rDCF: A Relay-Enabled Medium Access Control Protocol for Wireless Ad Hoc Networks

Hao Zhu, *Member*, *IEEE*, and Guohong Cao, *Member*, *IEEE*

Abstract—It is well known that IEEE 802.11 provides a physical layer multirate capability and, hence, MAC layer mechanisms are needed to exploit this capability. Several solutions have been proposed to achieve this goal. However, these solutions only consider how to exploit good channel quality for the direct link between the sender and the receiver. Since IEEE 802.11 supports multiple transmission rates in response to different channel conditions, data packets may be delivered faster through a relay node than through the direct link if the direct link has low quality and low rate. In this paper, we propose a novel MAC layer relay-enabled distributed coordination function (DCF) protocol, called rDCF, to further exploit the physical layer multirate capability. We design a protocol to assist the sender, the relay node, and the receiver to reach an agreement on which data rate to use and whether to transmit the data through a relay node. Considering various issues, such as, bandwidth utilization, channel errors, and security, we propose techniques to further improve the performance of rDCF. Simulation results show that rDCF can significantly reduce the packet delay, improve the system throughput, and alleviate the impact of channel errors on fairness.

Index Terms—IEEE 802.11, MAC, wireless networks.

1 INTRODUCTION

With the advantage of low cost and high data rate, IEEE 802.11-based wireless networks are becoming extremely popular. In order to improve the network performance, it is fundamentally important to design good media access control (MAC) protocols to efficiently utilize the limited spectrum [2], [6], [22], [24]. Two different MAC mechanisms are supported by the IEEE 802.11 standard [11]: One is called *distributed coordination function* (DCF), which is based on carrier-sense multiple access with collision avoidance. With DCF, the mobile nodes can spontaneously form an ad hoc network without any preinstalled infrastructure. Such networks can be quickly deployed in civilian and military environments, such as battlefield, disaster recovery, group conference, and wireless office; the other is called *point coordination function* (PCF), which is based on polling and is built on the top of DCF. The PCF protocol has not yet been commercialized [12].

IEEE 802.11 has physical-layer multirate capability [11], which means that data can be transmitted at a number of rates according to the channel condition. For example, when the signal-to-noise ratio (SNR) is high, i.e., error detection and recovery is not that important [9], an aggressive and efficient modulation scheme can be applied to increase the rate. When the SNR is low, a conservative and redundant modulation scheme should be applied to reduce the bit error rate. In practice, IEEE 802.11b supports transmission rates of 1, 2, 5.5, and 11 Mbps and IEEE 802.11a supports transmission rates of 6, 9, 12, 18, ..., 54 Mbps [9], [21].

To exploit the physical layer multirate capability, researchers have proposed various protocols. At the network layer, some channel state aware routing schemes [6], [2], [22] have been studied to improve the end-to-end throughput by taking into account the channel condition as one of the route selection metrics. However, due to the long latency of route updates and the high control overhead, these schemes cannot quickly react to dynamic channel condition and can not achieve high bandwidth utilization. At the MAC layer, [9], [14], [21] have been proposed to exploit the multirate capability. The basic idea of these schemes is to let the sender select a proper transmission rate according to the history of the successful transmissions or to let the receiver sense the channel condition before the transmission and notify the sender via a control packet (e.g., the clear-to-send (CTS) packet). However, these schemes only utilize the data rate of the direct link between the sender and the receiver. In many cases, data may be delivered much faster through multiple links that have high transmission rates than through the direct link with a low transmission rate.

In this paper, we propose a novel DCF-based MAC protocol called *relay-enabled DCF* (rDCF) to further exploit the multirate capability of IEEE 802.11. Based on the channel condition among mobile nodes, rDCF can intelligently apply multihop (mainly two-hop in this paper) data transmission to achieve higher transmission rate. Specifically, when the direct link between the sender and the receiver can only support a low transmission rate, but there exists a relay node such that both the links from the sender to the relay node and from the relay node to the receiver can support high transmission rates, the impending packet can be delivered from the sender to the receiver faster by

- *H. Zhu is with the Department of Electrical and Computer Engineering, Florida International University, 10555 W. Flagler St., Miami, FL, 33174. E-mail: hao.zhu@fiu.edu.*
- *G. Cao is with the Department of Computer Science and Engineering, The Pennsylvania State University, 354G Information Sciences and Technology Building, University Park, PA 16802. E-mail: gcao@cse.psu.edu.*

Manuscript received 20 Dec. 2004; revised 20 July 2005; accepted 1 Aug. 2005; published online 17 July 2006.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-0350-1204.

two-hop high speed transmission via the relay node. With rDCF, each mobile node senses the channel conditions among its neighbor nodes. Based on the collected channel conditions, if a node can become a relay node of its neighbors, it periodically advertises the relay information. When the sender sends the packet to the receiver, if it can find a relay node, a triangular handshake is formed among the sender, the relay node, and the receiver so that they can quickly agree on whether to perform relay and which rate to use according to the real-time channel condition. To deal with issues such as bandwidth utilization, time-varying channel condition, and security, we propose techniques to enhance the rDCF protocol. We carefully analyze the rDCF protocol and conclude that rDCF can improve system performance without incurring severely negative impacts and strict requirements of node density. We evaluate the rDCF protocol in various scenarios, and the simulation results show that rDCF can significantly reduce the packet delay, improve the system throughput, and alleviate the impact of channel errors on fairness.

The remainder of this paper is organized as follows: Section 2 describes the background and the related work. Section 3 gives the motivation of the work. The details of rDCF are presented in Section 4. Section 5 analyzes rDCF. Section 6 evaluates the performance of rDCF through simulations. Section 7 concludes the paper.

2 Background and Related Work

2.1 System Model

We consider a wireless network based on IEEE 802.11b that can support transmission rates of 1, 2, 5.5, and 11 Mbps. The wireless medium is shared among multiple contending mobile nodes, i.e., a single physical channel is available for wireless transmission. The DCF with request-to-send (RTS)/clear-to-send (CTS) handshake is used for medium access control since it has been shown that the RTS/CTS mechanism is effective to solve the hidden terminal problem [3] and to improve the system performance when the packet size is large [4]. According to the channel condition, a packet could be transmitted at different transmission rates. We assume that data packets can be transmitted at different transmission rates, but control packets (e.g., RTS, CTS, ACK) are transmitted with the base rate, which is 2 Mbps in this paper. For simplicity, we assume that each node transmits its packets using a constant transmission power. The wireless channel between the sender and the receiver is assumed to be almost symmetric. In this paper, we will not consider the motivation for nodes to relay. Many techniques [5], [17] can be used to address the motivation for relay.

Based on the distance, the sensing power, and the modulation scheme, a node can be in different range of the sender: the *transmission range* and the *carrier sensing range*.

- *Transmission range*: Within this range, the node can receive and correctly decode the packet.
- *Carrier sensing range*: Within this range, the node can sense the signal, but cannot decode the packet.

2.2 The IEEE 802.11 DCF Protocol

The standard DCF protocol is described in [11]. After a transmitting node senses an idle channel for a time period of a *distributed interframe space* (DIFS), it backs off for a time period which is chosen uniformly from the range of 0 to its contention window size (CW). After each successful data transmission, the window size is set to CW_{min}, which denotes the prespecified minimum contention window. After the backoff timer expires, the node sends an RTS to the receiver. If the receiver successfully receives the RTS, it replies a CTS after a time period of *short interframe space* (SIFS). When the sender receives the CTS, it transmits the impending packet. For the purpose of reliability, the receiver needs to reply an ACK after it receives the packet correctly. Any other node overhearing either the RTS or the CTS extracts the information contained in the packet and updates its *network allocation vector* (NAV), which indicates the time period reserved for data transmissions. Then, the node defers its transmission until its NAV expires. For each transmission failure, which may be caused by collisions or channel errors, a binary exponential backoff is applied to double the backoff window and the window size is bounded by the maximum contention window (denoted by CW_{max}).

2.3 Related Work

Kamerman and Monteban [14] designed the auto rate fallback (ARF) protocol to utilize the multirate feature of IEEE 802.11. In ARF, the sender adapts the rate of each data transmission based on the history of previous successful transmissions. Since ARF is a sender-initiated protocol, it does not work well when the channel condition becomes unstable. Holland el al. [9] proposed a receiver-based auto rate (RBAR) protocol. With the rate feedback by the receiver, RBAR can adapt the channel condition more promptly than ARF. Later, the opportunistic auto rate (OAR) scheme was proposed in [21]. OAR utilizes the fragment burst in IEEE 802.11 [11], which allows more than one packet to be transmitted when the sender is granted medium access. OAR outperforms RBAR only when the channel condition between the sender and the receiver can support a high transmission rate (say 11 Mbps). ARF, RBAR, and OAR only consider the channel quality between the sender and the receiver. When the channel quality between the sender and the receiver is poor, the performance of these schemes would significantly degrade.

The channel quality has been used as a metric for route selection in some routing protocols [2], [6], [7], [22]. A path with overall best channel condition is selected to improve the end-to-end throughput [2], [6], [22] or power efficiency [7]. However, compared to MAC layer relay, network layer relay has higher control overhead and may incur a long queuing delay. When the channel condition changes frequently, due to the slow response of the routing protocols, network layer relay cannot react quickly to exploit the opportunities to deliver data at a high transmission rate.

