An Emerging Concept for 4G+ Wireless Cellular Networks: Terminal Relaying

Dr. Ramy H. Gohary and Dr. Halim Yanikomeroglu
Dept. of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada
Email: {gohary,halim}@sce.carleton.ca.

I. Synopsis

The technological foundation of fourth generation (4G) cellular systems was outlined in the Long Term Evolution Advanced (LTE-A) standards which enable communication rates of 100 Megabits per second for mobile communications and 1 Gigabit per second for nomadic ones. In LTE-A the deployment of wireless intermediate communication-assisting nodes, known as relays, was standardized for the first time. Under ideal conditions it has been envisioned that the deployment of relays will enhance the network coverage and will enable reliable communication at the cell periphery. Despite the envisioned gains of LTE-A based systems, practical field tests indicate that slight deviations from the ideal conditions can result in significant loss in achievable rates and service coverage.

We identify relaying as one of the most effective and enabling components of the LTE-A standard. However, this standard suffers from a major weakness that results from the assumptions that the relaying nodes are static, their locations are known and their number is fixed. While these assumptions facilitate the design of the cellular system, they limit the scope of relay usability. As an alternative, we consider a cellular system in which the wireless terminals (WTs), in addition to transmitting their own signals, act as relays to assist other WTs. In such a system the number of relays scales with the number of WTs offering the system designer significantly more degrees of freedom. However, using WTs as relays presents a number of challenges, including: the mobility of the relaying WTs, the incidental manner in which the relays access the network, the resource blocks (RBs) to which each relay is entitled, the relaying mechanism used by each relay, and the way in which the power is allocated across frequency bands.

Acronyms

For brevity, the following acronyms will be used throughout: Additive white Gaussian channel (AWGN), amplify-and-forward (AF), base-station (BS), channel quality indicator (CQI), channel state information (CSI), co-channel interference (CCI), compress-and-forward (CF), decode-and-forward (DF), frequency hopping (FH), iterative water-filling (IWF), joint scheduling and routing (JSR), Long Term Evolution-Advanced (LTE-A), multiple-input multiple-output (MIMO), orthogonal frequency division multiple access (OFDMA), quadrature amplitude modulation (QAM), quality-of-service (QoS), resource block (RB), wireless terminal (WT), x-th Generation (xG).

II. Background

The steep rise in the number of WT users worldwide has been increasing over the past decade. This is essentially due to the involvement of electronic and mobile commercial activities, multimedia streaming and security applications, which place pressing demands on the data rates and coverage requirements of wireless networks. In order to meet the rising demands, the LTE-A standards, laid out for 4G cellular networks, have considered the deployment of fixed relays to assist the communication between source and destination nodes. Despite the envisioned advantages of relay deployment, field tests of LTE-A based systems indicated that the actual rates and coverage capabilities of these systems are well below the prospective ones. It is suspected that the main cause of these deficiencies is the limited number of fixed relays that can be deployed in practice and the difficulty of optimizing their locations [12].

To mitigate the difficulties encountered by systems based on LTE-A standards, communication resources must be procured at a rate that scales with the rising demands. One approach to do that is to enable the WTs to act as relays, thereby providing a means for increasing the data rates and improving the coverage of future wireless networks. In contrast with the fixed relays proposed in LTE-A, terminal relays are typically battery operated with limited processing and transmission capabilities [12]. The location of relaying terminals cannot be controlled; they may be non-uniformly distributed and may access the network in an incidental fashion [7]. Furthermore, relaying WTs may operate under different premises regarding the number of available transmit and receive antennas and available CSI. Hence, the goal is to address the different aspects of terminal relaying, to provide practical solutions and useful insights into its potentials and limitations.

III. Emerging research directions

A. Direction 1: Generalization relaying with advanced signal processing

In order for a relay-assisted wireless communication network to achieve high data throughputs, the relaying mechanism ought to be adapted to the underlying channel conditions.

