Forward focus: using routing information to improve medium access control in ad hoc networks

Brent Ishibashi*† and Raouf Boutaba

David R. Cheriton School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Canada N2L 3G1

Summary

Multihop packet forwarding is a vital process in an ad hoc network. All ad hoc networking protocols, but particularly routing and medium access control protocols, must work together in order for the network to be successful. However, current MAC protocols such as IEEE 802.11 do not consider this multihop nature at all. This work develops a modification to 802.11 that focuses on forwarding packets. Routing information is utilized to streamline the sharing of the medium, by allowing forwarding nodes to reuse an already-acquired channel. Using forward focus (FF), nodes are encouraged to participate in the forwarding process and are rewarded for doing so. Simulation-generated performance evaluations reveal that the result is a MAC protocol with improved efficiency and effectiveness.

Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: ad hoc networks; medium access control; cross layer; multihop; routing; forwarding

1. Introduction

Packet forwarding is a fundamental process in an ad hoc network. In a conventional wireless local area network (WLAN), a single transmission will deliver the packet to its destination—the network base station. In an ad hoc network no base station exists. Packets are sent in a multihop manner, with intermediate nodes forwarding the packet to its ultimate destination. Each intermediate node must make an additional transmission of the packet. This presents several problems. First is the determination of the path that the packet will take; this is the focus of ad hoc routing. Second, each intermediate node must consume valuable resources in order to forward the packet. Finally, the additional hops incur additional delay.

For each transmission, a node must utilize the wireless medium. In order to prevent all nodes from accessing the same channel simultaneously, a medium access control (MAC) protocol is used. Although a number of technologies exist, the IEEE 802.11 MAC has been commonly used for a large amount of ad hoc network research. However, 802.11 experiences several limitations in an ad hoc environment.

IEEE 802.11 was designed for use in WLANs. Although its operation allows it to be used in an ad hoc network, it does not directly support the properties of ad hoc networking. In fact, the very nature of ad hoc networking creates an environment of heavy traffic.
load (congestion) and high node density, two scenarios where 802.11’s performance suffers. While 802.11 has numerous attractive characteristics for its continued use, clearly it can be improved specifically for ad hoc networks.

This paper will address just that—improving IEEE 802.11 specifically for ad hoc networking. A MAC-level mechanism is used to improve the efficiency of the forwarding process. However, in order for the mechanism to work, it must have knowledge of the forwarding process itself. By making use of routing protocol information, the MAC improves its own efficiency, and in turn provides better service to the network layer.

Two approaches are described, both of which work to encourage intermediate nodes to forward packets as quickly as possible. The first, called immediate forwarding (IF), takes the policy of forwarding packets immediately after receiving them. It uses a MAC protocol mechanism to streamline this process, so that packets are rapidly relayed from source to destination. Unfortunately, while IF is shown to work in simple scenarios, it falters in a more realistic scenario. A second approach, called forward focus (FF), is therefore proposed. FF uses the same protocol mechanism as IF, but slightly alters the forwarding policy. By doing so, it encourages the forwarding of packets, rather than immediately forwarding them as in IF. It achieves improved performance by reducing the amount of medium time that is idly wasted by the protocol, as well as the need for repeated contention for the medium.

2. Medium Access Control in Ad hoc Networks

2.1. Ad hoc Networks

In an ad hoc network, nodes use wireless communication to exchange packets with other nodes. However, due to the range limitations of a wireless transmission, a node cannot send directly to all other nodes in the network. To allow a source to communicate with a more distant destination, packets are forwarded in a multihop manner. Other nodes in the network serve as routers for the source-destination pair. On receiving a packet from the traffic flow, these intermediate nodes relay the packet to the next node along the path, as shown in Figure 1(a).

The interconnection of neighbor nodes forms a mesh topology, as shown in Figure 1(b). However, because all nodes in the network are potentially mobile, the topology may be very dynamic. Frequent changes make determining and maintaining paths between sources and destinations within the mesh exceptionally challenging. A large volume of work has addressed the development of ad hoc routing protocols, designed to deal with the dynamic ad hoc environment.

An ad hoc routing protocol must find a balance between keeping current and correct routes, and avoiding excessive routing control overhead. Both proactive (table-driven) approaches such as DSDV [1], CCSR [2], or OLSR [3], and on-demand protocols such as AODV [4,5] or DSR [6], take steps to reduce routing overhead incurred by topology changes. As proactive protocols attempt to maintain complete routing information, they must react to all topology changes, although overhead can be reduced by delaying some of these reactions. On-demand protocols instead only build routes as they are needed; although this process is rather expensive, in most networks only a small number of the total possible paths are ever used.

Once a route is determined, packets can be sent between a particular source and destination. For a packet to be successfully delivered, it must be forwarded correctly by each intermediate node along the path. This requires the intermediate node to perform several operations. First, the node must participate in the operation...
of the routing protocol. This will allow the node to be included in the resulting path. Second, it must receive an incoming packet from the previous hop node along the path. After the packet is received, the next hop address must be determined, and then finally the packet can be retransmitted to the next node on the path.

Successful delivery of a packet requires that all nodes fulfill their responsibilities. However, an intermediate node expends valuable resources forwarding packets for other nodes. Handling a packet consumes power, processing, and wireless resources, in order to perform all the operations required to relay the packet on to the subsequent node. Despite this expenditure, the forwarding nodes receive no direct benefit from performing their duties.

Packets are only considered successful when they are delivered to their destination. The source and destination receive the benefit of the delivery, however the intermediate nodes do a large proportion of the work. This creates an interesting situation in an ad hoc network, where intermediate nodes are saddled with the responsibility and burden of forwarding packets for others, without directly receiving benefit from the process. Although they may receive reciprocal benefits from other nodes, this may or may not occur, regardless of their own actions.