In [24], a relay enabled PCF protocol, called rPCF, has been proposed to utilize the multirate capability via two-hop MAC layer relay. In rPCF, each mobile node reports the sensed channel condition to the access point. Based on

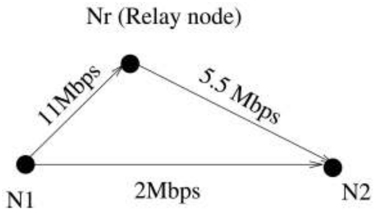


Fig. 1. The advantage of using the relay node.

the collected information, the access point decides and notifies the node at which rates to apply relay through the polling packet. Compared to rPCF, the design of rDCF is much more challenging: First, rDCF needs to operate in a distributed way and then it requires different techniques to coordinate the sender, the relay node, and the receiver in rDCF. Second, we need to consider the exposed terminal problem and the hidden terminal problem in rDCF, which does not exist in rPCF.

3 Motivations

3.1 Advantage of Two-Hop Relay

Since the channel condition varies with time and is location dependent [20], the multirate capability can be further exploited by enabling MAC layer multihop transmission. For example, as shown in Fig. 1, suppose N_1 needs to send data to N_2 and the channel of $N_1 \rightarrow N_2$ only supports a transmission rate of 2 Mbps. At the same time, the channel conditions of $N_1 \rightarrow N_r$ and $N_r \rightarrow N_2$ are much better, and they can support data rates of 11 Mbps and 5.5 Mbps, respectively. With a packet length of L, if the data can be transmitted along $N_1 \rightarrow N_r \rightarrow N_2$ at the MAC layer, the transmission delay is approximately $(\frac{1}{11} + \frac{1}{5.5})L$. Thus, the actual transmission rate is approximately equal to $\frac{5.5 \times 11}{5.5+11} = 3.7 Mbps$, which is much larger than 2 Mbps, when the packet is transmitted along $N_1 \rightarrow N_2$. Even after considering of the control overhead, when the packet size is not very small, the overall time to deliver the data packet can still be significantly reduced (see Section 5 for details). Although it is possible to have more than one relay node, considering the control overhead of the coordination among related nodes, we focus on two-hop MAC layer relay in this paper, which is sufficient in most cases.

There may be doubts on whether the relay mechanism will work since the channel conditions of $N_1 \rightarrow N_r$ and $N_r \rightarrow N_2$ may be unstable and then the actual transmission rate that can be achieved with relay could be lower than that with direct transmission. Fortunately, as stated in [21], when the node does not move very fast, i.e., less than 20 m/s, the coherence intervals [20], [21][1] are on the order of multiple packet transmission times. In most cases, since mobile nodes move fairly slowly (say less than 5 m/s) in ad hoc networks, it is feasible to exploit relay opportunities for each packet transmission (if there exists a suitable relay node) so that the performance of the system can be significantly improved.

1. The coherence interval is the average time interval during which the channel conditions are correlated.

3.2 MAC Layer Relay versus Network Layer Forwarding

As we mentioned in Section 2.3, the function of exploiting multirate capability can be performed through MAC layer relay or network forwarding. MAC layer relay is better than network layer forwarding in three aspects:

1. A packet relayed at the MAC layer do not have queuing delays, whereas a packet forwarded at the network layer would experience a long queuing delay if the relay node has many packets in the queue.
2. Because each network forwarding involves an RTS/CTS handshake plus an ACK at the MAC layer, the overall control overhead of forwarding a packet at the network layer is higher than that at the MAC layer.
3. Network layer forwarding may affect the bandwidth allocation of the relay node and, then, forwarding the packets of other nodes may affect the delivery of its own packets. In contrast, with the MAC layer relay, because each relayed packet does not enter the queue of the relay node, the MAC layer relay does not interfere with the node's transmission opportunity. This property is helpful for applying some rewarding schemes [5] to motivate the relay.

4 The Relay-Enabled DCF

In this section, we first present the basic protocol of rDCF and then propose techniques to enhance it. Finally, we discuss various impacts of relay and some implementation issues.

4.1 The Basic Protocol

4.1.1 The Service Advertisement

Similarly to most existing work [9], [21], we apply receiver-initiated channel condition measurement and let the receiver notify the sender of the transmission rate via CTS. With rDCF, each node promiscuously listens to all ongoing RTS and CTS packets. By extracting the piggybacked transmission rate in the CTS, a node knows the channel condition between the sender and the receiver. Meanwhile, it can measure the channel quality between the sender or the receiver and itself by sensing the signal strength of RTS or CTS, respectively. Since CTS packets do not have the MAC address of the packet sender, a node needs to infer the sender of the CTS according to the semantic of the CTS. In particular, suppose N_r overhears an RTS from N_i to N_j. If it overhears a CTS addressed to N_i after an SIFS, N_r can infer that the sender of the CTS is N_j.

For a given flow between a pair of sender and receiver, with the measured channel quality, if a node finds that the packets can be transmitted faster with the MAC layer relay, it adds the identity (e.g., MAC address) of the sender and the receiver into its willing list. In order to reduce the control overhead, we can limit the length of the willing list (e.g., 10 entries). Periodically, each node advertises its willing list to its one-hop neighbors. Some schemes, such as [23], can be used to improve the reliability of the broadcast. Once a node, say N_i, receives a willing list from N_r and finds that $N_i \rightarrow N_j$ is in the list, it adds N_r into its relay table

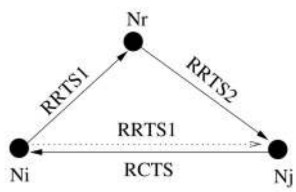


Fig. 2. An illustration of the triangular handshake.

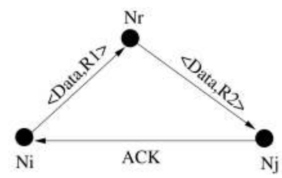


Fig. 3. An illustration of the MAC layer relay.

(note that it is possible that there are more than one relay nodes available for $N_i \rightarrow N_j$). As an optimization, the number of redundant service advertisements for a given flow can be reduced as follows: Before sending the advertisement, if N_r has overheard more than m advertisements containing $N_i \rightarrow N_j$ from other nodes, it knows that at least m other nodes have claimed to be the relay node for $N_i \rightarrow N_j$ and then deletes $N_i \rightarrow N_j$ from the willing list for the current advertisement.[2] In this paper, we set the value of m to be 3.

4.1.2 The Triangular Handshake

In the standard DCF protocol, the RTS/CTS handshake is required for each unicast packet transmission in order to prevent collisions. In [9], [21], this handshake is further utilized to probe the channel condition on a per-packet basis. Following these principles and considering backward compatibility to the standard DCF, we modify DCF and refer this new protocol as the basic protocol of rDCF. As shown in Fig. 2, the dashed line pointing to N_j means that N_j can overhear the packet. When a node N_i has a packet for N_j, it first searches the relay table using N_i as index. If N_i cannot find a relay node, the standard DCF with receiving-based rate feedback [9] is applied. Otherwise, N_i picks a relay node N_r and starts to coordinate the communication with N_r and N_j. Specifically, N_i sends a new packet, called *relay RTS* (RRTS1), to N_r. When N_r receives the RRTS1, it generates another relay RTS (RRTS2) and sends it to N_j. By sensing the signal strength of RRTS1, N_r and N_j individually determine the achievable transmission rate of $N_i \rightarrow N_r$, $N_i \rightarrow N_j$, denoted by R_1 and R_{dir}, respectively. According to the signal strength of RRTS2, N_j determines the transmission rate of $N_r \rightarrow N_j$, denoted by R_2. Based on R_1 (piggybacked by RRTS2), R_{dir}, and R_2, N_j replies CTS which piggybacks R_{dir} if the packet cannot be transmitted faster with relay. Otherwise, N_j replies a *relay CTS* (RCTS), which piggybacks R_1 and R_2, to N_i.

If N_i receives a CTS, it sends the data packet directly to N_j with the transmission rate of R_{dir}. If N_i receives a RCTS, as shown in Fig. 3, it sends the data packet to N_r with the transmission rate of R_1. After N_r receives the packet, it relays the packet to N_j with the transmission rate of R_2 after an SIFS. If the packet is correctly received by N_j, N_j replies with an ACK to N_i. If the transmission fails, the sender can detect it with a timeout mechanism similar to the standard DCF [11]. For example, when the sender sends an RTS to the receiver, it sets a timer to a proper duration in which it should receive the CTS from the receiver. If the sender does not receive the expected CTS when the timer expires, it knows that the transmission of RTS has been failed and tries to retransmit the RTS.

4.2 Enhancements of rDCF

The basic protocol of rDCF describes the basic mechanism to achieve relay-enabled DCF. However, considering the bandwidth utilization, the dynamic nature of wireless channels, and the impact of multirate transmissions, we propose techniques to further improve the performance of rDCF.

4.2.1 Dealing with Multirate Transmission

With IEEE 802.11 DCF, carrier sensing is performed using physical carrier sensing as well as virtual carrier sensing. As shown in Fig. 4a, when the data is transmitted with a fixed rate, the sender can easily calculate the duration of the packet transmission based on the packet length and the transmission rate. However, when the transmission rate can be adaptively changed, the sender cannot precisely calculate the length of the duration before sending the RTS since it does not know the transmission rate of the impending packet in advance. In the solution of [9], the sender chooses a data rate based on some heuristic, i.e., the most recent rate that was successfully used for transmission. This solution is not good enough for rDCF since the sender needs to estimate the transmission rates for both hops of the relay, and it may be difficult to get a precise estimate.

