



On Improving the Performance of IEEE 802.11 with Relay-Enabled PCF

HAO ZHU and GUOHONG CAO

Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

Abstract. Integrating wireless LAN (WLAN) techniques with the third generation cellular networks has become a promising way to improve the performance of wireless systems. As WLANs play an important role in such heterogeneous systems, the performance of WLANs becomes important to the whole system. It is well known that WLANs provide a physical layer multi-rate capability, and hence MAC layer mechanisms are needed to exploit this capability. In this paper, we propose a novel MAC layer relay-enabled point coordination function (PCF) protocol, called *r*PCF, to exploit the physical layer multi-rate capability. Since WLAN supports multiple data 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. To enable MAC layer relay, the access point needs to collect information about the channel conditions, and notify the mobile nodes which data rate to use and whether to transmit the data through a relay station. We design protocols to achieve this goal and refine these protocols to minimize the control overhead. Simulation results show that *r*PCF can significantly improve the system performance in terms of system throughput and transmission delay by adding only a negligible control overhead.

Keywords: IEEE 802.11, relay, rate adaptation, media access control

1. Introduction

With the advent of the third generation (3G) wireless infrastructure and the rapid growth of wireless communication technology such as IEEE 802.11, ubiquitous computing becomes possible. To increase the system capacity and improve the system performance, heterogeneous network technologies will likely be deployed together. The resulting network may use wireless LAN (WLAN) techniques for peer-to-peer communication or to serve as relay networks to extend the range and increase the capacity of 3G networks, which provide links to the wired Internet. The feasibility of such kind of heterogeneous systems has been studied in [19], and some wireless industry players have already proceeded toward this direction. For example, Nokia has released a new product (*Nokia D311* [11]) which incorporates GPRS [18] and 802.11b [5] into one PC card.

As WLANs play an important role in heterogeneous systems, the performance of WLANs becomes important to the whole system. Most of the WLAN cards today are based on IEEE 802.11 [8], especially IEEE 802.11b. The IEEE 802.11 standard specifies two different media access control (MAC) mechanisms: one is called *distributed coordination function* (DCF), which is based on carrier-sense multiple access with collision avoidance (CSMA/CA); the other is called *point coordination function* (PCF), which is based on polling. DCF is the basic MAC mechanism whereas PCF is built on top of DCF and provides contention-free media access. Due to the nature of contention-free, PCF can achieve higher throughput than the contention-based DCF and provide guaranteed service [10], which is important for real-time applications. Further, PCF could also be used for non-real-time services. Some efficient and robust polling schemes [4,16] have been

proposed to enhance the performance of PCF. In addition, it has been shown that PCF can achieve efficient power management [13], which is unlikely to be achieved under DCF due to the hidden terminal problem [3]. Consequently, the polling-based PCF will be an attractive option for future WLANs.

It is well known that IEEE 802.11a and 802.11b provide a physical-layer multi-rate capability [5]. With the multi-rate enhancement, data can be transmitted at a number of data rates according to the channel condition. When the signal-to-noise ratio is high, i.e., the channel demands for error detection and recovery is low, an aggressive and efficient modulation scheme can be applied to increase the data transmission rate. On the other hand, when the signal-to-noise ratio is low, the bit error rate of the transmission is reduced by using a robust and redundant modulation scheme, which decreases the data transmission rate. In practice, IEEE 802.11b supports data rates of 1, 2, 5.5, and 11 Mbps, and IEEE 802.11a supports data rates of 6, 9, 12, 18, . . . , 54 Mbps.

As multi-rate IEEE 802.11 enhancements are available at the physical layer, MAC layer mechanisms are required to exploit this capability. Several schemes have been proposed to exploit the multi-rate capability under DCF. Kamerman and Monteban [9] designed the auto rate fallback (ARF) protocol to utilize the multi-rate feature of IEEE 802.11, which has been implemented in current products. Later, Holland et al. [7] proposed the receiver-based auto rate (RBAR) protocol to serve the same goal. The key difference between ARF and RBAR is that ARF is a sender-based scheme and RBAR is receiver-based. With ARF, the sender attempts to exploit high transmission rates after consecutive transmission successes and return to low rates upon failure. With RBAR, the receiver measures channel conditions while receiving the request-to-send (RTS) packet. Then, it selects a possible high rate and

notifies the sender of the rate through the clear-to-send (CTS) packet. With the notified rate, the sender can send the data packet at an aggressive rate. Beyond RBAR, the opportunistic auto rate scheme has been proposed in [15]. With this scheme, the nodes with good channel conditions are granted media access for a period during which more than one packets are allowed to be transmitted. The ARF scheme can be used with both PCF and DCF. However, the other two schemes are based on DCF and cannot be directly applied to PCF. Furthermore, all the three proposed schemes only consider to exploit high transmission rate of the direct link between the sender and the receiver. In many cases, the rate can be further increased by transmitting the data packet through multiple links that have much higher rate than the direct link. Thus, it is desirable to design a PCF-based MAC protocol with the ability to further exploit the multi-rate capability.

In this paper, we propose a novel PCF-based MAC protocol, called *relay-enabled PCF (rPCF)*, to exploit the multi-rate capability of IEEE 802.11 WLANs. Based on the channel condition of each link, *rPCF* may use multi-hop (mainly two-hop in this paper) data transmission through MAC layer relay to improve the system performance. For example, if the direct link between the sender and the receiver has low quality and low rate, whereas both the links from the sender to the relay node and from the relay node to the receiver can support high data rate, performance can be improved if the sender sends the data to the relay node which fast forwards the data to the receiver at MAC layer. To enable the MAC layer relay, the access point needs to collect information about the channel conditions from each mobile node, and notify the mobile nodes which data rate to use and whether to use a relay station. We design protocols to achieve this goal and refine these protocols to minimize the control overhead. Extensive simulations are provided and used to evaluate the proposed solutions. Compared to PCF-based ARF, *rPCF* can significantly improve the system performance in terms of system throughput and transmission delay, with only a negligible control overhead.

The remaining of this paper is organized as follows. Section 2 describes the background and motivation of the work. The details of *rPCF* are presented in section 3. Section 4 evaluates the performance of *rPCF* through simulations and the simulation results are presented. Section 5 concludes the paper.

2. Background and motivation

2.1. The system model

The system discussed in the paper is a typical basic service set (BSS) where all mobile stations are in the coverage area of the *access point (AP)*. Each node, which refers to either a station or the AP, is equipped with a WLAN card such as [1] that enables the node to communicate with each other based on 802.11b, which supports data rates of 1, 2, 5.5, and 11 Mbps. The AP is a data bridge and is mounted in a fixed

position and connected to the Internet through the wired network. All stations are assumed to stay in the coverage area of the WLAN. The AP relies on the PCF protocol to control the media access within the BSS and grants access to each station by polling. The physical layer uses direct-sequence spread spectrum (DSSS). Following the specification of 802.11 [8], the direct peer-to-peer communication is enabled as long as two nodes are within the transmission range of each other. Similar to most existing work [15], we model the transmission rates of 2 Mbps, 5.5 Mbps and 11 Mbps for sending data in IEEE 802.11b. As specified in the standard [8], the rate for sending physical-layer headers is always set to be 1 Mbps. The control packets (i.e. poll, ACK, NULL, beacon, . . .) and the broadcast packets are required to be transmitted at the rate of 2 Mbps. Although most WLANs allow roaming or hand-off, for simplicity, we do not consider this in the paper.

