UCAN: A Unified Cellular and Ad-Hoc Network Architecture

Haiyun Luo*, Ramachandran Ramjee†, Prasun Sinha‡, Li (Erran) Li†, Songwu Lu*

*Computer Science Dept.
University of California
Los Angeles, CA 90095
{hluo,slu}@cs.ucla.edu

†Bell Labs
Lucent Technologies
Holmdel, NJ 07733
{ramjee,erranlli}@bell-labs.com

‡Computer and Information Science
Ohio State University
Columbus, OH 43210
prasun@cis.ohio-state.edu

ABSTRACT

In third-generation (3G) wireless data networks, mobile users experiencing poor channel quality usually have low data-rate connections with the base-station. Providing service to low data-rate users is required for maintaining fairness, but at the cost of reducing the cell's aggregate throughput. In this paper, we propose the Unified Cellular and Ad-Hoc Network (UCAN) architecture for enhancing cell throughput, while maintaining fairness. In UCAN, a mobile client has both 3G cellular link and IEEE 802.11-based peer-to-peer links. The 3G base station forwards packets for destination clients with poor channel quality to proxy clients with better channel quality. The proxy clients then use an ad-hoc network composed of other mobile clients and IEEE 802.11 wireless links to forward the packets to the appropriate destinations, thereby improving cell throughput. We refine the 3G base station scheduling algorithm so that the throughput gains of active clients are distributed proportional to their average channel rate, thereby maintaining fairness. With the UCAN architecture in place, we propose novel greedy and on-demand protocols for proxy discovery and ad-hoc routing that explicitly leverage the existence of the 3G infrastructure to reduce complexity and improve reliability. We further propose a secure crediting mechanism to motivate users to participate in relaying packets for others. Through extensive simulations with HDR and IEEE 802.11b, we show that the UCAN architecture can improve individual user's throughput by up to 310% and the aggregate throughput of the HDR downlink by up to 60%.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design; C.2.2 [**Computer-Communication Networks**]: Network Protocols

General Terms

Design, Performance

Keywords

3G Wireless Networks, Mobile Ad-Hoc Networks, Unified Architecture

1. INTRODUCTION

In recent years, wireless data networks have made tremendous progress with both the worldwide upgrade of cellular networks to support wide-area data access and the widespread deployment of IEEE 802.11-based local area networks. However, there are several important differences between the current wide-area and local-area wireless networks.

First, while wide-area wireless networks provide large cell coverage (up to 20 Km), the cell coverage in local-area wireless networks is limited (up to 250m for IEEE 802.11). Second, while wide-area wireless networks offer relatively low throughput (38.6 Kbps to 2.4 Mbps in the latest commercial deployment of 1xEV-DO), local-area wireless networks offer relatively high throughput (1-11Mbps for IEEE 802.11b, and up to 54Mbps for IEEE 802.11a and 802.11g). Third, while wide-area wireless networks operate in infrastructure mode with fixed base stations serving mobile users, local-area wireless networks can operate in ad-hoc mode where mobile clients relay packets for each other over multi-hop wireless links.

Although there has been extensive research to date on improving the performance of each of these two technologies in isolation, one open question that remains is whether they can be synergistically combined to leverage the advantages of each other. Our goal in this paper is to devise a new wireless networking paradigm that increases the throughput of wide-area wireless networks through opportunistic use of ad-hoc local-area wireless networks. We call such a model *UCAN*: the unified cellular and ad-hoc network.

One pre-requisite for the UCAN model is that each mobile device is equipped with two wireless interfaces. Fortunately, given the popularity of the IEEE 802.11b (Wi-Fi) interface, it is already being embedded in every mobile device and thus the device only needs a 3G interface card to operate in UCAN. The convergence of mobile phones and computers, such as walkie-talkie PC, also foresees the popularity of such wireless devices [14]. More recently, several companies such as GTRAN wireless [2] are offering integrated cards that implement both IEEE 802.11b and 3G wireless interfaces. Thus, if routing protocols can be made aware of both interfaces, they can improve performance significantly by selecting the best interface(s) to deliver packets to the mobile users.

The UCAN approach also helps us address one of the tough questions when the ad-hoc network model is applied to commercial use, i.e., why should a mobile user relay traffic for other users? In UCAN, a mobile user has strong incentives to relay traffic for other users, because, as we shall see in Section 8, relaying traffic for other users will also benefit the user in increasing his own throughput.

We believe that the vast majority of commercial network applications, which need high availability assurance, can only be supported through the managed infrastructure of wide-area wireless networks. In UCAN, we use the ad-hoc wireless connection exclusively to enhance the performance of a mobile user's access to the cellular infrastructure; in the absence of sufficient connectivity in the ad-hoc network, mobile users continue to access data through their wide-area network interface, albeit at a lower throughput.

We make two main contributions in this work. First, we propose a novel architecture that unifies cellular and ad-hoc networks opportunistically. Second, we devise a suite of protocols that enables the network architecture, including new proxy discovery and ad-hoc routing protocols (that leverage the managed infrastructure to decrease their complexity and overhead and increase their reliability), refined scheduling at the 3G base station (that balances throughput gain among users), and secure crediting (that provides strong motivation for autonomous users to serve as relays). Through extensive simulations with 1xEV-DO (HDR) and IEEE 802.11b (Wi-Fi), we show that these protocols can substantially benefit cellular networks in dense urban areas, improving individual user's throughput by up to 310% and the aggregate average HDR downlink throughput by up to 60%. In this paper we evaluate the performance of UCAN based on HDR and IEEE 802.11b. However, other 3G (e.g., 1xEV-DV) and IEEE 802.11 (e.g., IEEE 802.11a and IEEE 802.11g) technologies are also applicable in the UCAN architecture.

The rest of the paper is organized as follows. In Section 2, we review HDR and IEEE 802.11b technologies and survey the related work. In Section 3, we present data from measurements on a commercial 3G network and show throughput gains obtained using a single-hop IEEE 802.11b relay link to motivate the design of the UCAN architecture. In Section 4, we present the UCAN architecture. In Section 5, we describe two novel proxy discovery and ad-hoc routing protocols that exploit the 3G infrastructure to improve efficiency. In Section 6, we discuss enhancements to the scheduling algorithm at the 3G base station to improve throughput, while maintaining fairness. In Section 7, we present our secure crediting mechanism for motivating mobile users to serve as relays. In Section 8, we present extensive simulation results of the proxy discovery and routing protocols on a single cell that uses HDR and IEEE 802.11b wireless interfaces. We discuss related issues in Section 9. Section 10 concludes this paper.

2. BACKGROUND AND RELATED WORK

In this section we briefly review 3G HDR wide-area data and IEEE 802.11b networks, and discuss the related work.

2.1 HDR and IEEE 802.11b Networks

1xEV-DO (Evolution-Data Only), also known as HDR (High Data Rate), is an integral part of the CDMA2000 family of 3G standards. Designed for bursty packet data applications, it provides a peak data rate of 2.4Mbps and an average data rate of 600Kbps within one 1.25MHz CDMA carrier. HDR is commercially available in South Korea, Brazil, and upper-midwest US. HDR downlink has much higher data rate (2.4Mbps), compared with its uplink data rate of 153.6Kbps [3]. Users share the HDR downlink using time multiplexing with time slots of 1.67ms each. At any time instant, data frames are transmitted to one specific client, and the data rate is determined by the client's channel condition. The duration of transmission to each client is determined by the downlink scheduling algorithm. HDR uses a scheduling algorithm called Proportional Fairness Scheduling [3]. The scheduler serves the user with the highest ratio of the instantaneous downlink channel rate over the average throughput.

While HDR has the potential to provide "anywhere" "always-on" wide-area wireless Internet access, its peak downlink data rate of 2.4Mbps is relatively low compared with IEEE 802.11b links. In the past couple of years millions of people and business units have installed IEEE 802.11b (Wi-Fi) [1, 18] networks, making it the most popular local-area wireless data technology. IEEE 802.11b interfaces work in the license-free 2.4GHz ISM frequency band and provide data rate up to 11Mbps. The standard defines two modes. In the infrastructure mode each mobile client associates and communicates with an IEEE 802.11b access point. When an IEEE 802.11b access point is not available, IEEE 802.11b interfaces are able to communicate with each other on a peer-to-peer basis, namely the *ad-hoc* mode. Sources and destinations that are beyond immediate reach deliver data packets through multi-hop forwarding using an ad-hoc routing protocol, e.g., DSR [20] and AODV [23].

2.2 Related Work

The related work can be classified into different categories, depending on the traffic model (peer-to-peer versus infrastructure access), the relay model (dedicated/stationary versus mobile), and the number of interfaces (one versus two) used. Our design falls into the category of infrastructure access using mobile relays with two interfaces.

The majority of the work in the literature focuses on the ad-hoc network model that uses mobile clients as relays to route peer-to-peer traffic within the network [20, 23, 11]. However, given the lack of service availability guarantees due to potential network partitions, this model is typically used by niche applications in scenarios such as military communication and disaster relief. In UCAN, the ad-hoc routing component is much more efficient and reliable because of its explicit use of the cellular infrastructure, and the protocol complexity is also significantly lower.

There has been some work in the area of integrating the ad-hoc and infrastructure network models, but most of these projects involve the use of a single wireless interface for both the relay and infrastructure modes. For example, in [5], the authors allow GSM terminals to relay traffic to other terminals in order to improve coverage. In Opportunity Driven Multiple Access [25], the CDMA transmissions from a mobile host to the base station are broken into multiple wireless hops, thereby reducing transmission power. In [30], the channel pool is divided into a set of fixed channels and a set of forwarding channels so that data packets can hop from "hot" cells to "cold" cells using the forwarding channels in order to reduce delay and increase capacity. In [6], the authors consider a generic multihop wireless network where the mobile clients communicate with a mobile base station for Internet access, but the clients use only one interface. The authors in [16] also investigate a hybrid IEEE 802.11 network architecture with both DCF and PCF modes, again using one wireless interface. Thus, the total cell throughput achieved in their hybrid network is upper bounded by the throughput achievable in the cellular-only mode. In UCAN, since we use high-bandwidth wireless channels in ad-hoc mode (IEEE 802.11) to relay the traffic of the cellular network (3G), our hybrid network architecture exhibits significant cell throughput gains over the throughput achievable in the cellular-only mode. In [22], the authors propose a multihop cellular system where every mobile client participates in relaying traffic. The goal there is to reduce the number of base stations and use relay to increase coverage. However, the system increases overall capacity only when the communicating entities are in the same cell, a relatively uncommon occurrence.

