Distributed Resource Allocation for DS-CDMA-Based Multimedia ad hoc Wireless LAN's

Sanjay Lal, *Member, IEEE*, and Elvino S. Sousa, *Senior Member, IEEE*

***Abstract*—Power control in direct sequence-coded vision multiple access (DS-CDMA) systems and, more recently, power/rate allocation in multirate DS-CDMA based networks is an open and interesting research area which has attracted much attention. However, with a few exceptions, most researchers have emphasized centralized resource allocation algorithms for cellular systems where the base station keeps track of the requirements of the various users and is thus responsible for the management of network resources. Ad hoc wireless local area networks (WLAN's), on the other hand, are generally configured as peer-to-peer networks with no centralized hub or controller. Thus resource allocation has to be conducted in a distributed fashion. We address the issue of distributed resource management for multirate DS-CDMA based multimedia WLAN's by 1) presenting a distributed resource allocation protocol, known as distributed resource negotiation protocol (DRNP) that builds on the *RTS/CTS* bandwidth reservation mechanism provided by IEEE 802.111, and provides quality of service (QoS) guarantees through distributed control of resources in DS-CDMA based multimedia WLAN's and 2) investigating the performance of various resource allocation schemes within the context of DRNP, in terms of network wide metrics such as overall throughput and blocking rates.**

***Index Terms*— Code division multiaccess, pseudonoise coded communication, resource management, wireless LAN.**

I. Introduction

With the allocation of unlicensed personal communication system (PCS) bands intended for wireless local data communications, the future of wireless local area networks (WLAN's) within the context of PCS seems assured. Fueled by the explosive growth of portable computers in the last few years, researchers are contemplating new concepts such as ad hoc networking, nomadic access, and mobile computing, leading to the fusion of computers and communications in a ubiquitous computing environment [1]. Moreover, with the shift toward integrated multimedia networks such as asynchronous transfer mode (ATM) in the wireline arena, WLAN's in their traditional roles as extensions to the wired infrastructure are expected to bring the revolutionary capabilities of wireline multimedia networks such as ATM to the wireless arena (i.e., the ability to efficiently integrate disparate applications with varying service requirements into a single networking infastructure), in the form of multimedia WLAN's.

Manuscript received January 1997; revised January 1999. This work was supported in part by Communications and Information Technology, Ont., Canada. Parts of this paper were presented at IEEE Milcom'98 Conference, Boston, MA, October 1998.

S. Lal was with the Advanced Interworking Systems Group, Lucent Technologies, Holmdel, NJ 07733 USA. He is now with Tachion Network Systems, Eatontown, NJ 07724 USA.

E. S. Sousa is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ont., M55 3G4 Canada.

Publisher Item Identifier S 0733-8716(99)03081-4.

TABLE I
Typical Characteristics of Various Multimedia Application Classes

Media	Bit-Rate	Bursty	Loss Rate
Audio	4-64 Kb/s	medium	$< 10^{-2}$
Video	n x 64 Kb/s	low	$< 10^{-4}$
Image	> 10 Kbytes	high	~ 0
Data	variable	high	~ 0

A. Quality of Service (QoS) Parameters

Next generation multimedia wireless information networks (WIN's) are expected to support a wide variety of applications with varying service requirements. These service requirements are typically expressed through application dependent parameters for three quantities: bandwidth, loss, and delay [2]. Table I illustrates the minimum bandwidth and loss requirements for the various application classes expected to be supported by multimedia WIN's. For instance, voice communication can be classified as a low rate service that can tolerate relatively high loss rates due to the inherent redundancy of speech signals. In contrast, data traffic requires very low probability of loss/error to ensure the preservation of data integrity.

It is the role of the resource management scheme to map the service requirements of the various applications to network resources so that the QoS requirements of the various users are met, i.e., given certain QoS constraints, what resources does the application require from the network such that the QoS requirements are met? The number of research papers dedicated to answering this question in the context of ATM networks is a testament to the complexity of the problem [3]–[5].

In multimedia WIN's, the notion of resource management is inexorably tied to the medium access control (MAC) mechanism. For instance, in time division multiple access (TDMA) based systems, resource management involves the allocation of time slots, while transmitted power is the core resource in direct sequence-code division multiple access (DS-CDMA) systems. Thus, the issue of resource management must be addressed within the context of a multiple access mechanism. To this effect, we propose multirate DS-CDMA as the multiple access mechanism of choice for next generation multimedia WLAN's. We assume that the service requirements are expressed in terms of minimum bandwidth and maximum packet loss rates and address the issue of resource management in such networks in terms of power control as well as transmission bandwidth allocation.

