Cooperation in Wireless Communication Networks

Weihua Zhuang and Muhammad Ismail, University of Waterloo, Canada

Abstract

The concept of cooperation in wireless communication networks has drawn significant attention recently from both academia and industry as it can be effective in addressing the performance limitations of wireless networks due to user mobility and the scarcity of network resources. In this article, we aim to shed some light on potential benefits of such an approach and discuss its challenging issues. We focus on three cooperation scenarios, namely, cooperation to improve channel reliability through spatial diversity, cooperation to improve throughput through resource aggregation, and cooperation to achieve seamless service provision. Challenging issues which arise at different layers of the network protocol stack are discussed, with an emphasis on the medium access control, network, and transport layers. We also present some future research directions in this area.

INTRODUCTION

Cooperation is the process of working together, opposite of working separately in competition. Recently, such a concept has been adopted from social sciences and economics to constitute a major research area in wireless communication networks. The idea of employing cooperation in wireless communication networks has emerged in response to the user mobility support and limited energy and radio spectrum resources, which pose challenges in the development of wireless communication networks and services in terms of capacity and performance.

Generally, we can categorize three cooperation scenarios based on various studies in literature. In the first scenario, cooperation among different entities is employed to improve the wireless communication channel reliability through spatial diversity [1], [2]. In the second scenario, the system throughput is improved via aggregating the offered resources from cooperating entities [3], [4]. Finally, cooperation is used to achieve seamless service provision [5] - [8]. Early research on cooperation in wireless communication networks focuses on developing strategies at the physical layer to support such a cooperative transmission. However, such a cooperative operation introduces challenging issues at different layers of the network protocol stack. Some modifications to the networking protocol stack are required to achieve the objectives of cooperation. In fact, without proper modification of networking protocols at the higher layers, the achieved cooperation gain may not be significant.

In this article, we aim to address the topic of cooperation in wireless communication networks through answering the following questions: What are the potential benefits of employing cooperation? What are challenging issues that arise at the different layers of the protocol stack to support cooperation and how can we handle them? What are open research issues?

POTENTIAL BENEFITS OF COOPERATION

In this section, we discuss the potential benefits of employing cooperation in wireless communication networks. These potential benefits are the main motivations of adopting such an approach.

IMPROVED CHANNEL RELIABILITY

Mitigating Channel Impairments

The wireless communication channel suffers from several phenomena that decrease its reliability. These phenomena include path loss, shadowing, and fading. Cooperation in wireless networks can increase the reliability of the communications against the channel impairments. This improved reliability can be achieved by exploiting cooperative spatial diversity [1], [2]. When the channel between the original source and destination is unreliable, other network entities can cooperate with the source node to create a virtual antenna array and forward the data towards the destination. Hence, different transmission paths with independent channel coefficients exist between the source and destination nodes through the cooperating entities. As a result, the destination node receives several copies of the transmitted signal over independent channels. Based on this spatial diversity, the destination can combine the data received from these entities in detection to improve the transmission accuracy. This concept is illustrated in Figure 1(a) for a downlink transmission from a base station to a mobile terminal, where the source node transmits its data packets towards the destination node with the help of cooperating entities. In this context, a cooperating entity is a relay node with an improved channel condition over the direct transmission channel from the source to the destination. This relay node can be a mobile terminal or a dedicated relay station as shown in Figure 1(a).

Interference Reduction

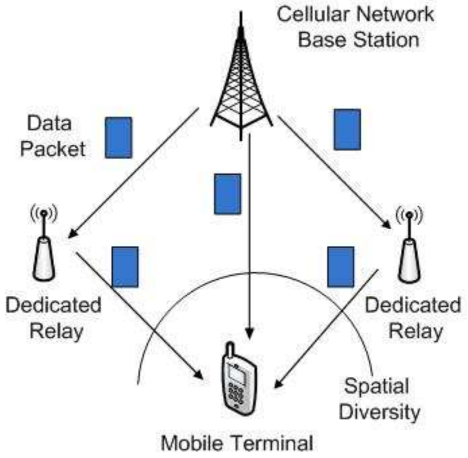
The broadcast nature of the wireless communication medium results in interference at the different nodes in the coverage area (interference region) of each other. Such interference reduces the signal to interference plus noise ratio (SINR) at the receiving nodes and hence degrades their detection performance. Thanks to the cooperation introduced by the cooperative relays, the transmitted power from the original source can be significantly reduced due to a better channel condition of the relaying links, which greatly reduces the interference region [9], as illustrated in Figure 1(b). This also helps to improve the energy efficiency of the communication system. In addition to reducing the interference region, cooperation can solve the hidden terminal problem and hence results in interference reduction [10].

Improved System Throughput

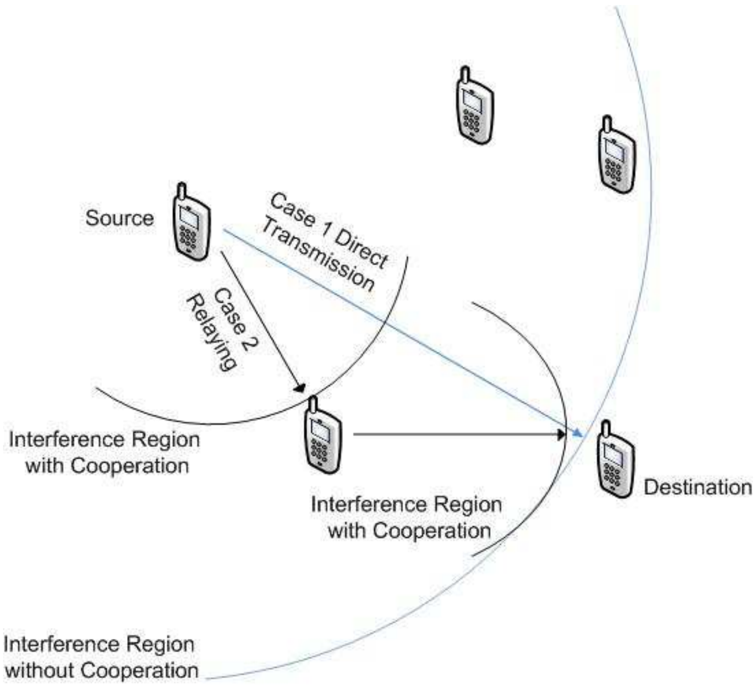
An improved system throughput can be a direct benifit from the enhanced wireless channel reliability through employing cooperative transmissions at the physical layer. In addition, co-operation can increase the achieved throughput through aggregating the offered resources from different cooperating entities [3], [4]. This is achieved through employing cooperative strategies at the network and transport layers. In this case, data packets are transmitted along multiple paths towards the destination. Different from the preceding cooperation scenario, the data packets transmitted through different paths are not the same copy of some transmitted signal. Instead, different transmission paths carry different data packets. This has the effect of increasing the total transmission data rate between the source and destination nodes. In this case, the cooperating entities can be mobile terminals, base stations or access points with sufficient resources (e.g. bandwidth), such that when these resources are aggregated, the total transmission data rate from the source to the destination can be increased. This strategy can support applications with a high required transmission rate. In Figure 2 for example, resources from the cooperating cellular network and wireless local area network (WLAN) are aggregated to provide a high data rate for the mobile terminal.