One system that uses two interfaces to integrate cellular and ad-hoc networks is the iCAR system [13]. However, the authors primarily focus on improving call blocking probability for circuit-like traffic by diverting traffic from congested cells to neighboring lightly-loaded cells. They use pre-deployed, dedicated stationary relays for this purpose, resulting in increased cost. Other techniques to improve the throughput of wide-area networks include increasing available spectrum, using multiple antennas [15] etc., but each of these approaches also incurs high cost. Our goal in this paper is to use the mobile clients themselves as relays to improve the data throughput of a single cell, thus incurring no additional equipment cost to the wide-area network operator. However, we do point out that, UCAN may work with these techniques in concert to further improve the system throughput.

3. MOTIVATION

In order to verify our hypothesis that one can use IEEE 802.11 relays to increase the throughput of 3G networks, we conducted a simple experiment. The testbed consists of a Windows laptop, a Linux-based relay device and a Linux server. The laptop has an integrated IEEE 802.11b interface and a Sierra Wireless AirCard 555 CDMA2000-1X PCMCIA card. The Linux-based relay has two interfaces, a Sierra Wireless CDMA2000-1X PCMCIA card and a Proxim IEEE 802.11b card. The Linux server is connected to the Internet via a T1 link (1.5Mbps) and acts as an FTP server. We subscribe to the Verizon Wireless CDMA2000-1X service that supports data rates of up to 144 Kbps.

We conduct experiments in two modes: no relay and relay. In the no relay mode, the laptop is placed in the lab and we download a one megabyte compressed file over the CDMA2000-1X network from the Linux server. We turn off data and header compression on the CDMA2000-1X link. We conduct multiple runs (13) of each download during the day (1-4pm) and compute the average and maximum throughput for the transfer of the one megabyte file.

In the relay mode, the laptop is placed in the same location in the lab but we place the Linux relay in the corridor, where the signal strength is higher. We configure the routing tables so that when the laptop is downloading the file in the relay mode, it uses the IEEE 802.11b interface connected to the Linux relay and the Linux relay uses its CDMA2000-1X link to perform the download. Again, we perform a number of runs and calculate the average and maximum throughput.

Mode	3G Signal Strength	Avg T'put	Max T'put
No Relay	-96 to -105 dBm	51.6 Kbps	95Kbps
Relay	-78 to -80 dBm	93.9 Kbps	130Kbps

Table 1: Throughput with/without relay

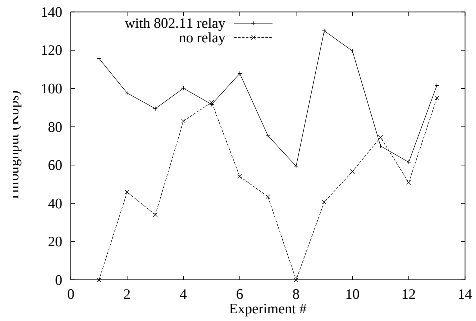

Figure 1: Experimental results with/without relay

The results of each run are shown in Figure 1 and summarized in Table 1. The runs with throughput of 0Kbps were cases where we couldn't complete the download as the 3G connection got disconnected. It is clear from the results that the IEEE 802.11b relay in the simple static configuration is able to significantly improve the average and maximum throughput by taking advantage of its better signal strength of the CDMA2000-1X channel. The average throughput values show typical performance while the maximum throughput values isolate the effects of background traffic loads on the commercial CDMA2000-1X network and gives us an idea of the best case throughput.

Since the simple static experiments show potential, we would now like to examine the general case where more mobile clients participate as relays for each other. We next describe our UCAN architecture and the protocols for the general case.

4. UCAN ARCHITECTURE

Although a large number of 3G wide-area and IEEE 802.11 local-area networking technologies apply, we present UCAN architecture in the specific context of 1xEV-DO, i.e. High Data Rate (HDR) [8], and IEEE 802.11b, i.e. Wi-Fi [18]. We choose these two technologies because of their support for high data rate and their popularity. We assume that each device in UCAN has dual wireless interfaces: HDR and IEEE 802.11b. It can be a portable computer with both 3G wireless modem and IEEE 802.11b PCMCIA card, or a PDA with both interfaces integrated in a single card [2].

The UCAN architecture is based on the key idea of opportunistic use of the IEEE 802.11 interfaces to improve the 3G wide-area cell throughput. Our design is based on three observations. First, only managed infrastructure of wide-area cellular network can provide the desired anywhere always-on availability assurance for the majority of Internet applications. Second, peer-to-peer communication offered by the IEEE 802.11b ad-hoc mode is cost-effective, given the popularity of IEEE 802.11b enabled devices. Fi-

Figure 2: UCAN Architecture

nally, IEEE 802.11b standard defines high-bandwidth wireless communication with data rate up to 11Mbps[1] within limited range, compared with existing 3G wide-area wireless technologies. Although IEEE 802.11b ad-hoc networks cannot be counted on by themselves due to the unreliable connectivity in the presence of client mobility and potential interferences in the license-free frequency band, opportunistic usage of their high-bandwidth links within a *local* area can greatly improve the users' quality of access to the wide-area cellular network.

Figure 2 shows the UCAN network architecture. For those mobile devices associated with the HDR base station, some of them may be actively receiving data packets from the Internet via the HDR downlink, while others may have their HDR interfaces in the dormant mode. Associated clients monitor the pilot bursts of the HDR downlink to estimate their current downlink channel conditions. At the same time, these devices turn on their IEEE 802.11b interfaces in ad-hoc mode, and run UCAN protocols. If a destination client experiences low HDR downlink channel rate (e.g., 38.6Kbps), instead of transmitting directly to the destination, the HDR base station transmits the data frames to another client (proxy client) with a better channel rate (up to 2.4Mbps). These frames are further relayed through IP tunneling via intermediate relay clients to the destination, using the high-bandwidth IEEE 802.11b links.

The above seemingly simple UCAN relay operation poses three main challenges:

- Given that clients are mobile, how does the HDR base station discover the proxy that has a good downlink channel rate while remaining connected with the destination client through the IEEE 802.11b ad-hoc network?
- Once an existing relay path in the IEEE 802.11b ad-hoc network breaks, or the existing proxy's channel rate decreases, how are the HDR base station and the destination client informed so that the discovery of a new proxy can be initiated, while avoiding bursty transient data frame losses?
- Given that individual mobile clients are autonomous, how are they motivated to turn on their IEEE 802.11b interfaces and consume their precious battery power to relay data frames for other clients?

[1]Both IEEE 802.11a and newly approved IEEE 802.11g provide data rate up to 54Mbps. Next standard IEEE 802.11n is expected to be at least 100Mbps and could reach 320Mbps.

The next three sections answer these questions in detail. We start with the design of proxy discovery and routing protocols. We then present the scheduling algorithm and the secure crediting protocol that balance throughput gain among clients and motivate autonomous users to participate in UCAN relay.

5. PROXY DISCOVERY AND ROUTING

In this section we present greedy and on-demand proxy discovery, and the route and proxy maintenance protocols. We start with an overview and then present the details.

5.1 Overview

When a mobile client that is actively receiving data frames from the HDR base station (henceforth called destination client) experiences low HDR downlink channel rate, it sends out a route request message using its IEEE 802.11b interface. This route request message is propagated through several intermediate mobile clients (henceforth called relay clients), according to the *proxy discovery* protocol, to reach a mobile client with high HDR downlink channel rate (henceforth called proxy client). This route request forwarding process installs routing information in each relay client to enable data frames to traverse the same path in the reverse order and reach the destination client. Thus, the proxy discovery protocol also serves as the route establishment protocol.

The proxy client then sends a proxy application message to the HDR base station through the HDR uplink. Accordingly the HDR base station updates the proxy table entry for the destination client. Starting from the next scheduled time slot for the destination client, the HDR base station transmits data frames to the proxy client.

When the proxy client receives a data frame from the HDR downlink, it checks the destination signature field of the frame [8] and forwards the frame to the destination client via its IEEE 802.11b interface, based on the routing information that is established during the route request propagation. We use IP tunneling to encapsulate the data frame in an IP packet.

The HDR downlink channel quality is a critical parameter that guides the proxy discovery. In UCAN, each mobile client maintains a moving average of its HDR downlink channel rate. The moving average filters out high-frequency variations of channel conditions due to fast fading and captures the signal strength due to the distance-based slow fading. The routing decisions are made based on this average downlink channel rate that remains stable in a relatively larger time scale (in seconds or more depending on the moving speed of the mobile client).

We devise two proxy discovery protocols: *Greedy* and *On-demand*. The greedy protocol is *proactive* in that all clients proactively maintain their immediate neighbors' average downlink channel rates. When the route request message is issued, it is *unicast* to the neighbor with the highest downlink channel rate. The message then traverses greedily through a set of relay clients with increasing downlink channel quality to the proxy client and then finally to the HDR base station. The on-demand protocol is *reactive*. When a mobile client issues a route request message, it *broadcasts* the message to all its neighbors within a given range. Those neighbors with high channel quality contend to serve as the proxy by sending application messages to the HDR base station. Thus, the two protocols may find different proxies. They also incur different overhead on the 802.11b network and the HDR uplink.

In the next two sections, we present the details of the greedy and on-demand proxy discovery and routing protocols. We then present the route and proxy maintenance mechanisms in the presence of client mobility and HDR channel rate variation.