2.2. The rate adaptation

The rate adaptation technique is widely used in wireless networks to adapt the coding schemes to the time varying channel quality. Generally, higher data rates are achieved by more efficient modulation techniques. However, highly efficient modulation techniques rely on denser modulation encodings and will result in high bit error rate. For example [7], given a signal-to-noise ratio (SNR) of 10 dB, a packet transmitted at 4 Mbps using 16-ary quadrature amplitude modulation could experience a bit error rate of 0.07, in comparison to a bit error rate of 4×10^{-6} when transmitted at 2 Mbps using the QPSK modulation. Since the performance of a modulation technique is measured by its ability to preserve the high accuracy of the encoded data, it is important to select the proper modulation scheme according to the current channel condition. In practice, there are many factors to affect the channel condition, and the wireless channel conditions may vary significantly. In particular, the signal observed at the receiver is related to the wavelength of the signal, the background noise, the interference from other ongoing transmissions, the spatial locations of the sender and receiver, the reflecting objects, and the orientation of the antenna of the sender and receiver [5,6]. The superposition of these factors could enhance or dramatically degrade the channel condition. As a result, SNR is widely used as the indicator of the channel condition. According to the value of SNR, the suitable modulation technique can be selected to achieve a balance between data transmission rate and the bit error rate.

2.3. The point coordination function (PCF)

PCF provides contention-free data transmission. In PCF, the AP acts in the role of the point coordinator, and it controls the medium access in a poll-and-response manner. The period during which PCF operates is called the *contention-free period (CFP)*. Before the CFP begins, the AP operates under DCF, but it makes use of the *priority inter-frame space (PIFS)* to seize the medium, and then sends out a beacon packet containing the duration of the CFP. Once the AP has announced

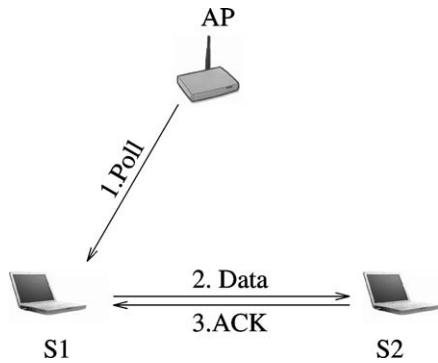


Figure 1. A typical polling procedure.

the start of a CFP, it may start transmitting data to the stations, or sending contention-free poll frames to the stations. During a CFP, a station can only transmit the impending data after being polled by the AP. If the polled station has data to send, it transmits one data packet. Otherwise, it will respond the AP with a NULL packet. When the AP sends a packet to a station, it expects to hear a packet within a *short inter-frame space* (SIFS) [13]. If no packet is heard within a SIFS, the AP will reclaim the medium and send its next poll packet after a PIFS period. When the CFP ends, the AP sends a contention-free end packet. As operating under the PCF, the AP maintains a polling list for selecting stations that are eligible to be polled. Each station entering the coverage area of the WLAN needs to perform an association procedure [8] so that it can be added to the polling list by the AP.

2.4. The motivation

With the explosive growth of WLANs, many real-time applications will be carried on WLANs in the near future. These real-time applications require guaranteed services, such as the minimum bandwidth, the maximum delay. This brings new challenges to the resource management of WLANs. Because PCF is able to provide contention-free medium access and guaranteed services to each flow, it is going to be widely adopted in future WLANs. With the use of PCF, it becomes desirable to design a PCF-based MAC protocol to exploit the multi-rate capability supported by IEEE 802.11 physical layer.

Due to the varying nature of the wireless channel conditions, the sender or the receiver must estimate the channel condition before selecting the proper data transmission rate. Under DCF, rate selection can be achieved by transmitting the selected rate on a per-packet basis with RTS/CTS exchanges. However, this may not work under PCF because there is no message exchange before the data transmission. For example, as shown in figure 1, after the AP sends a polling packet to S_1 , S_1 transmits a data packet to S_2 and S_2 replies an ACK packet then. During this process, there is no RTS/CTS-like message exchange between the sender and receiver. Another option is to let the sender estimate the channel condition according to the history of successful transmissions and select a suitable transmission rate to send the impending data packet. However, as shown in [7], this sender-based approach adapts the

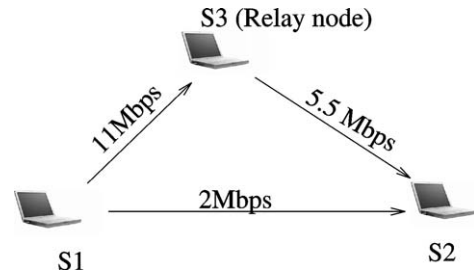


Figure 2. The advantage of using the relay node.

data transmission rate too conservatively, and cannot exploit the possible high rate in short time scale. As a result, the gain of the system performance is limited with such approaches. In WLANs, since mobile stations have computation and power constraints and all stations are in the coverage area of the AP, the data transmission rate selection can be done by the AP instead of the mobile stations. In particular, each station is required to sense the channel conditions of its links and then the AP can collect the channel conditions of all stations, select the proper rate, and piggyback the rate selection decision with the polling packet to the station being polled.

Since the channel condition is time varying and location dependent [18], the multi-rate capability can be further exploited by enabling MAC layer multi-hop transmission. For example, as shown in figure 2, suppose S_1 needs to send data to S_2 , and the channel of $S_1 \rightarrow S_2$ only supports a transmission rate of 2 Mbps. At the same time, the channel conditions of $S_1 \rightarrow S_3$ and $S_3 \rightarrow S_2$ are much better, and they can support data rates of 11 Mbps and 5.5 Mbps, respectively. Suppose the packet length is L , if the data can be transmitted along $S_1 \rightarrow S_3 \rightarrow S_2$ at the MAC layer, the transmission delay is approximately equal to $(\frac{1}{11} + \frac{1}{5.5})L$. Thus, the actual transmission rate is approximately equal to $\frac{5.5 \cdot 11}{5.5 + 11} = 3.7$ Mbps, which is much larger than 2 Mbps, when the packet is transmitted along $S_1 \rightarrow S_2$. There are three things needed to be mentioned: First, the relaying operation is performed at the MAC layer, which is different from the network layer forwarding. Second, we only consider MAC layer relay when the sender and receiver are within the transmission range of each other. Otherwise, network layer packet forwarding is required to deliver the packet. Note that MAC layer relay differs from network layer packet forwarding in that packets relayed at MAC layer do not have queuing delays. Third, the MAC layer relay does not affect the bandwidth allocation of the relay node. Since the polling opportunity of each node is controlled by the AP with PCF, the relaying operation does not interfere the polling opportunities of the relay node. In this paper, we focus on two-hop MAC layer relay, which is sufficient in most cases.

3. The proposed *r*PCF protocol

In this section, we present the details of the *r*PCF protocol. We first present a scheme to collect the channel condition, and then present the rate adaption and relay mechanisms. Finally, we discuss some implementation details.

Sender ID	Rate
0x100001	2.0
0x100014	5.5
⋮	⋮

Figure 3. The data structure of the receive table.

3.1. Collecting the channel condition

As mentioned in section 2.4, the AP is designed to perform the rate adaption for the mobile stations. As a result, the AP must be able to estimate the channel condition from the packet sender to the packet receiver. To achieve this goal, each station performs receiver-based channel condition measurement, and the AP collects the channel condition measurements from each station. In particular, each station is required to have a receive table, denoted by *recv_table*, to store the channel condition measurements of each flow. As shown in figure 3, each entry of the *recv_table* is a tuple: $\langle \text{sender_id}, \text{rate} \rangle$ and the content of each entry is determined as follows. When a station, say S_i , overhears a packet (by listening promiscuously), it measures the SNR when receiving the packet, and selects the proper data transmission rate based on the measured SNR. It also takes the MAC layer header of the packet and uses the source address as the *sender_id*. If the *sender_id* does not exist in the *recv_table*, a new entry is added. If the selected rate is different from the old one, the station needs to update the corresponding rate field. If there is any change in *recv_table*, the station should notify the AP by sending a *rate notification* (RN), which has a structure similar to that of the *recv_table*. In practice, a station only sends the notification when it sends out a data packet or an ACK. For example, when S_i receives a data packet or is polled but no data to send, it checks if any entry in the *recv_table* has been changed since the last time of sending the RN. If there is any change, S_i sends a RN piggybacked with the ACK or the NULL packet to the AP. Since the ACK or the NULL packet is transmitted at the rate of 2 Mbps, the AP can successfully extract the RN from the received packet.