Seamless Service Provision

Mobile users are more sensitive to call dropping than call blocking. Call dropping interrupts service continuity for different reasons depending on the networking scenario. Cooperative



(a)



(b)

Fig. 1. Cooperation to improve channel reliability: (a) Spatial diversity; (b) Interference reduction.

strategies at the link, network and transport layers can help to guarantee service continuity of an ongoing call [5], [6]. In Figure 3, when the service is interrupted along one path (Ch1), it still can be continued using another cooperative path (Ch2,Ch3). In this context, a cooperating entity can be a mobile terminal, base station or access point which can create a substitute path between the source and destination nodes.

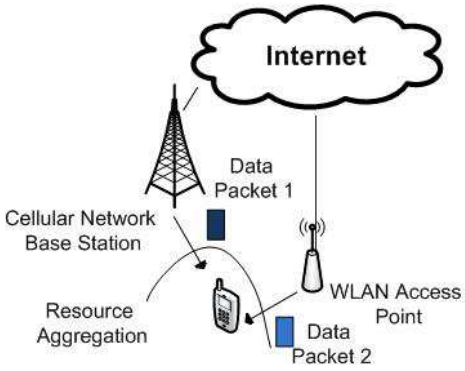


Fig. 2. Cooperative resource aggregation via a cellular network and a WLAN

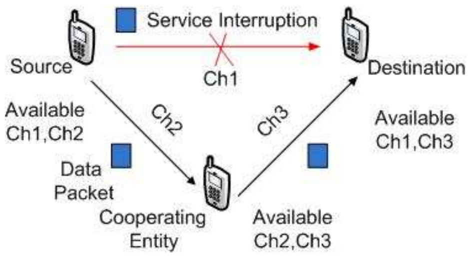


Fig. 3. Cooperation for seamless service provision

Operation Cost Reduction

Cooperation in wireless communication networks can reduce operation costs for both mobile users and service providers. For example, it can be achieved by improving the energy efficiency, which reduces the energy costs [11]. Another example is to extend network coverage area through relaying nodes. Such a solution is less expensive than deploying more base stations due to the high installation and maintenance costs.

Challenging Issues at Different Protocol Layers

While cooperation in wireless communication networks has various potential benefits, several challenging issues arise when such a cooperative operation is adopted. Modifications are required to the traditional roles played by different layers of the network protocol stack to support the cooperative strategy.

Cooperation can be employed to improve quality-of-service (QoS) through different means such as spatial diversity to improve channel reliability, resource aggregation to increase system throughput, and achieving seamless service provision. In the following, we discuss some challenging issues to achieve these objectives and modifications required at different layers of the protocol stack. In particular, we focus on issues at the medium access control (MAC), network, and transport layers, in addition to some discussion on the physical layer.

Challenges at the Physical Layer

In literature, there has been a great interest in developing different cooperative strategies to improve the reliability of a wireless communication channel through spatial diversity. These cooperative strategies include the amplify and forward (AF), decode and forward (DF), and coded cooperation (CC) methods. In these methods, the relay node either amplifies what it receives, as in AF, fully decodes, re-encodes, and retransmits the source message, as in DF, or integrates cooperative signaling with channel coding, as in CC. The methods are well explained in [1], [2], [12] and the transmission performance improvement is studied in terms of outage probability for a target data rate and symbol error rate in comparison with non-cooperative communication. However, the improvement is achieved at added complexity to the communication operation such as at the physical layer. One challenge that faces these strategies is that, for optimal decoding at the destination node, the channel coefficients between the source node and the cooperating entities should be known [2]. As a result, some techniques for exchanging and estimating the channel information must be considered in the implementation of these cooperative strategies. With user mobility, the channel coefficients become time variant. In this case, timely updates of channel coefficient information should be provided. Also, the cooperation adds hardware complexity at the destination nodes. When the destination node decodes a message from multiple transmitted packets from different cooperating entities, new modules should be added to the physical layer architecture. They include a sample buffer where symbols of previously received packets are stored and a combiner which merges the stored packets with the newly received packets for decoding [13]. If the newly received packets do not contain new coded symbols, as in both AF and DF with a repetition coding scheme, the sample buffer can store a linear combination of the previously received symbols regardless of how many copies are received. However, when new coded symbols are received, as in CC, the storage requirement of the sample buffer increases linearly with the number of received copies for that packet [13].

When cooperation is to improve the system throughput through resource aggregation or to achieve seamless service provision, one issue at the physical layer is that the nodes should be able to transmit (receive) different packets on multiple channels simultaneously. When these channels belong to different network technologies, a mobile terminal may be required to have multiple radio interfaces. In this case, multiple physical layers are employed in parallel to support multiple transmission technologies. With a single radio interface, discontiguous orthogonal frequency division multiplexing (D-OFDM) can be adopted as a physical layer transmission technique to enable the source (destination) node to transmit (receive) multiple packets on multiple frequency channels simultaneously [1], [14]. The source node modulates its data on the required channels by placing data and zeros in the appropriate inverse fast Fourier transform (IFFT) bins.