Power control in DS-CDMA systems [7], [8] and more recently power/rate allocation in multirate DS-CDMA based networks [9] is an open and interesting research area which has attracted much attention. However, with a few exceptions [10], most researchers have emphasized centralized resource allocation algorithms for cellular systems where the basestation keeps track of the requirements of the various users and is thus responsible for the management of network resources. WLAN's, on the other hand, are generally configured as peer-to-peer networks with no centralized hub or controller. Thus resource allocation has to be conducted in a distributed fashion. We address the issue of distributed resource management for multirate DS-CDMA based multimedia WLAN's by the following.

- Presenting a distributed resource allocation protocol that builds on the RTS/CTS bandwidth reservation mechanism provided by IEEE Standard 802.1113 and provides QoS guarantees through distributed control of resources in DS-CDMA based multimedia WLAN's.
- Investigating the performance of various resource allocation schemes in terms of network wide metrics such as overall throughput and blocking rates.

II. Network Model

We now present our model for an ad hoc multirate DS-CDMA based multimedia WLAN. A novel aspect of the architecture is a common receiver (C-R) based code protocol coupled with a dual receiver architecture. As with all multirate DS-CDMA networks, the transmission power and data rates are the core resources to be managed. We also define several quantities specific to the resource negotiation protocol to be presented later.

A. Asynchronous Multiprocessing Gain DS-CDMA

Although several asynchronous multirate CDMA schemes have been proposed [11], the scheme characterized by all users, regardless of their transmission data rates, using the entire system bandwidth is considered in this paper. Specifically, consider a terminal i, that generates an information bit stream of rate R_i. The information bits are spread by a pseudonoise (PN) code sequence with N_i chips per information symbol to obtain a transmission bandwidth of W. Thus the processing gain for terminal i is given by

$$G_i = \frac{W}{R_i}. \tag{1}$$

This scheme supports multiple data rates by varying the processing gain. Obviously, terminals transmitting at lower bit rates will have higher processing gains.

B. Packet Trains

In packet switched networks there is a tradeoff between packet sizes and protocol efficiency [12]. Generally, due to protocol overhead, it is desirable to maximize the data payload within a packet. However, in the presence of errors, retransmission of large packets leads to reduced protocol efficiency. We therefore propose using a packet train [13]. Instead of a single packet, a packet train consists of a large number of sequenced mini-packets, each suffixed with a CRC for error detection. At the receiver, the CRC's are used to detect errors within each of the mini-packets and a single negative acknowledgment (NACK) is sent back to the transmitter with the sequence numbers of the mini-packets that were received incorrectly. The transmitter then selectively retransmits the mini-packets that were incorrectly received.[1] As opposed to IEEE Standard 802.11, once a session has been set up, packet-trains do not require synchronization, source and destination addresses that must accompany individual packets. Moreover, when errors do occur, only the mini-packets that are damaged need to be retransmitted. Hence, packet trains allow us to reduce overhead by transmitting a relatively large payload, while avoiding the overhead caused by the retransmission of large packets by isolating the damage to a small portion of the payload.

C. Voice and Data Sessions

A data session is defined as the transmission of a packet train from the transmitter to the receiver. Although, a voice session will also involve the transfer of a packet train, due to delay constraints, it may not be desirable to negotiate for resources each time a burst occurs. We therefore negotiate for resources at the start of a voice session, and assume that transmission of voice packet trains is given priority over the data packets at the transmitter.

D. WLAN Topology

A single-hop, ad hoc, asynchronous DS-CDMA based multimedia WLAN with k terminals is considered. K is defined as the set of terminals in the network

$$K = \{1, 2, 3, K, k\}. \tag{2}$$

The terminals are assumed to be distributed randomly. Also, it is assumed that their positions are either fixed or slowly varying.[2] Thus, diversity techniques, such as multiple antennas, are used to compensate for the flat-slow fading of the indoor channel [14]. Thus the received signal strength is influenced mainly by path loss and shadow fading. We define a path loss matrix as

$$H = \{h_{ij}\} \qquad i, j \in K, \quad i \neq j \tag{3}$$

where h_{ij} is the path loss from terminal i to terminal j.