Our approach. We designed a new carrier sensing scheme for rDCF, which is shown in Fig. 4b. Instead of estimating the possible transmission rates and calculating the duration of the data transmission, the sender first calculates the duration of the RTS and CTS transmissions only.[3] The duration can be precisely calculated since all control packets (e.g., RTS, CTS, ACK, ...) are transmitted at the base rate, say 2 Mbps. After the sender receives CTS or RCTS, it calculates the durations of the data packet (if necessary) and the ACK based on the piggybacked transmission rate(s). In this way, our scheme can guarantee that other nodes within the transmission range of the sender and the receiver would defer medium access for exactly the data packet transmission time. Compared to the standard approach, our approach can achieve better bandwidth utilization in some situations. For example, suppose a CTS is lost at the sender due to collision or channel error, since the standard approach has longer duration piggybacked in the RTS than our approach, the neighbor nodes of the sender would defer for a longer time period in the standard DCF, which reduces the bandwidth utilization. Table 1 lists the duration for each packet used in rDCF. In the table, σ is the maximum propagation delay, $DATA(L, r)$ is the time

2. N_r will add $N_i \rightarrow N_j$ before sending the next advertisement if it finds that packets can be transmitted faster through its relay.

3. In case of relay, it needs to calculate the duration of RRTS1, RRTS2, and RCTS transmissions.

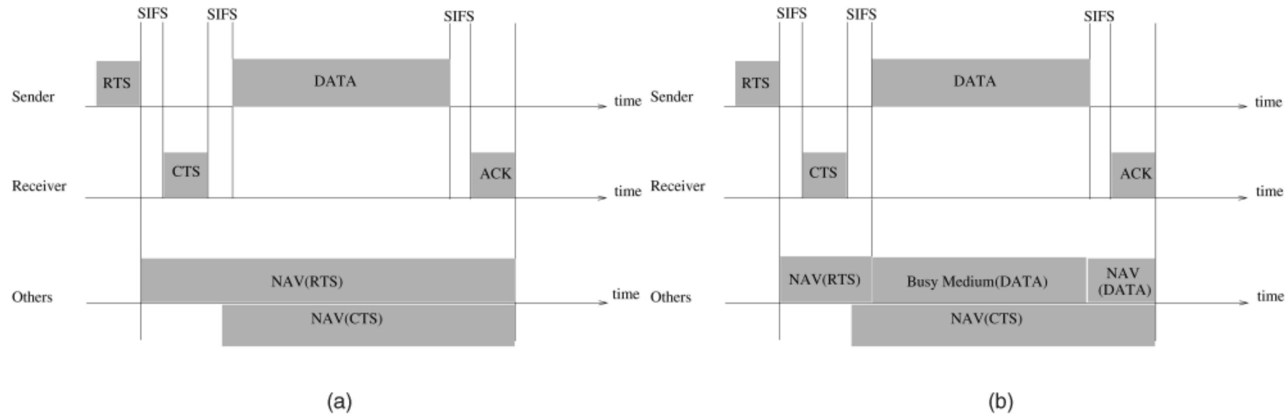


Fig. 4. The comparison of two different carrier sensing schemes. (a) The standard scheme. (b) The new scheme.

needed to transmit the packet with length of L at rate r. Note that the calculation of each duration includes the transmission time of both the PHY layer header and the MAC layer header. $Data_{dir}$ refers to the data packet with direct transmission and $Data_1$ is the data packet sent from the sender to the relay node. Other unlisted packets have a duration of 0.

Besides the impact on virtual carrier sensing, different transmission rates also result in different transmission ranges. For a given receiving power level, the packet transmitted with higher rate may have a higher bit error rate. As shown in Fig. 5, suppose N_i and N_j are far away from each other and the channel quality can only support 2 Mbps. N_j may not be able to decode a packet if N_i sends the packet at the rate of 5.5 Mbps. In this case, N_j is out of the transmission range of N_i. Based on this fact, when the sender sends data at a high rate, some one-hop neighbors may stay within its carrier sensing range but cannot extract the information of the duration piggybacked in the MAC header. To deal with such problems, we adopt the *reservation-sub-header* (RSH) in [9]. Specifically, an RSH is inserted preceding the data frame and is sent at the same or lower rate compared to RTS. Differently from [9], as shown in Fig. 6 and Fig. 7, our RSH does not need to include the MAC addresses of the sender and the receiver because the revised carrier sensing scheme would not incur any incorrect medium reservation of RTS. As a result, the overhead of our RSH is smaller than that in [9]. Since RSH is transmitted at a low rate (2 Mbps in this paper), all one-hop neighbor nodes can extract the duration in the RSH and update their NAV values accordingly.

TABLE 1
The Calculations of the Duration in rDCF

Packet Type	The Duration
RTS	$CTS + \sigma + 2SIFS$
CTS	$DATA(L, R_{dir}) + 2\sigma + 2SIFS + ACK$
RRTS1	$RRTS2 + RCTS + 2\sigma + 3SIFS$
RRTS2	$RCTS + DATA(L, R_1) + 2\sigma + 3SIFS$
RCTS	$DATA(L, R_1) + DATA(L, R_2) + 3\sigma + 3SIFS + ACK$
$Data_{dir}$	$ACK + \sigma + SIFS$
$Data_1$	$DATA(L, R_2) + ACK + 2\sigma + 2SIFS$

4.2.2 Dealing with Dynamic Channel Condition

The channel condition may change frequently in wireless networks [20], which may have significant impacts on the performance of rDCF. In order to alleviate the impacts of dynamic channel conditions, it is desirable to adaptively decide when to perform relay according to the channel conditions.

We design a heuristically randomized algorithm with very low computation complexity as follows: Each relay node in the relay table of N_i is associated with a *credit* ranging in $[0.0, 1.0]$. To exploit successful relays, each time when N_i finds a relay node for the receiver N_j, N_i chooses the one with the largest credit. After selecting the relay node, N_i generates a random number in $[0.0, 1.0]$ and sends RRTS1 to the chosen relay node if the credit is greater than or equal to the random number. Otherwise, N_i applies DCF and sends RTS to N_j. When a node N_r successfully relays a packet for N_i, which is indicated by receiving the ACK, the credit of N_r is increased by 0.1. When a relay via N_r fails, the credit is decreased by 0.1. When N_i receives that willing list from N_r and finds itself in the list, the credit of N_r is enhanced by 0.5.

Some types of transmission failures can be detected and recovered quickly in rDCF to reduce the cost of failures. As shown in Fig. 2, suppose N_i has a packet for N_j and finds the relay node N_r. We add two optimizations to the basic protocol as follows:

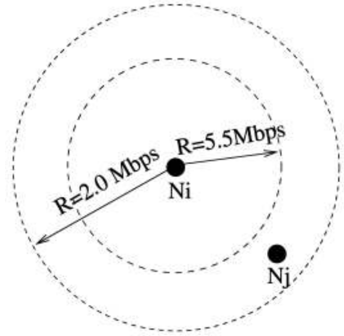


Fig. 5. An illustration of different transmission ranges.

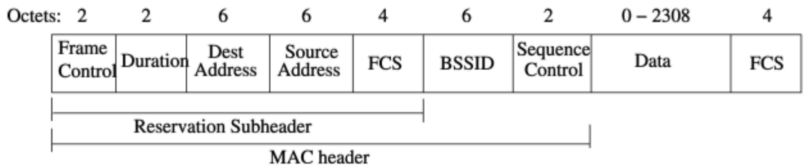


Fig. 6. Data packet frame format in [9].

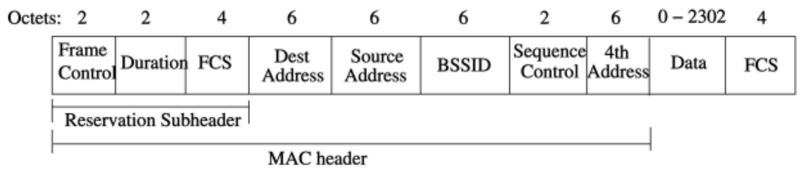


Fig. 7. Data packet frame format in our scheme.

- If RRTS1 from N_i to N_r is lost, which may be caused by collisions, N_j can detect it if no packet is overheard after $SIFS + \sigma$ when the transmission of RRTS1 is finished. Then, it replies with a CTS to N_i;
- If the data packet sent from N_i to N_r is lost, N_i can detect it if no packet is overheard after $SIFS + \sigma$. Then, N_i backs off based on the binary exponential backoff protocol for retransmission.