This work is supported by Research In Motion
978-1-4577-0069-9/11/$26.00 ©2011 Crown

For instance, for degraded channels in which the channel between the source and the relay is less noisy than the channel between the source and the destination, DF is the optimal relaying mechanism. In this mechanism, the relay decodes the signals received in each data block and transmits an assisting signal to enable the destination to unravel the uncertainty about the signal transmitted by the source in the previous block [1]. For general non-degraded channels, the DF mechanism can be highly suboptimal [3]. In fact, it was shown in [3] that in some cases simple AF relaying in which the relay transmits a linearly-processed version of its received signals can outperform sophisticated DF techniques.

In addition to maximizing the network throughput, when a sufficient number of relays is available, these relays can improve the QoS of the WTs and extend the coverage of the BSs. Using the WTs as relays provides a number of relays that scales with the demand. However, it introduces several design challenges. For instance, the optimal relaying strategy of each terminal may differ over time depending on the channel conditions. Furthermore, the modulation scheme, the data rate and the power allocated to each relaying terminal will also have to be adapted to the channel conditions. In particular, using a particular strategy at each relay terminal can be highly suboptimal, not only from a network throughput perspective, but also from computational and practical perspectives. For instance, DF imposes a stringent constraint on the relay to completely decode its received signals, CF relaying invokes a high computational complexity, and AF relaying does not distinguish between desired and noise components of the received signals.

In comparison with AF and DF, CF is more flexible and can be tailored to various channel conditions. In particular, in CF, the relay sends a compressed channel-dependent description of the received signal to the destination [1]. The flexibility of the CF relaying strategy can be enhanced by performing signal processing on the compressed description. For instance, by processing a compressed description of the received signal, the relay might be able to reduce the noise component in the signal it forwards to the destination, and hence alleviates the main drawback of AF relaying. Possible contributions in this direction are outlined in the following points.

- Generalization of DF-CF relaying schemes: Although this generalization was initially proposed in [1], its applicability and usefulness have not been explored. The study will be intended to analyze the impact of the correlation and power allocation among different signals at the source and relay. Our current study shows that there are operation regions in which the generalization outperforms both DF and CF. This occurs when the relay-destination link is better than the source-relay and the source-destination links; a typical scenario in terminal relaying.
- Design of CF quantization codebooks: The performance of the CF relaying scheme in practice hinges on the structure of the quantization codebook. To explore the gains that can be extracted from CF relaying, one can study the design and performance of three classes of codebooks, viz., codebooks that minimizing the distortion between relay received signal and quantization output, codebooks that maximize the mutual information between source and relay output signals, and codebooks that directly maximize the the mutual information between source and destination signals.
- Joint optimization of input covariance and relay precoder: Our preliminary results suggest that preprocessing of the source and relay transmitted signals can have a fundamental impact on the achievable rate of AF-based relaying systems. Hence, one approach is to develop efficient methodologies for designing the input covariance at the source and the precoder at the relay. The joint design of these modules will rely on a branch-and-bound approach in which the Karush-Kuhn-Tucker optimality conditions will be used to decompose the feasible set into non-overlapping regions. For each region, efficient algorithms will be developed for obtaining jointly optimal covariance-precoder pairs.

B. Direction II: Distributed power allocation in relay networks

In emerging cognitive radio communication systems multiple transmitter-receiver pairs share a common spectrum of several orthogonal narrowband tones. Each WT has a certain power budget and is interested in maximizing its own throughput. Treating the interference caused by other WTs as additive noise, each WT updates its power allocation across tones in order to maximize its throughput. However, by updating its power allocation the interference pattern observed by other WTs in the network change, which causes the WTs to update their power allocations accordingly.

For scalar Gaussian channels, the IWF algorithm [19] can be autonomously applied by the WTs to allocate their powers in a way that maximizes their instantaneous capacities. In IWF, each WT distributes its power across tones such that the sum of the power it allocates to any tone and the noise-plus-interference (NI) level on this tone is a constant, typically referred to as the water-level. Convergence analysis for this algorithm were developed for synchronous scalar [9] and asynchronous MIMO [13] systems. The case in which a malicious jammer attempts to reduce the system utility was studied in [4] and the case in which the NI levels available at the transmitters are inaccurate was studied in [5]. Surprisingly, it was shown in [5] that uncertainty about the NI levels can lead the wireless terminals to behave in a less selfish manner and to collaborate unintentionally to increase the network throughput.