E. Spread-Spectrum Modeling

The chip rate for all terminals is fixed and the total bandwidth W, is used by all terminals. Additionally, we assume that all supported transmission bit rates $R_1, R_2, R_3, \cdots, R_n$

[1] The selective repeat ARQ is used only in cases where the ARQ is feasible. For instance, for voice applications the mini-packets that were received incorrectly may just be discarded.

[2] Specifically, it is assumed that the channel characteristics are more or less constant for the duration of the packet train.

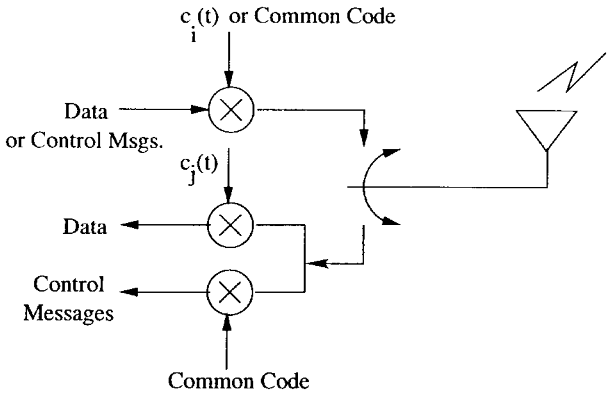

Fig. 1. Dual receiver architecture.

are multiples of the lowest rate, R_1, and that the bit rates are restricted to $R_i^{2(i-1)}$ bits/s $(i = 1, 2, \cdots, n)$, so that the processing gain varies by a factor of two. It is assumed that there is a set of spreading codes that may be used by the terminals

$$C = \{c_1, c_2, K, c_m\} \qquad m \gg k. \tag{4}$$

Given a set of spreading codes there must be a protocol which dictates how they are utilized. This is called the spreading code protocol and can be classified as common code, receiver or transmitter based and hybrid common transmitter and common receiver. For the system being presented, a dual receiver architecture with the hybrid common/receiver (C-R) code protocols is considered [15], [16]. This is illustrated in Fig. 1. The common code is essential to the operation of the resource negotiation protocol presented in the next chapter. It allows all terminals in the network to track network activity and thus implicitly gain knowledge about the state of the system. This is imperative if QoS guarantees are to be sustained. Moreover, although the dual receiver architecture does add additional complexity to the terminal hardware, it does allow terminals to track network activity while receiving. This again limits the loss of critical state information. Since each code in a CDMA system may be thought of as a channel, the common PN code forms a single channel. This is called the common control channel (CCCH), since it is used for control messages and is shared by all terminals. Similarly, each receiver code, C_i, forms a data channel, denoted as DCH_i.

Each terminal is equipped with two receivers and a transmitter. One receiver is always synchronized to the common code while the second receiver is synchronized to the unique receiver code assigned to each terminal. A terminal cannot transmit and receive simultaneously.[3] A terminal that is not transmitting, monitors the CCCH. Moreover, the dual receiver architecture allows a terminal to receive control and data messages simultaneously.

Specifically, let $K^t \subset K$, denote the set of terminals that are currently transmitting. Then, $K^c = (K^t)^c$ is the set of terminals monitoring the CCCH. Obviously, $K^t \cup K^c = \emptyset$. Moreover, $K^r \subseteq K^c$, defines the set of terminals that are currently receiving data frames. A session between terminals, $i \in K^t$, and $j \in K^r$, is denoted by $\{i, j\}$. Also, $K^a = K^t \cup K^r$, is defined as the set of currently active terminals. Moreover, it is assumed that a transmitter can communicate with only a single receiver at a time and vice versa.

[3]This is true of all single-channel radio networks. When a terminal is transmitting, any received signal is comparatively much weaker than the transmitted signal. Hence, the received signal is essentially blocked out.

The CCCH is assumed to have a slotted structure with each slot equal to the transmission time of a control message. All control messages are assumed to be of the same size. The data rate on the CCCH is fixed. All messages are transmitted on the CCCH with the same power, i.e., no power control is used. Moreover, it is assumed that control messages are transmitted with a high enough transmission power so that they can be received by all terminals K^t. Since the CCCH is a single channel, it is vulnerable to collisions. Due to the relatively light traffic on the CCCH, a simple multiple access scheme such as slotted ALOHA is proposed. Moreover, due to the critical nature of the control messages it is assumed that some form of forward error correction (FEC) is used for messages on the CCCH.