The rate notification tunneling. When S_i is polled and has data packet to send, the situation becomes more complex due to different transmission rates. It is possible that the data packets are transmitted at a high rate (e.g., 11 Mbps) from S_i , whereas the channel quality of $S_i \rightarrow \text{AP}$ can only support a low rate (e.g., 2 Mbps). Hence, the AP may not be able to overhear the data packet from S_i , and cannot extract the piggybacked RN_i correctly. Without the up-to-date channel condition from S_i , the AP may under-estimate or over-estimate the channel capacity of the links related to S_i , and the system performance will be degraded. To deal with this problem, we use the following solution: suppose station S_i sends a data packet to another station S_j , it piggybacks its RN_i (if exists) with the packet. When S_j receives the packet, it checks the transmission rate as follows.

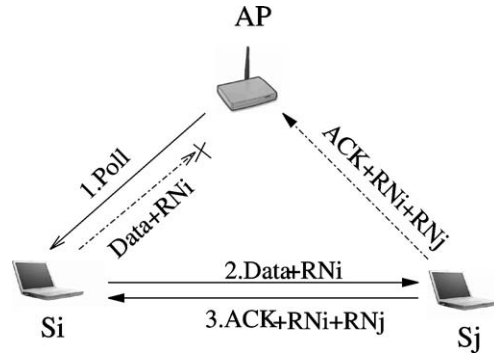


Figure 4. An example of the rate notification tunneling.

- If the transmission rate is equal to the lowest data rate (e.g., 2 Mbps in this paper) or there is no piggybacked rate notification, it replies an ACK piggybacked with RN_j if there is any change in its *recv_table*.
- If the transmission rate is greater than the lowest rate, and RN_i has been piggybacked with the data packet, S_j sends S_i an ACK which piggybacks both RN_i and RN_j (if exists).

This process has been illustrated in figure 4. In figure 4, the dotted line pointing to the AP means that the AP can successfully overhear the packet. We can see that if RN_i is transmitted at a high data rate, it will be tunneled to the AP through the ACK from S_j . Since the ACK is transmitted at the rate of 2 Mbps, the AP can overhear (by listening promiscuously) and extract RN_i from the ACK.

The sender-initiated NACK mechanism. S_j may not be able to receive the packet from S_i correctly under some situations. Since the AP may use out-of-date channel condition to estimate the data transmission rate of $S_i \rightarrow S_j$, it is possible that the AP asks S_i to send the data packet to S_j with a high data rate, which is higher than what the channel can support. As a result, S_j cannot extract the piggybacked RN_i , and cannot tunnel it to the AP. To address this problem, we propose a new acknowledgment mechanism as follows:

- If S_i sends the data packet to S_j and cannot hear the ACK within a SIFS, which means the transmission failed (otherwise, S_j would reply the ACK), S_i generates a NACK and sends out the NACK after a SIFS. If it also has RN_i to send, the RN_i is piggybacked with the NACK.
- If the AP does not hear an ACK after the transmission of the data packet within a SIFS, it will send another poll provided that it does not hear any packet within a PIFS.

This process has been illustrated in figure 5. With the rate notification and the sender-initiated NACK mechanisms, RN_i can always be delivered to the AP in time if all transmissions are successful at the lowest rate. When the transmission failure is detected, there are two options: First, S_i can reduce the rate of $S_i \rightarrow S_j$ to the next lower rate and the AP polls it again for retransmission; Second, the AP just skips over S_i and polls the next station in the polling list. In this paper, we select the first one.

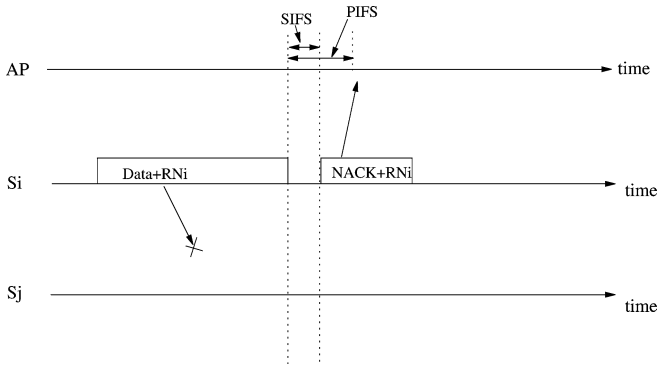


Figure 5. An illustration of the sender-initiated NACK mechanism.

3.2. The rate adaption and relay mechanism

In this section, we describe how the AP finds the proper data rate and how to achieve the MAC layer relay mechanism.

The rate selections. Before the AP polls a station S_i , it needs to determine the data transmission rate for S_i according to the collected channel conditions from other stations. With the estimated channel conditions related to S_i , the AP finds the proper rate for all possible receivers of S_i because it does not know to which node S_i will send. Suppose the AP stores the collected information into a rate table. where an entry, denoted by \hat{r}_{ij} , is the estimated transmission rate from node i to node j (A node refers to a station or the AP). The AP can search the rate table and find out the set of potential receivers of S_i , denoted by \mathcal{R}_i , according to:

$$\mathcal{R}_i = \{\forall j \in BSS \mid \hat{r}_{ij} \neq NULL\}, \quad (1)$$

where BSS represents all nodes in the basic service set. However, $\hat{r}_{ij} \neq NULL$ only implies that node j is within the transmission range of node i , which does not necessarily mean that node j will be the receiver of node i . Since finding and sending all possible transmission rates of the nodes in \mathcal{R}_i may incur unnecessary computation and transmission overheads. In order to get a more accurate receiver set of each station, the AP relies on a learning process as follows: The AP maintains a flow table, denoted by FT , which contains the address pair of the sender and receiver of each flow that has been active in the system. When the AP overhears a data packet, and the address of the sender and receiver has not been registered in the FT , the AP adds the new address pair into the FT . With the FT , \mathcal{R}_i can be represented by:

$$\mathcal{R}_i = \{\forall j \in BSS \mid \hat{r}_{ij} \neq NULL \wedge (i, j) \in FT\}. \quad (2)$$

With \mathcal{R}_i , before the AP polls S_i , it looks up the rate table and finds the proper transmission rates for S_i . Since the way of transmission can be one-hop or two-hop, the data rate set of S_i , denoted by DR_i , is computed according to:

$$DR_i = \left\{ \max\left(\hat{r}_{ij}, \max\left(\frac{\hat{r}_{ik} \cdot \hat{r}_{kj}}{\hat{r}_{ik} + \hat{r}_{kj}}\right)\right) \mid \forall j \in \mathcal{R}_i \wedge \forall k \in BSS \wedge \hat{r}_{ik}, \hat{r}_{kj} \neq NULL \right\}. \quad (3)$$

Dest ID	1st hop rate	(optional)	(optional)
		Relay ID	2nd hop rate
0x100002	5.5	0x100004	11.0
⋮			

Figure 6. The data structure of the data transmission notification.

Intuitively, for each flow whose sender is S_i , the AP selects the highest transmission rate among all possible options the transmission (one-hop or two-hop). The computational cost of the selections is at $O(n^2)$ where n is the number of the nodes in the BSS. Since n is a moderate number per BSS, the cost is affordable to APs.

The relay mechanism. After selecting the data transmission rates for S_i , the AP needs to notify S_i of what rate to use and the way of the transmission. This data transmission notification can be piggybacked with the poll message. Figure 6 shows one example of such data rate notification. If the data rate is achieved by one-hop transmission, the relay ID and the second hop rate will not be needed. If the channel conditions do not change very frequently, the content of two consecutive DR_i s does not have too much difference. In order to reduce the transmission overhead, the AP only sends the difference between the current DR_i and the previously computed DR_i . Meanwhile, each station also has a transmit table which caches the received data transmission notifications. When S_i extracts the data transmission notification from the poll message, it updates its transmit table. Before it sends a data packet, it checks if there is an entry in the table whose destination ID matches the address of the packet receiver. If an entry is found, S_i can get the correspondent transmission mode (one hop or two hop) and data rate. Otherwise, it transmits the data packet at the lowest data transmission rate.