Consider the case in which a terminal relay cooperates with a particular WT pair, but competes with the other WTs in the system. After estimating the NI levels on each tone, the WTs and the relays can update their power allocations in order to maximize their individual throughputs. To apply the IWF philosophy to terminal relay networks, the following routes can be investigated:

- Distributed power allocation algorithm: In IWF, the power allocation scheme that maximizes the instantaneous capacity of each wireless terminal pair is a generalized version of the standard water-filling algorithm [2]. For relay-assisted networks, one can develop a distributed power allocation strategy that maximizes its achievable rate. Such a strategy will also determine the role of each relay terminal on each tone. For example, the relay terminal may use a DF relaying strategy on some tones and an AF strategy on other tones.

- Game theoretic analysis: In practical communications, it is desirable for iterative algorithms to converge because after convergence, the WTs do not need to feed back the instantaneous NI levels and the relays and sources do not need to change their modulation schemes and rate allocations. To analyze the convergence of the prospective power allocation scheme, the competition between the terminal-assisted WT pairs can be modelled as a non-cooperative game wherein each terminal-assisted WT pair represents a player with a power-dependent set of feasible strategies. This representation will be used to derive bounds on the error norms in subsequent iterations of the algorithm. By assessing the evolution of the bounds, one can derive conditions under which the power allocation game has a unique Nash equilibrium to which the relay-assisted WTs are guaranteed to converge.
- Generalization to asynchronous operation and MIMO relay networks: The assumption that the WTs operate in a synchronous fashion simplifies the analysis, but is not realistic for future open-spectrum communication systems. Hence, it is interesting to extend prospective power-allocation algorithms to the case in which the WTs operate asynchronously. This algorithm can also be extended to the case in which the sources, relays and destinations have multiple antennas [13], [18].

C. Direction III: Autonomous medium access control

In terminal relaying systems the channels undergo rapid variations and the number of relays is large, which renders conventional RB assignment strategies [10], [6] cumbersome an inefficient. Hence, a more practical approach for terminal relaying is to autonomously assign the RBs to the WTs in the systems.

Most of the existing RB assignment techniques are extensible to relay-assisted communication scenarios. For instance, for OFDMA-based systems with fixed relays, the network throughput can be improved by using a static RB assignment technique known as Fractional Frequency Reuse [17]. Being static, this technique is not suitable for terminal relaying schemes in which the number and location of relaying terminals and the distribution of the wireless terminals vary with time. A more flexible RB assignment scheme is proposed in [6]. When this scheme is applied in relay-assisted networks, each relay chooses its initial assignments randomly and independently of other relays. Based on the CQIs of the wireless channel, the assignments are updated, either in a centralized or a distributed manner. In the absence of centralized coordination and CQIs, this strategy may result in undesirable instances at which one RB is used by many WTs while other RBs are not used. Such instances result in high, yet avoidable, intra-cell interference.

To address the RB assignment problem in OFDMA-based cellular systems with terminal relays, one can develop schemes that enable intra-cell interference to be minimized by reducing the number of hit occurrences; that is, occurrences at which an RB is assigned to multiple WTs. To suit terminal relaying applications, the developed schemes must be not only autonomous, but also computationally efficient with favourable spectral characteristics. Instead of deterministically partitioning the available RBs among the relays, the RBs must be assigned in an evolutionary fashion. In particular, each relay will be entitled to use as many RBs as needed; a feature that enables the relays to efficiently adapt to bursty traffic and nonuniform WT distributions without coordination. To develop the prospective algorithms, one can consider the following research directions:

- Generation of assignment patterns: In the absence of CQIs at the relays, good RB assignments should minimize the number of hit occurrences; that is, the occurrences at which one RB is assigned to multiple WTs when other RBs are not used. To generate such patterns, the set of available RBs can be endowed with a group structure that facilitates the parametrization of assignment sequences. If the group possesses a cyclic structure, one can be able to obtain optimal group-based assignment sequences by examining the group generators only.
- Hit source identification and avoidance: Suppose that each relay adopts an RB assignment pattern and that the relays are able to detect hits occurrences. In this case, if the relays are able to determine the relay with which a hit has occurred, they will be able to assign RBs to incoming WTs in a way that avoids future hits. The first step is hence for the relays to identify the hit sources. This can be done if the relays have access to the assignment sequences of other relays. It may be possible to design the assignment patterns, not only to minimize hit occurrences, but also to facilitate the identification of hit sources.
- Directly and indirectly identifiable hit sources: Upon successful completion of the previous two research directions, one can consider problems related to incorporating fading in the hit source identification and avoidance schemes. In particular, one can consider situations that arise when some hits cannot be detected due to fading and the impact of these occurrences on the identification of the hit sources and subsequent resource assignment.

D. Direction IV: Joint routing and scheduling for terminal relay networks

For a wireless cellular network to concurrently meet several design objectives, the scheduling of the transmissions of the WTs must ensure optimal exploitation of the channel conditions. For instance, channel-opportunistic schedulers that maximize the network throughput have been proposed in [15]. However, this class of schedulers does not ensure a minimum QoS for WTs with poor channel conditions; e.g., cell-edge WTs. To alleviate this drawback, scheduling techniques that account for fairness among the WTs have been developed [14]. Since both throughput and fairness are desirable objectives, they can be traded off to meet various WT requirements.

In addition to scheduling, the performance and coverage of cellular networks can be further improved by using relays; using relays it is possible to achieve spectrally efficient communication between source-destination pairs that are otherwise virtually disconnected. Despite the potential benefits of relays, their introduction in the wireless network gives rise to routing design challenges. For instance, for proper exploitation of available relays, an appropriate set of relays must be selected to serve each source-destination pair and an appropriate set of source-destination pairs must be selected to be served by each relay. Such selection problems can be formulated as NP-hard integer optimization problems. Preliminary results addressing these challenges appeared for particular relay-assisted network configurations in [16], [8].

To complement recent work in this area [11] in which only a few relays with fixed locations are available, one can investigate the design of JSR algorithms for wireless communication systems with terminal relays. Due to their mobility, their large number and the incidental manner in which they access the system, terminal relays present several fundamental challenges to the network design problem. The following set of challenges are of interest.

- Autonomous distributed JSR: In fixed-relay networks, the BS can acquire CSI from all nodes and hence, can make centralized decisions for JSR optimization. In terminal relay networks, the number of nodes is large, which renders providing the BS with CSI cumbersome and impractical. To address this difficulty, one can design distributed JSR schemes that maximize the network throughput without requiring feedback from the nodes. The prospective schemes will be based on efficient combinatoric optimization formulations, and relaxed and randomized versions thereof.
- Collaboration-based distributed JSR: For wireless networks with a moderate number of relays, the relays may be able to exchange a small number of bits with neighbouring nodes. To determine which information should be communicated, one can examine the Lagrange dual of the jointly optimal design of centralized scheduling and routing. It is anticipated that the coupling of the centralized design can be captured by the Lagrangian dual variables, which can be exchanged to achieve a close-to-optimal performance. one can also study the trade-off between the proposed autonomous and collaboration-based JSR approaches.
- Fairness-aware JSR: Fairness is a crucial aspect that must be considered in the joint design of scheduling and routing algorithms. In systems with fixed relays, global fairness can be approached using centralized techniques. However, in systems with terminal relays approaching global fairness is not feasible because nodes can only communicate with immediate neighbours; cf. the previous item. In this case, local fairness becomes a more appropriate design objective, which can be captured by maximizing the harmonic mean rate of each wireless terminal. The appropriate parameters to be exchanged between nodes can be determined using dualization techniques.