2.2. Medium Access Control

By relying on wireless communication, ad hoc networks inherit a number of characteristics from the air interface. Relatively low bandwidth and high error rates limit the capacity, the medium can support. However, even more importantly, a node must share the medium with all the other nodes within range of one another. This means that this already scarce resource must be divided between all of these nodes.

When using a shared medium, only a single transmission can occur on a particular channel at a particular time. If two (or more) nodes transmit packets simultaneously, a collision will occur. If this happens, the two signals interfere with each other and neither can be correctly received. Therefore, it is critical that only one node transmit at a time within the range of the receiver.

For this reason, a MAC protocol is used to define rules for how a node can access the medium. However, avoiding collisions is not the only requirement of a good MAC. It must also permit all nodes fair use of the medium. Efficiency is also important, as the MAC must not unnecessarily waste valuable bandwidth through protocol overhead.

While a large number of wireless MAC protocols exist, not all are appropriate for ad hoc networking. Several, such as the global system for mobile communications (GSM) [7] and HiperLAN/2 [8], use time-slotted approaches for assigning medium access. While this provides contention-free access, the decentralized topology of an ad hoc network makes coordinating this approach difficult.

Technologies such as Bluetooth [9] (now included in IEEE 802.15 [10]) and CDMA have been suggested for use in ad hoc networks. Bluetooth, designed as a wireless cable replacement, is well-suited for ad hoc scenarios. However, design of Bluetooth has made most ad hoc networking solutions very technology dependent. CDMA, while very successful in cellular systems, suffers from several issues when applied to an ad hoc network. First, unique codes must be established for each node, and distributed to the other nodes [11,12]. Second, CDMA suffers from a near-far problem: large differentials in received power can prevent the successful reception of the data. In a cellular network this is avoided by balancing the received power at the access point, however this is difficult in a non-centralized network.

Because of this distributed nature, carrier-sense multiple access (CSMA)-based MAC protocols have received the bulk of the attention for ad hoc networks. Nodes sense the medium and defer if the medium is busy; only when the medium is sensed as idle can a node attempt to transmit. However, if all nodes attempted to use the medium as soon as it became idle, a collision would almost certainly occur. Therefore, statistical methods are used in order to improve the chances of only one transmission occurring at a particular time. This collision avoidance mechanism forces nodes to wait random intervals before attempting to transmit.

Either the sender or the receiver can contend for the medium. It has been shown in References [13,14] that a receiver-based protocol can in fact make more efficient use of the medium. This is due to the fact that the receiver is better aware of its own channel condition; the sender’s channel condition in order to interfering with other transmissions. However, it seems more natural for the sender to initiate a transmission, as it is the one that knows that there is data to be sent. For this reason, existing technologies such as HiperLAN [15], and IEEE 802.11 [16,17], along with a number of predecessors [18–20], rely on sender-based contention mechanisms. Both 802.11 and HiperLAN support both infrastructure and ad hoc modes, however 802.11 has been adopted much more rapidly [21].
Many other MAC improvements have also been proposed [22–24]. Protocols have been developed to conserve power, a large concern in the mobile environment [25–28]. Many realized that adjusting transmission power levels could also have other benefits: controlling topology [29–31] and better spatial reuse of the medium resulting in improved throughput [32]. Transmitting at lower power could reduce the number of direct neighbors a node was forced to compete with. Lower powers cause less interference, allowing more exchanges to occur simultaneously.

References [33–35] make use of busy-tone schemes. The transmitting node uses a small secondary channel to transmit a signal indicating its use of the primary channel. This busy tone can then be used in order to improve channel reuse, avoid collisions, and allow collision detection. The secondary channel does require the subdivision of the channel, reducing the bandwidth dedicated to data transmission.

Part of the problem with these schemes is that they require a more complex transceiver. Eventually, technological improvements will likely make them, and others, a reality. For example, work on antenna technology has produced several concepts that could be useful to ad hoc networks. First, directional antennae allow a node to focus its transmissions in a particular direction [36,37]. Beam-forming can increase a node’s range and decrease the number of nodes it interferes with, compared to an omnidirectional antenna. Multiple-in, multiple-out (MIMO) technologies could also be used in an ad hoc MAC. This would allow nodes to send and/or receive multiple transmissions simultaneously, improving the chances of successfully reaching a node [38].

3. IEEE 802.11 MAC

With the exception of work focused specifically on developing new MAC protocols for ad hoc networks, almost all upper-layer work has relied on IEEE 802.11. The reason for this: 802.11 technology is relatively simple and inexpensive. 802.11 has been widely accepted and deployed in WLAN situations. Because of this, the 802.11 MAC protocol has been extensively studied [21,39] and is implemented in most common simulation packages. Although 802.11 is not ideal, it does present an interesting baseline for comparison and experimentation.

Although different versions are available (a, b, g), the versions differ primarily at the physical layer. Two mechanisms are contained within the specification. The point coordination function (PCF) uses a time-slotted access mechanism suitable for centralized systems, while the distributed coordination function (DCF) is a CSMA-based (carrier sense multiple access) method with collision avoidance (CA). While intended as a WLAN protocol, the DCF also allows direct node-to-node communication in ad hoc mode.

3.1. Protocol Description

The basic CSMA/CA mechanism of 802.11 consists of a contention period, where the medium is idle, followed by the transmission of data, then a positive acknowledgement. A short interframe space (SIFS) is also included as the minimum separator between two consecutive packet transmissions.