5.2 Greedy Proxy Discovery

In greedy proxy discovery, neighboring mobile clients within one-hop IEEE 802.11b transmission range periodically exchange their average downlink channel rates by broadcasting a neighborhood advertisement message (NBADV). Thus, each mobile client proactively maintains a table of its neighbors' IDs (e.g., IP addresses) and their most-recently advertised average HDR downlink channel rates. When a destination client decides to look for a proxy client, it unicasts a route request message (RTREQ) to the neighbor with the best HDR downlink channel rate. The destination client also sets the TTL field of the RTREQ message to control the propagation range, and therefore the length of the ad-hoc relay path.

```
recvGreedyRTREQ(pkt, w)//u gets RTREQ from w
createRouteEntry(pkt.src,w);
v = getBestChanNbr();
if ((v.channelRate > channelRate)
   AND (pkt.TTL > 0)) then
      pkt.TTL = pkt.TTL-1;
       // u forwards RTREQ to its best neighbor
      unicast(pkt,v);
else
       // u declares to BS that it is the proxy
      declareProxyToBS(u,pkt.src);
endif
```

Figure 3: Greedy proxy discovery at node u

The processing of RTREQ message at each relay node is shown in Figure 3. On receipt of a RTREQ message, the mobile client inserts an entry into its routing table for the destination client and sets the next-hop relay as the client from which it receives the RTREQ. If the RTREQ TTL is still larger than zero, the client further forwards it to the neighboring node with the best HDR downlink channel rate. If the RTREQ TTL reaches zero or the client does not have any neighbor with a better HDR downlink channel rate, the client constructs and sends a proxy application message to the HDR base station via the HDR uplink. The HDR base station updates its proxy table entry for the destination client and sets the proxy client accordingly. Figure 4 presents an example showing destination client A using greedy proxy discovery to discover proxy client D.

Greedy proxy discovery protocol relies on the existence of a greedy path to reach a proxy client with high HDR downlink channel rate. However, such a greedy path may not always locate the proxy with the best channel rate. As we can see in Figure 4, client E actually has the best HDR downlink channel rate among clients 2-hop away from the destination client A. The greedy proxy discovery is unable to find it due to the local minimum at client C.

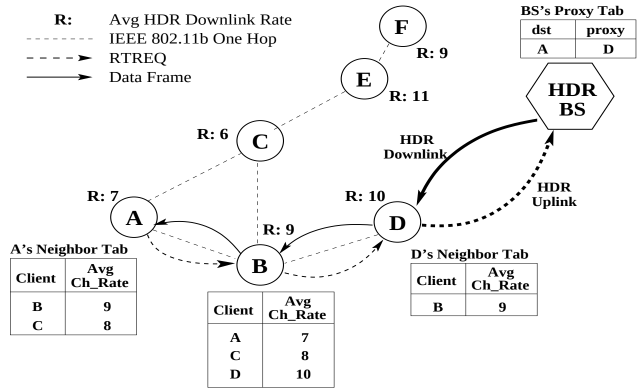

Figure 4: Greedy Proxy Discovery and Routing: Destination client A initiates a RTREQ to client B that finally reaches client D. Since client D has no neighbors with better HDR downlink channel rate, it sends a proxy application message to HDR BS using the HDR uplink. HDR BS updates its proxy table and sends proxy client D data frames destined for client A. Proxy client D forwards these data frames through intermediate relay client B to destination client A.

5.3 On-demand Proxy Discovery

In on-demand proxy discovery, mobile clients do not proactively maintain their neighborhood information. Instead, the destination client reactively *floods* a RTREQ message within a certain range. The RTREQ message carries the destination client's average HDR downlink channel rate and a sequence number that is incremented every time the destination client initiates a new round of proxy discovery.

```
recvOnDemandRTREQ(pkt, w)//u gets RTREQ from w
(seqNo,hop) = getSeqNoAndHopCount(pkt.src);
if (pkt.seqNo > seqNo) OR
   ((pkt.seqNo == seqNo)AND(pkt.hop < hop)) then
      createRouteEntry(pkt.src,w);
      if pkt.channelRate < channelRate then
         pkt.channelRate = channelRate;
          // u declares to BS that it can be a proxy
         declareProxyToBS(u,pkt.src);
         if pkt.TTL > 0 then
            pkt.TTL = pkt.TTL-1;
             // u broadcasts RTREQ to its neighbors
            broadcast(pkt);
         endif
      endif
endif
```

Figure 5: On-demand Proxy Discovery at node u

The processing of a RTREQ message in on-demand proxy discovery is shown in Figure 5. Whenever a mobile client receives a RTREQ message, it compares the sequence number with the largest RTREQ sequence number it has seen for the destination client. It drops the RTREQ message if the sequence number is smaller, or if the sequence numbers are equal but the hop number is no smaller. Otherwise, the client updates its routing table for the destination client, and compares its own average HDR downlink channel rate with

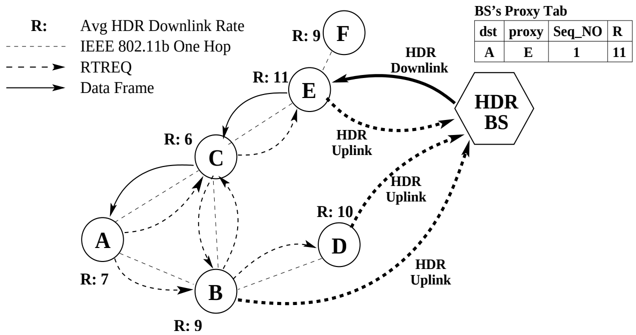

Figure 6: On-demand Proxy Discovery and Routing: Destination client A floods a RTREQ with TTL set to 2. Client B, D and E forwards the RTREQ messages via the HDR uplink to apply to be the proxy client for A. HDR BS chooses E that has the best average downlink channel rate as the proxy client.

the HDR downlink channel rate carried by the RTREQ message. If its own HDR downlink channel rate is higher, the client writes its own channel rate into RTREQ, and forwards a copy of the RTREQ message to the HDR base station to request for becoming a proxy client. Moreover, the client decrements the RTREQ message's TTL, and further broadcasts the RTREQ message if the TTL is still positive. See Figure 6 for an example.

When it receives a RTREQ from a mobile client applying to be the proxy of a certain destination, the HDR base station first compares the sequence number of the RTREQ with the sequence number of the proxy table entry. If the sequence number of RTREQ is larger, HDR base station uses the new proxy client and updates the proxy table entry. If equal, HDR base station chooses the new proxy only if its HDR downlink channel rate is no smaller and the path length is no larger than the existing proxy client. The RTREQ message is discarded if its sequence number is smaller.

Compared with the greedy scheme, on-demand proxy discovery is able to locate the proxy client with the best HDR downlink channel rate (assuming that broadcast is reliable). The cost that comes with this optimality is the larger overhead on the HDR uplink, since usually multiple clients apply to the HDR base station to be the proxy. We could avoid the overhead of multiple HDR uplink messages by making the proxies reply to the destination client first, and allowing the client to select the best proxy. However, this approach has several drawbacks. First, there will be significant contention on the IEEE 802.11b link as replies from candidate proxies converge on the destination client. Second, it will be hard to gauge a suitable timeout value for the destination client to stop waiting for more replies. Third, the delay in discovering the proxy will be significantly higher due to the contention and the need for timeouts; the channel conditions or routes may change by then, resulting in proxy discovery thrashing. Our approach of making the proxies independently apply to the HDR base station provides a simple solution to these problems, albeit at the cost of a few messages transmitted on the HDR uplink.

5.4 Route and Proxy Maintenance

In the UCAN model, changes in the ad-hoc relay arises in three cases:

- The HDR downlink channel rate can change substantially between advertisements due to client movement or fading, resulting in routing loops.
- Mobile clients (destination, relays, or proxies) can move out of range from the ad-hoc relay path, resulting in route breakage.
- The HDR downlink channel rate of the proxy client can decrease over time and even become lower than that of the destination client.

In this section, we address these issues by leveraging the uniqueness of the UCAN model, i.e., the availability of a central coordinator (base station). Different from all existing ad-hoc network routing protocols [20, 23, 11], the always available HDR uplink and downlink, although with low bandwidth, allows us to devise simple yet effective solutions to all the above problems.

5.4.1 Routing Failures and Recovery

A relay path breaks when the proxy, relay or destination client moves out of range. When the next-hop relay client is out of reach, the IEEE 802.11b MAC layer calls a callback function to inform the client of such failures. The client then reports this routing failure to the HDR base station. The routing failure messages reset the proxy table entry for the destination client and new data frames will be sent to the destination client directly using the HDR downlink. This way, it only takes one transmission of a single routing failure message via the HDR uplink to recover from the routing failures. From the destination client's perspective, consistent direct transmission via the HDR downlink implies the failures of the previous relay route. If its current downlink channel rate is still unsatisfactory, the destination client can simply issue another round of proxy discovery to establish a new proxy client at the base station.

The above simple route failure-recovery mechanism reacts very fast to the routing failures. Although on-the-fly data frames that are queued at the old proxy client and the intermediate relay clients have to be dropped, we can rely on the radio link protocol (RLP) running between the HDR base station and the destination client to retransmit these lost data frames.

5.4.2 Proxy Maintenance

Besides route breakage that causes proxy re-discovery, the HDR downlink channel quality of the proxy client may degrade with time due to mobility, channel fading and/or interferences. At the same time, the average downlink channel rate of the destination client may increase. In this case, using proxy client can result in lower throughput than direct transmission to the destination client. Proper proxy maintenance mechanisms have to be designed to ensure superior HDR downlink channel quality of the proxy client, compared with the destination client's own.