4.2.3 Dealing with Some Security Issues

Compared with the one-hop transmission, the relay may suffer from security attacks. Without necessary security enhancements, the performance of rDCF could be significantly degraded by malicious relay nodes. In the following, we summarize some possible attacks and give some solutions:

- **False advertisement**: The malicious node N_r may periodically broadcast a willing list which contains all addresses of its neighbors no matter whether a packet can be transmitted faster by using relay or not. Since the relay is performed in the MAC layer, the sender can precisely calculate the maximum transmission delay of the impending packet. Consequently, if the actual transmission delay is greater than the maximum delay, N_i knows that the relay node has sent false advertisements and marks it as a malicious node.
- **Malicious dropping**: The malicious node N_r could drop some packets to be relayed through it. In this case, both the sender N_i and the receiver N_j can act as watchdogs [18] to detect the malicious dropping. In particular, after overhearing the data packet sent from N_i to N_r, if the sender or the receiver does not hear any packet after $SIFS + \delta$, it can infer that the packet is dropped at N_r. It is possible that the packet is dropped at the relay node due to collision or channel error. However, since the relay happens after the triangular handshake, the possibility of collision or corruption at the relay node is very low. Based on this fact, if the frequency of droppings is greater than a threshold, the sender and the receiver can identify that N_r is a malicious node.
- **Packet integrity**: Instead of dropping, the malicious node N_r could modify the packet from N_i before relaying it to N_j. One solution is to authenticate each relay node before using it. Several schemes can be used to achieve this goal in ad hoc networks [10]. If there is no such security mechanism in an ad hoc network, this problem can be addressed with message digest techniques (e.g., MD4 [15]). When N_j receives a packet, it generates the digital digest of the packet and piggybacks the digest with the ACK packet, which is relied directly to the sender N_i. When N_i receives the ACK, it extracts the piggybacked digital digest and verifies the packet integrity by comparing the piggybacked digital digest with the one generated from the packet being acknowledged. If these two digital digests are not equal, N_i knows that the integrity of the packet is violated by N_r and marks N_r as the malicious node. This solution incurs several bytes of overhead for each ACK.[4] However, if data packet size is large, this overhead can be very small.

4.3 Impacts of Relay

In multihop ad hoc networks, the relay node may have some impacts on the system performance. In this section, we discuss some issues caused by relaying packets, and show that these impacts are very small in most cases through analysis.

4.3.1 The Impact on Spatial Reuse

As packets are being relayed, rDCF may have impacts on the spatial reuse of the network. As shown in Figs. 8a and 8b, any pair of nodes connected by a solid line can hear each other. With the standard DCF, f_1 and f_2 can simultaneously transmit data since they don't contend with each other for the medium. When N_r relays packets for flow f_1, N_3 has to defer its transmissions in order to avoid collisions, which may cause exposed or hidden terminal problems [3], [4]. At a first glance, if N_r always relays packets for f_1, the performance of f_2 may be significantly affected. After looking into the carrier sensing mechanism of IEEE 802.11, we can see that the impact is quite small in most cases.

4. With MD4 or MD5, the length of each message digest is 16 bytes.

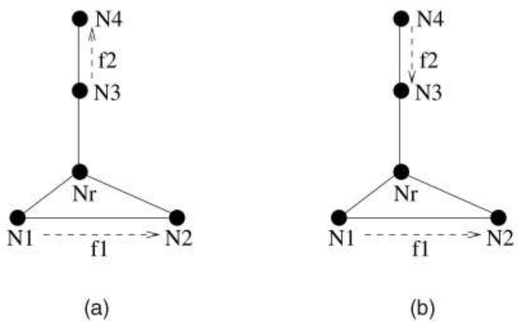


Fig. 8. An illustration of the impact of rDCF on spatial reuse. (a) Exposed terminal. (b) Hidden terminal.

Suppose N_r relays a packet for f_1 at time t. For an exposed terminal problem, there are two cases:

- **Case 1**: N_3 is in the transmission range of N_r at t, which means that it can extract the transmission duration. N_3 can defer medium access for the exact time period of the ongoing data transmission, and then start to contend for the medium again. As a result, in the long run, N_3 and N_1 have similar opportunities to access the channel.
- **Case 2**: N_3 is within the carrier sensing range of N_r so that it cannot extract the transmission duration. In this case, N_3 resumes contending for the medium only when the medium is idle for an extended interframe space (EIFS), which is equal to 364 μs [11]. As a result, N_3 may defer the medium access to sometime later after N_1 receives the ACK. Since the expect transmission time of (ACK + the post backoff[5] + $DATA(L, R_{1\to r})$) is greater than EIFS, where $R_{1\to r}$ is the transmission rate of $N_1 \to N_r$, we can see that N_3 would not be starved and can eventually obtain the medium access.

When N_3 transmits a packet to N_4, N_r sets its NAV to be either the data transmission time from N_3 to N_4 or EIFS (when a collision happens). When N_1 sends packets to N_r at this time, N_r will not send RRTS2 to N_2 since its NAV has not expired. In this case, the receiver applies the optimization technique in Section 4.2.2 and the impending packet of N_1 is served with DCF.

For the hidden terminal problem, the impact of relay could be greater since the sender of f_2 will double its current contention window size and back off again. However, similarly to the exposed terminal problem since N_3 does not always sense busy medium, this impact would not significantly affect the performance of f_2.

We also analyze the extended sensing area caused by N_r. As shown in Fig. 9, the extended sensing area S is N_r's sensing area, which does not overlap with the sensing areas of N_1 and N_2. It is not difficult to see that, for a given distance (d) between N_1 and N_2, the size of S increases as $d_1 + d_2$ increases. To meet the criteria of relay, $d_1 + d_2 \leq D_{5.5} + D_{11}$ should hold, where $D_{5.5}$ and D_{11} are the maximum transmission range of 5.5 Mbps and 11 Mbps, respectively. By setting d_1 and d_2 to be $D_{5.5}$ and D_{11}, respectively, we can calculate the upper bound of S.

5. After receiving the ACK, the sender is required to back off for a random period between 0 and CW_{min}.

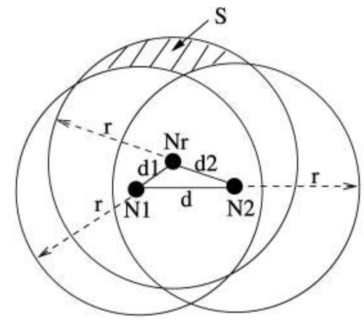


Fig. 9. An illustration of the extended sensing area.

TABLE 2
The Impact of Relay on the Sensing Area

d (meters)	210	220	230	240	250
Upper bound of increased sensing area (%)	11.5	10.5	9.2	8.2	7.2

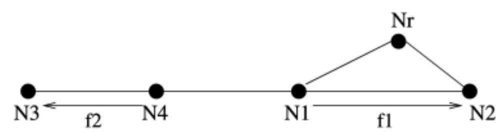


Fig. 10. An illustration of the impact of the hidden relay node.

We give some numerical results on the upper bound of the increased sensing area as a function of d. Following ns-2 [8], we set r, $D_{5.5}$, and D_{11} to be 550 m, 200 m, and 100 m, respectively. d changes from 210 m to 250 m. The numerical results are shown in Table 2. As can be seen, compared to the total sensing area of the sender and the receiver, the increased sensing area is small.

4.3.2 The Impact of Hidden Relay

Based on the location of the relay node, some node may be able to hear from the sender, but unable to hear from the relay node. For example, as shown in Fig. 10, N_4 can hear from N_1 but cannot hear from N_r. This may cause collisions at N_1 since N_4 may not defer medium access for the period of one data transmission when N_r relays a packet for f_1. In the following, we analyze this impact and show that it is very small. Suppose N_1 sends a packet to N_r at time t, there are two cases:

- **Case 1**: N_4 can extract the duration from the MAC header and defer medium access accordingly. Since the duration is equal to the time needed for relaying the data packet, N_4 would not contend for the medium before N_1 gets the ACK.
- **Case 2**: N_4 cannot extract the duration from the MAC header and then sets its NAV to be EIFS. With DCF, EIFS can be used to guarantee that the sender can receive the ACK. However, it may not always hold in rDCF. Since EIFS could be smaller than $DATA(L, R_{r\to 2}) + ACK + DIFS$, N_4 may send a packet to N_3 before N_1 receives the ACK (note that N_4 does not sense the signal of the packets sent by

N_r and N_2). As a result, it is possible that the packet sent by N_4 collides with the ACK at N_1.

If the impending packet is a unicast one, N_4 would back off for a period after EIFS and the probability for Case 2 to happen becomes very small. The claim has been verified through simulations (see Section 6.4). In other cases, the occurrence of Case 2 is bounded. As stated in Section 4.2.2, N_1 also reduces the credit of N_r by 0.1 since the previous relay operation has been failed. As Case 2 keeps happening, the credit of N_r will be eventually small enough so that the probability of choosing N_r to relay is very small.

4.4 Implementation Issues

In this section, we describe how rDCF can be incorporated into IEEE 802.11. The MAC layer header and the format of the MAC frame used for unicast is shown in Fig. 7. Similarly to the standard [11], each MAC frame has four address fields to indicate the BSS identifier (BSSID), source address (SA), destination address (DA), and the fourth address. These addresses may appear in different order and in different types of frames. In order to support rDCF, some minor modifications to the standard 802.11 frames are required: Each relay related data or control frame (e.g., RRTS1) uses all four address fields in the order of SA, DA, BSSID, and the fourth address. The first and second hop relay can be differentiated by the subtype value[6] in the frame control field. With SA, DA, and the fourth address fields, the addresses of the sender, the relay node, and the receiver can be stored in each frame. In order to identify the piggybacked transmission rates, we append an 8-bit *rate tag* to the frame if necessary. The tag is divided into two 4-bit fields, which can be used to represent two transmission rates. Since many functions of DCF (e.g., RTS/CTS, rate adaptation) are implemented in firmware [12], these modifications can be easily done.