According to the standard of IEEE 802.11, each data packet only has two addresses: one for the sender and one for the receiver. In order to support two-hop MAC layer transmission, the address of the relay node needs to be added into the MAC layer header. With three different addresses, the relay mechanism is designed as follows. If a station receives a poll packet and finds out that the packet to send needs a two-hop transmission, it switches the addresses of the receiver and the relay node so that the packet is transmitted to the relay node first. Meanwhile, it sets the subtype of the packet to a pre-specified value, which indicates that the packet should be delivered in the MAC layer relay manner. When the relay node receives the packet, it checks the subtype of the packet and finds out that the packet needs to be transmitted with two-hop. Then, the relay node sets its address as the source address saves the address of the original sender in the header and sends the packet to the original receiver. If the packet is correctly transmitted, the receiver replies an ACK directly to the original sender (note that the address of the original sender is stored in the packet header). If the packet is corrupted in

either hop, the sender of that hop is able to know the error by overhearing if there is any packet transmitted within a SIFS, and reacts similarly to one-hop transmissions. Figure 7 shows how the relay mechanism works. To achieve high data rate, no ACK is transmitted from the relay node (S_2) to the sender, and no ACK is transmitted from the receiver to the relay node. The acknowledgment is achieved explicitly by overhearing.

3.3. Implementation issues

In this section, we describe how r PCF can be incorporated into IEEE 802.11. The general IEEE 802.11 frame format is shown in figure 8. In addition, figure 9 shows the format of the MAC frame used for sending unicast data packet and ACK. As shown in figure 8, the frame control field, shown in

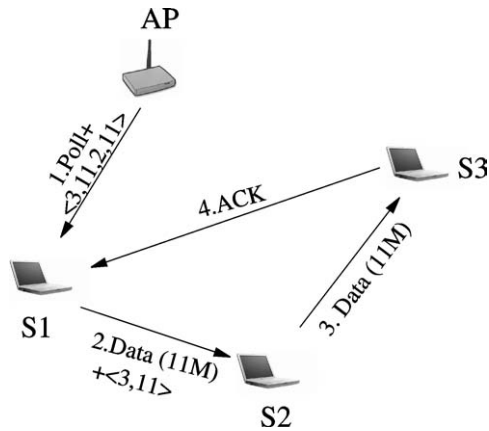


Figure 7. An example of data transmission with relay.

figure 8 is used to carry frame identification information such as the type of the frame (e.g., control or data), the protocol version, and the power/order management information. The duration field contains the time needed for the packet transfer; the four address fields can be used to indicate the BSS identifier (BSSID), source address (SA), destination address (DA), transmitting station address, and receiving station address. These addresses may appear in different order in different type of frames. The sequence control is a sequence number that is used to detect duplicate frames. The FCS is the frame check sequence. The standard formats of data and ACK frames are shown in figure 9, which may not have all the general fields but have the same order of the set of fields. The format of poll frames is almost identical to that of ACK frames, except that poll frames contains BSSID field.

In order to support r PCF, some minor modifications to the standard 802.11 frames are required. Following the design rationale, the modifications are listed as follows:

1. A new data frame format is defined to support MAC layer relay. As shown in figure 10, each data frame uses all four address fields in the order of SA, DA, BSSID, and RA (relay address). The first and second hop relay can be differentiated by the subtype value in the frame control field. For data type, which is indicated by type value of 10 (binary), subtype value can be selected from the reserved ones from 1000 to 1111. For the first hop transmission, the sender sets the destination address to be the address of the relay node, and stores the address of the original receiver in the fourth address field. For the second hop transmission, the relay node changes the subtype of the packet, and puts the

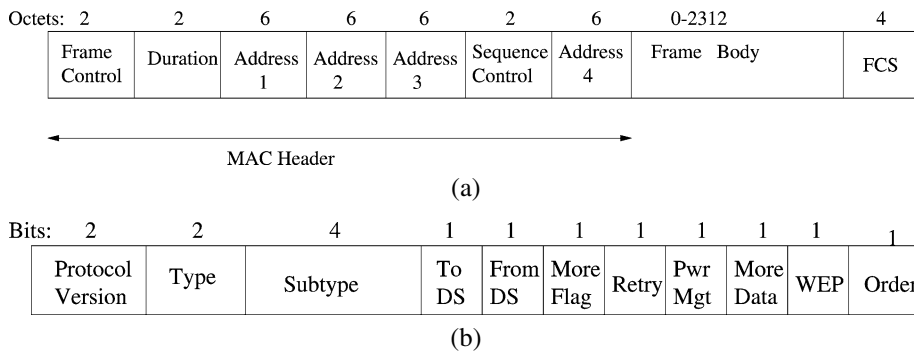


Figure 8. The general 802.11 MAC frame formats. (a) The general packet frame format. (b) The frame control field format.

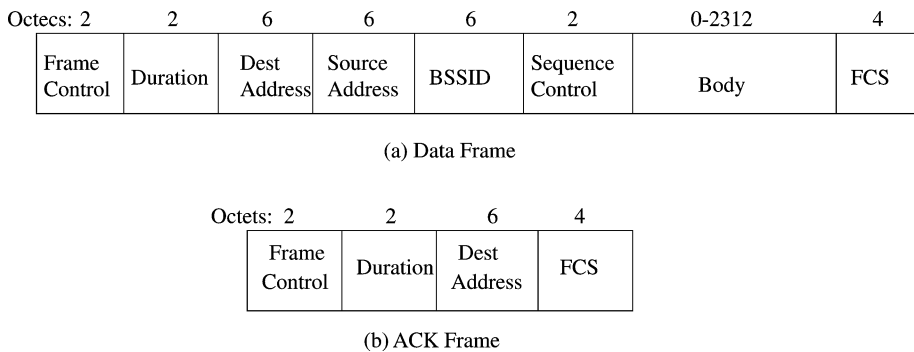


Figure 9. The standard 802.11 MAC data and ACK frame format.

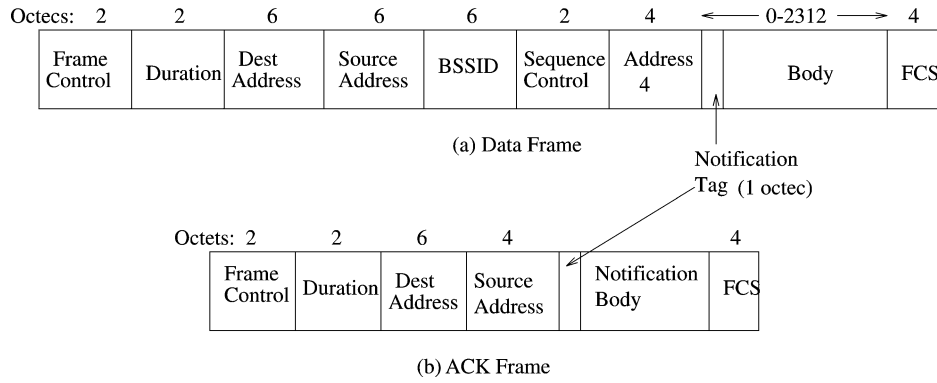


Figure 10. The modified data and ACK frame format.

address of the original sender in the fourth address field. When the packet receiver gets the second hop packet, it picks up the original source address from the fourth address field and sends an ACK to the packet sender.

2. One feature of *r*PCF is that the receiver-based channel condition measurement is performed on each station and the results of the measurement is reported to the AP through the rate notifications piggybacked with data or ACK transmitted from each station to the AP. Meanwhile, before sending poll, the AP makes rate adaptation for each station based on the collected channel conditions, and notifies the station about the transmission mode and correspondent rates through data transmission notifications piggybacked with the poll. In order to identify the piggybacked notifications, as shown in figure 10, an 8 bits *notification tag* is added right behind the *last* address field of the frame. The value of the tag indicates the number of rate notifications or data transmission notifications piggybacked with the frame. For data transmission notifications, one-hop and two-hop notifications can be differentiated by the different subtype value.
3. In order to collect and estimate the channel condition of its neighbors, each station needs to know the address of the packet sender of each packet transmitted. Thus, the standard IEEE 802.11 ACK frame format has been changed to include the source address.