One approach is to periodically send out RTREQ messages to look for potentially better proxy clients. However, there are two problems with this approach. First, it increases the routing overhead over the ad-hoc network and

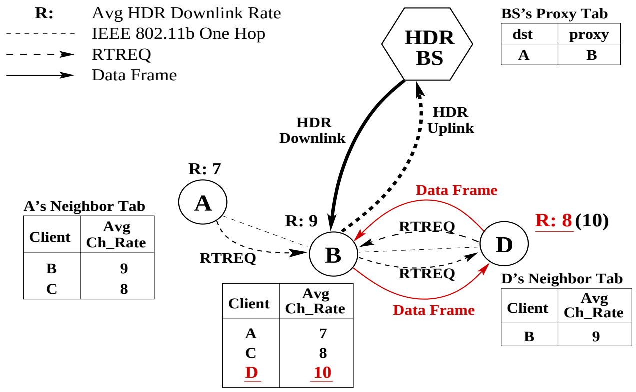

Figure 7: Proxy Discovery: Routing Loop

consumes energy. Second, it will be difficult to correctly gauge the right value of the periodic interval for sending RTREQ messages as it depends on the moving speed of the clients along the relay path and their wireless channel characteristics.

We take a reactive approach by letting the proxy client statistically piggyback its own updated average downlink channel rate in the data frames that are forwarded toward the destination client. If the proxy client's average downlink channel rate goes below certain threshold compared with the destination client's own, the destination client simply issues another round of proxy discovery to locate a better proxy. Once the new proxy reports itself to the HDR base station, the proxy table entry will be updated and the base station will start to use the new proxy client thereafter.

5.4.3 Routing Consistency and Loops

For greedy proxy discovery, a routing loop may form if a mobile client's average HDR downlink channel rate changes substantially between neighborhood advertisements. In the example shown in Figure 7, client D's average HDR downlink channel rate decreases from 10 to 8 after its last neighborhood advertisement. The neighborhood table maintained at client B is therefore *inconsistent* with client D's updated average HDR downlink channel rate. When it receives the RTREQ message forwarded by client B, client D sends it back to client B because from D's neighborhood table client B has a larger average HDR downlink channel rate. If destination client A sets the RTREQ TTL as 3, on receipt of the RTREQ bounced-back from client D, client B will apply to the HDR base station as proxy because the RTREQ TTL is zero. A routing loop then appears. When data frames are sent over the HDR downlink to the proxy client B, they will loop between client B and D until their TTL hits zero. Moreover, because IEEE 802.11b broadcast messages are not protected by the RTS-CTS channel reservation, NBADV broadcast collisions may also lead to stale neighborhood table entries and therefore routing loops.

Note that the sequence number used in on-demand proxy discovery can potentially be used to detect routing loops. In the above example, when the RTREQ is bounced back from D, client B can compare the sequence number in the RTREQ message with the maximum RTREQ sequence number it has seen for the destination client. Equal or smaller sequence number in the incoming RTREQ message (equal in this example) signals potential routing loops. However, this approach requires each client maintain a sequence number for potentially all other clients in the network, raising a scalability issue.

In UCAN, we piggyback the complete relay path in the RTREQ messages to eliminate potential routing loops, at the cost of the growing size of the RTREQ messages. However, RTREQ messages are infrequent compared to data packets, and the length of UCAN relay path is usually limited to 2 to 4 hops. Furthermore, as we shall see in Section 7, the presence of the complete path in the RTREQ messages also enable secure crediting, an important mechanism to motivate users in participating as UCAN relays. Note that the sequence number is still necessary for on-demand proxy discovery, because the base station uses it to differentiate the most recent round of proxy discovery messages from old ones that are simply delayed.

Our solution is different from source routing in that we only carry the complete relay path in the RTREQ messages. The relay path is not included in data frames, thereby the per data packet overhead of source routing is avoided.

6. UCAN SCHEDULING ALGORITHM

In this section we investigate the impact of UCAN relay on HDR scheduling fairness. HDR adopts the proportional fairness scheduling [3, 29] to schedule clients at each time slot. Specifically, let $R_i(t)$ be the instantaneous downlink channel rate for mobile client i at time t, and $T_i(t)$ be client i's average throughput in a past time window. $T_i(t)$ is maintained as a moving average $T_i(t+1) = (1-1/w)T_i(t) + 1/wT_i(t)$ where w is the window size. A proportional fairness scheduler schedules the client k with the minimum $\frac{T_k(t)}{R_k(t)}$ at every time slot. This algorithm leverages multi-user diversity in the instantaneous downlink channel rates. A client is scheduled when its downlink channel rate is high in order to improve the overall downlink throughput, while the short-term fairness in terms of the clients' throughput is also considered.

6.1 Throughput Gain Balance

In UCAN, packets may be transmitted to a proxy client over the 3G downlink, and we cannot simply use the $R_i(t)$ and $T_i(t)$ of either the final destination or the proxy client to run the the proportional fairness scheduling. However, in order to maintain compatibility with the scenarios where no ad-hoc relay is enabled, and preserve the fairness and throughput optimality of the proportional fairness scheduling, we would like to still use the same scheduling criterion, i.e., $\frac{T_i(t)}{R_i(t)}$.

Therefore, we need to choose the scheduling metric in the presence of proxy forwarding. For the average throughput $T_i(t)$, a straightforward way is to update $T_i(t)$ of the destination client with the amount of bits that the base station transmits for it, either directly or through a proxy. The question remains as how to set $R_i(t)$: we could use the downlink channel rate of either the proxy or the destination client.

At a first glance, the proxy client's downlink channel rate seems to be a reasonable choice, because it is the data rate that is actually used for the HDR downlink transmission. However, compared with the case where no proxy relay is enabled, this choice places the destination client in an advantageous position in scheduling: given a certain throughput

Client	No Relay	UCAN Relay Schd w/ Prxy's R		UCAN Relay Schd w/ dst's R	
	t'put	t'put	% increase	t'put	% increase
A	2/2	2/2	0%	4/3	33%
B	1/2	2/2	100%	2/3	33%
A+B	3/2	2	33%	2	33%

Table 2: Scheduling w/ proxy client's downlink channel rate v.s. scheduling w/ destination client's downlink channel rate: throughput and ratio (over the no-relay throughput).

the destination client will have a smaller scheduling metric $\frac{T_i(t)}{R_i(t)}$ because the proxy client's downlink channel rate is larger than the destination client's own. This increased rank for the destination client may negatively affect the proxy client's motivation to participate in relay. On the other hand, if we use the destination client's own downlink channel rate in computing $\frac{T_i(t)}{R_i(t)}$, the increased scheduling rank of the destination client can be eliminated. The increased HDR downlink channel utilization can then be shared among the destination and the proxy clients.

We use a simple example for an illustration. Suppose client A has a constant downlink channel rate of 2, and client B has a constant downlink channel rate of 1. Recall that under the proportional fairness algorithm, $T_i(t)/R_i$ has to be equal for A and B. Without relay, the ratio of the channel rates of A and B is 2:1. Therefore their throughput ratio will be 2:1 as well. For every 2 slots, A will be scheduled in 1 slot (with throughput = A.rate×1slot/2slot = 2/2) and B will in the other slot (with throughput = B.rate×1slot/2slot = 1/2). The aggregate throughput is 1/2+2/2=3/2.

With relay, the BS will always transmit to A (either as a destination or a proxy) using its superior rate of 2, resulting in an increase of the aggregate throughput from 3/2 to 2. If we use the proxy's rate to calculate $T_i(t)/R_i$, then throughput of A and B will all be 1. Note that all of the increase in HDR downlink channel utilization goes to client B (from 1/2 to 1), and there is no improvement on the proxy client (A)'s own throughput. On the other hand, if we use the client destination's rate to calculate $T_i(t)/R_i$, then the throughput ratio of A to B will remain 2:1, the same as the case without relay, and the throughput gain will be distributed between A and B proportionally.

The individual and aggregate throughput values for various cases are shown in Table 2. From the last column in the table, we can see that using the destination client's own downlink channel rate for scheduling balances the throughput gains among the destination client, the proxy client, and the aggregate HDR cell throughput. This property serves as strong motivation for the network operator to enable UCAN relay for increased aggregate downlink utilization. It also serves as motivation for the proxy and intermediate relay clients in terms of perceived throughput increase for their own downlink flows. For those clients that are not actively receiving packets from the base station, we further propose a secure crediting mechanism in Section 7 to provide extra incentives so that all clients are encouraged to participate in UCAN relay.

6.2 Exploiting Increased Diversity

Note that proxy clients are established based on their *average* downlink channel rate. Due to fast fading, a large average downlink channel rate may not always lead to a large *instantaneous* downlink channel rate that is actually used in downlink transmission. If the instantaneous downlink channel rate of the destination is larger than that of the proxy for a specific time slot, the HDR base station can send the data frames directly to the destination client. This provides a way to take advantage of increased channel diversity [29] because of our UCAN relay. Although beneficial for throughput, this mechanism may result in occasional out-of-order packets as they arrive at the destination client. However, the packet re-ordering is limited by the fundamental difference between the maximum HDR downlink channel rate and the IEEE 802.11b wireless link bandwidth. Furthermore, the reliable link protocol (RLP) of the HDR system will ensure that packets are delivered to the IP layer in correct sequence. In our simulations, a small buffer of three packets can eliminate almost all the out-of-order packets effectively.

To further leverage the increased channel diversity in UCAN architecture, HDR base station can transmit a data frame to the intermediate relay clients with the best instantaneous downlink channel rate along the ad-hoc relay path from the proxy to the destination client. This allows *all the clients in the ad hoc relay path to serve as proxies, thereby further increasing the channel diversity.* Note that the base station will need the complete relay path information, which is already available at the proxy client for the purpose of loop detection (see Section 5.2). Thus, all we need is to forward the complete path information to the HDR base station. While forwarding the path information incurs extra overhead on the 3G uplink, this overhead is limited by the short length of the relay path that is usually around 3 hops, as we will show in our simulations.

7. SECURE CREDITING

Although it is clear from the previous section that clients who are actively receiving data from the HDR base station are motivated to participate in the UCAN relay with perceived throughput increase for their own downlink flows, extra incentive has to be provided to encourage other clients who are associated with the HDR base station but not actively receiving. We design secure crediting as part of the UCAN architecture for this purpose. In essence, all the intermediate clients along an ad-hoc relay path, including both the proxy and the relay clients, are awarded *credits.* These credits can be redeemed in the form of shared revenue, or increased priority in the future call admission, packet scheduling and/or network traffic engineering. We leave the details of the credit accounting [12, 26], and focus on the identification of legitimate intermediate clients along the data relay path.

