better than that in the RS SOLICIT message. To resolve possible collisions, the link gain is partitioned into a finite number of intervals, each of which corresponds to one particular backoff value. The potential RS will defer its response for the corresponding backoff time. To gain higher priority, RSs with higher link gains have shorter backoff times. After the requesting MH receives the first RS RESPONSE message (i.e., from the idle MH with the highest link gain), it broadcasts an RS CANCEL message that contains the ID of the chosen RS to cancel other pending RS RESPONSE transmissions from other MHs. The BS also receives the RS CANCEL message and transmits data to the corresponding MH via the chosen RS.

After the relay selection protocol decides on the RSs (if any) for the active MHs, the rate/power control algorithm will be activated. In particular, the rate/power control algorithm attempts to increase transmission rates for all MHs in such a way that network throughput is maximized. It is assumed that link gains from all BSs to each MH are estimated by the corresponding MH and fed back to a controller in the network to perform the rate and power control. If an MH acts as an RS for another, the end-to-end performance for the corresponding connection is limited by the BS-RS link only (i.e., transmission rate on the link from the RS to the corresponding MH is assumed to be higher than that on the BS-RS link). The proposed rate/power control algorithm is described below. We denote the power vector at iteration t by $\mathbf{P}(t)$ where its bth element is $P_b(t)$. Also, we denote the temporary transmission power vector and the power increase at iteration t when transmission mode of MH i is increased by one, by $\mathbf{P}(i, t)$ and $\Delta\mathbf{P}(i, t)$, respectively.

Algorithm: Power and Rate Control for Minimum Total Power (MinP)

- Initialize transmission mode one for all active connections. Check whether this rate allocation is feasible or not (i.e., check the condition $\lambda < 1$). If yes, calculate the corresponding transmission power vector $\mathbf{P}(0)$ and check whether the maximum power constraint is satisfied (i.e., check the condition $P_b(0) < P_M$). If yes, go to step 2. Otherwise, declare an outage.
- For each active MH i, temporarily increase its transmission mode to the next one ($m_i := m_i + 1$) and check the rate feasibility condition. If yes, calculate the corresponding power vector $\mathbf{P}(i, t)$ for this iteration t and check the power constraint. If yes, calculate amount of power increase as $\Delta\mathbf{P}(i, t) = \Sigma_{b=1}^{B} [P_b(i, t) - P_b(t - 1)]$. If there exists at least one feasible rate update, go to the next step. Otherwise, keep transmission modes for all MHs the same as at the beginning of this step and finish.
- Find the MH that achieves minimum power increase, that is, $i^* = \arg\min \Delta\mathbf{P}(i, t)$. Update transmission mode for this MH i^* while keeping transmission modes of other MHs as at the beginning of step 2. Update the power allocation vector as $P(t) := P(i^*, t)$.
- Return to step 2 until no further rate increase is possible.

In each iteration this algorithm basically increases transmission rate for one MH, which results in minimum increase of network power. With minimum increase of network power in each iteration, the MinP algorithm aims to maximize the achievable network throughput. If there are M active MHs in each cell, the MinP algorithm requires $O(M^2B^2K)$feasibility checks and power calculations in the worst case where B is number of cells in the network, and K is the number of transmission modes. The power/rate control algorithm should be executed every time the link gains change; therefore, this complexity may be too high for many practical applications.

To reduce the computational complexity, we modify the searching criteria in steps 2 and 3 of the MinP algorithm as follows. In each iteration we search for the MH using the criterion $i^* = \arg\max g_{ic}/R_{m_i}$. The motivation for this metric is as follows. Since power increase for each mode advancement is higher for higher transmission mode (i.e., higher transmission rate), we should give MHs with low transmission rate higher priority. At the same time, we should favor the MHs with higher channel gains, which could potentially result in smaller power increase.

For beamforming implementation, synchronization of simultaneous transmissions from multiple relay nodes for coherent summation of their signals at the receiving side is a very challenging task.

The average data rate is obtained by averaging over all active MHs in cell zero and using 500 simulation runs. To keep the throughput gain conservative, we require channel gain from an RS to a requesting MH be larger than that from the BS to the RS.