The QoS requirements for a session are expressed through the maximum packet error rate (PER) which can be mapped into an equivalent E_b/N_0 signal-to-interface ratio (SIR) requirement [17]. The matrix Γ is defined as

$$\Gamma = \{\gamma_{ij}\} \qquad i, j \in K, \quad i \neq j \tag{5}$$

where γ_{ij} represents the minimum SIR requirement for $\{i, j\}$. The transmission data rate requirement for $\{i, j\}$ represented by the matrix, which is defined as

$$\Psi = \{\psi_{ij}\} \qquad i, j \in K, \quad i \neq j \tag{6}$$

where ψ_{ij} is the minimum transmission data rate for $\{i, j\}$. The maximum physical transmission power for a terminal is given by Θ. The transmitted power allocated for a session $\{i, j\}$ is represented by the matrix P

$$P = \{p_{ij}\} \qquad i, j \in K, \quad i \neq j \tag{7}$$

where p_{ij} represents the transmitted power from terminal i to terminal j. The rate matrix R is defined similarly

$$R = \{r_{ij}\} \qquad i, j \in K, \quad i \neq j \tag{8}$$

where r_{ij} represents the allocated data rate for $\{i, j\}$. Background additive white Gaussian noise (AWGN) with one-sided power spectral density η_0 is assumed. Using the Gaussian approximation for the multiple access interference (MAI) [18], [19], the interference matrix Z is defined as

$$Z = \xi_{ij} = \left\{ \sum_{\{l,m\} \neq \{i,j\}} p_{lm} h_{ij} + \eta_o W \right\} \tag{9}$$

where ξ_{ij} is the interference experienced by $\{i, j\}$. The SIR matrix can be defined as

$$\Omega = \omega_{ij} = \left\{ \frac{W h_{ij} p_{ij}}{r_{ij} \xi_{ij}} \right\} \tag{10}$$

where ω_{ij} is the SIR for the $\{i, j\}$ session. Given that $\{i, j\}$ has been allocated a SIR of $\omega_{ij} > \gamma_{ij}$, the additional interference that can be sustained by $\{i, j\}$, is denoted by δ_{ij}. We thus have

$$\frac{W}{r_{ij}}\frac{h_{ij}p_{ij}}{(\xi_{ij}+\delta_{ij})}=\gamma_{ij}. \tag{11}$$

Solving for δ_{ij} the maximum sustainable interference (MSI) matrix Δ, is defined as

$$\Delta=\delta_{ij}=\left\{\frac{Wp_{ij}h_{ij}}{r_{ij}\gamma_{ij}}-\xi_{ij}\right\} \tag{12}$$

where δ_{ij} represents the maximum additional interference that the receiver j can sustain given that the interference is ξ_{ij} and $\{i, j\}$ has been assigned a transmission power of p_{ij} and a data rate of r_{ij}.

III. Distributed Resource Negotiation Protocol

Using the model outlined above, we now present a distributed resource allocation protocol known as distributed resource negotiation protocol (DRNP) that can be used to allocate resources on a per session basis.

A. Global Versus Incremental Resource Management

Resource management can be performed on a global or incremental (per session) basis. The global resource management (RM) entails renegotiation of resources each time a session is activated or leaves the system. In the incremental scheme, resources are allocated only once per session. This is analogous to the packing problem, where certain boxes of various sizes are to be fit into a container. Here the volume of the container is the resource that needs to be allocated. In the global scheme, boxes that are already packed can be moved around to make room for incoming boxes, while in the incremental scheme, boxes once packed are static. Incoming boxes then have to fit around the boxes already there. For instance, assume that $\{i, j\}$ has been allocated certain resources. Another session $\{m, l\}$ now requests resources from the network. In the global case, the network will allocate resources to $\{m, l\}$ as well as reallocate the resources allotted to $\{i, j\}$ so that some optimization criterion (such as the minimization of total transmission power, or maximization of total throughput) is met. However, in the incremental allocation case, the network allocates resources to $\{m, l\}$, while preserving the resources allotted to $\{i, j\}$.

Global RM has the advantage of being extremely flexible as it is able to reevaluate resource allocation decisions as needed. Although, this does allow it to utilize resources very efficiently, there is a price to be paid for this flexibility. Global RM schemes are highly computationally intensive and entail a lot of protocol overhead in a distributed environment as they require that the entire state of the system be known at each decision interval. This makes global RM schemes well suited to centralized implementations, such as in cellular systems, where the base station can monitor all activity and make reallocation decisions as required. There is another fundamental limitation to global allocation protocols. They require that a terminal be able to receive resource allocation messages while it is transmitting. This requires some form of duplexing[4] and is impossible in the single channel architecture being proposed. Although less efficient than global allocation schemes, incremental schemes are better suited to single channel based distributed implementation. DRNP is one such incremental resource allocation scheme.