The contention period is designed to give every node an opportunity to access the medium. In order to attempt access, each node must first wait until it senses an idle period of at least the distributed interframe space (DIFS). However, if nodes attempted to transmit immediately following every idle DIFS, collisions would be virtually assured, with the resulting transmissions wasted. Therefore, a number of mechanisms are included for the purpose of avoiding and reducing the cost of collisions.

Any time a node wishes to send, it senses the medium. If the medium is idle, and continues to be idle for the period of a DIFS, it can begin its packet transmission. The point coordination function (PCF) uses a time-slotted access mechanism suitable for centralized systems, while the distributed coordination function (DCF) is a CSMA-based (carrier sense multiple access) method with collision avoidance (CA). While intended as a WLAN protocol, the DCF also allows direct node-to-node communication in ad hoc mode.

The basic CSMA/CA mechanism of 802.11 consists of a contention period, where the medium is idle, followed by the transmission of data, then a positive acknowledgement. A short interframe space (SIFS) is also included as the minimum separator between two consecutive packet transmissions.

The contention period is designed to give every node an opportunity to access the medium. In order to attempt access, each node must first wait until it senses an idle period of at least the distributed interframe space (DIFS). However, if nodes attempted to transmit immediately following every idle DIFS, collisions would be virtually assured, with the resulting transmissions wasted. Therefore, a number of mechanisms are included for the purpose of avoiding and reducing the cost of collisions.

Any time a node wishes to send, it senses the medium. If the medium is idle, and continues to be idle for the period of a DIFS, it can begin its packet transmission. The point coordination function (PCF) uses a time-slotted access mechanism suitable for centralized systems, while the distributed coordination function (DCF) is a CSMA-based (carrier sense multiple access) method with collision avoidance (CA). While intended as a WLAN protocol, the DCF also allows direct node-to-node communication in ad hoc mode.

The contention period is designed to give every node an opportunity to access the medium. In order to attempt access, each node must first wait until it senses an idle period of at least the distributed interframe space (DIFS). However, if nodes attempted to transmit immediately following every idle DIFS, collisions would be virtually assured, with the resulting transmissions wasted. Therefore, a number of mechanisms are included for the purpose of avoiding and reducing the cost of collisions.

Any time a node wishes to send, it senses the medium. If the medium is idle, and continues to be idle for the period of a DIFS, it can begin its packet transmission. The point coordination function (PCF) uses a time-slotted access mechanism suitable for centralized systems, while the distributed coordination function (DCF) is a CSMA-based (carrier sense multiple access) method with collision avoidance (CA). While intended as a WLAN protocol, the DCF also allows direct node-to-node communication in ad hoc mode.
not receive a response to its transmitted data, the contention window size is doubled and chosen backoff times also increase. In congestion, the contention window quickly reaches a maximum. If node demands exceed this maximum, frequent collisions will occur.

The basic packet exchange follows this contention period (Figure 2(a)). Once a node is allowed to transmit by the contention procedure, in the basic exchange it transmits its data packet (DATA). If it is a broadcast packet, there is no way of knowing if the packet is received by anybody. For all other packets, the receiver returns a positive acknowledgement (ACK), if the data packet is received error-free. The packet exchange is completed when the initial sender receives the ACK packet, thereby knowing that the packet has been successfully received.

Unfortunately, the chance of a collision remains. As collisions are undetectable to the sender, they continue to transmit their data packets. Only when the ACK packet is not received does the sender become aware that an error has occurred. In fact, the error could be due to a number of reasons: a collision occurred, the packet was received with errors, something happened to the ACK, or even the destination was no longer within range. However, all the sender can do is attempt a retransmission of the packet. This results in the wasted expenditure of power and bandwidth.

In order to reduce the expense of collisions, 802.11 optionally utilizes a request-to-send (RTS) and clear-to-send (CTS) packet sequence. RTS and CTS packets are very short MAC protocol packets that can be exchanged prior to the actual DATA-ACK exchange. As the RTS-CTS time is short, collisions can be resolved far more quickly than if the entire DATA packet were transmitted. Although the RTS-CTS mechanism results in some additional overhead, the savings in the case of a collision is sufficient to justify use of RTS-CTS, except for very small data packets. A packet exchange using RTS and CTS packets is depicted in Figure 2(b).

The RTS-CTS mechanism on its own does not actually reduce the probability of collisions [40]. Instead, 802.11 utilizes them to implement additional passive collision avoidance, in addition to active carrier-sense and backoff measures. Carrier-sense alone is insufficient for preventing many collisions, because of the hidden terminal problem. Due to the limitations in sensing and transmitting ranges, situations frequently exist where two nodes attempt to reach a common receiver. While each can reach the receiver, they are out of range of the other sender.

In order to alleviate this problem, 802.11 implements virtual carrier sensing, using the network allocation vector (NAV) in conjunction with the RTS-CTS mechanism. Both the RTS and CTS packet include a duration field, indicating how long the rest of the packet exchange should take. Any node that receives either packet takes this information and sets their NAV timer to the indicated duration. This timer prevents other nodes from attempting a transmission until after the timer expires (and the packet exchange is complete), regardless of whether or not the medium is sensed as idle. This prevents a node that is hidden from the sender from interfering at the receiver, as it should have received the CTS packet from the receiver.

3.2. Improvements to 802.11

While 802.11 has been frequently used for testing ad hoc networking, a large number of works have revealed significant performance issues and problems [41–45]. As discussed earlier, some research has focused on finding different technologies altogether. However, 802.11’s popularity has lead to numerous attempts to improve the protocol’s ad hoc capabilities.

Part of the problem with 802.11 is its operation in congested conditions. Congestion causes frequent collisions and long backoff times. Many evaluations of the 802.11 technology have shown that network throughput is quite poor when faced with many contending nodes and heavy traffic loads, even in WLANs. In an ad hoc network, potentially higher network densities and the required multihop retransmissions of each packet aggravate these problems. Congested conditions are the norm, rather than the problem case [46].