Challenges at the MAC Layer

Cooperation in wireless communication networks requires many unique features in the MAC layer design. The MAC layer plays a vital role on answering some questions related to such a cooperative strategy. These questions include: when to use cooperation, whom to cooperate with, and how to select the cooperating entities, and in the case of distributed selection of cooperating entities, how to reduce the interference and solve the hidden and exposed terminal problems in different cooperation scenarios [15] - [17]. For resource aggregation and seamless service provision, the questions related to when to use cooperation and whom to cooperate with are answered by either the MAC layer to define a layer 2 solution or the network layer to define a layer 3 solution.

- *When to Use Cooperation*

  Cooperation in wireless communication networks may not be beneficial or even necessary. This would be the case when cooperation gain in terms of throughput and energy efficiency for example is too small to compensate for its cost (complexity, signaling overhead, etc.). Hence, it is crucial to develop adaptive MAC protocols that use cooperation only when it is needed.

  Regardless of the cooperation scenario, control messages need to be exchanged among the source node, destination node, and a set of potential cooperating entities for selecting the cooperating entities and for coordinating the operation. The signaling overhead should be considered in making a decision on whether or not to use cooperation [16].

Whether or not cooperation is beneficial depends on the cooperation scenario. When cooperation is employed to improve the channel reliability, the decision is based on the achieved throughput of the source node. Hence, the achieved transmission rate with cooperation should be compared with the achieved transmission rate with direct transmission [15]. It has been proven in [15] that cooperation is beneficial only when the source-destination link has a low transmission rate and/or the payload length is sufficiently large compared to the signaling overhead for cooperation. Other factors should be considered for cooperation to improve throughput via resource aggregation. In this case, cooperation is employed when the resources available by direct transmission are not sufficient to satisfy the required service quality [3], [4]. For seamless service provision, cooperation however is imperative when the source and destination nodes are unable to communicate through the direct link [5], [6].
In general, the cooperation decision at the MAC layer depends on the instantaneous measurements for the channel quality and the achieved throughput. Hence, cross-layer design between the physical and MAC layers is required [15] - [17].

- *Optimal Selection of Cooperating Entities*

  Another vital role played by the MAC layer protocol is to select the optimal cooperating entities among all available entities. In this context, three factors affect the design of a cooperative MAC protocol: the number of the cooperating entities (single versus multiple selection), the employed mechanism for the cooperating entities selection (centralized versus distributed), and how to stimulate the optimal entities to cooperate.

  1) *Single versus Multiple Entities*

     A cooperative MAC protocol design should first deal with whether a single or multiple entities will be selected for cooperation. In a case of single entity cooperation, only the best cooperating entity is selected [15] - [17] such as in terms of effective transmission rate between the source and destination nodes. This has the attractive feature of simplicity of the selection operation. However, it may fail to meet the required QoS. To enhance the service quality, multiple entities can be involved in the cooperation [3], [18], [19]. Two issues should be considered from a MAC layer perspective when multiple cooperating entities are involved. First, the interference range will be enlarged proportionally to the number of cooperating entities, and it may affect the spatial frequency reuse in wireless networks [15]. In this case, the objective should be to minimize the number of selected cooperating entities and reducing the interference range, while satisfying the required QoS constraints [19]. The second issue is that more control signaling overhead is required for selecting and coordinating multiple cooperating entities [15]. This issue can be addressed through clustering of cooperating entities in order to keep the cooperation overhead manageable. It is also worth mentioning that, when cooperation is employed to improve channel reliability, having a small number of relay nodes over a large distance may be more effective than a large number of closely located relay nodes with correlated channel coefficients. Hence, in this case, spatially uncorrelated channel coefficients need to be considered in selecting multiple cooperating relay nodes.

2) *Centralized versus Distributed Entity Selection*

The selection of cooperating entities can be centralized or distributed. In a centralized selection, a central controller is responsible for selecting the optimal cooperating entities [18], [19]. Such a centralized mechanism often requires significant signaling overhead in terms of feedback messages from the cooperating entities to the central controller. The overhead can be significant when the channel state information from the potential cooperating entities changes with time. Such an approach has the advantage that the central controller has a global view of the network for an optimal selection. As a result, it can achieve a better performance gain than the distributed counterparts [20]. Yet, in some scenarios, it is infeasible to have such a central controller [3], [15] - [17]. Hence, without a central controller, how to find the optimal cooperating entities effectively and efficiently is vital to a practical MAC protocol. The first issue is how to identify the cooperation capabilities of the possible cooperating entities. A utility function can be defined to measure the ability of a potential cooperating entity. The utility function can incorporate several metrics including throughput and power consumption [21], [17]. The challenge is how to select the entity with the best cooperation capability in a distributed manner. A busy tone signal can be transmitted by each potential cooperating entity, whose duration is proportional to its cooperation ability. Hence, the best cooperating entity is the one having the longest busy tone signal [17]. However, one drawback of this approach is that the long duration busy tone consumes resources in terms of time and spectrum. It is more appropriate to make the optimal cooperating entity the fastest to win the channel. As a result, a timer based selection can be employed [15]. In this case, each possible cooperating entity maintains a timer and sets an initial value for the timer inversely proportional to its cooperation capabilities. Hence, the best cooperating entity exhausts its timer earlier than the other entities. Such a timer can represent the back-off time in a contention based MAC mechanism.

3) *Cooperation Incentives*

Whether or not to cooperate is a two-way decision: while the source and destination nodes decide whether or not cooperation is needed and try to select the best cooperating entities, these cooperating entities must have some incentives that stimulate them to participate. When different entities belong to different users/operators, they can choose not to cooperate. One reason is that the cooperative operation of a node results in an increase in its resource consumption (e.g. power, bandwidth, buffer space, which are likely to be limited). Hence, the assumption that all nodes will unconditionally participate in the cooperative operation is impractical. The design of the MAC protocol to select an optimal cooperating entity should result in a win-win situation for the source, destination, and cooperating entities. In this case, the allocated resources from the cooperating entities (which reflect their cooperation capabilities and affect the selection decision) should be based on some incentive scheme. In literature, cooperation incentive schemes can be reputation based or remuneration based [22] - [24]. A very effective tool to guarantee a win-win situation is provided through game theory. In this sense, all involved nodes will have an incentive (payoff) to participate in the cooperation process [23] - [25].