4. The performance evaluations

4.1. Propagation models

The quality of the wireless channel places fundamental limitations on the performance of wireless networks. Unlike wired channels which are stationary and predictable, the quality of the wireless channel quality may be quite random due to many factors, such as interference, fading, relative location between the sender and receiver. Several propagation models [14] have been proposed to estimate the signal strength for an arbitrary pair of sender and receiver. There are two categories of propagation models: one is the *large-scale* propagation model which is used for characterizing the mean signal

strength of the sender and receiver; the other is the *small-scale* propagation model which focuses on characterizing the rapid fluctuations of the received signal strength over a short time duration (on the order of seconds).

We use the Friis free space propagation model and the 2-ray ground propagation model as the large-scale propagation model to simulate the situation when the wireless channel condition is stable. The Friis free space propagation model is defined as follows:

$$P_r(d)_{\text{Friis}} = \frac{P_t G_t G_r \lambda}{(4\pi)^2 d^2 L}, \quad (4)$$

where d is the distance between the sender and the receiver, P_t and P_r represent the transmit power and the receive power (in Watts), G_t and G_r are the transmit and receive antenna gains, λ is the carrier wavelength (in meters), and L is the system loss factor. In this paper, P_r is equal to 15 dBm, G_t , G_r , L are set to be 1.0 and λ is equal to $3 \times 10^8 / 2.4 \times 10^9 = 0.125$ m (note that 802.11b operates on 2.4 GHz). The Friis model works well only when the sender and receiver have a clear, unobstructed line-of-sight propagation path between them, and hence the Friis model cannot work accurately alone. The 2-way ground model is based on geometric optics, and it considers both the direct path and the ground reflected propagation path between the sender and the receiver. Under 2-way ground model, given a distance d from the sender, the received power can be expressed as

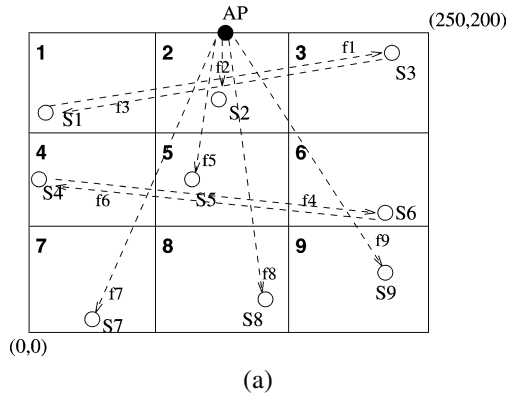
$$P_r(d)_{2\text{-way}} = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4}, \quad (5)$$

where h_t and h_r are the transmit and receive antenna heights, and equal to 1.0 meters in this paper. In our simulations, we combine these two models by using the Friis model at short distance and the 2-way ground model at long distance. In particular, a crossover distance d_{cross} is defined as

$$d_{\text{cross}} = \frac{4\pi h_t h_r}{\lambda}. \quad (6)$$

If $d > d_{\text{cross}}$, $P_r(d)$ is calculated by equation (5); otherwise, $P_r(d)$ is obtained through equation (4).

When there is multi-path fading or relative movement between the sender and receiver, the channel condition between them may change frequently. The movement of the mobile



ID	AP	S1	S2	S3	S4	S5	S6	S7	S8	S9
x	125	13	112	223	9	96	224	35	96	237
y	200	139	163	192	105	92	70	11	36	59

(b)

Figure 11. The experiment setup. (a) The experiment setup. (b) The coordinates of each node.

nodes causes a Doppler shift in the frequency during the communication between the sender and receiver. 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 [14] to simulate the fading channel conditions. The Ricean distribution is given by:

$$p(r) = \frac{r}{\sigma^2} e^{-(r/(2\sigma^2)+K)} I_0(2Kr), \quad (7)$$

where K is the distribution parameter representing the line-of-sight component of the received signal, σ^2 is the variance of the background noise, r is the received power, and $I_0(\cdot)$ is the modified Bessel function of the first kind and zero order [14]. For $K = 0$, the Ricean distribution reduces to the Rayleigh distribution, in which the signal is transmitted only by reflecting. We will use the Ricean distribution in our experiments.

4.2. The simulation setup

Our simulation is based on ns-2 extension [12,17]. The experiment setup is shown in figure 11(a). As shown in figure 11(a), one AP and nine stations exist in a 250×200 flat area. The coordinates of the nodes are listed in figure 11(b), and the distance between any two nodes can be calculated. We assume the stations are evenly distributed in the area. We divide the area into 9 subareas indexed from 1 to 9, and the mobile station with the same index resides in the correspondent subarea. Each station has a flow, starting either from another station or from the AP. If the connection is from one station to another, the flow number is the same as the index of the sender (e.g., f_1 represents the connection from $S_1 \rightarrow S_3$). If the connection is between the AP to a station, the flow number is the same as the index of the receiver (e.g., f_2 represents the connection from $AP \rightarrow S_2$). Similar to [15], the distance threshold for 11 Mbps, 5.5 Mbps, and 2 Mbps are 100 m,

200 m, and 250 m, respectively. The thresholds for different data rates are chosen based on the distance ranges [13]. For simplicity, all flows in the system are assumed to have the same type of traffic source with the same rate. There are two kinds of traffic configurations: (1) each sender always has data to send; (2) each sender has CBR traffic with the rate of 20 packet/second. The packet length is set to be 1000 bytes and the simulation time is assumed to be 500 seconds.

We compare r PCF with PCF-based ARF [9], which is a sender-based rate adaption scheme. The protocol works as follows. If ACKs for two consecutive data packets are lost, the sender reduces the transmission rate to the next lower data rate and starts a timer. If ten consecutive ACKs are received, the transmission rate is raised to the next higher data rate and the timer is cancelled. If the timer expires, the transmission rate is raised if an ACK is received for the next packet; otherwise, the rate is lowered again and the timer is restarted. Since the simulation results of [7] show that the throughput of ARF does not change much when the timeout is larger than 60 ms, we set the timeout value to be 80 ms in our simulations.

4.3. Simulation results

We evaluate the impacts of different factors on the system performance, which is measured by the system throughput and the average packet transmission delay. We also evaluate the system overhead of the proposed solution.

4.3.1. Performance under stable links

In this subsection, we assume the channel conditions are stable, and the received power is calculated only by the Friis and 2-ray ground propagation models. Further, all stations are assumed to be static. In this way, we can maximize the performance of different schemes. The simulation results are shown in figure 12. From figure 12(a), we can see that r PCF improves the system throughput by more than 70%. The performance gain is mainly due to the MAC layer relay. For example, the transmission rate for f_1 is limited to 2 Mbps under ARF due to the long distance between S_1 and S_3 . Under r PCF, the packets of f_1 can be relayed by the AP. Because the data rates of $S_1 \rightarrow AP$ and $AP \rightarrow S_3$ are 5.5 Mbps and 11.0 Mbps, respectively, the actual data rate of f_1 is approximately 5.5 Mbps. As shown in figure 12(c), under ARF, the delays of f_1 , f_3 , f_4 , and f_6 are much higher than the rest of the flows. The reasons are as follows. Let us take f_1 as the example, ARF raises the transmission rate of f_1 to 5.5 Mbps after S_1 receives the ACKs for ten consecutive packets. Then, this data rate raising will cause f_1 's next packet to fail since the distance between S_1 and S_3 is too far to support 5.5 Mbps. In addition, ARF reduces the transmission rate of f_1 back to 2.0 Mbps only when S_1 misses two ACKs for two consecutive data packets. This leads to extra delay of f_1 due to re-transmissions. Furthermore, the extra delay also affects the delay of other flows (e.g., f_2) since the AP polls each station in a round robin way. This also explains why the delay of f_2 under ARF is still greater than that under r PCF al-