Two problems arise in the crediting subsystem. One is the *deletion of legitimate clients* and the other is *addition of extra clients.* Client A on an ad-hoc relay path may intentionally add another client B that is not actually forwarding the data frames, so that client B can earn credits without contributing to the packet relay. The addition of extra clients discourages the network operators from enabling UCAN relay. On the other hand, a malicious client may intentionally remove a legitimate client from a relay path in order to gain a larger share of the credits with some credit assignment strategy. The deletion of legitimate clients discourages honest relay clients and defeats the very purpose of crediting.

Our solution is to piggyback a single keyed Message Authentication Code (MAC) in the RTREQ message as it propagates to the base station from the destination client[2]. The MAC authenticates the relay path so that the base station can precisely keep track of the number of data frames that are relayed by each proxy and relay client. To this end, each client negotiates a secret key with the base station. Depending on specific 3G system, this secret key can be derived from the already established secret between the client and its home network registration center, resulting in no extra key management overhead. In the rest of this section, we present the detailed design for both the greedy proxy discovery and the on-demand proxy discovery protocols. Our protocol for on-demand proxy discovery is similar with Ariadne [17] that is designed for DSR.

7.1 Greedy Proxy Discovery

In greedy proxy discovery, RTREQ messages are *unicast* from the destination to the proxy through relay clients. Each client along the relay path can encode both the upstream and the downstream clients in the MAC. Take the following example for an illustration. Destination client D is building a relay path through relay client C and B, and proxy client A. The propagation of RTREQ is described in Table 3. $(M)_K$ denotes the computation of the keyed digest of the message M using key K.

D:	$MAC_D = (RTREQ, [DC])_{K_D}$
D→C:	$(RTREQ, D, MAC_D)$
C:	$MAC_C = (RTREQ, [DB], MAC_D)_{K_C}$
C→B:	$(RTREQ, MAC_C, D, C)$
B:	$MAC_B = (RTREQ, [CA], MAC_C)_{K_B}$
B→A:	$(RTREQ, MAC_B, D, C, B)$
A:	$MAC_A = (RTREQ, [BA], MAC_B)_{K_A}$
A→BS:	$(RTREQ, MAC_A, D, C, B, A)$

Table 3: Secure Crediting Procedure for greedy RTREQ forwarding from destination client D to base station BS. Note that only invariant entries of the RTREQ are encoded in the MAC computation.

Because client A, B, C and D share secret keys (K_A, K_B, K_C, K_D) with the base station respectively, the base station can easily verify the authenticity of the path by repeating the above process of MAC computation. If the verification fails, the base station solicits every client to submit their MACs, e.g., MAC_A, MAC_B, MAC_C, in order to detect the cheating client(s).

A client cannot be added into or removed from the path by any single relay or proxy client (e.g., client C, B or A) without triggering a MAC verification failure. Even the destination client D cannot add another client, because its downstream client C only encodes in its MAC the upstream client from which it receives the RTREQ message. Proxy client A cannot add any client before or after itself, as the former will trigger a MAC verification failure at the base station and the latter will be detected by the base station immediately on receipt of the proxy application message through the HDR uplink.

7.2 On-demand Proxy Discovery

In on-demand proxy discovery, RTREQ messages are *broadcast* to all neighbors, and a client cannot encode the downstream client in the keyed MAC. In this case, a relay or proxy client may add another client before itself without triggering a MAC verification failure. However, unlike the greedy proxy discovery protocol, the increased path length will leave the longer path in a disadvantageous position as multiple candidate proxy clients compete at the HDR base station (see Section 5.3). Adding extra clients into the path will likely lead to the rejection of that path by the HDR base station.

Note that a client cannot be removed without triggering a MAC verification failure. In the previous example client B can remove its upstream client C only if it can receive the RTREQ message that is transmitted from D to C. In that case, a shorter path, i.e., D→B→A→BS, exists and should be used instead.

Note that the above mechanisms do not handle the case where two or more *consecutive* clients on a relay path conspire to add another client in the middle. The addition of a forged client will result in bursty data frame losses whenever the base station transmits data frames to the forged relay client (see Section 6.2) and, therefore, is subject to detection by other traffic analysis tools.

8. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the UCAN architecture in improving the downlink channel utilization in an HDR cell. We first present the models, metrics and methodology for our evaluation in Section 8.1, and two important optimization techniques in Section 8.2. We then present the simulation results, investigating the impact of a wide range of parameters such as client locations, mobility, and density. We start with the simplest scenarios with a single destination client in Section 8.3, and then move to the scenarios with multiple destination clients in Section 8.4. We compare the performance and overhead of our greedy and on-demand proxy discovery protocols under different settings, and compare with the scenarios where no UCAN relay is enabled.

8.1 Model, Metrics and Methodology

We implement the UCAN architecture and protocols in the *ns-2* simulator. The HDR downlink channel is modeled according to the published experimental data in [3, 8]. The HDR downlink channel quality is determined by both slow fading and fast fading[3]. Slow fading is modeled as a function of the client's distance from the HDR base station, as shown in Figure 8. Fast fading is modeled by Jakes' Rayleigh fading [19] as shown in Figure 9. The combined E_c/N_t for both slow and fast fading is then mapped to a table of supported data rate with 1% error [8]. Figure 10 presents a snapshot of HDR downlink instantaneous channel rates, and the average rate over a long time period for clients with different distances from the base station.

[2]Note that the MAC computation is very efficient resulting in negligible extra delay, e.g., $5 \sim 20\mu$s on a P-III Portable PC. It adds 20 bytes to the RTREQ messages with SHA-1 or 16 bytes with MD5.

[3]We do not model the zero mean shadow fading.

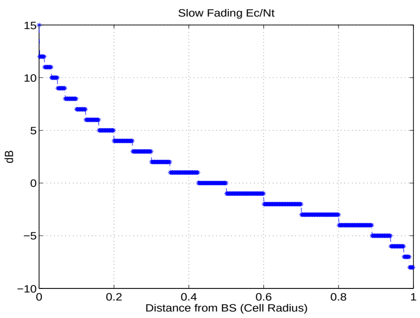

Figure 8:
HDR Downlink Slow-Fading E_c/N_t

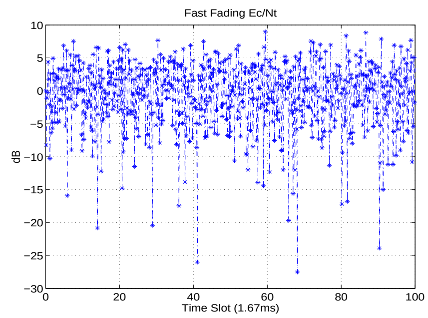

Figure 9:
Rayleigh Fast-Fading E_c/N_t

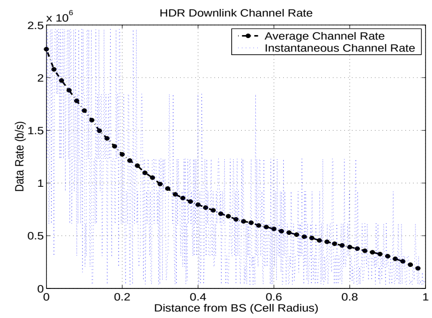

Figure 10: HDR Downlink Instantaneous and Average Channel Rate

Two observations can be made from Figures 8-10. First, the average HDR downlink channel rate degrades rapidly as a mobile client's distance from the HDR base station increases, with an average channel rate of around 600Kbps. There exists a lot of room for UCAN ad-hoc relay to improve the downlink channel utilization, especially for those clients that are located close to the edge of the cell where average channel rate is only around 100-200Kbps. Second, the channel rate varies with large amplitude in small time scale (1 time slot of 1.67ms). Each mobile client has to maintain a moving average of its instantaneous channel rate based on which relatively stable relay paths can be established. The large and rapid channel variation also justifies our scheduling algorithm (see Section 6.2) that explicitly leverages the increased channel diversity under our UCAN architecture where multiple relay and proxy clients are associated with each single destination client.

We use the IEEE 802.11b implementation in ns-2 version *2.1b9a* where 11Mbps data rate is supported at 115-meter communication range. The radio propagation model for IEEE 802.11b uses the Two-Ray Ground reflection model [24]. The mobility of clients is set according to the random waypoint model [20]. The mobile client starts at a random location, waits for a certain pause time, and randomly choose a new location and moves with a random speed chosen from zero to the maximum speed parameter. We set the pause time to be 3 seconds, and vary the maximum speeds to investigate the impact of client mobility. All the mobile nodes are within a square cell of $886{\times}886\text{m}^2$ with the HDR base station located in the center, approximating a 500-meter radius circular cell. These mobile clients share the HDR downlink using time multiplexing, with slot size of 1.67ms [3].

We simulate a certain number of TCP/FTP or UDP/CBR flows, each of which originates at the HDR base station and ends at a mobile client. The packet size is set to 1024 bytes for both TCP and UDP flows. The total load of the CBR flows is set to $1.01{\times}2.457$Mbps, making the HDR downlink overloaded even at its peak channel rate. For the IEEE 802.11b interface, we set the transmission and the receiving power consumption to be 1.089W and 0.792W respectively, based on Agere's short antenna type II extended PCMCIA IEEE 802.11b wireless LAN card [4]. We do not count the idle energy consumption as it is largely dependent on traffic patterns. Each simulation runs for 100 seconds, and each data point presented in the figures is the average over 5 random scenarios.

We use three metrics to evaluate the performance of our UCAN relay protocols. We compare the **maximum**, **minimum** and **aggregate throughput gains** for data flows in order to evaluate the effectiveness of our UCAN relay in improving the aggregate HDR downlink channel utilization as well as individual flow's throughput. **Routing overhead on HDR uplink**, and **energy consumption** on the IEEE 802.11b interface are analyzed to compare the greedy and on-demand proxy discovery protocols in terms of the overheads. Before we present the results, we describe two optimization techniques to decrease the energy consumption and overhead in proxy discovery.