5 Analysis of rDCF

5.1 Throughput Gain

In this section, we analyze the throughput gain of rDCF over the single rate DCF (operating at 2 Mbps). For simplicity, we assume the channel condition is ideal (i.e., no hidden terminals and capture [4]). The cases with dynamic channel condition are studied through simulations (see Section 6).

Theorem 1. *Let CW_{min} denote the minimum contention window size (in the number time slots) and assume that each node applies the binary exponential backoff scheme with the maximum backoff stage m (i.e., $CW_{max} = 2^m * CW_{min}$). For a fully connected topology with n flows, which are continuously backlogged, the ratio between the throughput of rDCF and that of DCF, denoted by γ, follows:*

$$\gamma = \frac{(1-P_{tr})\sigma + P_{tr}P_sT_s^{DCF} + P_{tr}(1-P_s)T_c}{(1-P_{tr})\sigma + P_{tr}P_sT_s^{rDCF} + P_{tr}(1-P_s)T_c} \quad (1)$$

and

$$P_{tr} = 1 - (1-\tau)^n, \quad (2)$$

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{tr}}, \quad (3)$$

$$\begin{aligned} T_s^{DCF} = &RTS + CTS + ACK + DATA(L, R_b) \\ &+ 3SIFS + 4\delta + DIFS, \end{aligned} \quad (4)$$

$$\begin{aligned} T_s^{rDCF} = &RRTS1 + RRTS2 + RCTS + ACK \\ &+ DATA(L, R_1) + DATA(L, R_2) \\ &+ 5SIFS + 6\delta + DIFS, \end{aligned} \quad (5)$$

$$T_c = RTS + DIFS + \delta, \quad (6)$$

where L is the packet length, R is the base rate (i.e., 2 Mbps), and R_1 and R_2 are the average transmission rate of the first hop relay and the second hop relay. The value of τ can be obtained by solving the following equation [4]:

$$\begin{aligned} &2(2(1-\tau)^{n-1} - 1) - (2(1-\tau)^{n-1} - 1)(CW_{min} + 1) + \\ &(1 - (1-\tau)^{n-1})CW_{min}(1 - (2^m(1 - (1-\tau)^{n-1})^m)\tau = 0. \end{aligned} \quad (7)$$

Proof. Since we do not consider capture, as shown in Fig. 4, the carrier sensing scheme of rDCF is exactly equivalent to that of DCF. In rDCF, for each node other than the sender and the receiver of the packet being transmitted, it defers its transmission for the same time period as in DCF, no matter whether it relays the packet or not. With the fact that rDCF and DCF have the same backoff scheme, we can see that the process of contending for the channel at each node in rDCF is the same as that in DCF. Consequently, the time spent on contention in rDCF is the same as that in DCF.

Since the analysis model in [4] has been proven to be valid for various CSMA/CA-based MAC protocols provided that collision avoidance follows binary exponential backoff, it should also be valid for rDCF. Following the results of [4] and calculating the average time the channel is sensed busy under DCF and rDCF, which are T_s^{DCF} and T_s^{rDCF}, respectively, the theorem holds. □

With (1), we show the numerical results of the throughput gain as the function of packet length L. We also validate our analysis through simulations. We assume that $n = 5$, $CW_{min} = 32$, $m = 4$, and each flow has a relay node which provides $R_1 = 5.5Mbps$ and $R_2 = 11.0Mbps$. As shown in Fig. 11a, the results between analysis and simulation are quite close. We can see that the gain increases as L increases. In particular, when L is too small (say less than 400 bytes), rDCF performs worse than DCF. The reason is that, when L is too small, the reduced transmission time by relaying data packet cannot overcome the extra control overhead of rDCF (e.g., RRTS2 packets). We also show the throughput gain under various rate distributions of R_1 and R_2 by fixing the packet length to be 1,000 bytes. We first assume R_2 is 11 Mbps and show the

6. The subtype value can be selected from the reserved ones between 1,000 and 1,111 (binary).

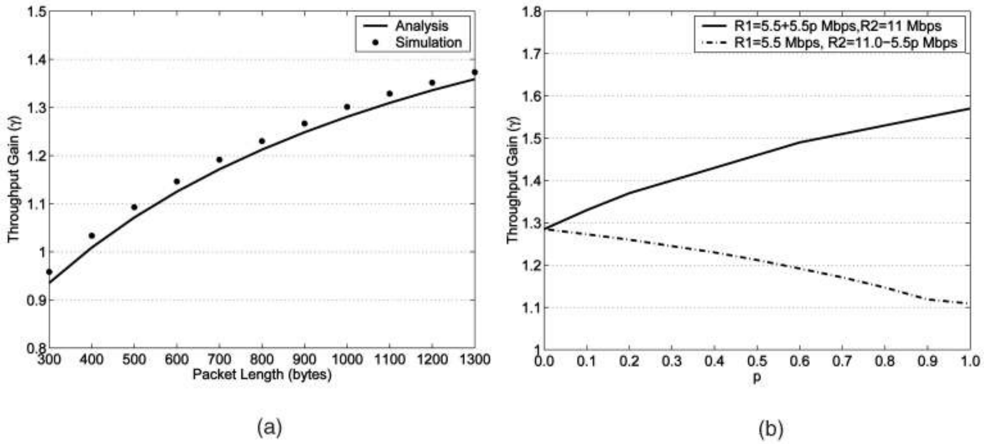


Fig. 11. Throughput gain of rDCF. (a) Throughput gain: analysis versus simulation. (b) Throughput gain under different rate distributions.

impact of R_1 on the throughput gain. As shown in Fig. 11b, when R_1 reaches 11 Mbps, the throughput gain can be 1.57. Then, we fix R_1 to be 5.5 Mbps and decrease the value of R_2 from 11 Mbps to 5.5 Mbps. As can be seen, the throughput gain decreases as R_2 becomes smaller. The reason that slow relays do not occur in our work is as follows: If the packet cannot be delivered much faster than the direct transmission, the receiver will let the sender transmit the packet through the direct link.

5.2 Node Density Requirement

In this section, we analyze the node density requirement of having a proper relay node for a given flow. For simplicity, we only consider path loss [20], which means that the signal strength mainly relies on the distance between the sender and the receiver for a given transmission power. Based on the results shown in Fig. 11, we assume that packet length is fixed and has been chosen to be a proper value (say 1,000 bytes).

Lemma 1. *Given a sender and a receiver whose coordinates are* (x_s, y_s) *and* (x_r, y_r)*, respectively, suppose the distance thresholds for 2 Mbps, 5.5 Mbps, and 11 Mbps are* D_2*,* $D_{5.5}$*, and* D_{11}*, respectively. In order to significantly improve the system performance through relays, the regions in which a relay node should be available, denoted by* S*, is obtained by:*

$$S = \{(x,y)|x = x_s + Xcos\alpha - Ysin\alpha,\\ y = y_s + Xsin\alpha + Ycos\alpha\}, \tag{8}$$

where $\alpha = sin^{-1}(\frac{y_r - y_s}{d})$, $d = \sqrt{(x_r - x_s)^2 + (y_r - y_s)^2}$*, and*

$$(X,Y) \in \left\{(x,y)|d - D_{5.5} \le x \le \frac{d^2 + (D_{11}^2 - D_{5.5}^2)}{2d},\right.\\ \left. -\sqrt{D_{5.5}^2 - x^2} \le y \le \sqrt{D_{5.5}^2 - x^2}\right\}, \tag{9}$$

$$\cup \left\{(x,y)|\frac{d^2 + (D_{11}^2 - D_{5.5}^2)}{2d} \le x \le D_{11},\right.\\ \left. -\sqrt{D_{11}^2 - x^2} \le y \le \sqrt{D_{11}^2 - x^2}\right\}, \tag{10}$$

$$\cup \left\{(x,y)|d - D_{11} \le x \le \frac{d^2 + (D_{5.5}^2 - D_{11}^2)}{2d},\right.\\ \left. -\sqrt{D_{11}^2 - x^2} \le y \le \sqrt{D_{11}^2 - x^2}\right\}, \tag{11}$$

$$\cup \left\{(x,y)|\frac{d^2 + (D_{5.5}^2 - D_{11}^2)}{2d} \le x \le D_{5.5},\right.\\ \left. h - \sqrt{D_{5.5}^2 - x^2} \le y \le \sqrt{D_{5.5}^2 - x^2}\right\}. \tag{12}$$

Proof. From the throughput gain analysis, in order to achieve fast relay for each packet, as shown in Fig. 12, the location of the relay node should follow:

$$\begin{cases} d_1 + d_2 > d \\ \max(d_1, d_2) \le D_{5.5} \\ \min(d_1, d_2) \le D_{11}. \end{cases} \tag{13}$$

Suppose $d_1 \le d_2$, the maximum area of the region in which the relay node should reside is the overlap of the circle centered at the sender with the radius of D_{11} and the circle centered at the receiver with the radius of $D_{5.5}$ (depicted in Fig. 12). With the knowledge of algebra and geometry, the region can be defined with (9) and (10).

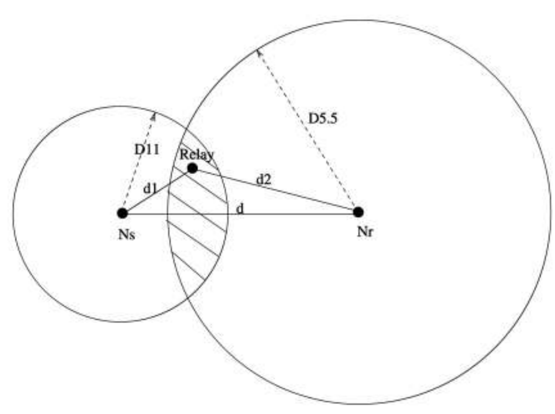


Fig. 12. The location requirement for the relay node.