IV. Concluding remarks

Cooperative communication is an emerging technology in wireless communications. Exploiting the typical availability of a large number of unused mobile terminals, the research directions proposed will make key contributions to finding desirable cooperation schemes between relaying terminals. These directions are in line with the tremendous efforts in the research community around the world to pave the road for future machine cooperation and communication.

References

[1] T. M. Cover and A. A. El-Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, IT-25(5):572–584, Sept. 1979.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[3] A. A. El Gamal, M. Mohseni, and S. Zahedi. Bounds on capacity and minimum energy-per-bit for AWGN relay channels. *IEEE Trans. Inf. Theory*, 52(4):1545–1561, Apr. 2006.

[4] R. H. Gohary, Y. Huang, Z.-Q. Luo, and J.-S. Pang. A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer. *IEEE Trans. Signal Processing*, 57(7):2660–2674, July 2009.

[5] R. H. Gohary and T. J. Willink. Robust IWFA for open-spectrum communications. *IEEE Trans. Signal Processing*, 57(12):4964–4970, Dec. 2009.

[6] C. Koutsimanis and G. Fodor. A dynamic resource allocation scheme for guaranteed bit rate services in OFDMA networks. In *Proc. IEEE Int. Conf. Commun.*, Beijing, 2008.

[7] G. Kramer, I. Marić, and R. D. Yates. Cooperative communications. *Found. Trend. Commun. Inf. Theory*, 1(3–4):271–425, 2006.

[8] M. Lott, M. Weckerle, W. Zirwas, H. Li, and E. Schulz. Hierarchical cellular multihop networks. In *Proc. Euro. Pers. Mobile Commun. Conf.*, Glasgow, Apr. 2003.

[9] Z.-Q. Luo and J.-S. Pang. Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines. *EURASIP Appl. Signal Processing*, 2006(1):1–10, Apr. 2006.

[10] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Commun. Mag.*, 42(9):80–89, Sept. 2004.

[11] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim. Fairness-aware radio resource management in downlink OFDMA cellular relay networks. *IEEE Trans. Wireless Commun.*, 9(5):1628–1639, May 2010.

[12] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, E. Kim, and Y.-C. Cheong. An overview of radio resource management in relay-enhanced OFDMA-based networks. *IEEE Commun. Surv. Tutorial*, 12(3):422–438, 2010.

[13] G. Scutari, D. P. Palomar, and S. Barbarossa. Asynchronous iterative waterfilling for Gaussian frequency-selective interference channels. *IEEE Trans. Inf. Theory*, 54(7):2868–2878, July 2008.

[14] Z. Shen, J. G. Andrewa, and B. L. Evans. Adaptive resource allocation in multiuser OFDM system with proportional rate constraints. *IEEE Trans. Wireless Commun.*, 4(6):2726–2737, Nov. 2005.

[15] G. Song and Y. G. Li. Cross-layer optimization for OFDM wireless networks—Part I: Theoretical framework. *IEEE Trans. Wireless Commun.*, 4(2):614–624, Mar. 2005.

[16] V. Sreng, H. Yanikomeroglu, and D. Falconer. Capacity enhancement through two-hop relaying in cellular radio systems. In *Proc. IEEE Wireless Commun. Ntwk Conf. (WCNC)*, pages 881–885, Orlando, FL, Mar. 2002.

[17] A. L. Stolyar and H. Viswanathan. Self-organizing dynamic fractional frequency reuse in OFDMA systems. In *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, Phoenix, 2008.

[18] B. Wang, J. Zhang, and A. Høst-Madsen. On the capacity of MIMO relay channels. *IEEE Trans. Inf. Theory*, 51(1):20–43, Jan. 2005.

[19] W. Yu, G. Ginis, and J. M. Cioffi. Distributed multiuser power control for digital subscriber lines. *IEEE J. Select. Areas Commun.*, 20(5):1105–1115, June 2002.