8.2 Two Optimization Techniques

8.2.1 Frame Aggregation

The limited HDR downlink channel rate results in small data frames on the IEEE 802.11b ad-hoc network. For example, a proxy client with average downlink channel rate of 600Kbps will receive data frames of only 128 bytes each. Simply encapsulating and relaying each data frame in an IP packet may result in congestion and collision over the IEEE 802.11b network due to the high per-packet overhead for small packets. Figure 11 shows the throughput comparison between the HDR downlink and the IEEE 802.11b multihop wireless path. Due to space limitations, we do not present details of the per-packet overhead of IEEE 802.11b DCF. From Figure 11 we can see, if we relay each small data frame as it is, then the throughput over the 802.11 network limits the maximum length of the relay path to 1 hop only. If we use relay hops of length greater than one, we find that the IEEE 802.11b ad hoc network becomes the bottleneck resulting in reduction of the HDR downlink throughput. On the other hand, if we limit the proxy client to be 1 hop away from a destination, then significant throughput improvement is unlikely due to the limited choice of proxy clients.

We solve this problem using an optimization technique called *Frame Aggregation*. That is, a proxy client does not relay small data frames unless its IEEE 802.11b interface queue (IFQ) is empty. On receiving a data frame from its HDR downlink, the proxy client first tries to aggregate the frame with an existing frame waiting in the queue, as long as these two frames are for the same destination client. Each proxy client sets a threshold packet length. A data frame is transmitted only if its length is longer than the threshold or the IEEE 802.11b IFQ is empty. Based on the results of the

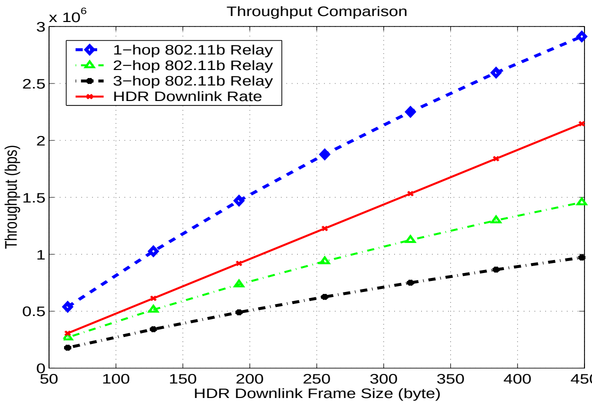

Figure 11: One HDR Frame in a Pkt

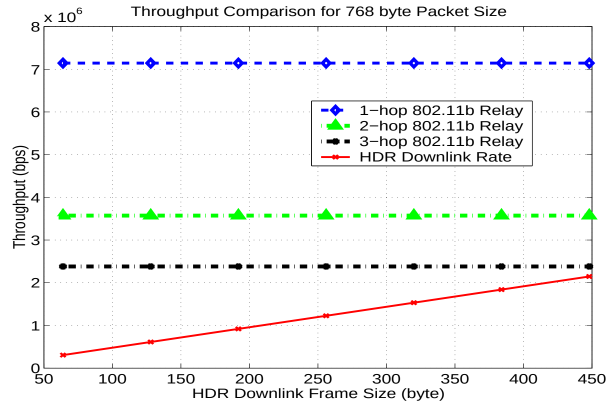

Figure 12: 768-byte Aggregated Pkt

analysis, as shown in Figure 12, *a threshold of 768 bytes in frame aggregation allows us to increase the relay path length from 1 to 3 hops without causing the IEEE 802.11b network to become the bottleneck.*

8.2.2 *Scoped Neighborhood Advertisements*

Greedy proxy discovery requires the maintenance of one-hop neighboring clients' average HDR downlink channel rates in order to forward the RTREQ message greedily. That is, each mobile client has to periodically broadcast its average HDR downlink channel rate. For mobile clients that are far away from destination clients, the bandwidth and energy consumption on the neighborhood advertisements are wasted.

Our scoped neighborhood advertisement seeks to avoid this excessive overhead. A destination client periodically advertises its average HDR downlink channel rate through IEEE 802.11b broadcast, but it sets the TTL of its NBADV messages to be the same as the TTL used in its RTREQ messages. Other mobile clients set the TTL of their NBADV messages to be $\text{TTL}_{max}-1$, where TTL_{max} is the maximum TTL in the NBADV messages advertised by their neighbors. A mobile client will periodically send NBADV messages if and only if $\text{TTL}_{max} > 0$. This way, we confine the range of mobile clients that are involved in the periodical advertisements, thereby saving the energy expenditure of mobile clients that are far away from destination clients.

In the example shown in Figure 4, if destination client A limits its relay path length to be 2 hops (by setting the TTL of its RTREQ messages as 2), only neighboring clients that are within 2 hops away may possibly be involved in the ad-hoc relay. Client F, for example, will be exempted from neighborhood advertisement as it is beyond the 2 hop range from A. We found through our simulations that *scoped neighborhood advertisements can substantially decrease the energy consumption for the greedy proxy discovery by up to 70%*, especially when a small number of destination clients are distributed in a dense HDR cell.

8.3 Single Destination Client Scenarios

In this section, we start with a simple scenario of a single static destination client receiving packets from the HDR base station. We fix the location of the destination client to be 400m away from the HDR base station, i.e., $0.8R$ where $R = 500$m is the radius of the simulated HDR cell. Considering the maximum communication range of 115m for IEEE 802.11b at 11Mbps data rate, we limit the maximum number of UCAN relay hops to be 3, by setting $TTL = 3$ for the RTREQ messages. That is, in the best case, we can discover a proxy client 100m away from the HDR base station with an average channel rate of 1.25Mbps (see Figure 10). We vary the mobile client's (other than the destination client) maximum moving speed from 0, 2, 5, 10 to 15m/s, and experiment with different client densities by placing 35, 50, 80 and 100 clients, including the destination client, in the cell.

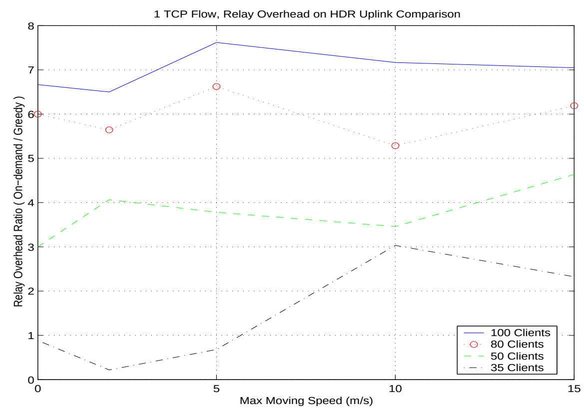

Figure 15: 1 TCP Flow, HDR Uplink Overhead

8.3.1 *Throughput Gain*

Figures 13 and 14 show the throughput gains over the scenario without UCAN relay for one UDP flow and one TCP flow respectively, under different client mobility and density settings. As we can see for UDP flow with modest mobility (2m/s) and high client density (100 clients), our UCAN relay can achieve a throughput gain of around 310% with an aggregate throughput of 1.18Mbps that is within 94% of the optimal 1.25Mbps. For both greedy and on-demand proxy discovery, the throughput improvement is robust under a certain client density as the client moving speed increases, demonstrating the effectiveness of our route maintenance and recovery from route breakage due to node mobility (see Section 5.4). In general, on-demand proxy discovery has a slightly higher throughput gain than greedy proxy discovery, as the RTREQ flooding approach is able to locate the proxy

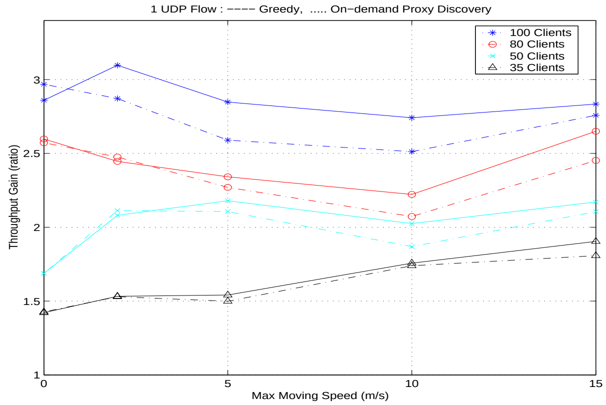

Figure 13: Throughput Gain for 1 UDP Flow. Greedy: Dashed-line, On-demand: Solid-line

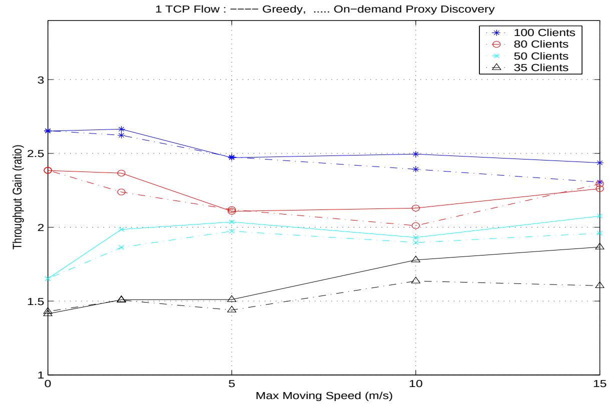

Figure 14: Throughput Gain for 1 TCP Flow. Greedy: Dashed-line, On-demand: Solid-line