Parameters	**Value**
CDMA bandwidth, W	5 MHz
Number of BSs, B	9
Number of transmission modes, K	6
Channel gain factor, K_0	10^8
Mobile active probability, P_a	0.2
Path loss exponent, κ	5
Standard deviation of log-normal fading, σ	8 dB
Channel orthogonal factor, η	0.5
Transmission rate of mode one, R_1	64 kb/s
AWGN noise power, N_0	10^{-5}mW
Connection BER requirement	10^{-3}
Maximum BS transmission power, P_M	100 mW

Table 1. *Simulation parameters.*

Therefore, we choose the metric that is the ratio between channel gain and transmission rate.

Now the algorithm works as follows. We temporarily update transmission mode for user i^* (i.e., $m_i^* := m_i^* + 1$) and perform feasibility check, power calculation, and power constraint check. If the rate increase for MH i^* passes both feasibility and power constraint checks, we retain this transmission mode update; otherwise, we decrease the transmission mode for this MH and remove MH i^* from the list of potential MHs in subsequent iterations. We repeat this procedure until no further rate increase is possible. We refer to this algorithm as *maximum channel gain and transmission rate ratio* (MGR). This algorithm requires $O(MBK)$feasibility checks and power calculations in the worst case, which is much lower than that of the MinP algorithm.

PERFORMANCE EVALUATION: NUMERICAL RESULTS

We evaluate throughput performance for the resource allocation framework with and without relaying. We consider a cellular network with nine rectangular cells where cell zero is in the center and eight other cells are immediate neighbors of cell zero. Channel gain is modeled as $g_{ic} = K_0.d_{ic}^{-\kappa}.10^{X_{ic}/10}$, where κ is the path loss exponent, d_{ic} is the distance from mobile host i and BS c, X_{ic} is a Gaussian random variable with zero mean and standard deviation σ, and K_0 is a factor capturing different system and transmission effects such as antenna gain and carrier frequency. Each cell in the network accommodates the same number of MHs whose locations are generated randomly in the cell. MHs employ M-ary quadrature amplitude modulation (M-QAM) with $K = 6$ transmission modes where the transmission rate for mode k satisfies $R_k = kR_1$. Each MH is active with probability $P_a = 0.2$. The SINR threshold corresponding to a desired BER level for transmission mode k is approximated as $\Gamma_k \approx -(2^k - 1) \times \ln(5 \times \mathrm{BER})/1.6$. The system and the channel parameters are summarized in Table 1. Here, we do not consider a power constraint for the radio operating on the ad hoc frequency band (i.e., an MH can serve as an RS if it is an immediate neighbor in the same cell of the active MH).

The average data rate is obtained by averaging over all active MHs in cell zero and using 500 simulation runs. To keep the throughput gain conservative, we require that the channel gain from an RS to a requesting MH be larger than that from the BS to the RS. The average data rates achieved by each MH in cell zero for both MinP and MGR algorithms with and without relaying are shown in Fig. 4. Here, NR and R stand for without and with relaying, respectively. Thus, R-MinP, for example, stands for relaying using MinP for rate/power control. This figure shows that significant throughput enhancement can be achieved by the multihop transmission scheme compared to direct transmission. In addition, the relative throughput enhancement due to multihop implementation increases with the number of users in each cell, although the data rate achieved by each MH decreases. This is because the network becomes more congested (i.e., more interference) when the number of users in the network increases. With higher user density, each active MH can find good RSs more easily, and therefore could achieve higher throughput in the end. It is also evident that the MGR algorithm trades implementation complexity for throughput, and the throughput enhancement of the relaying scheme for the MGR algorithm is less significant than that achieved by the MinP algorithm.

The throughput performances of different schemes for different values of cell size D are summarized in Table 2. The relative throughput improvements of relaying are also calculated for both MinP and MGR algorithms. It can be observed that when cell size increases, the throughput achieved by non-relaying schemes (especially the MinP algorithm) decreases significantly. When relaying is employed, throughput remains stable for both MinP and MGR algorithms for different values of cell size. This can be interpreted as follows. For larger cell size and nonrelaying implementation, BS power has be to increased to meet the desired QoS performance (i.e., desired BER level) for MHs at the cell edge that adversely impacts other MHs in the networks. In contrast, by relaying traffic through other RSs, transmission power at the BS can be significantly reduced from the non-relaying case, which ultimately improves system throughput. In addition, the relative throughput improvement due to relaying implementation is very significant for the MinP algorithm: more than 100 percent for $D = 2000$ m. The relative throughput enhancement for the MGR algorithm is almost 17 percent for $D = 2000$ m, which is much smaller than that achieved by the MinP algorithm.