- *Solving the Hidden and Exposed Node Problems*

  A cooperating entity not only receives a packet from the source, but also helps to transmit the packet to the destination. In a distributed networking scenario, how to schedule the transmissions from the cooperating entities and their neighbors to avoid collisions is a challenging issue that should be addressed by the MAC protocol [17]. A ready-to-help (RTH) and clear-to-help (CTH) handshake can be employed to notify all the neighbors of cooperating entities to stop contention and hence avoid collisions [9], [15].

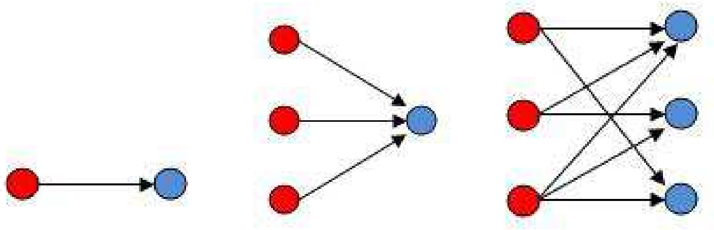
Challenges at the Network Layer

While the MAC layer protocol at different nodes is responsible for deciding whether or not cooperation is needed and for selecting the optimal cooperating entities among all available nodes, the network layer should define a routing protocol to deliver the data packets between the source and destination nodes using all cooperating entities.

When cooperation is employed in order to improve the channel reliability, several challenges need to be addressed by the network layer routing protocol, as discussed in the following.

New Link Definition - Traditionally, an end-to-end route specifies a sequence of intermediate node transmissions, originating from the source node and terminating at the destination node. Hence, the route in this definition is a series of point-to-point links. This is not the case in the context of cooperative communication, in which a cooperative link needs to be defined [13]. The cooperative link is composed of a set of transmitting nodes employing coordinated actions to deliver messages to a set of receiving nodes. Hence, a cooperative link can be viewed as a multi-terminal link as compared to the point-to-point link in a traditional network. In this case, a route from the source to the destination becomes a sequence of one or more cooperative links [13]. In literature, two types of cooperative links are defined. One is a multiple input single output (MISO) link, where a set of transmitting nodes coordinate their transmissions to a single receiving entity [26] - [27]. The other is a multiple input multiple output (MIMO) link, where a set of transmitting nodes coordinate their transmission to a set of receiving nodes [28]. The link definitions are illustrated in Figure 4. It is obvious that a cooperative MIMO link includes the cooperative MISO link as a special case, where the receiving set consists of only a single node. With this new link definition, the routing protocols at the network layer should be reinvestigated. The network layer routing protocol faces the challenge of constructing an optimal route (following some design criteria) which consists of a sequence of point-to-point, point-to-multipoint, cooperative MISO, and cooperative MIMO links.

Optimality versus Complexity - In cooperative communication, the routing problem can be viewed as a multi-stage decision making process [28]. At each stage, the decision is to select the transmitting and the receiving sets of nodes. A link cost can be defined between the transmitting and receiving sets. This link cost can be the total power needed for transmissions from the



(a) (b) (c)

Fig. 4. Link definitions in cooperative networks: (a) Point-to-point; (b) MISO; (c) MIMO.

transmitting to the receiving sets as in [27] - [28]. It can also incorporate the residual energy of the transmitting nodes. The optimal decision should minimize the summation of the links' cost along the route from the source node to the destination node. It is shown in [28] that, while this dynamic programming problem can develop an optimal route selection, it suffers from two shortcomings. On one hand, such a formulation has a high computational complexity. This is due to the fact that the optimal route can be a combination of the different link types. In fact, the optimal selection of a cooperative route can be computational intractable for an arbitrary network size, as its complexity grows exponentially with the number of available cooperating nodes [28]. Further, the optimal route selection is difficult to implement in a distributed manner because of the global information needed in its calculation. Hence, the challenge that faces a cooperative routing protocol designer is to develop a heuristic technique that selects a sub-optimal cooperative route, while reducing the associated computational complexity. In literature, some heuristic cooperative routing algorithms are based on the selection of a non cooperative shortest path route, while allowing the last l nodes within each hop along that route to cooperatively send the message to the next node [26]. It however does not take into consideration the effect of cooperative communications while constructing the route. To address this issue, a distributed cooperative routing algorithm is proposed in [26] using a cascade of any number of two building blocks, namely direct transmission (point-to-point link) and cooperative transmission (MISO link) blocks, without considering the general case of MIMO links.

Multi-flow Throughput - Another challenge in developing a cooperative routing protocol is the existence of multiple traffic flows in the network. When several nodes cooperatively transmit, network throughput may be affected as cooperation increases the interference in the presence of

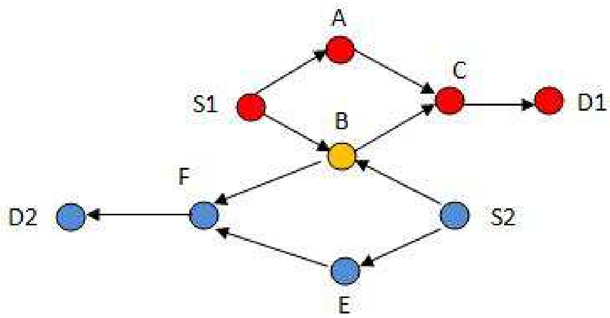


Fig. 5. Multi-flow interference [27]

multiple flows. This is illustrated in Figure 5, where interference takes place at node B between the flows from S1 to D1 and S2 to D2. As the number of flows increases in the network, the probability of collision will increase and hence the throughput will be significantly reduced. It is shown in [28] that, when there are 10 flows in the network, 40% savings in energy cost can be achieved while 50% of network throughput is lost. This issue is addressed in [27] where the cooperative routing protocol takes into consideration that links selected for different flows should not interfere with each other, using a contention graph approach, without including the cooperative MIMO links.