The 802.11 MAC lacks any notion of guaranteed quality-of-service and congested conditions mean
over-provisioning is not an option [47]. Several methods have been proposed for creating a notion of priority and achieving the differentiation of service classes [48]. These include adjusting the size of the contention window, changing the interframe spacing, and altering the maximum MAC frame lengths.

802.11 has also been shown to favor particular nodes, because of the way the contention window is reset following successful exchanges. Shrinking the contention window leads to shorter backoff times. This allows a node that has just completed a transmission to quickly be allowed to attempt to send another packet. Although all nodes eventually have the opportunity to access the medium, nodes are sometimes forced to wait a long time before they can attempt to send a packet. If a collision occurs, they will likely be forced to backoff again. Meanwhile, successful nodes may have the opportunity to send multiple times.

While this leads to a fairness issue in conventional networks, it can cause even larger problems in ad hoc networks. When nodes are not able to send packets for a length of time, they run the risk of affecting the network’s links and routes. If a neighbor is not detected regularly, then it is removed from the neighbor list. Additionally, repeated collisions or non-responses to RTS packets can also result in neighbor removal.

Part of the problem in achieving efficient use of the wireless medium is its variable nature. Significant effort is used to establish control of the channel, however only one packet is sent. Later, when another attempt is made, the receiver may be unavailable or prevented from replying by another transmission. In Reference [49], after a node acquires the medium for transmitting to the receiver, it sends its packet and waits for an ACK. However, because the link has already been established as usable, the source can attempt to transmit subsequent packets (Figure 3). To prevent the node from holding the channel indefinitely, a limit is placed on the length of the burst of packets that can be sent.

The 802.11 protocol includes a considerable amount of wasted bandwidth within the packet exchange. A MAC header is added to every DATA packet. RTS and CTS packets (if used), and the ACK also consume bandwidth. Additionally, the packet exchange also includes interframe spaces.

Within an ad hoc network, this overhead is multiplied. Network throughput only counts the data packet once when it is finally delivered successfully. However, several packet exchanges may be required to deliver the packet to its destination. For example, in a four-hop path, four RTS, four CTS, and four ACK packets must be used. In addition, four transmissions of the DATA packet, and its MAC header only contribute to the throughput of one DATA packet.

4. Requirements for Creating an Ad hoc MAC

On its own, the 802.11 MAC suffers from a number of deficiencies when used in an ad hoc environment. What it does provide is a well-known and widely available technology that can be used for studying ad hoc networks. Higher level protocols, particularly routing, can be investigated, although researchers must consider how 802.11 will affect their operation. Additionally, it can serve as a base protocol for exploring mechanisms at the MAC layer.

In developing a MAC protocol for ad hoc networks, the characteristics of the ad hoc network must be considered. From these characteristics, specific requirements arise.

4.1. Minimize Overhead

In any network, protocol overhead should be kept as small as possible. However, small amounts of additional overhead can sometimes yield greatly improved overall performance. This is particularly true in high-bandwidth networks, where additional overhead can be added without consuming a significant amount of the total available. In an ad hoc network, bandwidth is scarce; a slight increase in overhead consumes a proportionately higher percentage of the available resources.

The 802.11 protocol includes a considerable amount of wasted bandwidth within the packet exchange. A MAC header is added to every DATA packet. RTS and CTS packets (if used), and the ACK also consume bandwidth. Additionally, the packet exchange also includes interframe spaces.

4.2. Reduce Contention and Collisions

802.11 does not perform well when many nodes are simultaneously contending for the channel, as frequent
collisions and long backoff times result in the wastage of the channel for long periods. Unfortunately, this is exactly the scenario that exists in an ad hoc network. Within the protocol, the probability of collisions can be reduced by increasing the backoff window, however this results in lower efficiency. Conversely, shorter backoff times result in a higher probability of collision; in this case, the channel is wasted due to the transmissions being lost, rather than the medium being idle. Therefore, an ad hoc efficient MAC must attempt to reduce the frequency and the length of contention. This must occur without increasing the occurrence of collisions, although ideally it should reduce these as well. Even more importantly, it must still permit or encourage different nodes to transmit regularly. Without spreading the access around, the negative interactions with the routing layer will override any efficiency gains.

4.3. Fairness
MAC-level fairness in an ad hoc network is more than just dividing the available bandwidth equally. The responsibilities and rewards for nodes in an ad hoc network are fundamentally different from a WLAN. Terminals using a WLAN contend for the medium in order to send their own traffic directly to the base station. Ad hoc nodes must transmit forwarded packets, in addition to their own traffic. Whenever a node accesses the medium to forward a packet for another node, it gains no direct benefit from it. In fact, accepting a packet to forward will delay the node’s sending and receiving of its own traffic. However, it is a responsibility that a node must accept, in order for the network to function. Although this responsibility has been considered at higher levels, it has not been considered at the MAC level.

802.11 operates on a premise of providing all nodes with equal opportunity to access to the medium. However, this is not necessarily fair in an ad hoc network. As nodes are required to fulfill their forwarding responsibilities, they must transmit packets that are not their own. An equal opportunity scheme penalizes a node that is carrying a large volume of forwarded traffic.

An ad hoc MAC should instead consider the unique responsibility of forwarding places on a node. It should facilitate the forwarding process, making it as inexpensive as possible for a node to perform its duties. By reducing the cost to nodes, or rewarding them for their participation, the MAC can encourage nodes to forward packets.