TABLE 3
The Lower Bound of Node Density

d (meters)	200	210	220	230	240	250
Upper bound of ρ (node/m^2)	3.56e-5	4.13e-5	4.87e-5	5.89e-5	7.35e-5	9.57e-5

Following the same reason, when $d_1 \geq d_2$, the region can be defined with (11) and (12). By combining these two areas and applying coordination transformation (i.e., rotation and shift), we derive (8) and conclude the proof. □

Theorem 2. *Suppose all the nodes are randomly distributed in a region with the density of ρ $(node/m^2)$. For a sender and a receiver to be able to find at least one proper relay node, ρ should satisfy:*

$$\rho \geq \frac{1}{4}\left[\int_{\frac{d^2+D_{11}^2-D_{5.5}^2}{2d}}^{D_{11}} \sqrt{D_{11}^2 - x^2}dx + \int_{\frac{d^2+D_{5.5}^2-D_{11}^2}{2d}}^{D_{5.5}} \sqrt{D_{5.5}^2 - x^2}dx - I\left(D_{11} \geq \frac{d}{2}\right)\frac{1}{4}\int_{\frac{d}{2}}^{D_{11}} \sqrt{D_{11}^2 - x^2}dx\right]^{-1}, \tag{14}$$

where $I(x)$ is 1 if x is true, and is 0 otherwise.

Proof. We first calculate the area of the region derived from Lemma 1. Following Fig. 12 and without loss of generality, we assume $d_1 \leq d_2$. The area is maximized when d_1 and d_2 are equal to D_{11} and $D_{5.5}$, respectively. As shown in Fig. 12, the region is defined by (9) and (10) and its area, denoted by $A(d_1, d_2)$, follows:

$$A(d_1, d_2) = 2\left(\int_{\frac{d^2+d_1^2-d_2^2}{2d}}^{d_1} \sqrt{d_1^2 - x^2}dx + \int_{\frac{d^2+d_2^2-d_1^2}{2d}}^{d_2} \sqrt{d_2^2 - x^2}dx\right). \tag{15}$$

Considering symmetry, the area of the region defined by (11) and (12) is equal to $A(d_1, d_2)$. When these two regions overlap, the area of the overlap is equal to $A(d_1, d_1)$, which is calculated by replacing d_2 (in (15)) with d_1 since we assume $d_1 \leq d_2$. Thus, the total area of the region is equal to $2A(d_1, d_2) - A(d_1, d_1)$. Letting d_1 and d_2 be equal to D_{11} and $D_{5.5}$, respectively, and, observing the fact that the overlap exists only when $d_1 \geq \frac{d}{2}$, the maximum area of the region, denoted by A_{max}, is equal to $2A(D_{11}, D_{5.5}) - I(D_{11} \geq \frac{d}{2})A(D_{11}, D_{11})$. Since all the nodes are randomly distributed, the expected number of nodes in this region is equal to $\rho * A_{max}$. By solving $\rho \geq \frac{1}{A_{max}}$, we conclude the proof. □

Some numerical results are given in Table 3 to show the lower bound of ρ. As can be seen, the required node density increases as the distance between the sender and the receiver increases. However, the overall density requirement is quite loose. For example, when d is 250 m, we require only one node in each 10^4 m^2 area, which is not difficult to satisfy.

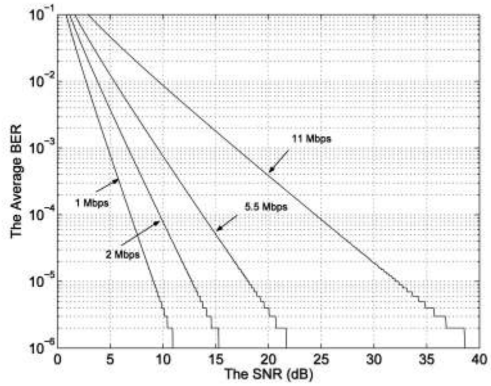


Fig. 13. The BERs under different transmission rates.

6 Performance Evaluations

6.1 The Propagation Model

When the wireless channel is assumed to be stable, we use the propagation model in ns-2 [8], which combines the Friis free space propagation model and the two-ray ground propagation model [20]. Basically, when the sender and the receiver are close, the Friis free space model is applied so that the path loss exponent is 2. Otherwise, the two-ray ground propagation model and the path loss exponent become 4.

When there is multipath fading or relative movement between the sender and receiver, the channel condition between them may change frequently. The frequency of this change depends on the relative speed of the mobile node with respect to its surroundings. We use the Ricean fading model [20] to simulate the fading channel conditions. The Ricean distribution is given by:

$$p(r) = \frac{r}{\alpha^2} e^{-(\frac{r}{2\alpha^2}+K)} I_0(2Kr), \tag{16}$$

where K is the distribution parameter representing the line-of-sight component of the received signal, α^2 is the variance of the background noise, r is the received power, and $I_0(.)$ is the modified Bessel function of the first kind and zero order [20].

When a node receives or overhears a packet, it determines whether the packet is corrupted according to the packet length, the SNR, and the corresponding bit error rate (BER). With the BER of BPSK given by [16] and the approximate BER performance using different modulation techniques in [1], we have the BERs at different transmission rates shown in Fig. 13. The probability that p can be successfully received, denoted by P_{succ}, is calculated by:

$$P_{succ} = (1 - BER(\gamma))^L, \tag{17}$$

where $BER(\gamma)$ is the BER with the SNR of γ and L is the packet length.

6.2 The Simulation Setup

Our simulation is based on ns-2 and its extensions [19], [8]. Similarly to [21], the distance thresholds for 11Mbps, 5.5Mbps, and 2Mbps are 100m, 200m, and 250m, respectively. The thresholds for different data rates are chosen based on the distance range. The mean period for service

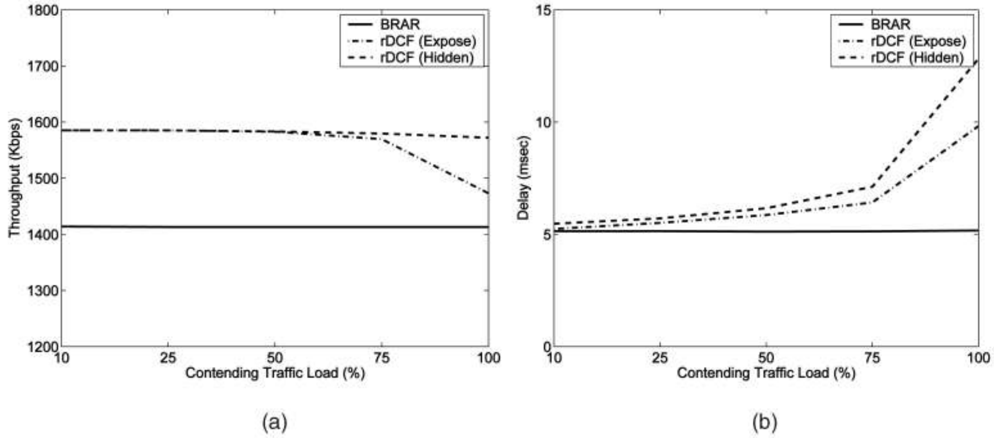


Fig. 14. The impact of rDCF on spatial reuse. (a) Throughput of Flow 1. (b) Delay of Flow 2.

advertisements is 1.0 second. The data packet length is set to be 1,000 bytes and the simulation time is set to be 100 seconds. Based on the analytical results in Section 5, we set the packet size threshold for relay to be 400 bytes. We run each case five times and use the average as the simulation result.

We compare rDCF with the state-of-the-art protocol called *receiving-based auto rate* (RBAR) [9]. It has been shown that RBAR outperforms the standard DCF and the sender-based rate adaption protocol called auto rate fallback (ARF). We do not compare rDCF with the opportunistic auto rate (OAR) protocol since OAR degrades to RBAR when the link quality between the sender and the receiver is poor. The RBAR protocol works as follows: The receiver measures the channel quality based on the signal-to-noise ratio of the arriving RTS packet. Then, it sets the transmission rate according to the highest feasible value allowed by the channel condition and piggybacks the rate with the CTS packet. After receiving the CTS, the sender sends out the data packet with the piggybacked transmission rate.

We use throughput and delay to measure the performance. The throughput is the total amount of data (in bits) delivered divided by the simulation time. The packet delay is the time interval from the packet entering the sender's queue to the time being delivered to the receiver. Note that the control overhead is also counted in the measurement.

6.3 Simulation Results

6.3.1 Impacts on Spatial Reuse

In this experiment, we evaluate the impacts of rDCF on the spatial reuse and assume the channel condition is stable. The topologies used are shown in Figs. 8a and 8b, under which the performance results are denoted as rDCF (Exposed) and rDCF (Hidden), respectively. The channel quality between the sender and the receiver of each flow can only support 2 Mbps. N_r and N_3 are within carrier sensing range of each other. The *contending traffic load* (CTL) (in percentage of the saturation throughput) of flow 1 (flow 2) increases as the aggregated traffic of flow 2 (flow 1) as well as the flows that are spatially close to flow 2 (flow 1) increases, and vice versa. In particular, the traffic of each flow contributing CTL is set to be 10 percent of the saturation throughput. In addition to flow 2's traffic, the CTL of flow 1 is increased as more flows that are spatially close to flow 2 are admitted.