Similarly, when cooperation is to improve throughput by resource aggregation or to achieve seamless service provision, new challenges are introduced at the network layer which needs to implement multi-path routing [29]. In multi-path routing, the route between the source and destination nodes is composed of several transmission paths, each originating from the source node and terminating at the destination node. Some unique issues associated with the multi-path routing are discussed in the following.

Cost of Route Establishment and Maintenance - Multi-path routing can be classified to three categories. The first one is a node disjoint route, which has no nodes or links from different paths in common. The second category is a link disjoint route, which has no links in common, but may have nodes in common. The third category is a non-disjoint route, which can have nodes and links from different paths in common. The three categories are illustrated in Figure 6, where in Figure 6a paths SAD, SBD, and SCD have no links or nodes in common, in Figure 6b paths SABCD and SBD have node B in common, and in Figure 6c paths SAD and SABD

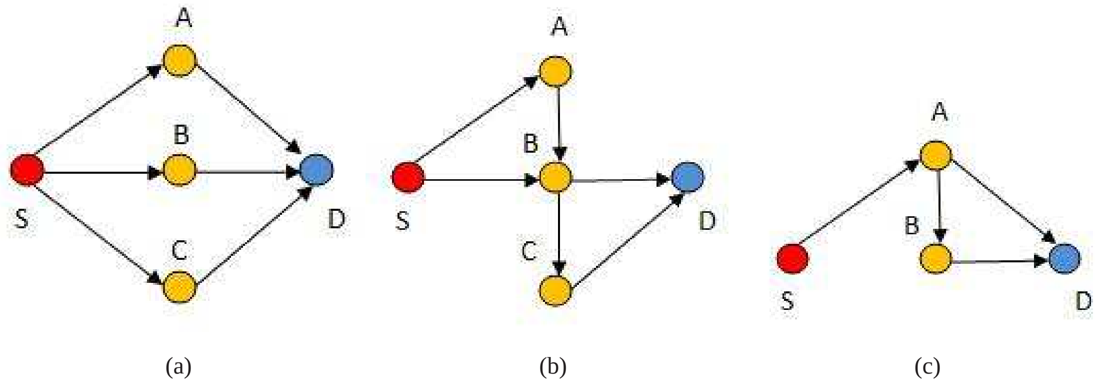


Fig. 6. Multi-path routing categories [30]: (a) Node disjoint; (b) Link disjoint; (c) Non disjoint.

have node A and link SA in common. While disjoint routes offer the most aggregate resources and provide the highest degree of fault tolerance, they are the most difficult to find due to the restrictions that they impose on the optimal route [30]. In addition to these restrictions, other QoS metrics need to be considered while establishing an optimal route. For example, the aggregated bandwidth should satisfy the required bandwidth while the delay of the longest path satisfies some delay constraint. This problem is shown to be NP-hard [31]. In the route establishment, the construction of this multi-path route should not consume much energy, which is very necessary for energy limited nodes [32]. As a result, heuristic algorithms should be developed with the consideration of the tradeoff among optimality, complexity, and energy consumption as in [31], [32]. Another important issue is related to the multi-path route maintenance. Route discovery can be triggered each time one of the paths fails or only after all routes fail. While initiating a route discovery every time a path fails can protect the QoS from degradation, it incurs a large amount of overhead. Also, waiting until all paths fail for route re-discovery may result in service interruption and throughput degradation. A good compromise is to perform route discovery when N paths fail, where N is less than the total number of available paths [30]. However, how to determine an optimal value of N to balance the QoS degradation and the amount of signaling overhead required for route re-discovery needs further studies.

Delay Differences - In multi-path routing, different paths may have different end-to-end delays, which causes delay jitter at the destination. This problem is commonly referred to as the differential delay problem [33]. With differential delays, packets can arrive at the destination out of order. As a result, the received packets from different paths need to be buffered for re-ordering. The limited buffer size at the destination node should be considered while establishing the optimal multi-path route. The multi-path routing algorithm should be carefully designed to optimize over both the achievable bandwidth and the differential delays. This issue is addressed in [33] for a wired network, where the routing algorithm has an exponential complexity in the order of the number of the paths. An extension to a wireless environment with a time varying link quality and network topology due to user mobility is required.

Multi-path Interference - When choosing an optimal multi-path route, it is necessary to select paths that are as mutually independent as possible to ensure minimal interference among these paths. While multi-path routing can aggregate resources from different paths, the achieved throughput can be severely degraded if these paths are in the interference range of each other. Multiple metrics are proposed in [30] to calculate the relative degree of independence among a set of paths. One metric which can be used for both disjoint and non-disjoint routes is the coupling metric. The coupling between two paths is defined as the average number of nodes that are blocked from receiving data along one of the paths when a node in another path is transmitting [30].

Challenges at the Transport Layer

When cooperation is employed to improve the channel reliability through spatial diversity, most of the work required to support such a cooperative operation is taken care of at the physical, MAC, and network layers. Hence, the transport layer is not much involved in supporting this cooperation. However, cooperation to improve the throughput through resource aggregation with multi-path routing needs to address the following challenges at the transport layer.

Multi-homing Capabilities - In resource aggregation, multi-path routing is employed. As a result, several IP addresses are used by the source and destination nodes. However, the traditional transport layer protocols, transmission control protocol (TCP) and user datagram protocol (UDP), are not designed to handle several IP addresses for a single node. TCP allows binding to only one network address at each end point [34]. A simple approach to deal with this issue is to use multiple TCP sockets between the end points, and use the application layer to perform striping of data packets among the paths. However, it is shown in [35] that this approach not only fails to achieve the aggregate data rate, but can result in the effective aggregate data rate being lower than the data rate of the slowest connection. As a result, this issue is better addressed by a transport layer protocol with a multi-homing feature. Multi-homing refers to the capability of a transport layer protocol to establish communication between end points with multiple IP addresses, and to make use of multiple paths between these end points [36]. In this context, stream control transmission protocol (SCTP) [37] is designed to support this multi-homing capability. SCTP multi-homing defines a transport layer connection as an association and allows for binding this association to multiple IP addresses at each end of the association. As a result, the source node is able to communicate with the destination node through different destination addresses [35].