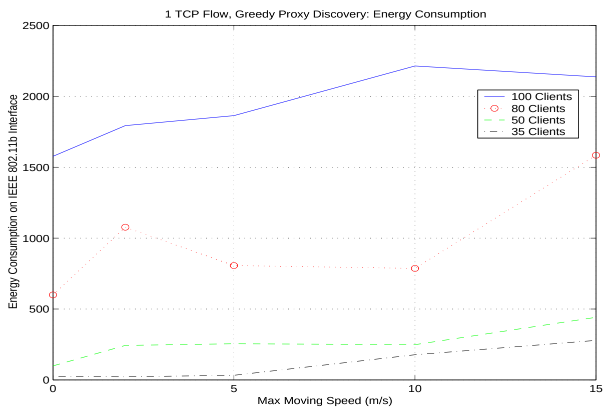

Figure 16: 1 TCP Flow, Energy Consumption for Greedy Proxy Discovery

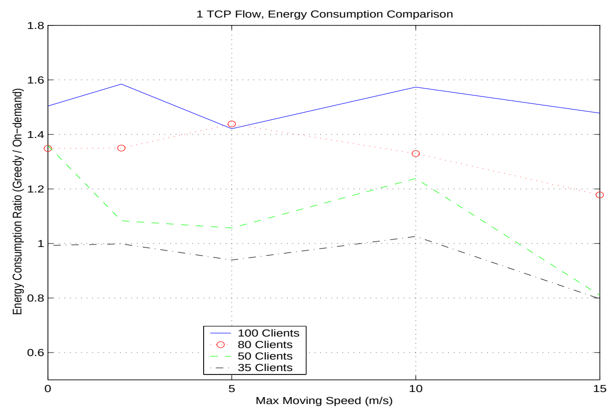

Figure 17: 1 TCP Flow, Energy Consumption Comparison

client with the best downlink channel within 3 hops. The throughput gain increases as the client density increases because of the increased average length of the relay path that leads to the proxy client with higher downlink channel rate. However, even at the low density with 35 clients in the cell, our UCAN relay is still able to achieve a 50% throughput increase for both TCP and UDP flows.

8.3.2 Overhead on HDR Uplink

Figure 15 compares the overhead of greedy and on-demand proxy discovery protocols on the HDR uplink due to the candidate proxy's application messages. Recall that in the greedy protocol, only one client generates a proxy application to the HDR base station while in the on-demand protocol, multiple clients independently generate proxy applications. We can clearly see from Figure 15 that on-demand proxy discovery has substantially larger overhead on the HDR uplink (up to 7 times) compared to the greedy proxy discovery. This overhead increases as the client density increases due to the increased number of candidate proxy clients. Thus, *while on-demand provides slightly higher throughput gains than the greedy approach, these gains come at the cost of substantial overhead on the HDR uplink*.

It is interesting to note that the ratio of overhead of on-demand to greedy approach remains almost constant with respect to client mobility speeds. The reason is that the ratio of the number of candidate proxy applications to the HDR base station remains the same for each route breakage, and high client mobility results in the same frequency of re-routing for both greedy and on-demand proxy discoveries. Thus, high client mobility speed results in higher absolute but similar relative overhead for both approaches.

8.3.3 802.11 Energy Consumption

The energy consumption of the IEEE 802.11b interfaces for greedy proxy discovery protocol is shown in Figure 16. The total energy consumption increases as the client density increases because more clients are involved in RTREQ and NBADV propagation. The higher frequency of re-routing due to route breakage as a result of higher client moving speed also increases the total energy consumption. Figure 17 shows the ratio of the energy consumption of greedy and on-demand proxy discovery protocols. As we can see from the figure, greedy proxy discovery consumes more energy for the scenarios of high client density and low mobility speed. However, *energy consumption for the greedy proxy discovery is only between 0-50% higher than on-demand approach due to our scoped NBADV optimization* (see Section 8.2.2).

It is interesting to note that the greedy proxy discovery protocol consumes less energy than the on-demand proxy

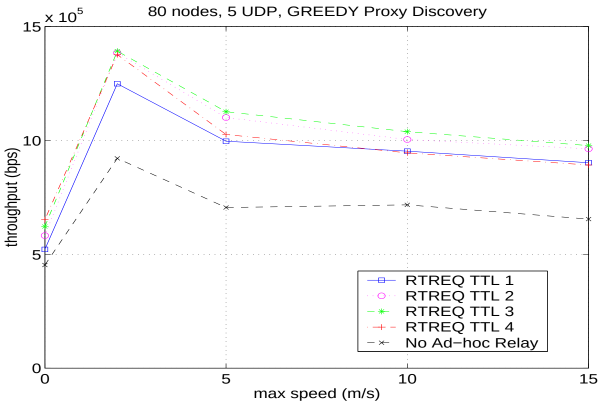

Figure 18: 5 UDP Flows, Throughput for Greedy Proxy Discovery

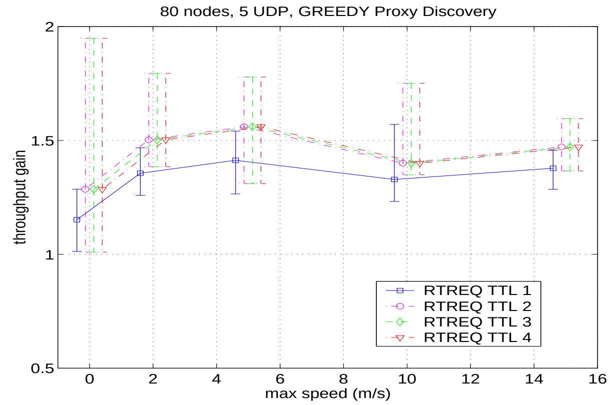

Figure 19: 5 UDP Flows, Throughput Gain for Greedy Proxy Discovery

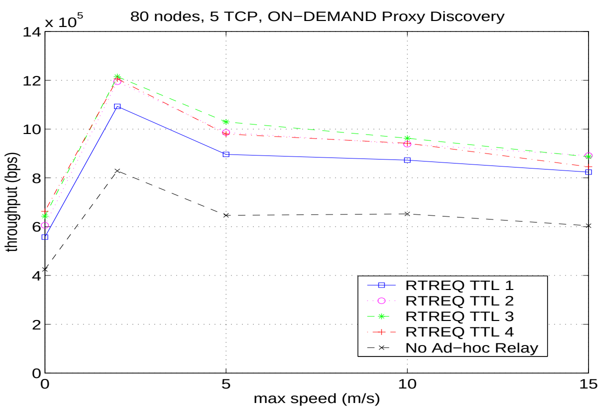

Figure 20: 5 TCP Flows, Throughput for On-demand Proxy Discovery

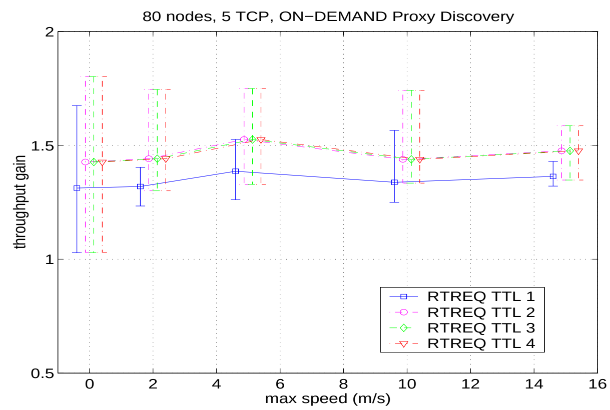

Figure 21: 5 TCP Flows, Throughput Gain for On-demand Proxy Discovery

discovery when the client moving speed is high for the scenarios with 35 and 50 clients. The reason is that high mobility results in frequent re-routing, and on-demand proxy discovery's RTREQ flooding consumes more energy than greedy proxy discovery's unicast forwarding of the RTREQ messages.

8.4 Multiple Destination Client Scenarios

In this section, we investigate the interactions between multiple competing flows and the throughput gain for each individual flow. We use a population of 80 clients in a 500-meter HDR cell. Five randomly chosen destination clients set up TCP or UDP connections with the base station. All clients (including the destination clients) are mobile during the simulations and we vary the moving speed from 0, 2, 5, 10 to 15 m/s. We also study the impact of the relay path length by setting the TTL of the RTREQ messages from 1 to 4.

Figures 18 and 19 show the throughput gain for greedy proxy discovery protocol with 5 random UDP flows. It is clear the aggregate throughput of these 5 UDP flows increase as we increase the relay path length from 1 to 3. For 4-hop relay path, it achieves highest throughput at low mobility (< 5m/s), but lower throughput than 2- or 3-hop relay with high client mobility. The reason is that longer relay path is subject to more route breakage in the presence of high node mobility, in which case the highest throughput gain is achieved with 3-hop relay. Figure 19 shows the maximum and minimum throughput gains for individual flows, and aggregate throughput gain compared with the case where the UCAN relay is disabled. As we can see the minimum throughput gains for individual flows are all larger than 1. This result verifies that our scheduling algorithm (see Section 6) proportionally distributes the increased downlink channel utilization to all active flows.

However, the aggregate throughput gains are around 30-60%, lower than the one destination client scenarios (see Section 8.3). This is due to a higher base aggregate throughput, i.e., without UCAN relay, of 500-700Kbps as compared with the scenarios of one single destination client where the base throughput is only around 340Kbps. The reason is that the HDR proportional fairness scheduler does a reasonable job of exploiting user diversity given that there are 5 backlogged flows all the time. Similar conclusions can be drawn for multiple TCP flows with on-demand proxy discovery protocol as shown in Figures 20-21.

9. DISCUSSION

In this section we discuss other issues relevant to the UCAN architecture.

Frugal Usage of HDR Links: It is tempting to push the routing overhead to the omni-present HDR links. However, aggressive usage of the HDR links is against our goal of improving the HDR throughput through IEEE 802.11b ad-hoc relay. Based on the difference in link bandwidth between these two types of technologies, it is desirable to use the IEEE 802.11b links and conserve HDR link resources. In our protocols, HDR up/downlink is used only when small overhead is incurred and high performance return is achieved.

Base Station Pull v.s. Client Push: We take the client-push approach to simplify the design and implementation at the base station. The base station only maintains a proxy table for each destination mobile client. Whenever a client is scheduled, the base station sends the frame to client's proxy with a higher downlink channel rate. The proxy will be the destination client itself initially and also during the route recovery period. It is left to the destination to initiate the proxy discovery and the intermediate mobile clients to update the proxy table and routing changes at the base station. This way, we distribute the overhead and complexity among mobile clients, and push most of the overhead away from the base station.