However, incremental schemes do have the tendency of being unfair. For instance, continuing the example above, once $\{i, j\}$ is active, the $\{m, l\}$ session is essentially constrained by the resources allocated to $\{i, j\}$. In certain situations, this can seriously degrade the performance of the network. For instance, if $\{i, j\}$ is allocated the minimum required transmission power, the $\{m, l\}$ session will not be able to transmit, as the MAI introduced by $\{m, l\}$ will violate the QoS for $\{i, j\}$. This issue is analyzed in detail in the next section which deals specifically with resource allocation policies.

B. Resource Allocation List (RAL)

The resource allocation list is an extension of the network allocation vector (NAV) used in 802.11 WLAN's and the power constraint list (PCL) introduced by Whitehead [20]. Each terminal maintains a database which encodes its knowledge about other ongoing sessions in the network. For illustration purposes, let $\{i, j\}$ be a currently active session. The set of third party terminals with respect to $\{i, j\}$ can be defined as

$$K_{ij}=K-\{i\}-\{j\}. \tag{13}$$

Let K_{ij}^c be the set of third party terminals with respect to $\{i, j\}$ that can track the CCCH. A terminal $l\in K_{ij}^c$, encodes information about the $\{i, j\}$ session in its database RAL_l, as a record containing the following fields.

- The source (i) and destination (j) addresses for $\{i, j\}$. This can be thought of the index or key field of the database. All entries are accessed through this field.
- The estimated path loss to the source h_{li}, and destination h_{lj}.
- The maximum sustainable interference (MSI) for $\{i, j\}$, δ_{ij}.[5]
- The estimated duration for the $\{i, j\}$ session, τ_{ij}.[6]

RAL_l contains a similar record for every active session in the network. Using h_{lj} and δ_{ij}, l can determine its maximum transmission power such that the interference at j does not exceed δ_{ij}. This is given by

$$\mu_{lj}=\min\left\{\frac{\delta_{ij}}{h_{lj}},\Theta\right\}\qquad i\in K^t,\quad j\in K^r,\quad l\in K_{ij}^c. \tag{14}$$

We now define π_l as the maximum transmission power for l, such that the MSI constraints for all active sessions are preserved, as

$$\pi_l=\min_{\{x, y\}\in RAL_l}\{\mu_{ly}\}. \tag{15}$$

[4] There are two forms of duplexing, frequency division duplexing (FDD) and time division duplexing (TDD). Duplexing includes isolation of the send and receive channels either by separating them in frequency (FDD) or in time (TDD).

[5] This information is advertised by j and is explained in the next section.

[6] The receiver j, calculates the duration of the transmission from i, by using the size parameter in the RTS message and the allocated data rate. This is explained in detail in the following sections.

C. Data and Voice Session Negotiation

As stated earlier, voice and data sessions are treated differently when negotiating for resources. Data sessions negotiate on a per packet train basis while voice sessions are allocated resources once in the beginning of the session. Due to the difference in duration between a voice session (typically 90 s) and a data session 30 (typically <1 s), a voice session can potentially block many data sessions, even though there may not be any voice activity. We therefore assume that a transmitter can establish data sessions during the silent periods of the voice sessions. Hence a transmitter can potentially establish two sessions at a given time, although a receiver can only participate in a single session at any given time.

D. Control Message Formats

Control messages are broadcast on CCCH and are used to setup/tear down sessions, etc. The following convention is used in the presentation of the control message formats.

- Physical layer entities such as synchronization and error detection/correction fields have been excluded. Only the data structures specific to the DRNP are presented.
- Each message is prefixed by the source and destination addresses.
- Each message contains a unique entry identifying the message type and is denoted by the respective message name in the presentation of message formats that follows.