Simultaneous Transmissions - While the SCTP standard allows for multi-homing capabilities, it does not support simultaneous transmission of data packets to multiple destination addresses. Instead, the SCTP standard defines a primary path for transmission of data packets while the other secondary paths are used only for retransmission of lost data packets or as a backup for the primary path. Hence, resource aggregation through the simultaneous use of multiple paths is not achieved. Several extensions of the SCTP protocol along with other protocols [34] - [36], [38] are proposed to address the challenging issues that arise when multiple paths are used for simultaneous transmissions of data packets, as discussed in the following.

1) *Path Assignment*

   With multiple paths, a key question now is how to assign transmission paths for the available data packets, i.e. how to determine the path for transmitting each packet. This assignment can be based on the bandwidth availability of the paths. It can be performed using information such as the congestion window and the round trip time (RTT) of each used path [35], [38]. Reassignment of data packets from a path (whose congestion window is reduced) to another path with sufficient resources should be handled by the transport layer [38]. Also, in order to improve the reliability of packet retransmission, the transport layer protocol should be able to retransmit the data packet on a different path from the one used for the original transmission [38]. Similarly for timed out paths, packets should be assigned to a different path from the one which they were originally timed out on [38]. Choosing the transport layer protocol to perform these tasks is due to the fact that the transport layer has the most accurate information of the used end-to-end paths. As shown in [35], it is not efficient to perform such path assignment in the application and link layers.

2) *Packet Reordering*

When the multiple paths used for simultaneous transmissions have different characteristics in terms of delay and bandwidth, the received packets arrive out of order at the destination. With the SCTP standard or TCP based protocols, out-of-order arrivals of data packets can unnecessarily initiate selective acknowledgment (SACK) transmission which are sent to the source node to report the missing packets. This can increase the network load, hence unnecessarily wastes network resources. Moreover, when duplicate SACKs are received at the source node with the same missing packets, they are interpreted as loss of these packets. As a result, unnecessary retransmission of these data packets takes place, which further wastes the network resources. In addition, the source node would reduce its congestion window, unnecessarily degrading the system throughput. Hence, the issue of packet reordering needs to be addressed at the transport layer to avoid these side effects. To simplify the handling of path related congestion control, path related sequence numbers and time stamps can be used [38]. This results in a separation of flow control from congestion control at the transport layer and can avoid degrading the system throughput. In this case, both the source and destination end points use an association buffer to hold the data chunks regardless of their transmission paths. The flow control is based on this association buffer. On the other hand, the source node has unique congestion control for each path [35], [38]. Also, the destination node should be able to differentiate between the reordering due to difference in the delay of the paths and that due to packet losses, in order not to unnecessarily generate SACKs to report out-of-order arrivals of data packets [34], [38]. To avoid the fast retransmission of out-of-order packets, the counter for data packets in the retransmission queue should increase only when packets with a higher sequence number are acknowledged for the path over which the data packets had been sent [34]. Hence, retransmission takes place when this counter reaches a predefined threshold.

Different challenges arise at the transport layer when cooperation is employed to achieve seamless service provision. When an original path becomes unavailable, an alternative cooperative path is used [5]. This can be viewed as a handoff from the original path to the alternative cooperative path. As here we are interested in cooperation for seamless service provision, we consider the situation where the connection is established on the new cooperative path before the original path becomes unavailable. That is, we consider a make before break connection for seamless service provision. When the new cooperative path has different characteristics in terms of bandwidth and delay from the original path, several issues should be considered for the transport layer protocol. We can distinguish four scenarios for such a handoff from the original path to the alternative cooperative path [39]. In the first scenario, the handoff takes place from a low delay path to a high delay path. The retransmission time out (RTO) timer will not be able to adapt to such a sudden increase in the RTT after the handoff. As a result, a TCP sender retransmits packets unnecessarily, which consumes the network resources. Also, the TCP sender reduces its congestion window, leading to an under-utilization of the new path. This issue is known as spurious RTOs [39]. The second scenario involves a handoff from a high delay path to a low delay path. This induces a problem of packet reordering, as the packets sent through the fast new link reach the receiver sooner than some packets sent through the slow path. In the third scenario, a handoff occurs from a high bandwidth delay product (BDP) path to a low BDP path. The TCP sender continues to transmit packets on the new path at the same high rate as the old path. This may result in buffer overflow of the new path and consequent packet losses. In the last scenario, a handoff is performed from a low BDP path to a high BDP path. The TCP sender will not be able to quickly adapt to the high sending rate capability of the new path. As a result, an inefficient utilization of the new path is expected. In literature, there exist different transport layer protocols to address these issues. Most of the proposed protocols are based on cross-layer design with the MAC layer, as they require the knowledge of the handoff event. We refer the interested reader to [39] for a discussion on the proposed solutions in literature to address the issues.

FUTURE DIRECTIONS

While there exist various works in literature to exploit the potential benefits of cooperation and address its challenge issues, many open issues remain to be further investigated.

Cooperation Overhead - While cooperation in wireless communication networks can improve network performance, such a cooperative communication can incur a considerable overhead. This overhead includes signaling and network control overhead for cooperating entities selection and coordination, additional required resources such as radio bandwidth for relay transmission and energy consumption at the cooperating entities. Another form of cooperation overhead is the incurred delay of the whole communication process which includes the time consumed in selecting the cooperating entities and establishing the cooperative paths. Finally, this cooperation overhead also includes the overall added complexity to the communication and networking process. The cooperation overhead affects the decision of whether or not cooperation should proceed. In literature, only signaling overhead for cooperating entities selection and coordination is considered in the decision process [17], [21]. Other forms of cooperation overhead should be appropriately modeled and taken into account in the cooperation decision. This is essential especially when cooperation is employed in a distributed manner. In this case, the analysis of the convergence rate towards the optimal selection of cooperating entities should be carried out, which affects the cooperation overhead in terms of potential performance loss due to time delay, control signaling, and complexity. Also, the overhead introduced by different network protocols at the MAC layer, network layer and transport layer should be modeled and considered in the cooperation decision. For example, the signaling together with additional time delay and complexity overhead introduced for establishing and maintaining cooperative routes for end-to-end information delivery should be considered in the cooperation decision. In addition to perform appropriate modeling for the cooperation overhead from different layers' perspectives and considering it in the cooperation decision, effective techniques for reducing this cooperation overhead should be developed. Cooperation overhead reduction will have a great impact on making cooperation in wireless communication networks feasible in practice.