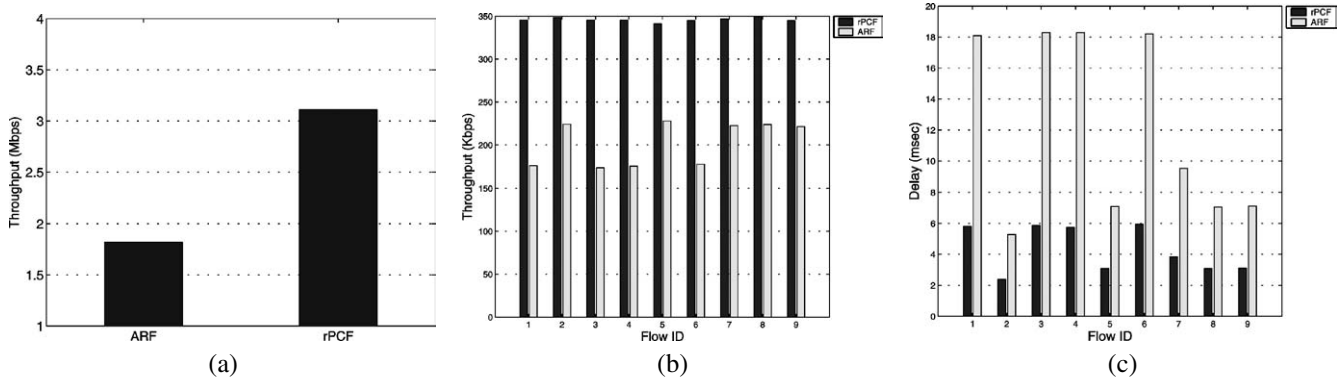


Figure 12. Performance comparisons between r PCF and ARF. (a) System throughput. (b) Throughput of each flow. (c) Delay of each flow (rate = 20 pkt/s).

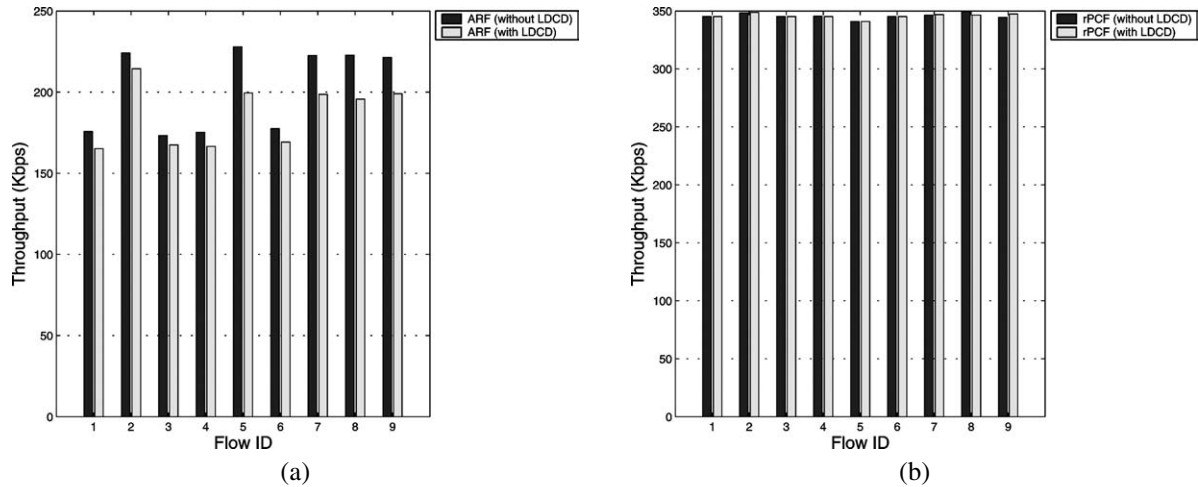


Figure 13. Impacts on throughput. (a) Throughput difference of ARF. (b) Throughput difference of r PCF.

though the transmission rate of f_2 is 11.0 Mbps under both schemes.

4.3.2. Impacts of location dependent channel degradation

The wireless channel condition can be affected by several factors. One of them is the surrounding environment, and the channel quality can be significantly degraded when the surrounding interference or noise is high. As a result, the location dependent channel degradation (LDCD) commonly exists in wireless networks. To evaluate its impacts, we assume that the channel condition of $AP \rightarrow S_5$ is poor, and it can only support a data rate of 2 Mbps. All stations are assumed to be static. The simulation results are shown in figures 13 and 14. As shown in figure 13, the system throughput of ARF is reduced. However, the throughput is not affected under r PCF. Under r PCF, after S_5 sensed the poor channel condition between the AP and itself, it notifies the AP of such change. When the AP needs to send a data packet to S_5 , it can find that the direct link between them can only support 2 Mbps, but a higher data rate can be achieved if the packet is relayed by S_2 . Since the rates of $AP \rightarrow S_2$ and $S_2 \rightarrow S_5$ are both 11.0 Mbps, with the relay of S_2 , the actual transmission is approximately 5.5 Mbps, which is approximately the same as the rate can be support if the channel condition of $AP \rightarrow S_5$ is good. For the same reason, as shown in figure 14, we can

Table 1
The control overhead of r PCF.

Type	Size (bits)
RN	52
One-hop TN	52
Two-hop TN	104
NACK	304

find that the packet delay increases under ARF, but remains the same under r PCF.

4.3.3. Impacts of the line-of-sight parameter K

In the previous two subsections, we assume the channel condition is stable. Since the channel condition is stable and the stations do not move, the AP and the mobile stations only send a very limited number of rate notifications (RNs) and data transmission notifications (TNs), and the overhead of r PCF is negligible. Since the wireless channel condition may change frequently, the overhead of r PCF, which includes transmitting RNs, TNs, and NACKs, should be evaluated. To evaluate this overhead, we change the line-of-sight parameter K . A large K means a better channel quality while a small K means a poor channel quality. The sizes of RN, TN, and NACKs are listed in table 1. We collect the generated RNs, TNs and NACKs, and add them together to get the control

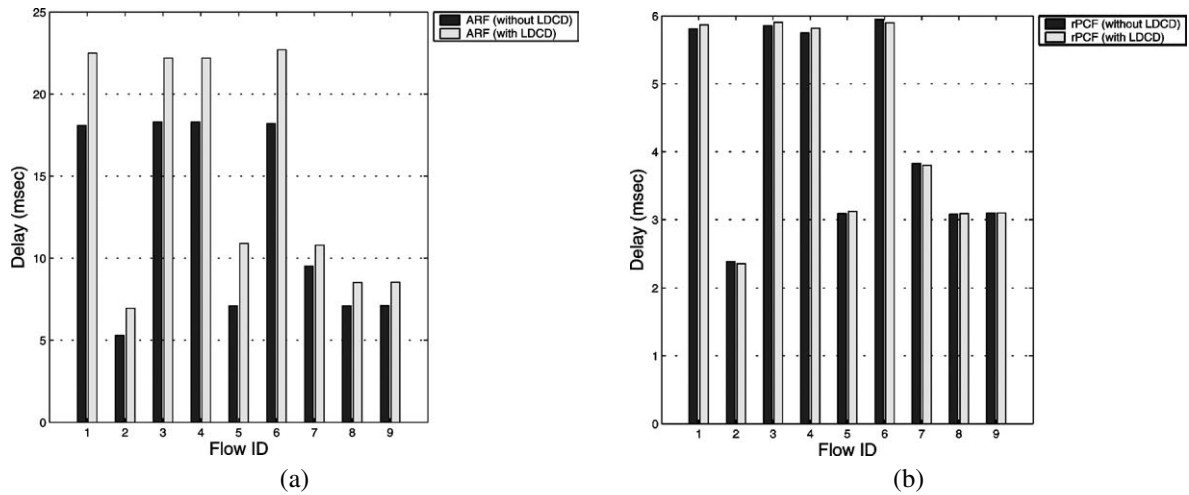


Figure 14. Impacts on delay. (a) Delay difference of ARF (rate = 20 pkt/s). (b) Delay difference of r PCF (rate = 20 pkt/s).