Variable Data Rate and Transmission Range in the 802.11 Network: We use constant data rate and communication range for the IEEE 802.11b network in the simulations presented in this paper. However, these parameters may adapt to the channel conditions at the sending or receiving mobile clients as defined in the standard [18]. Based on the discretion of the destination client and proxy clients, UCAN relay can be temporarily disabled if the IEEE 802.11b ad-hoc network is experiencing low data rate due to heavy interferences from adjacent ad-hoc networks, or heavy congestion due to contention with other peer-to-peer traffic. Note that the 3G base station does not need to be involved in this adaptation and remains simple.

HDR Uplink Proxy: While the focus of this paper is on improving the downlink performance of the 3G network since most applications primarily involve downloading, a similar proxy-based approach can also be adapted for improving uplink performance. One of the issues in the design of an uplink proxy is the need for the base station to provide feedback of the uplink signal strength back to the mobile clients, which could result in overheads on the 3G downlink channel.

Co-located HDR BS and IEEE 802.11 AP: There may be cases where the HDR BS and IEEE 802.11 AP are co-located. The question arises as to when to use the IEEE 802.11 channel and when to use HDR channel? We can extend our protocol to this case easily. All we need to do is to let IEEE 802.11 AP operate in ad hoc mode and advertise a very high (virtual) 3G channel quality. Those mobile clients who discover the IEEE 802.11 AP as a proxy will utilize the IEEE 802.11 AP for transferring data from the network, resulting in reduction of load on the 3G downlink.

Interaction with Peer-to-Peer Traffic: We focus on communications between an entity outside the HDR cell and a mobile client inside the cell as this type of traffic dominates in typical use. However, there may be other applications running in the IEEE 802.11 network, whose traffic are confined within the cell. Intensive research efforts have been made on routing and packet forwarding for this type of peer-to-peer traffic. In UCAN, these designs can be made much more reliable with substantially smaller overhead and lower complexity by leveraging the HDR uplink/downlink.

HDR Scheduling and End-to-end Delay Although we have chosen to use the proportional fairness scheduling algorithm in UCAN, other HDR scheduling algorithms such as [7, 27] that take both instantaneous channel rate and the queuing delay into account, can also be adapted to work in UCAN. The performance of such HDR scheduling algorithms, including proportional fairness scheduling algorithm, has been analyzed in terms of stability, mean throughput, mean number of active users and mean delay in [10, 9, 21, 28]. Given the same number of active users, UCAN improves the mean delay and mean throughput of the cell by relaying through proxy clients, assuming a packet reaches the destination client before the next frame finishes its transmission along the HDR downlink.

Multiple Cell Relay In this paper we only consider a single HDR cell where the proxy, relay and destination clients are located. It is possible to extend the design to allow ad-hoc relay across multiple adjacent cells, i.e., through a proxy client that is located at a different cell. However, this will require collaborations among multiple neighboring base stations, and the involvement of the upper layer entities in the 3G network architecture (e.g., RNC) so that packets can be dispatched to the cell where the proxy client is located. To maintain its simplicity in design and operations, UCAN in its current form does not consider this option.

Application Scenarios In Section 3 we motivate the UCAN architecture in a scenario where higher throughput can be achieved for the destination client located in a lab room through the proxy client located in the corridor. In general UCAN architecture is useful for both indoor and outdoor environments. For indoor environment, some devices, such as the ones near the glass windows, may have considerably better 3G data rates than other indoor devices. For outdoor environments, because IEEE 802.11 link has larger communication range compared with indoors, users in places like shopping plazas and residential areas can easily connect to other users in search for proxies. The UCAN architecture will improve the throughput of all the devices using IEEE 802.11 relay.

10. CONCLUSION

In this paper we present UCAN, a novel network architecture that unifies cellular and ad-hoc networks opportunistically to enhance the cell throughput. We propose two new proxy discovery and routing protocols that leverage the managed infrastructure to decrease the complexity and overhead, and increase the reliability. We then refine the HDR scheduling to balance throughput gain among clients. Finally we devise a secure crediting mechanism, providing strong motivation for the mobile clients to serve as relays. We found that the on-demand proxy discovery protocol delivered the highest throughput gains (up to 310% for individual user's throughput and up to 60% for aggregate cell throughput) but also resulted in high overhead on the 3G uplink. The greedy proxy discovery protocol delivered gains close to the on-demand approach and resulted in much lower overhead on the 3G uplink; however, the energy consumption of the greedy protocol was up to 50% higher than that of the on-demand protocol. Thus, one can deploy the on-demand or the greedy protocol, depending on the relative importance of the uplink bandwidth versus energy.

Acknowledgment

The authors thank Girish Chandranmenon for help in setting up the experiment testbed and the anonymous reviewers for their constructive criticisms.

11. REFERENCES

[1] 802.11 HotSpot. http:www.80211hotspots.com.

[2] GTRAN Dual-Mode 802.11/CDMA Wireless Modem. http://www.gtranwireless.com/newsevents/pressreleases_20020422.htm.

[3] 1xEV: 1x EVolution IS-856 TIA/EIA Standard - Airlink Overview. QUALCOMM Inc. White Paper, Nov. 2001.

[4] Short Antenna Type II Extended PCMCIA IEEE 802.11b Wireless LAN Card. http://www.agere.com/client/docs/DS02305.pdf, Aug. 2002.

[5] G. N. Aggelou and R. Tafazolli. On the Relaying Capacity of Next-generation GSM Cellular Networks. *IEEE Personal Communications Magazine*, 8(1):40–47, Feb. 2001.

[6] I. Akyildiz, W. Yen, and B. Yener. A New Hierarchical Routing Protocol for Dynamic Multihop Wireless Networks. In *Proceedings of IEEE INFOCOM*, volume 3, pages 1422–1429, 1997.

[7] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijaykumar, and P. Whiting. CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions. Technical report, Bell Laboratories, April 2000.

[8] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi. CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users. *IEEE Communications Magazine*, 38:70–77, Jul. 2000.

[9] S. Borst. User-Level Performance of Channel-Aware Scheduling Algorithms in Wireless Data Networks. In *Proceedings of IEEE INFOCOM*, volume 1, pages 321–331, 2003.

[10] S. Borst and P. Whiting. Dynamic Rate Control Algorithms for HDR Throughput Optimization. In *Proceedings of IEEE INFOCOM*, volume 2, pages 976–985, 2001.

[11] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva. A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols. In *Proceedings of ACM MOBICOM*, pages 85–97, 1998.

[12] L. Buttyan and J. P. Hubaux. Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks. *ACM/Kluwer Mobile Networks and Applications (MONET)*, 8(5), Oct. 2003.

[13] S. De, O. Tonguz, H. Wu, and C. Qiao. Integrated Cellular and Ad Hoc Relay (iCAR) Systems: Pushing the Performance Limits of Conventional Wireless Networks. In *Proceedings of Hawaii International Conference on System Sciences*, pages 3931–3938, 2002.

[14] Economist.com. The Fight for Digital Dominance. http://www.economist.com/displaystory.cfm?story_id=1454300, Nov. 21, 2002.

[15] G. J. Foschini and M. J. Gans. On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas. *Wireless Personal Communications*, 6(3):311–335, Mar. 1998.

[16] H.-Y. Hsieh and R. Sivakumar. On Using the Ad-hoc Network Model in Wireless Packet Data Networks. In *Proceedings of ACM MOBIHOC*, pages 36–47, 2002.

[17] Y.-C. Hu, A. Perrig, and D. B. Johnson. Ariadne: A Secure On-demand Routing Protocol for Ad Hoc Networks. In *Proceedings of ACM MOBICOM*, pages 12–23, 2002.

[18] IEEE Computer Society. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4GHz Band. IEEE standard 802.11b, 1999.

[19] W. C. Jakes. *Microwave Mobile Communication*. Wiley, 1974.

[20] D. B. Johnson and D. A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. In *Mobile Computing*, volume 353, pages 153–181. Kluwer Academic Publishers, 1996.

[21] H. Kushner and P. Whiting. Convergence of Proportinal Fair Sharing Algorithms under General Conditions. Technical report, Bell Laboratories, 2002.

[22] Y.-D. Lin and Y.-C. Hsu. Multihop Cellular: A New Architecture for Wireless Communications. In *Proceedings of IEEE INFOCOM*, volume 3, pages 1273–1282, 2000.

[23] C. E. Perkins and E. M. Royer. Ad-hoc On Demand Distance Vector Routing. In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, pages 90–100, 1999.

[24] T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[25] T. Rouse, I. Band, and S. McLaughlin. Capacity and Power Investigation of Opportunity Driven Multiple Access (ODMA) Networks in TDD-CDMA Based Systems. In *Proceedings of IEEE ICC*, volume 5, pages 3202–3206, 2002.

[26] N. B. Salem, L. Buttyan, J. P. Hubaux, and M. Jakobsson. A Charging and Rewarding Scheme for Packet Forwarding in Multi-hop Cellular Networks. In *Proceedings of ACM MOBIHOC*, pages 13–24, 2003.

[27] S. Shakkottai and A. L. Stolyar. Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR. In *Proceedings of International Teletraffic Congress (ITC)*, pages 793–804, 2001.

[28] A. Stolyar. Mult-User Throughput Dynamics under Proportinal Fair and other Gradient-Based Scheduling Algorithms. Technical report, Bell Laboratories, 2002.

[29] D. Tse. Multiuser Diversity in Wireless Networks: Smart Scheduling, Dumb Antennas and Epidemic Communication. In *IMA Workshop on Wireless Networks*, 2001.

[30] X. Wu, S.-H. Chan, and B. Mukherjee. MADF: A Novel Approach to Add an Ad-hoc Overlay on a Fixed Cellular Infrastructure. In *Proceedings of IEEE WCNC*, volume 2, pages 549–554, 2000.