Mobility - Supporting user mobility is one of the most prominent features of wireless communications. However, the impact of node mobility on the cooperative framework is not fully studied. Node mobility can greatly affect the cooperation decision and the selection of the cooperating entities. When cooperation is used to improve channel reliability through spatial diversity, the mobility of nodes can increase the correlation between the channel coefficients of the cooperating entities which reduces diversity gain. Hence, it is desired to select cooperating entities with independent channel coefficients. Also, with source and/or destination node mobility, the channel condition between them changes, leading to different decisions on whether or not cooperation is necessary for reliable communication. Similarly, in cooperation for resource aggregation or seamless service provision, mobility of cooperating nodes can affect the availability of resources. As the selection of cooperating entities depends on the network topology, the optimal set of the entities changes as nodes move. Further, the issue of mobility is not fully addressed in the context of cooperative routing. Most of the cooperative routing algorithms are originally proposed for static networks such as sensor networks [26] - [28]. They cannot be applied for mobile networks such as vehicular ad hoc networks (VANETs). The user mobility should be incorporated in developing the networking protocols to minimize cooperation overhead and, therefore, to maximize cooperation gain.

Cross-layer Design - In literature, cross-layer design for cooperative communication is mainly introduced between the physical and MAC layers. Cooperation decision at the MAC layer is based on channel measurements or estimates done at the physical layer [15] - [17]. In fact, cooperation calls for a tighter integration of the transport, network, MAC and physical layers. The cooperation decision should be based on cross-layer design among these layers, as it depends on the amount of cooperation overhead seen by these layers. Cross-layer design between the network and MAC layers can be beneficial to resolve the multi-flow throughput degradation issue with cooperative routing. Through appropriate coordination between the MAC and network layers, cooperative paths can be determined to avoid or minimize interference among multiple flows in the network. On the other hand, cross-layer design between the transport and network layers can help to address the packet reordering issue when cooperative resource aggregation is employed through multi-path routing. The multi-paths which constitute the route between the end nodes can be selected by the network layer based on the packet reordering requirements of the transport layer, based on the bounds of the differences among the RTTs of the selected paths. However, cross-layer design should be carefully studied as they can lead to negative effects on system performance [40].

New Cooperative Routing Protocol Metrics - Most of the cooperative routing algorithms are designed based on minimization of the total transmitted power [26] - [28]. This is because these algorithms are originally designed for power constrained networks, such as mobile ad hoc networks (MANETs) and sensor networks. As a result, the link cost captures only the transmitted power or the residual energy of the cooperating entities, which may not be very important for networks such as VANETs. New performance metrics should be considered for the link cost, including link lifetime, delay, and bandwidth. Multi-objective link cost which combines these metrics should also be considered.

Business Models - Accounting mechanisms and billing systems should be designed for a successful commercial deployment of the cooperative strategies in wireless networks. Business models need to facilitate different levels of cooperation. Three levels of cooperation can be distinguished, cooperation among service providers [11], cooperation among service providers and mobile users [3], and cooperation among mobile users [4]. Without appropriate business models for benefits of mobile users and service providers, it can be impossible to employ the cooperative techniques in practice.

CONCLUSION

Potential benefits of cooperation in wireless communication networks include: improved reliability and throughput, reduced service cost and energy consumption, and supporting seamless service provision. The benefits come with various challenging issues at different layers of the network protocol stack. Some approaches from literature to address the challenges are introduced. Further research is necessary to characterize cooperation overhead at different layers, taking account of user mobility. Also, cross-layer design among the transport, network, MAC and physical layers should be studied to improve cooperation performance gain. New metrics should be considered for cooperative routing. Finally, appropriate business models should be developed for successful commercial deployment of the cooperative wireless communications and networks.

REFERENCES

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. Information Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[2] A. Nosratinia, T. E. Hunter, and A. Hedayat, “Cooperative communication in wireless networks,” *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74-80, Oct. 2004.

[3] M. Ismail and W. Zhuang, “A distributed multi-service resource allocation algorithm in heterogeneous wireless access medium,” *IEEE J. Select. Areas Communications*, vol. 30, 2012.

[4] Q. Zhang, J. Jia, and J. Zhang, “Cooperative relay to improve diversity in cognitive radio networks,” *IEEE Communications Magazine*, vol. 47, no. 2, pp. 111-117, Feb. 2009.

[5] K. Letaief and W. Zhang, “Cooperative communications for cognitive radio networks,” *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878-893, May 2009.

[6] H. Choi and D. Cho, “On the use of ad hoc cooperation for seamless vertical handoff and its performance evaluation,” *Mobile Networks and Applications*, vol. 15, no. 5, Oct. 2010.

[7] W. Song and W. Zhuang, “Multi-service load sharing for resource management in the cellular/WLAN integrated network,” *IEEE Trans. Wireless Communications*, vol. 8, no. 2, pp. 725-735, Feb. 2009.

[8] W. Song, W. Zhuang, and Y. Cheng, “Load balancing for cellular/WLAN integrated networks,” *IEEE Network*, vol. 21, no. 1, pp. 27-33, Jan./Feb. 2007.

[9] T. Zhou, H. Sharif, M. Hempel, P. Mahasukhon, W. Wang, and T. Ma, "A novel adaptive distributed cooperative relaying MAC protocol for vehicular networks," *IEEE J. Select. Areas Communications*, vol. 29, no. 1, pp. 72 - 82, Jan. 2011.

[10] I. Akyildiza, B. Lo and R. Balakrishnana, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40-62, March 2011.

[11] M. Ismail and W. Zhuang, "Network cooperation for energy saving in green radio communications," *IEEE Wireless Communications*, vol. 19, 2012.