The RTS message is used by a transmitter, $i \in K^c$, to initiate a session with another terminal, $j \in RAL_i$. Its format is

$$RTS_{ij} = \{i, j, RTS, \{\gamma_{ij}, \psi_{ij}\}, \pi_i, \sigma_{ij}\}$$

where σ_{ij} is the size (in octets) of the packet train to be transmitted. The ESR message is used by the transmitter i to signify the end of a session

$$ESR_{ij} = \{i, j, ESR\}.$$

The CTS message is issued by the recipient j of the RTS message if it can support the QoS requested by j within the data rate and power constraints. Its format is

$$CTS_{ji} = \{j, i, CTS, \{\gamma_{ij}, \psi_{ij}\}, p_{ij}, r_{ij}, \tau_{ij}\}$$

where τ_{ij} is the duration of the session. The ESA_{ji} message is sent to the transmitter j, by the receiver i, in response to the ESR_{ij} message. Its format is

$$ESA_{ji} = \{j, i, ESA, NACK\}$$

where $NACK$ contains information about the mini-packets that were received incorrectly. The primary reject (PREJ) is issued by the receiver j if it is unable to support the QoS requested by the transmitter, i. The format of the $PREJ_{ji}$ message is

$$PREJ_{ji} = \{j, i, PREJ, \{\gamma_{ij}, \psi_{ij}\}, \pi_i\}.$$

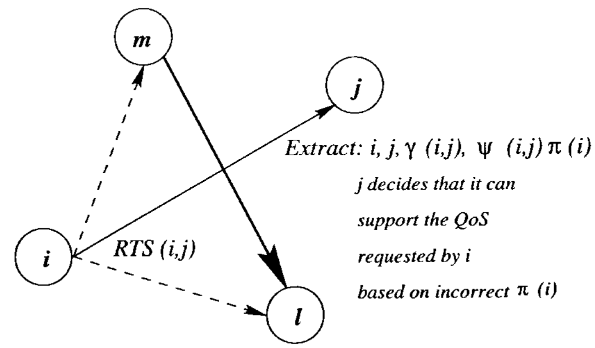

Fig. 2. Out of data RAL_i leads to an incorrect value for π_i to be advertised.

1) Secondary Reject (SREJ): Recall that a terminal cannot transmit or receive simultaneously. This implies that a terminal is essentially deaf while it is transmitting. Thus, even if no other form of message loss occurs, a terminal's RAL will not be updated while it is transmitting. An out of date RAL can cause a receiver to allocate resources to a session that would degrade the QoS of the currently active sessions in the network. This scenario is illustrated in Fig. 2, where i has out of date information. When it tries to establish a session with j, an incorrect value for π_i will be advertised in RTS_{ij}. The receiver j has no way of validating π_i and may allocate a transmission power that causes ω_{ml} to fall below γ_{ml}.

The secondary reject (SREJ) mechanism allows third-party receivers[7] to interrupt the setup of sessions that violate their QoS guarantees. Continuing the scenario outlined above, the dual receiver architecture allows l to track the CCCH while it is receiving data. If the MAI introduced by $\{i, j\}$ causes ω_{ml} to drop below γ_{ml}, l issues a $SREJ_{li}$ message as

$$\mathrm{SREJ}_{li} = \{l, i, \delta_{ml}\} \qquad m \in K^t.$$

It is imperative that the SREJ message reach i before it starts transmitting. To help accomplish this, i waits for a certain number of slots (τ_{DATA}) before initiating transmission. The third party terminals wishing to transmit a SREJ to i, randomly choose a slot for transmission of the SREJ message.[8] However, the SREJ mechanism does suffer from certain fundamental limitations.

- A SREJ message may arrive after the timeout period. This is a drawback for all timeout-based synchronization schemes and is unavoidable.
- The SREJ messages are themselves vulnerable to loss through collisions with other traffic on the CCCH. Standard collision resolutions mechanisms involving ARQ are not feasible in this case due to the delay overhead they cause.

In principle, the probability that a SREJ message is lost can be minimized by making τ_{DATA} as large as possible. Realistically, the gain in terms of QoS guarantees is far outweighed by the overhead through delay caused by excessively large values

[7] As opposed to PREJ, where the intended receiver denies access to the network.

[8] Note that there is an implicit assumption here that transmission delay is much greater than propagation delay. This is quite valid in WLAN environments as terminals are confined to a limited geographical area.

for τ_{DATA}. Ideally, the timeout period should be varied with network load, although this again might entail too much computational overhead, although it does form an interesting avenue for research. The effect of message loss involving SREJ is investigated later.