4.4. Support Routing
An ad hoc-specific MAC could address one final issue of conventional WLAN MAC protocols. The dynamic topology and routing is the fundamental challenge in ad hoc networks. All other protocols should work in support of the routing protocol.

However, issues in the 802.11 MAC can often cause link and routing failures. With mobility, the ad hoc topology is already extremely fragile. Every effort must be made to avoid any additional link instability. Therefore, steps must be taken to eliminate any MAC-induced failures.

If the MAC impacts the routing process, then perhaps the MAC should consider the routing protocol in its operation. By making the MAC aware of the forwarding process, the MAC could contribute to the maintenance of established routes. If packets are regularly passed along the entire length of the route, the route will not be allowed to expire. This requires a mechanism to make sure that each individual link is used, so that nodes remain aware of their neighbors and links do not timeout. By ensuring that repeated attempts to contact a neighboring node are not blocked, links can be maintained for their full usable lifetime.

5. Immediate Forwarding
The immediate forwarding (IF) concept directly addresses forwarding in an ad hoc network. As multihop traffic is fundamental to an ad hoc network, the MAC works to streamline the delivery of these packets. As packets arrive at an intermediate node, they are immediately identified and forwarded as soon as possible.

A MAC-level mechanism is used in conjunction with the policy of immediately forwarding packets. This mechanism attempts to re-use the channel without forcing re-contention. The receiver is granted an opportunity to initiate a new packet exchange without having to wait or compete with other nodes. The forwarding nature of the ad hoc network is utilized in order to spread the medium usage around. Additionally, because the mechanism is forwarding based, it is hoped that the mechanisms will better maintain the overall path.

5.1. Mechanism
As previously suggested, the receiving node can immediately utilize the medium after it transmits the acknowledgement to the initial sender (S1). In order to prevent two nodes from dominating the medium, the receiver (R1) may only use this mechanism if R1
is not the ultimate destination for the packet it has just received (i.e., R1 is serving as an intermediate node for the packet). As R1 does not gain any benefit from receiving the initial packet, it will be required to forward the packet, R1 can be allowed to transmit a packet. We describe this as FF; as reuse of the medium is guided by, and aims to promote, the forwarding responsibilities of the nodes.

After sending the ACK, the R1 must wait for the expiry of a SIFS, in order to separate the packets, then may re-use the medium. No other node within range can attempt to use the medium, as normally the first slot following the reception of an ACK packet does not begin until the completion of the period of one DIFS. Therefore, R1 is guaranteed to still control the medium, allowing the immediate reuse of the medium, free from a contention period. If the R1 is the destination of the received packet, or if the packet is a broadcast packet (in which case there is no ACK sent), then the medium is released to contention in the normal manner.

Even if another node’s backoff timer indicates it can transmit immediately, it must wait for an idle DIFS, in addition to the propagation time required for the ACK packet to arrive from R1. If this mechanism is successful, R1 can transmit after only a SIFS period. The medium is idle for only the SIFS period, rather than at least a DIFS, even without considering additional backoff time, and the possibility of a collision. In both cases the result is two packets being sent by two different senders. However, different nodes are involved in the second packet exchange, as illustrated in Figure 4(a) and 4(b).

5.2. IF Procedure

In the IF procedure, R1 attempts to immediately forward the packet received from S1. The protocol is designed to facilitate the forward action, making it as easy as possible for the packet to be forwarded on to its ultimate destination. It is hoped that, by attempting to get the data packet to its destination as quickly as possible, the MAC will better support the routing and transport protocols. The protocol is designed to deliver the packets as efficiently as possible, improving overall throughput, and minimizing the delay experienced by each packet.

Following the reception of the data packet, the node R1 determines the destination of the packet by checking the IP address contained in the packet header. If the packet is destined for the node, the packet is stripped of its MAC header and is sent up the protocol stack, as it would be ordinarily. Control and broadcast packets are also not eligible for immediate forwarding, and are therefore handled normally (control packets were omitted to simplify the MAC-routing interaction). If the packet’s IP destination does not match the node’s address, the packet is held within the MAC protocol. The packet is then acknowledged in the typical manner. While the packet is being acknowledged, the node must determine the next hop destination for the packet. Using an interface with the routing protocol, the MAC consults the routing tables to find the appropriate route entry. If an active route can be found, an attempt can be made to forward the packet. Otherwise, if no entry is found or the route is invalid, the packet is sent upwards. The medium is then released as in 802.11.

If the packet is selected for forwarding, the normal operations must occur on it. First, all routing information in the route tables must be updated, ensuring that the routing protocol takes any information it needs from the packet. This is done in order to ensure that routes do not become prematurely stale. Second, the packet itself must be updated. IP header information is left intact, with the exception of time-to-live information, which must be decremented as usual. A new MAC header must also be generated, preparing the MAC information for transmission in the typical manner. An RTS packet must also be prepared, to initiate the following packet exchange.

Following the completion of the ACK transmission, R1 must delay for a period of one SIFS. This is the mandatory separator between consecutive packets.
Remember that all other nodes that either receive the ACK or sense the medium as busy, are forced to wait at least until the completion of a DIFS before they can attempt to access the medium. Therefore, the medium should still idle following the SIFS.

The node can then commence the transmission of the RTS packet. If successful, the RTS’s target (R2) would respond with a CTS in the typical manner. The packet exchange would then continue as in 802.11. If the node does not receive a CTS, the immediate forwarding attempt is abandoned and the medium will return to a contention state, as it would following any other failed RTS. The RTS is discarded, the data packet is stripped of the MAC header and the packet is sent up to the link layer. No retransmits are attempted, to avoid interfering with other packets waiting in the queue for too long.