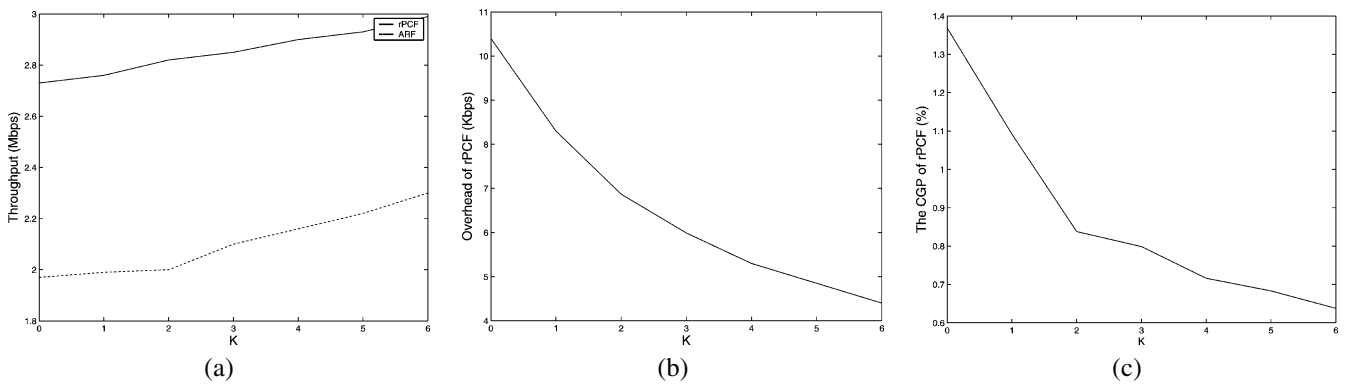


Figure 15. Impacts of the line-of-sight parameter K . (a) Throughput of ARF and r PCF. (b) Overhead of r PCF. (c) The CGP of r PCF.

overhead in terms of bits. Then, using these control overhead (in bits) divided by the simulation time, we get the control overhead in terms of data rate (bps). Since ARF does not have any control overhead, we introduce a new metric called *cost gain percentage* (CGP) to evaluate the control overhead of r PCF fairly. The CGP (in percent) is the ratio of the control overhead (bps) to the throughput gain of r PCF (compared to ARF).

The simulation results are shown in figures 15 and 16. From figure 15(a), we can see that the system throughput of both r PCF and ARF increase as K increases. As K increases, the impact of fast fading becomes smaller, and the channel condition becomes better. As a result, the system throughput improves. From the figure, we can also see that the impact of K on the throughput of ARF is higher than that on r PCF. This is due to the fact that the number of mis-adaptations increases under ARF when the channel condition becomes unstable. However, under r PCF, each station frequently reports the channel conditions to the AP, so that the AP can perform rate adaption and notify correspondent stations in time. As a result, stations can adjust the transmission rates in time and reduce chance of using wrong data rates. As shown in figure 15(b), the control overhead of r PCF drops quickly as K increases. From the results shown in figure 15(c), we can see that the control overhead is very small. For example, the CGP

is only about 1.3% when $K = 0$, and drops to 0.6% when $K = 6$. As shown in figures 16(a) and (b), we can see that the impact of K on the packet delay of ARF and r PCF. For the same reason, the impact of K on r PCF is less than that on ARF. Meanwhile, the packet delay of each flow under r PCF is much smaller than that under ARF.

4.3.4. Impacts 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 power. Second, due to Doppler shift in frequency of the received signal, it may reduce the channel coherence time period, which is the interval during which the channel condition keeps stable. In this subsection, we evaluate the impact of mobility on the performance of r PCF. Each station is assumed to move randomly within the correspondent subarea with the same speed, and the mean speed changes from 0 to 5 m/s. Similar to [15], the value of K is fixed to be 5. The simulation results are shown in figures 17 and 18. As shown in figure 17(a), the system throughput of r PCF and ARF drops as the the mean moving speed increases. As the moving speed increases, the channel condition changes frequently. When the channel condition becomes unstable (e.g., when the mean speed is 5 m/s), it is difficult for ARF to adjust the transmission

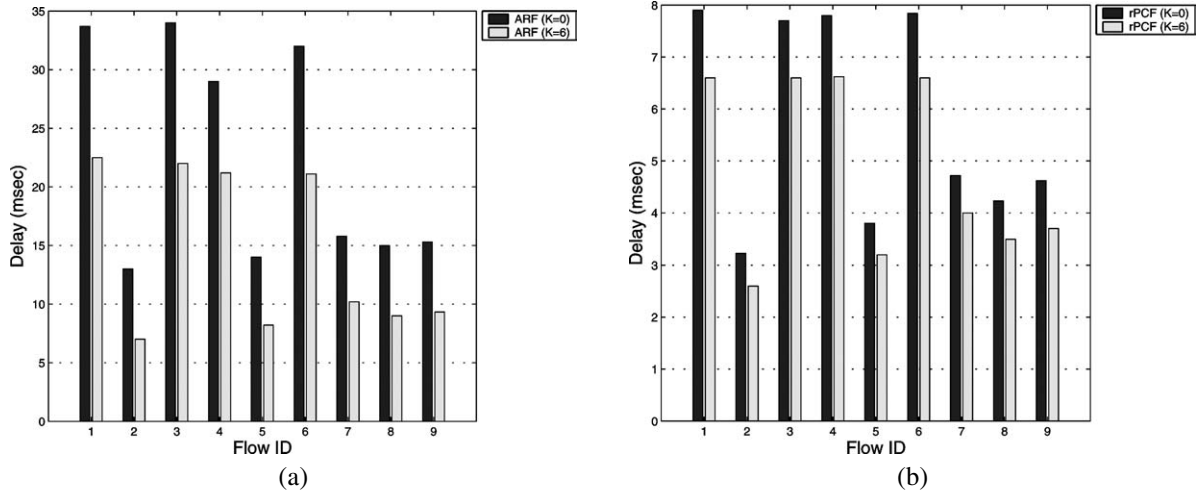


Figure 16. Impacts of the line-of-sight parameter K . (a) Delay difference of ARF (rate = 20 pkt/s). (b) Delay difference of $rPCF$ (rate = 20 pkt/s).

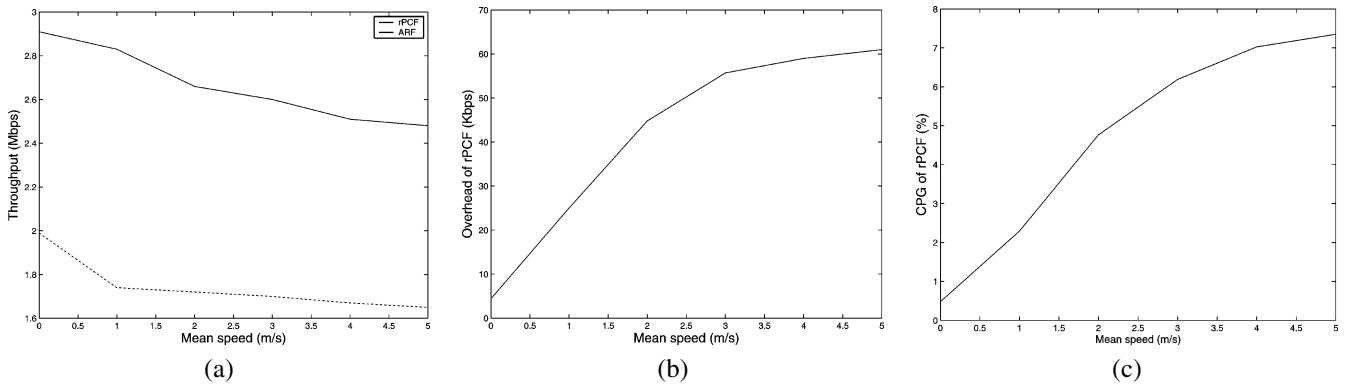


Figure 17. Impacts of mobility. (a) Throughput of ARF and $rPCF$. (b) Overhead of $rPCF$. (c) The CGP of $rPCF$.