The SREJ mechanism introduces another tradeoff by requiring that an active receiver transmit a SREJ. This entails that the incoming data stream be disrupted for the duration of the transmission. Although, the damage is limited to a few mini-packets, depending on the state of the ongoing session, transmitting a SREJ may cause more harm than good. This is especially true if the session in question is near completion. DRNP in its current form uses a hard decision metric to decide whether a SREJ message should be sent out. Nevertheless, it is clear from the above discussion that DRNP's performance can be improved by the following.

- Using some form of soft decision metric, i.e., the severity of the damage caused by the errant session must be taken into account. If the drop in SIR is minimal, it may be preferable to allow the session in question to proceed.
- Taking the status of the current data transfer into consideration. If near completion, the transmission of the SREJ may cause more damage than the incoming session.

2) Update Maximum Sustainable Interference: The update maximum sustainable interference (UPD_MSI) message is broadcast by all third party receivers, $l \in K_{ij}^r$ when the CST_{ij} message is received. This message updates the interference margin ml where $m K^t$, due to the addition of $\{i, j\}$ to the network. Its format is

$$\mathrm{UDP\_MSI}_{lx} = \{l, x, \{m, l\}, \delta_{ml}\}$$

where x is a network specific broadcast address.[9] Since UPD_MSI messages are transmitted by one or more third party receivers, they are particularly vulnerable to loss through collision. Thus the same random slot mechanism used for SREJ is applicable here as well. In fact, UPD_MSI and SREJ messages are extremely similar, i.e., they are both used to update third party RAL's. Thus, in the current incarnation of DRNP, the time window τ_{DATA} is shared by both SREJ and UPD_MSI messages.

E. Functional Specification[10]

Each terminal executes the following processes: 1) dispatch; 2) transmitter; 3) receiver; and 4) third party, that communicate through the RAL and a shared *STATE* variable. Fig. 3 is a block diagram showing the highest level of process interaction. All messages are received by the dispatch process and are routed to the appropriate processes.

The *STATE* variable is a global variable used as a form of interprocess communication (IPC) by all the processes. At any given time, it contains the current status of the terminal, which can be one of the following. 1) **IDLE** (not involved in a session); 2) **TX** (has a packet train to transmit); 3) **RX** (after receiving an RTS message); 4) **WAIT_FOR_CTS_PREJ** (issued a RTS message and waiting for a response from the receiver); 5) **WAIT_FOR_SREJ** (CTS message has been received); and 6) **WAIT_FOR_ESA** (after issuing an ESR). Each "*wait-for*" state has an associated timeout, denoted by the τ variable suffixed with the name of the corresponding state. For instance, the timeout variable for the **WAIT_FOR_SREJ** state is denoted by $\tau_{\mathrm{WAIT\_FOR\_SREJ}}$. For illustration purposes, we assume that terminal $i \in K^c$ is attempting to establish a session with terminal j. Terminals $l, m \in K_{ij}$ are two third-party terminals with respect to $\{i, j\}$.

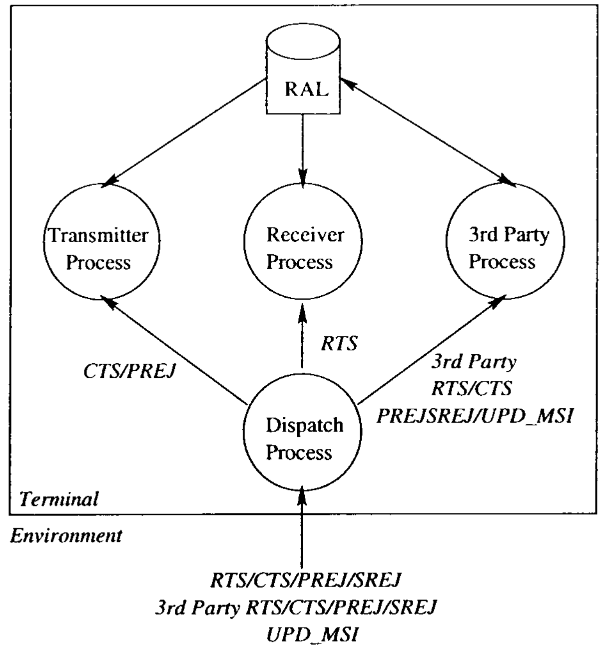

Fig. 3. Top level process architecture.