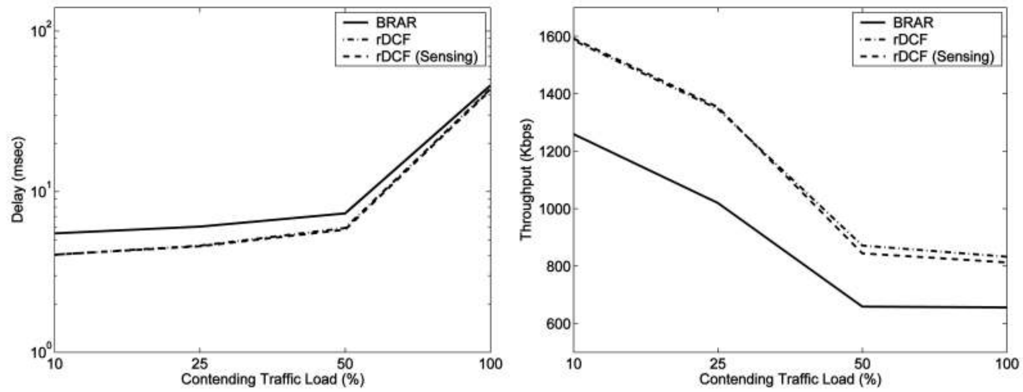
We first evaluate the impacts of CTL on the throughput of flow 1. Suppose flow 1 is backlogged. As shown in Fig. 14a, when the CTL of flow 1 is not high (e.g., 50 percent), the throughput of flow 1 under rDCF is not affected and is much higher than that under RBAR. In case of rDCF (Expose), when the CTL of flow 1 is high (i.e., over 75 percent), the throughput of flow 1 decreases. As discussed in Section 4.3.1 since N_r frequently defers the medium access of flow 2, many data packets are transmitted with direct transmissions. In case of rDCF (Hidden), the impact of flow 1's CTL is very small since N_3 only sends short packets (i.e., CTSs and ACKs). Note that N_3 and N_r are within the carrier sensing range of each other.

We then evaluate the impacts of CTL on the delay of flow 2. The rate of flow 2 is fixed to be 160 Kbps (or 20 pkt/sec). As shown in Fig. 14b, in case of both rDCF (Expose) and rDCF (Hidden), when the CTL of flow 2 is not high (i.e., less than 50 percent), its impact on flow 2's delay is quite small. When the CTL of flow 2 is very high (i.e., near 100 percent), the delay of flow 2 increases. The reason of the prolonged delay has been discussed in Section 4.3.1 and the result conforms our claim that flow 2 would not be starved.

6.4 The Impact of Hidden Relay

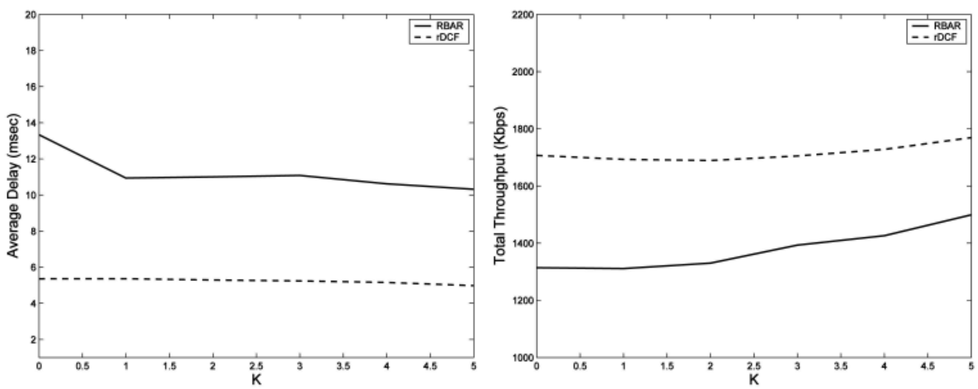
We study the impact of hidden relay in this section. The topology has been shown in Fig. 10. We assume that the channel is stable. By default, in rDCF, we assume N_4 can extract the duration of each data packet sent by N_1. rDCF (Sensing) denotes the situation that N_4 cannot extract the duration field. As stated in Section 6.4, the impact of hidden relay does not exist in the default rDCF, but it exists in rDCF (Sensing).

We evaluate the impact of CTL on the delay of flow 1. The rate of flow 1 is fixed to be 160 Kbps. As shown in Fig. 15a, when the CTL of flow 1 is low, because of relay, the delay of flow 1 in rDCF and rDCF (Sensing) is much smaller than that under RBAR. As the CTL of flow 1 increases, the delay of flow 1 under rDCF and rDCF (Sensing) increases and becomes close to that under RBAR. Since N_4 and N_1 can hear each other, they compete for



(a) (b)

Fig. 15. The impact of hidden relay on rDCF. (a) Delay of flow 1. (b) Throughput of flow 1.



(a) (b)

Fig. 16. The performance comparison between RBAR and rDCF under different K. (a) Delay. (b) Throughput.

medium access. Thus, as the CTL of flow 1 increases, N_1 takes more time to contend for the medium. From the figure, we can also see that the delay of flow 1 under rDCF (Sensing) is almost the same as that under rDCF, which shows that the impact of hidden relay on the delay of flow 1 is almost negligible.

We then examine the impact of CTL on the throughput of flow 1, which is always backlogged. As shown in Fig. 15b, the throughput of flow 1 under rDCF and rDCF (Sensing) is always greater than that under RBAR. Only when the CTL of flow 1 is high (say more than 50 percent), can we see the difference between rDCF and rDCF (Sensing). As expected in Section 6.4, due to collisions caused by N_4, the throughput of flow 1 under rDCF-S is less than that under rDCF. However, the throughput difference is small, which shows that the impact of hidden relay on the throughput of flow 1 is not a big issue.

6.5 Fully Connected Topology

In this section, we study the performance of rDCF in a fully connected topology where nodes can hear each other. We put 20 nodes in the area ($220\text{m} \times 220\text{m}$). Among them, 10 nodes act as either the sender or the receiver of the five flows. To examine the effectiveness of relay, we assume the average channel condition between the sender and the receiver of each flow can only support 2 Mbps. The remaining 10 nodes are randomly distributed in the area. We use the Ricean propagation model to emulate the dynamic channel condition and evaluate the impacts of the line-of-sight parameter K and the mobility.

6.5.1 Impact of K

The channel condition could be quite dynamic due to various factors. One important factor is the line-of-sight parameter K. A large K means a good channel quality while a small K means a poor channel quality. We first set the rate of each flow to be 160 Kbps and evaluate the packet delay under rDCF and RBAR. As shown in Fig. 16a, the delay under rDCF is much smaller than that under RBAR and the impact of K on rDCF is smaller than that on RBAR. We then evaluate the system throughput under rDCF and RBAR by letting all the flows always backlogged. As shown in Fig. 16b, under rDCF and RBAR, the system throughput increases as K increases since the system-wide channel condition becomes better when K is larger. Compared to RBAR, rDCF can have much higher system throughput (at least 25 percent more). The performance gain is mainly due to the high transmission rate achieved by the MAC layer relay.

After looking at the throughput of each flow, we found that the impact of channel errors on fairness can be significantly reduced by rDCF. Fig. 17 shows the throughput

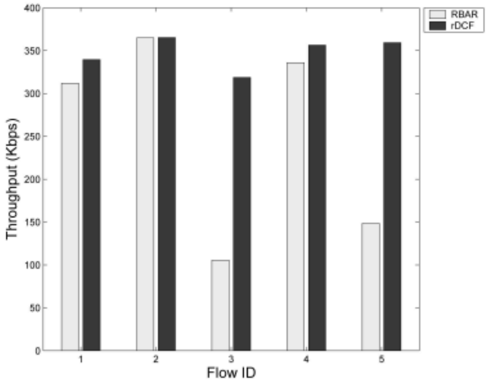


Fig. 17. The fairness comparison between RBAR and rDCF.

of each flow when K = 0. As can be seen, under RBAR, the throughput of flow 3 and flow 5 is much less than that of flow 1, flow 2, and flow 4. The reason is that the distance between the sender and the receiver of flow 3 and flow 5 is longer than that of other flows. As a result, the accumulated time period when the channel condition is poor becomes larger, which causes more packets of flow 3 and flow 5 being lost due to channel errors. Consequently, due to the binary exponential backoff, the accumulated backoff time of flow 3 and flow 5 becomes more than other flows. However, as shown in the figure, this unfairness does not exist under rDCF. The reason is that most packets from flow 3 and flow 5 can be delivered via relay, where both the channel conditions between the sender and the relay node and between the relay node and the receiver are more stable than the direct link. As a result, the number of transmission failures due to channel errors can be significantly reduced by using relay.