The data packet is maintained within the MAC at R1 until an ACK is received. At this point, the packet can be discarded, as it has been successfully forwarded. If the data packet is transmitted and no ACK is returned, the forwarding process is deemed to have failed, and again the packet is sent upwards. The packet will then be handled as determined by those protocols.

The design of this protocol allows for the chaining of immediate forwarding processes at adjacent nodes. At each intermediate node, the packet is immediately sent on to the next intermediate and so on, until the packet reaches the destination. Clearly, if this is accomplished for the entire path, considerable savings can be achieved in terms of cost of delivering the packet.

Figure 5 displays the packet sequence for a four-node (three-hop) chain of immediate forwards.

6. IF Performance Evaluation

6.1. Simulation Details

A simulation was constructed in ns-2 [51], in order to evaluate the effectiveness of the proposed IF mechanism. The IF mechanism was compared to the standard implementation of the 802.11 MAC included in the simulator package.

AODV was used as the ad hoc routing protocol. Due to the interactions between the mechanism and the routing protocol, it was necessary to allow the MAC protocol to access some information within the routing protocol. For this reason, the mechanisms were implemented within the simulator to operate specifically with AODV.

Two main performance measures are used throughout the evaluation. Throughput is measured as the amount of traffic successfully delivered to its destination per time unit. Typically, an average is used, dividing the total amount of data delivered by the total simulation time. End-to-end delay is also measured, for all successfully delivered packets. This is the amount of time elapsed between the source sending a packet and the destination receiving it. Again, an average is typically used.

6.2. Linear Topology

In order to illustrate the effect of the IF mechanism on a multihop flow, a set of simple, linear topologies were used. These are the same type of topologies used in Reference [44]. For the purpose of these evaluations, the lines of nodes ranged from 2 nodes (a direct link) to 10 nodes (requiring 9 hops). Examples of these topologies are illustrated in Figure 6. The nodes have been numbered from 1 to n, with node 1 at one end of the chain, and node n at the other.

All of the nodes were spaced 200 m apart. Due to the nominal transmission range of 250 m, this created the situation where each node could directly reach both the node immediately preceding it on the chain, as well as the node immediately following it. No nodes could reach more than these two other nodes (with the end nodes only reachable by their one adjacent node). Therefore, all traffic must travel hop-by-hop down the line.

Fig. 5. Three-hop chain of immediate forwards.

Fig. 6. Three, seven, and ten node linear topologies.
For each $n$-node topology, a constant bit rate (CBR) traffic source agent was created at node 1. This traffic source generated 512 byte packets (4 kb), a packet size which has been shown to be an effective choice for ad hoc networks [52]. The interval between packets was adjusted, in order to vary the offered load in the network.

A sink was created at node $n$. The source and sink were connected using the user datagram protocol (UDP), so that traffic would be delivered from node 1 to node $n$. In all cases, this flow was created at $t = 5$ s, and allowed to continue until $t = 105$ s (for a flow length of 100 s). The simulation was ended at 110 s, in order to allow the nodes some time to finish delivering packets after the termination of the traffic source. A total of five runs were made for each combination of nodes and offered load.

Figure 7(a) and (b) shows the average throughput achieved in the linear network as a function of path length. Note that both offered loads could be supported over a direct link (path length of two nodes). However, as path length increases, the average throughput begins to decrease. For a low offered load (328 kbps), this decrease is minimal for the IF protocol, but quite drastic for 802.11. At the higher load (1148 kbps), IF performance falls along with 802.11, only slightly outperforming 802.11 in short paths. However, in the longer paths, IF throughput plateaus, achieving almost identical throughput (310 kbps) as in the 328 kbps load case.

In Figure 8 throughput is instead plotted against offered load. In the three-node path, IF outperforms 802.11, but only marginally. This is due to the fact there is only minimal contention between the three nodes. 802.11 operates fairly efficiently, as the sending node and the intermediate node take turns using the medium.

IF gains a slight improvement, due to the reduction of the DIFS period to a SIFS period at the intermediate node. However, in the longer path, a much larger difference is visible.

Figure 9(a) and (b) suggests a possible explanation: packet delivery in the 802.11-based network is extremely erratic, while IF sustains a steady flow. Within the 802.11 throughput trace of a single test case, a number of periods exist where throughput goes to zero, such as at $t = 32$ s. Frequently, these indicate route failures, during which time packets are delayed or dropped. By avoiding this, IF achieves considerably better results.

Figure 10(a) and (b) presents the delay characteristics of 328 kbps and 1148 kbps flows. One thousand one hundred forty-eight kilobits per second was chosen as a heavy load, but within the channel bandwidth; it is sustainable only over a direct link. Three hundred twenty-eight kilobits per second represents a lower load, at just above the plateau level observed at longer paths using IF.

In the low load scenario, short paths experience very low delay. This is because the network is successfully
carrying the full flow. When the load exceeds the network’s capabilities, the delay increases sharply. This is primarily due to the effects of queuing. As the load exceeds capacity, queues fill up. As packets arrive at the queue faster than they are removed, the queue remains full, and most of the packets experience roughly the same delay. Therefore, for IF, when the load exceeds a certain threshold, delay increases only marginally for each additional hop—the transmission time at each hop. Most of the delay is caused in the initial queue. In 802.11 however, contention occurs at each intermediate node and additional queuing delay is added.

6.3. Static Topology

The linear network simulations demonstrated the benefits of the basic concept, however it is essentially an ideal scenario for the IF and EF mechanisms. The stream experiences no outside interference— all nodes are focused solely on delivering packets from node 1 to node n. Therefore, random topologies were created in order to examine the performance in a more realistic topology. In this scenario, multiple traffic streams were generated, creating interference between the streams, and increased contention between nodes.