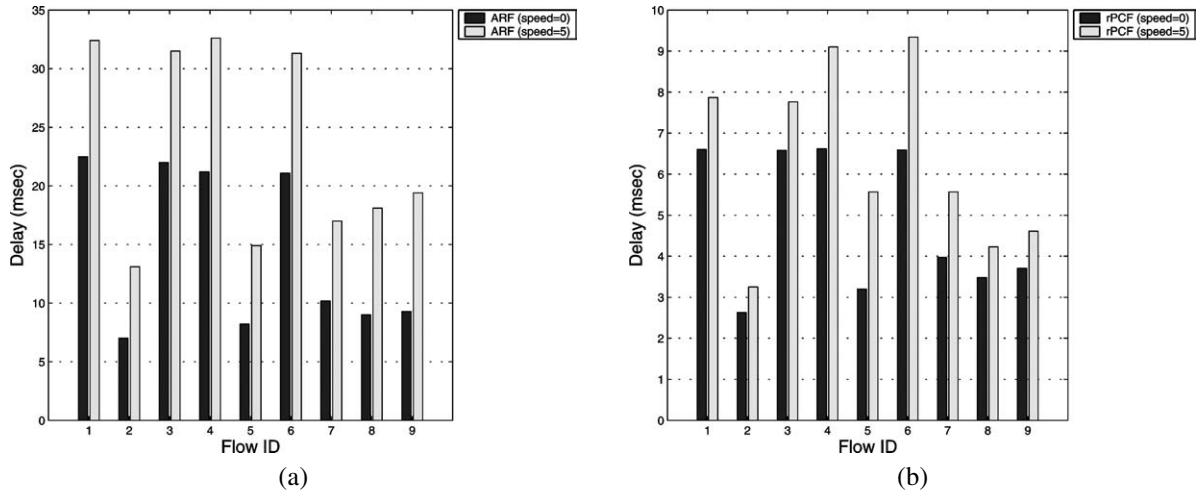


Figure 18. Impacts of mobility. (a) Delay difference of ARF (rate = 20 pkt/s). (b) Delay difference of $rPCF$ (rate = 20 pkt/s).

rate to a high rate, and hence most of the data are transmitted at a high rate of 2 Mbps. For $rPCF$, the increased control overhead such as NACKs, RNs and TNs also reduces the system throughput as the moving speed increases. This is also verified in figure 17(b). As shown in figure 17(c), the control overhead is still very small. For example, when the mean speed is 5 m/s, the CGP is only about 7.7%. From

figures 18(a) and (b), we can see that the mobility increases the packet delay of $rPCF$ and ARF. However, the impact on ARF is more than that on $rPCF$ because the net increase of the delay under ARF is larger than that under $rPCF$. As we can see, the packet delay of each flow is increased by more than 9 milliseconds under ARF, but only less than 3 milliseconds under $rPCF$. Again, this can be explained by the

fact that *r*PCF can quickly react to channel condition updates.

5. Conclusions and future work

In this paper, we presented a novel relay-enabled PCF protocol, called *r*PCF, to exploit the physical layer multi-rate capability. Since IEEE 802.11b supports multiple data 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. Many protocols and their optimizations are proposed to help the AP collect the up-to-date channel information and notify the mobile node what data rate to use and whether to use a relay station. Simulation results showed that *r*PCF outperforms PCF-based ARF in terms of throughput and transmission delay with a negligible control overhead. In the future, we can still improve the performance of *r*PCF by further reducing the control overhead. For example, the channel estimate and rate adaptive schemes used in [2] may be able to reduce the number of unproductive rate notifications and misadaptations when the channel condition changes frequently. Although we only discussed two-hop transmission, the idea can be generalized to multi-hop data transmissions. Besides *r*PCF, we also will design the protocols to support the MAC layer relay mechanism for DCF.

Acknowledgements

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

References

- [1] Agere Systems Inc., Wireless LAN products and technologies, <http://www.agere.com/client/wlan.html>.
- [2] K. Balachandran, S.R. Kadaba and S. Nanda, Channel quality estimation and rate adaptation for cellular mobile radio, *IEEE Journal on Selected Areas in Communications (JSAC)* (July 1999) 1244–1256.
- [3] V. Bharghavan, A.J. Demers, S. Shenker and L. Zhang, A media access protocol for wireless LANs, in: *ACM Sigcomm '94* (1994) pp. 212–225.
- [4] A. Ganz, A. Phonphoem and Z. Ganz, Robust superpoll with chaining protocol for IEEE 802.11 wireless LANs in support of multimedia applications, *Wireless Networks* (2001).
- [5] J. Geier, *Wireless LANs*, 2nd ed. (Sams Publishing, 2002).
- [6] A. Hills, Large-scale wireless LAN design, *IEEE Communication Magazine* (November 2001).
- [7] G. Holland, N. Vaidya and P. Bahl, A rate-adaptive MAC protocol for multi-hop wireless networks, in: *ACM Mobicom 2001* (July 2001).
- [8] IEEE, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Spec, IEEE 802.11 Standard (1999).
- [9] A. Kamerman and L. Monteban, WLAN-II: A high-performance wireless LAN for the unlicensed band, *Bell Labs Technical Journal* (Summer 1997).
- [10] A. Kospel, J.P. Ebert and A. Wolisz, A performance comparison of point and distributed coordination function of an IEEE 802.11 WLAN in the presence of real-time requirements, in: *7th International Workshop on Mobile Multimedia Communication (MoMuC2000)* (October 2000).
- [11] Nokia Inc., WLAN and GPRS data is now in the cards for on-the-go connectivity, http://press.nokia.com/PR/200203/852460_5.html.
- [12] R. Punnoose, P. Nikitin and D. Stancil, Efficient simulation of Ricean fading within a packet simulator, in: *IEEE VTC 2000* (2000) pp. 764–767.
- [13] D. Qiao, S. Choi, A. Soomro and K.G. Shin, Energy-efficient PCF operation of IEEE 802.11a wireless LAN, in: *IEEE INFOCOM'02* (June 2002).
- [14] T.S. Rappaport, *Wireless Communications: Principle and Practice* (Prentice Hall, Englewood Cliffs, NJ, 1996).
- [15] B. Sadeghi, V. Kanodia, A. Sabharwal and E. Knightly, Opportunistic media access for multirate ad hoc networks, in: *ACM Mobicom 2002* (July 2001).
- [16] O. Sharon and E. Altman, An efficient polling MAC for wireless LANs, *IEEE/ACM Transactions on Networking* (August 2001).
- [17] VINT group, UCB/LBNL/VINT Network simulator – ns (version 2), <http://mash.cs.berkeley.edu/ns>.
- [18] B.H. Walke, *Mobile Radio Networks: Networking, Protocols and Traffic Performance*, 2nd ed. (Wiley, 2002).
- [19] H. Wu, C. Qiao, S. De and O. Tonguz, Integrated cellular and ad hoc relaying systems: iCAR, *IEEE Journal on Selected Areas in Communications (JSAC)* (October 2001) 2105–2115.



Hao Zhu received his B.S. degree from Xian Jiaotong University, Xian, China and his M.S. degree in computer science and engineering from the Institute of Software, Chinese Academy of Sciences, Beijing, China. He is currently working towards the Ph.D. degree at the Pennsylvania State University. His research interests include resource management, QoS and power efficient protocols in wireless networks.
E-mail: hazhu@cse.psu.edu



Guohong Cao received his B.S. degree from Xian Jiaotong University, Xian, China. He received the M.S. degree and Ph.D. degree in computer science from the Ohio State University in 1997 and 1999, respectively. Since Fall 1999, he has been an Assistant Professor of computer science and engineering at The Pennsylvania State University. His research interests are mobile computing, wireless networks, and distributed fault-tolerant computing. He currently leads several projects on resource management and data dissemination in mobile environments. He has served as an editor of *IEEE Transactions on Mobile Computing* and *IEEE Transactions on Wireless Communications*, and has served on the program committee of various conferences including ICDCS and INFOCOM. He was a recipient of the Presidential Fellowship at the Ohio State University in 1999, and a recipient of the NSF CAREER award in 2001.
E-mail: gcao@cse.psu.edu