CONCLUSIONS

We have presented an overview of multihop cellular technologies in this article. Potential gains of different multihop cellular architectures and the related research challenges have been described. In particular, we have argued that for 3G CDMA cellular networks, network capacity is interference limited; therefore, routing and resource allocation algorithms should take interference into account. We have also pointed out how the emerging cooperative diversity technique can be exploited in multihop cellular networks as well as related research issues. Finally, we have presented a resource allocation framework for an out-of-band two-hop relaying scheme. Numerical results have confirmed that significant throughput gain can be achieved through multihop implementation of CDMA cellular networks.

REFERENCES

[1] R. Pabst *et al.*, "Relay-Based Deployment Concepts for Wireless and Mobile Broadband Radio," *IEEE Commun. Mag.*, vol. 42, no. 9, Sept. 2004, pp. 80–89.

[2] H. Wu *et al.*, "Integrated Cellular and Ad Hoc Relaying Systems: iCAR," *IEEE JSAC*, vol. 19, no. 10, Oct. 2001, pp. 2105–15.

[3] Y. Lin *et al.*, "Integrated Radio Resource Allocation for Multihop Cellular Networks with Fixed Relay Stations," *IEEE JSAC*, vol. 24, no. 11, Nov. 2006, pp. 2137–46.

[4] T. Rouse, S. McLaughlin, and I. Band, "Congestion-based Routing Strategies in Multihop TDD-CDMA Networks," *IEEE JSAC*, vol. 23, no. 3, Mar. 2005, pp. 668–81.

[5] J. Cho and Z. J. Haas, "On the Throughput Enhancement of the Down-Stream Channel in Cellular Radio Networks through Multihop Relaying," *IEEE JSAC*, vol. 22, no. 7, Sept. 2004, pp. 1206–19.

[6] K. -D. Lee and V. C. M. Leung, "Slot-Splitting TDD-CDMA for Cellular Wireless Networks," to appear, *IEEE Trans. Wireless Commun.*, 2007.

[7] D. Zhao and T. D. Todd, "Cellular CDMA Capacity with Out-of-Band Multihop Relaying," *IEEE Trans. Mobile Comp.*, vol. 5, no. 2, Feb. 2006, pp. 170–78.

[8] R. Bhatia *et al.*, "ICAM: Integrated Cellular and Ad Hoc Multicast," *IEEE Trans. Mobile Comp.*, vol. 5, no. 8, Aug. 2006, pp. 1004–15.

[9] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity — Part I: System Description — Part II: Implementation Aspects and Performance Analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, Nov. 2003, pp. 1927–48.

[10] A. Bletsas and A. Lippman, "Implementing Cooperative Diversity Antenna Arrays with Commodity Hardware," *IEEE Commun. Mag.*, vol. 44, no. 12, Dec. 2006, pp. 33–40.

[11] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Network: Efficient Protocols and Outage Behavior," *IEEE Trans. Info. Theory*, vol. 50, no. 12, Dec. 2004, pp. 3062–80.

[12] S. Wei, D. L. Goeckel, and M. C. Valenti, "Asynchronous Cooperative Diversity," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, June 2006, pp. 1547–57.

[13] "Opportunity Driven Multiple Access," Tech. rep. 25.924, 3GPP, v. 1.0.0., Dec. 1999.

[14] S. V. Hanly, "Congestion Measures in DS-CDMA Networks," *IEEE Trans. Commun.*, vol. 47, no. 3, Mar. 1999, pp. 426–37.

[15] R. G. Cheng, S. M. Cheng, and P. Lin, "Power Efficient Routing Mechanism for ODMA Systems," *IEEE Vehic. Tech.*, vol. 55, no. 4, July 2006, pp. 1311–19.