6.5.2 *Impact of Mobility*

Mobility affects the channel condition in two ways. First, it changes the node's location, which may affect the value of K and the strength of the received signal strength. Second, due to a Doppler shift in the frequency of the received signal, it may reduce the channel coherence time period. We evaluate the impact of mobility on the performance of rDCF. Similarly to [9], each receiver of a flow keeps moving back and forth. More specifically, it moves toward the sender until the distance between them is equal to 200m and then moves back until the distance between them is 250m. Similarly to [21], K is fixed to be 5 and the simulation time long enough to make the average time spent at each distance independent of the node velocity. As shown in Fig. 18a, the delay under rDCF is significantly less than that under RBAR. With relay, rDCF outperforms RBAR because it can have higher transmission rate when the sender and the receiver are far away from each other. For the same reason, as shown in Fig. 18b, the total throughput under rDCF is much better than that under RBAR.

TABLE 4
The Performance Comparison under the Multihop Topology

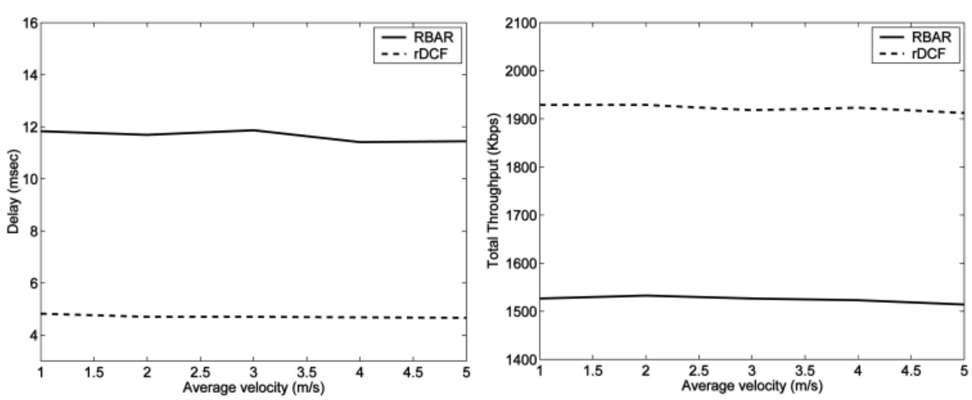
	Avg. Delay (msec)	Avg. Throughput (Kbps)
RBAR	272.4	213.5
rDCF	67.6	319.7

6.6 Multihop Topology

We evaluate the performance of rDCF under multihop topology in which 50 nodes are randomly distributed in a rectangular area of $1,000\text{m} \times 600\text{m}$. All nodes are assumed to move around the area randomly and the mean moving speed is 3 m/s. The line-of-sight parameter K is set to be 5. We simulate five flows in the system and the routing protocol is the dynamic source routing (DSR) [13]. The average end-to-end delay (when the flow rate is 160 Kbps) and the end-to-end throughput (when the flow is always backlogged) are shown in Table 4. As can be seen, compared to RBAR, rDCF has a significantly shorter delay since the impact of channel error has been largely relieved via relay. With the same reason, rDCF also achieves much better throughput than RBAR.

7 Conclusions

In this paper, we presented a novel relay-enabled DCF protocol, called rDCF, to exploit the physical layer multi-rate capability. According to the channel condition, data can be transmitted with different rates and some data packets



(a) (b)

Fig. 18. The performance comparison between RBAR and rDCF under different velocities. (a) Delay. (b) Throughput.

may be delivered faster through a relay node than through the direct link if the direct link has low quality and low rate. The basic protocol of rDCF is proposed to help the sender, the relay node, and the receiver coordinate to decide what data rate to use and whether to use a relay node. Considering the bandwidth utilization, the dynamic nature of wireless channels and security, we also propose several techniques to enhance the performance of rDCF. Simulation results showed that rDCF outperforms the receiver-based auto rate (RBAR) protocol in terms of throughput and delay in various scenarios. We can further improve the performance of rDCF, for example, by considering power efficiency in ad hoc networks and the use of directional antennas. We also plan to conduct formal analysis on the impacts of relay and find optimal ways to adjust the credit of each relay node.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the anonymous referees whose insightful comments helped us to improve the quality of the paper. This work was supported in part by the US National Science Foundation (CAREER CCR-0092770 and ITR-0219711).

REFERENCES

[1] M.S. Alouini and A. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels," *Wireless Personal Comm.*, pp. 119-143, May 2000.
[2] B. Awerbuch, D. Holmer, and H. Rubens, "High Throughput Route Selection in Multi-Rate Ad Hoc Wireless Networks, Wireless On-Demand Network Systems," *ACM/Kluwer Mobile Networking and Applications (MONET)*, Jan. 2004.
[3] V. Bharghavan, A.J. Demers, S. Shenker, and L. Zhang, "A Media Access Protocol for Wireless LANs," *Proc. ACM Sigcomm Conf.*, pp. 212-225, Sept. 1994.
[4] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE J. Selected Areas in Comm.*, pp. 535-547, Mar. 2000.
[5] L. Buttyan and J.P. Hubaux, "Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks," *ACM/Kluwer Mobile Networks and Applications (MONET)*, Oct. 2003.
[6] D.S.J. DeCouto, D. Aguayo, J. Bicket, and R. Morris, "A High-Throughput Path Metric for Multihop Wireless Routing," *Proc. ACM Mobicom Conf.*, pp. 134-146, Sept. 2003.
[7] J. Gomez, A.T. Campbell, M. Naghshineh, and C. Bisdikian, "Conserving Transmission Power in Wireless Ad Hoc Networks," *Proc. IEEE Int'l Conf. Network Protocols (ICNP)*, Nov. 2001.
[8] VINT Group, "UCB/LBNL/VINT Network Simulator—ns (Version 2)," http://mash.cs.berkeley.edu/ns, 2004.
[9] G. Holland, N. Vaidya, and P. Bahl, "A Rate-Adaptive MAC Protocol for Multihop Wireless Networks," *Proc. ACM Mobicom Conf.*, pp. 236-251, July 2001.
[10] Y. Hu, A. Perrig, and D.B. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks," *Proc. ACM Mobicom Conf.*, pp. 12-23, Sept. 2002.
[11] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Spec," IEEE 802.11 Standard, 1999.
[12] Agere Systems Inc., "WaveLAN 802.11b Chipset for Standard Form Factors," http://www.agere.com/client/docs/PB03025.pdf, 2002.
[13] D. Johnson and D. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Network," *Mobile Computing*, pp. 153-181, 1996.
[14] A. Kamerman and L. Monteban, "WLAN-II: A High-Performance Wireless LAN for the Unlicensed Band," *Bell Labs Technical J.*, Summer 1997.
[15] C. Kaufman, R. Perlman, and M. Speciner, *Network Security: Private Communication in a Public World*, second ed. Prentice Hall, 2002.
[16] UCLA Parallel Computing Lab, "GloMoSim," http://pcl.cs.ucla.edu/projects/glomosim/, 2004.
[17] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "UCAN: A Unified Cellular and Ad-Hoc Network Architecture," *Proc. ACM Mobicom Conf.*, pp. 353-367, Sept. 2003.
[18] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating Routing Misbehavior in Mobile Ad Hoc Networks," *Proc. ACM Mobicom Conf.*, 2000.
[19] R. Punnoose, P. Nikitin, and D. Stancil, "Efficient Simulation of Ricean Fading within a Packet Simulator," *Proc. IEEE Vehicular Technology Conf.*, pp. 764-767, Sept. 2000.
[20] T.S. Rappaport, *Wireless Communications: Principle and Practice.* Prentice Hall, 1996.
[21] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic Media Access for Multirate Ad Hoc Networks," *Proc. ACM Mobicom Conf.*, pp. 24-35, Sept. 2002.
[22] Y. Seok, J. Park, and Y. Choi, "Multirate Aware Routing Protocol for Mobile Ad Hoc Networks," *Proc. IEEE Vehicular Technology Conf.*, pp. 22-25, Apr. 2003.
[23] K. Tang and M. Gerla, "MAC Reliable Broadcast in Ad Hoc Networks," *Proc. IEEE Military Comm. Conf. (MilCom)*, pp. 1008-1013, Oct. 2001.
[24] H. Zhu and G. Cao, "On Improving the Performance of IEEE 802.11 with Relay-Enabled PCF," *ACM/Kluwer Mobile Networking and Applications (MONET)*, vol. 9, pp. 423-434, 2004.

Hao Zhu received the BS degree from Xian Jiaotong University, Xian, China, and the MS degree in computer science and engineering from the Institute of Software, Chinese Academy of Sciences, Beijing, China. He received the PhD degree from the Pennsylvania State University and joined Florida International University as an assistant professor in 2004. His research interests include resource management, QoS, power efficient protocols, and medium access control protocols in wireless networks. He is a member of the IEEE.

Guohong Cao received the BS degree from Xian Jiaotong University, Xian, China. He received the MS and PhD degrees in computer science from the Ohio State University in 1997 and 1999, respectively. Since then, he has been with the Department of Computer Science and Engineering at the Pennsylvania State University, where he is currently an associate professor. His research interests are wireless networks and mobile computing. He has published about 100 papers in the areas of sensor networks, cache management, data dissemination, resource management, wireless network security, and distributed fault-tolerant computing. He is an editor of the *IEEE Transactions on Mobile Computing* and the *IEEE Transactions on Wireless Communication*, a co-guest-editor of the *ACM/Kluwer Mobile Networking and Applications* special issue on heterogeneous wireless networks, and has served on the program committees of many conferences. He was a recipient of the Presidential Fellowship at the Ohio State University in 1999 and a recipient of the US National Science Foundation CAREER award in 2001. He is a member of the IEEE.