[12] T. E. Hunter, S. Sanayei, and A. Nosratinia, "Outage analysis of coded cooperation," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 375-391, Feb. 2006.

[13] G. Kramer, I. Maric, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*, vol. 1, no. 3, pp. 271-425, Aug. 2006.

[14] R. Rajbanshi, A. M. Wyglinski, and G. J. Minden, "An efficient implementation of NC-OFDM transceivers for cognitive radios," *Proc. 1st Intl. Conf. Cognitive Radio Oriented Wireless Net. Commun.*, pp. 1-5, June 2006.

[15] H. Shan, W. Zhuang, and Z. Wang, "Distributed cooperative MAC for multihop wireless networks," *IEEE Communications Magazine*, vol. 47, no. 2, pp. 126-33, Feb. 2009.

[16] H. Shan, H. Cheng, and W. Zhuang, "Cross-layer cooperative MAC protocol in distributed wireless networks," *IEEE Trans. Wireless Communications*, vol. 10, no. 8, pp. 2603-2615, Aug. 2011.

[17] H. Shan, P. Wang, W. Zhuang, and Z. Wang, "Cross-layer cooperative triple busy tone multiple access for wireless networks," *Proc. IEEE GLOBECOM*, pp. 1-5, Dec. 2008.

[18] Y. Li, P. Wang, D. Niyato, and W. Zhuang, "A dynamic relay selection scheme for mobile users in wireless relay networks," *Proc. IEEE INFOCOM Workshop*, 2011.

[19] Q. Du and X. Zhang, "QoS-aware base-station selections for distributed MIMO links in broadband wireless networks," *IEEE J. Select. Areas Communications*, vol. 29, no. 6, pp. 1123 - 1138, May 2011.

[20] Y. Yang, H. Hu, J. Xu and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100-105, Oct. 2009.

[21] Q. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," *Proc. IEEE NOMS*, pp. 144-151 , April 2008.

[22] I. Stanojev, O. Simeone, U. Spagnolini, Y. Bar-Ness, and R. L. Pickholtz, "Cooperative ARQ via auction-based spectrum leasing," *IEEE Trans. Communications*, vol. 58, no. 6, pp. 1843-1856, June 2010.

[23] Y. Yi, J. Zhang, Q. Zhang, T. Jiang, and J. Zhang, "Cooperative communication-aware spectrum leasing in cognitive radio networks," *Proc. IEEE DySPAN*, pp. 1-11, April 2010.

[24] F. Jiang, H. Tian and P. Zhang, "A user cooperation stimulating strategy based on cooperative game theory in cooperative relay networks," *EURASIP J. on Wireless Communications and Networking*, pp. 1-11, Aug. 2009.

[25] T. Chen, L. Zhu, F. Wu, and S. Zhong, "Stimulating cooperation in vehicular ad hoc networks: A coalitional game theoretic approach," *IEEE Trans. Vehicular Technology*, vol. 60, no. 2, pp. 566 - 579, Feb. 2011.

[26] A. Ibrahim, Z. Han, K. Liu, and S. Zhong, "Distributed energy-efficient cooperative routing in wireless networks," *IEEE Trans. Wireless Communications*, vol. 7, no. 10, pp. 3930-3941, Oct. 2008.

[27] J. Zhang and J. Zhang, "Cooperative routing in multi-source multi-sestination multi-hop wireless networks," *Proc. IEEE INFOCOM*, pp. 2369-2377 , April 2008.

[28] M. Dehghan and M. Ghaderi, "Energy effcient cooperative routing in wireless networks," *CPSC Technical Report*, 2009. In Performance Tools and Applications to Networked Systems, pp. 209-234, 2004.

[29] P. Thulasiramanz, J. Chen, and X. Shen, “Multipath routing and max-min fair QoS provisioning under interference constraints in wireless multihop networks,” *IEEE Trans. Parallel and Distributed Systems*, Vol. 22, No. 5, pp. 716-728, 2011.

[30] S. Mueller, R. Tsang, and D. Ghosal, “Multipath routing in mobile ad hoc networks: issues and challenges,” *Performance Tools and Applications to Networked Systems*, pp. 209-234, 2004.

[31] F. Wang, Z. Wang, Y. Li, and L. Zeng, “Reliable multi-path routing with bandwidth and delay constraints,” *Proc. 2010 Int’l Conf. Multimedia Technology*, pp. 1-6 , Oct. 2010.

[32] L. He, “Efficient multi-path routing in wireless sensor networks,” *2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, pp. 1-4 , Sept. 2010.

[33] X. Chen, M. Chamania, A. Jukan, A. Drummond, and N. da Fonseca, “QoS-constrained multi-path routing for high-end network applications,” *Proc. IEEE INFOCOM Workshops*, pp. 1-6 , April 2009.

[34] T. Chen, L. Zhu, F. Wu, and S. Zhong, “Concurrent multipath transfer using SCTP multihoming over independent end-to-end paths,” *IEEE/ACM Trans. Networking*, vol. 14, no. 5, pp. 951 - 964, Oct. 2006.

[35] H. Hsieh and R. Sivakumar, “A transport layer approach for achieving aggregate bandwidths on multi-homed mobile hosts,” *Wireless Networks*, vol. 11, no. 2, pp. 99-114, Jan. 2005.

[36] A. Jungmaier and E. Rathgeb, “On SCTP multi-homing performance,” *Telecommunication Systems*, vol. 31, no. 2, pp. 141-161, Feb. 2006.

[37] R. Stewart, Q. Xie, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang and V. Paxson “Stream control transmission protocol,” *RFC - 2960*, Oct. 2000.

[38] A. Abd El Al, “LS-SCTP: a bandwidth aggregation technique for stream control transmission protocol,” *Computer Communications*, vol. 27, no. 10, pp. 1012-1024, June 2004.

[39] L. Daniel, “TCP performance with vertical handoff,” *PhD Thesis, University of Helisinki*, Nov. 2008.

[40] V. Kawadia and P. R. Kumar, “A cautionary perspective on cross-layer design,” *IEEE Wireless Communications*, vol. 12, no. 1, pp. 3-11, Feb. 2005.