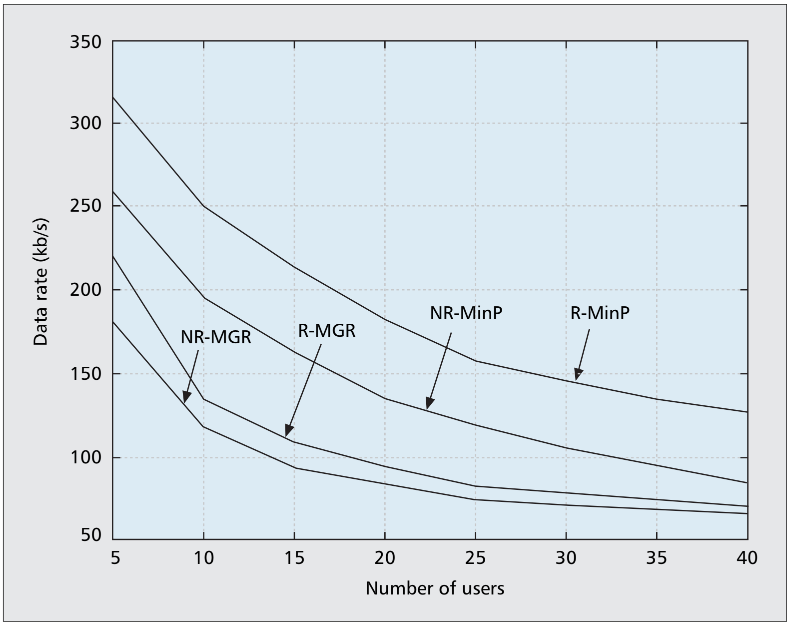

Figure 4. *Variation in average data rate per MH with the number of MHs in each cell with and without two-hop relaying for cell size (side to side) of* D *= 1000 m.*

		No relaying	With relaying	Improvement
D = 500 m	MinP	114.26 kb/s	143.87 kb/s	25.92 percent
	MGR	72.89 kb/s	74.71 kb/s	2.50 percent
D = 1000 m	MinP	101.45 kb/s	143.54 kb/s	41.49 percent
	MGR	70.43 kb/s	75.91 kb/s	7.78 percent
D = 2000 m	MinP	66.58 kb/s	138.87 kb/s	108.58 percent
	MGR	64.24 kb/s	75.03 kb/s	16.79 percent

Table 2. *Throughtput performance of different schemes* (M *= 30).*

BIOGRAPHIES

LONG LE [S'04, M'07] received a B.Eng. degree with highest distinction from Ho Chi Minh City University of Technology in 1999, an M.Eng. degree from the Asian Institute of Technology (AIT) in 2002, and a Ph.D. degree from the University of Manitoba in 2007. He is now a postdoc fellow in the Department of Electrical and Computer Engineering at the University of Waterloo. He was awarded the university gold medal in the undergraduate program, a Keikyu scholarship, a University of Manitoba graduate fellowship, an Edward R. Toporeck graduate fellowship in engineering, a University of Manitoba students' union scholarship, and IEEE student travel awards for IEEE WCNC 2003 and IEEE ICC 2005. His current research interests include network coding, link and transport layer protocol issues, cooperative diversity and relay networks, stochastic control, and cross-layer design for wireless communication networks.

EKRAM HOSSAIN [S'98, M'01, SM'06] is an associate professor in the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, Canada. He received his Ph.D. in electrical engineering from the University of Victoria, Canada, in 2000. He was a University of Victoria Fellow and also a recipient of the British Columbia Advanced Systems Institute (ASI) graduate student award. His current research interests include design, analysis, and optimization of wireless networks and cognitive radio systems. Currently he serves as an Editor for *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, *IEEE Wireless Communications*, *KICS/IEEE Journal of Communications and Networks*, *Wireless Communications and Mobile Computing Journal* (Wiley InterScience), and *International Journal of Sensor Networks* (IJSNet) (Inderscience Publishers). He served as a guest editor for special issues of *IEEE Communications Magazine* (Cross-Layer Protocol Engineering for Wireless Mobile Networks, Advances in Mobile Multimedia) and *IEEE Wireless Communications* (Radio Resource Management and Protocol Engineering for 802.16). He was a recipient of the Lucent Technologies, Inc. research award for his contribution to the IEEE International Conference on Personal Wireless Communications (ICPWC), 1997.