A square two-dimensional network area was used, with side-lengths of 1000 m. Fifty nodes were randomly distributed within this area, using a uniform distribution. They remained in those positions for the duration of the simulation. This would create a static network with typical average path lengths between three and four hops, with the longest paths reaching up to about 10 hops. In order to avoid scenarios where nodes were unreachable, any disconnected network topologies were discarded.

Ten traffic streams were created. The streams were again CBR traffic, transported by UDP in 512-byte packets. Again, the packet interval was varied to control the offered load. In order to ensure that results were not impacted by a source or destination becoming
overloaded by multiple flows, in this scenario the sources were selected as nodes 1–10, and the destinations were nodes 11–20. As the nodes’ locations were chosen randomly, this numbering is in fact arbitrary.

The simulation time was reduced to 70 s in this scenario. In order to avoid overwhelming the network by a large burst of routing packets, the 10 flows were started at 0.1 s intervals, from $t = 4.1$ s to $t = 5.0$ s. All flows were exactly 60 s in duration, again leaving 5 s after the end of the traffic generation before the end of the simulation. For each level of offered load, 10 independent test cases were simulated using different generated topologies, and the results were averaged to provide the final results.

Figure 11 illustrates the throughput in a static random network. The performance of IF is obviously quite disappointing, trailing well behind the original 802.11 protocol.

The failure of IF is fairly easily explained. Contention and interference play a significant role in overall network performance. IF immediately forwards data packets, however this comes at the expense of other packets in the intermediate nodes’ queues. Often, these packets are control packets (routing messages) that are critical to the other flows.

Ordinarily, the priority queue feeding the MAC layer prioritizes control packets over other types. Therefore, a node’s routing packets will reach the MAC before data packets. However, in IF, the forwarded packets skip the interface queue, putting control packets at a disadvantage.

7. Forward Focus

As a result of IF’s failure, a second protocol has been designed. Based on the same forwarding focused mechanism, this protocol encourages nodes to receive packets for forwarding, and rewards them by granting them immediate access to the channel. However, rather than re-transmitting the packet it just received, the intermediate node is allowed to attempt whatever packet the IFQ has provided to the MAC for transmission. This preserves the ordering of packets as in 802.11, while encouraging the node to participate in the forwarding process.

7.1. FF Procedure

On reception of a data packet, the node again determines the destination from the IP header. However, in this protocol the packet is stripped of its MAC header and sent upward, regardless of the result. Therefore, with the exception of checking the destination, this protocol handles the incoming data packet entirely in the normal manner.

If the IP address of the incoming packet does not match the node’s address, and the packet is not a broadcast packet, the node is encouraged to begin a packet transmission. Following the ACK packet, the node again delays for a SIFS. After this period elapses, the transmission can be attempted. In this protocol it does not matter what kind of packet is to be sent. Whatever type of packet is present at the MAC can be attempted, whether it is a short data packet (no RTS), a data packet requiring an RTS-CTS exchange, or even a broadcast packet. In any case, the packet is transmitted in the normal manner, as if the medium had been idle.

An FF packet sequence with RTS is identical to the IF sequence shown in Figure 4(b), except a different data packet is transmitted in the second transmission. Figure 12 shows the sequence without the RTS-CTS.

![Fig. 12: Forward focus sequence without RTS-CTS](https://example.com/ff_sequence_no_rts_cts.png)
8. FF Performance Evaluation

A simulation of FF was created in the same manner as IF. In running the simulations, FF performed almost identically to IF in the linear topology. Therefore, the IF curves in Figures 9(a), 9(b), and 8 almost exactly (within 0.1%) match those of FF. This is not surprising, as with a single flow, the only packet in an intermediate node’s queue is the packet to be forwarded. Therefore, the two protocols are almost equivalent.

8.1. Static Topology

This changes when the static topology is considered. As seen in Figure 13(a), FF does not suffer the same fate as IF. Instead, FF continues to outperform the standard 802.11 MAC protocol in terms of overall throughput. FF avoids the problems of IF by not changing the order of packet transmissions. In fact, waiting control packets benefit by gaining access to the medium upon receiving a packet to be forwarded.

Interestingly, FF does not see the same delay advantage in the static topology scenario as in the linear topology. In fact, as seen in Figure 13(b), FF’s average delay actually exceeds 802.11’s as loads increase. The delay is quite severe, quickly rising to over 2 s, at less than a megabit per second total offered load.

The explanation for this is not immediately evident. Increased delay should typically result in longer interface queue lengths and more dropped packets. The answer appears to involve the way the nodes handle route failures: each time a route fails, intermediate nodes empty their queue of packets using that route. The longer a packet remains undelivered, the greater the chances are that the route will break and the packets will be dropped. In 802.11, the packets experiencing lengthy delay tend to be dropped, resulting in a shorter average delay. However, in the EF protocol, a few of these path failures are avoided, due to the better coordination of the medium. This results in more packets being delivered, including some of the ones that have suffered considerably greater delay.

8.2. Mobile Networks

Finally, a fully dynamic network was simulated. This scenario was intended to include all of the challenges of the ad hoc environment. Both mobility and changing traffic flows were included, creating a very dynamic network topology and set of routes. In this situation, the mobility of the nodes and the multiple traffic flows, were randomly generated.

Starting positions were chosen similarly to the static scenario. A random waypoint mobility model was then used to generate the mobility patterns. Destinations and speed were selected with a uniform distribution. The maximum speed was 10 m per second. Wait time was set to 0 s, meaning that the nodes were always in motion.

Traffic flows again used CBR traffic sources, however they were generated randomly. For each node, a pause time was randomly generated. At the completion of that pause time, the node chose a random destination node (not itself), and started a flow with a randomly chosen duration. The node then repeated the process for the duration of the simulation. An average pause time of 40 s, and an average flow length of 20 s were chosen, so that on average, only about one-third of the nodes should be transmitting at a time.

Packet rate was consistent for all flows in each network case, but was varied in order to adjust offered load. The tested rates were considerably lower than the ones used in the static case. This allowed for an increased number of flows and the
heavier impact of routing overhead on the network as a whole.

For this scenario, simulations of 100 s were performed. A total of 18 runs were made for each offered load level, using all combinations of three traffic patterns, three mobility patterns, and two random seeds. Represented data points in the figures are averaged results from all of the runs.

Throughput for the mobile network scenario is shown in Figure 14(a). The FF mechanism shows almost identical performance as 802.11. Only through the mid-load region does FF exhibit a small increase in throughput; above and below this region, 802.11 and FF perform almost identically. This increase may be due to EF being slightly more efficient than 802.11. However, the difference is insufficient to claim a significant improvement in throughput when using the FF mechanism.

Interestingly, FF does show a considerable reduction in the delay experienced by the packets. Figure 14(b) shows that FF packets are delayed less than 802.11 packets at medium and high load levels. There is almost a 14% reduction in delay at the throughput maximum. FF gains the most benefit at just under the 300 kbps offered load. In this range, FF's throughput appears to slightly exceed that of 802.11, however the average packet delay is decreased by over 25%. Note that FF starts to gain its delay advantage just before this point, just as PDR starts to fall.

The scenarios that were used are fairly difficult for an ad hoc network. The mobility traces averaged approximately 700 link changes in the 100 s simulation, due solely to mobility. With approximately half of these changes link breakages, three links break per second (on average), somewhere in the network. Clearly, this would cause a considerable number of route failures within the network.

In each scenario between 80 and 90 flows were created in each traffic pattern. As the probability of identical routes is fairly low, this means that the network was performing on average almost a route discovery per second. This also created a network where there were, on average, approximately 16 flows active at one time. However, this number varied over the simulation time, and frequently the network contained many more than that.

This suggests why there is little difference between the packet delivery rates given by 802.11 and by FF. In the static topology simulations, route discovery was limited to initial period, as well as any route failures caused by interference effects. Therefore, through the rest of the simulation, the network was primarily focused on delivering data packets. However, in the mobile scenario routing occurred throughout, due to both the creation of new flows, and the failure of existing routes. This caused two things: first, the network was dominated by routing operations and forwarded packets make up a small percentage of the total packet exchanges; and second, severely delayed packets were dropped at intermediate nodes, when the route eventually broke. However, FF did gain an advantage by improving efficiency as the network approaches its capacity.

9. Conclusions

This paper has focused on directly considering the unique properties of an ad hoc network in the development of a MAC protocol. As a result, the MAC both
supports the forwarding process and uses the forwarding process to improve the own efficiency. A simple mechanism based on the IEEE 802.11 protocol has been used to show the validity of this approach.

The initial concept was to streamline the forwarding process by immediately forwarding packets at intermediate nodes. However, although this approach was shown to work in principle, it failed under more realistic scenarios. This failure emphasizes the importance of control traffic in an ad hoc network. In order to react effectively to changes in the network topology, all control traffic must be handled as quickly and efficiently as possible. Any delays in the propagation of these packets can lead to significant deterioration in the performance of the network.

The initial promise and subsequent demise of IF led to a slight modification of the initial concept. Rather than immediately re-transmitting a multihop packet, the multihop nature was instead used to streamline the problem of medium access control. The resulting mechanism still focuses on forwarding packets, encouraging the process, but without forcing the immediate retransmission of the data packet. FF was shown to have the same benefits as IF, but also delivered benefits under more realistic topologies.

Several conclusions can be drawn from the success of FF. First, the simulation results indicate that this mechanism can provide real benefits to an ad hoc network, with minimal change to the 802.11 protocol. Second, it shows that providing simple mechanisms in the MAC protocol to directly support the traffic patterns found in an ad hoc network can result in better service to the network. Third, by focusing on forwarding, the MAC and routing protocols can be made to work together, rather than interfering with each other. Improved interaction has yielded performance improvements; increased integration may provide further benefits.

Most importantly, development of protocols for ad hoc networks should focus on utilizing their unique features to their advantage. The ad hoc environment is unquestionably difficult, and properties such as the wireless medium and multihop forwarding create some severe challenges. However at the same time, they also yield some potential benefits. In developing protocols, increased attention should be paid to reaping those benefits, rather than merely minimizing the difficulties.

References

17. Part II: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: higher-speed physical layer extension in the 2.4 GHz band. IEEE Standard 802.11b, September 1999.


33. Hoar JD, Deng J. Dual busy tone multiple access (DTBMA)—a multiple access control scheme for ad hoc networks. IEEE Transactions on Communications 2002; 50(6): 975-985. DOI: 10.1109/TCOMM.2002.1010617


Authors’ Biographies

Brent Ishibashi received his B.Sc. degree from the University of Guelph (Canada) and his M.Math degree from the School of Computer Science of the University of Waterloo (Canada). He is now working towards a Ph.D. degree, also at the University of Waterloo. His research focuses on resource management in multihop wireless networks, including ad hoc networks and infrastructure-based wireless meshes.

Dr. Raouf Boutaba is an associate professor in the School of Computer Science of the University of Waterloo. Previously, he was with the Department of Electrical and Computer Engineering of the University of Toronto. Prior to academia, he founded and directed the telecommunications and distributed systems division of the Computer Research Institute of Montreal. Dr Boutaba also serves on numerous journal editorial boards and conferences committees, including IFIP Networking, NOMS, ICC, and Globecom.