The dispatch process is essentially a message router that receives messages from other terminals, and depending on the state of the terminal, routes them to the appropriate processes. It is designed to relay information to the other processes only when they require it. It makes the presentation and implementation of the other processes in DRNP more manageable as these processes do not have to deal with error conditions caused by the arrival of an unexpected message. The dispatch process algorithm is presented as Algorithm A in Fig. 4.[11] A flow diagram depicting the dispatch process algorithm is presented in Fig. 5.

1) Transmitter Process: The transmitter process is triggered by the arrival of a packet train for transmission and is responsible for the setting up tearing down sessions originated by a terminal. The backoff count parameter β dictates the maximum number of times a session can request resources from network before being discarded. If the terminal is currently idle, i.e., STATE $=$ IDLE, the transmitter process initiates session setup by issuing a RTS message to the receiver. If the receiver responds with a CTS message, the transmitter waits for τ_{DATA} slots before assuming that the

[9] This address may be similar to Ethernet which uses a MAC level broadcast address to signify all tranceivers in a segment.

[10] Presenting the protocol in this way serves a dual purpose; it simplifies presentation as well as proposed a method of implementation.

[11] The algorithm for the various processes are presented in *pseudo*-C.

```
Dispatch Process
Loop (Forever)
 if (message.destination == my address)
 {
  switch (message.type)
   case (RTS):
    if (STATE == IDLE)
     STATE = RX;
     Transmitter Process (message);
   case (CTS):
    if (STATE == WAIT_FOR_CTS)
     Receiver Process(message)
   case (PREJ):
    if (STATE == WAIT_FOR_CTS_PREJ);
     Transmitter Process (message);
   case (SREJ):
    if (STATE == WAIT_FOR_SREJ)
     Transmitter Process (message);
   case (ESR):
     if (STATE == RX)
      Receiver Process (message);
   case (ESA):
    if (STATE == WAIT_FOR_ESA)
     Transmitter Process (message);
   default:
    Delete Message(message)
 }
 else
  3rd Party Process(message)
```

Fig. 4. Algorithm for dispatch process (algorithm A).

session setup has been successful. If an SREJ message is received within τ_{DATA}, the transmitter goes into backoff mode and attempts to setup a session again at a later time. Session tear down is initiated by the transmitter i by issuing the ESR_{ij} message to the receiver j and then waiting for the ESA_{ji} message from the receiver. Algorithm B (see Fig. 6) illustrates the transmitter algorithm for terminal $i \in K^c$, wanting to initiate a session with terminal j. Additionally, it is assumed that $m, l \in K_{ij}$. The corresponding flow diagram is presented in Figs. 7 and 8.

2) Receiver Process: The receiver process is triggered by the arrival of a RTS message. It is responsible for determining whether the QoS requested by the transmitter can be supported. If it can, the receiver issues a CTS message containing the allocated transmission power and date rate to the transmitter. Otherwise a PREJ message is issued. If an ESR message is received from the transmitter, an ESA message is issued. The receiver process algorithm is presented in Algorithm C (see Fig. 9). Fig. 10 illustrates the flow diagram for the receiver algorithm as executed by terminal j.

3) Third-Party Process: Third-party terminals monitor the RTS/CTS/PREJ and ESR/ESA handshakes relating to $\{i, j\}$ and use the information to update their RAL. The third party process is the all purpose message handler for all third party messages and is responsible for updating the RAL. It is also responsible for implementing the timeout mechanism for all third party sessions contained in a terminal's RAL. In our example, this is denoted by $\tau_{i,j}, \{i,j\} \in RAL_e$. Thus if no ESA is received within the timeout period, the corresponding third party session is deemed complete and is removed from the RAL. This is necessary as loss of ESA messages can result in resources being allocated even when the sessions using them have left the system. In the following section it is assumed that $l \in K_{ij}$. Algorithm D in Fig. 11 depicts the third party algorithm as executed by l. Fig. 12 illustrates the flow diagram for the third party algorithm as executed by terminal l.

F. Protocol Validation of DRNP

Spin is a protocol modeling tool developed at Bell Labs [21] that can be used to simulate communication protocols. A model of DRNP was created using the *Promela* language and simulated using the Spin package. The model was used for conformance testing as well protocol validation using lossy channels. The results indicated that DRNP was free of deadlock situations and could recover from almost every conceivable combination of message loss on the CCCH.

IV. Resource Allocation Policies

A. MSI

The MSI is a local parameter advertised by a session, that effects the global performance of the network. In general, the MSI for a session $\{i, j\}$, ij, is given by

$$\delta_{ij} = \frac{W p_{ij} h_{ij}}{r_{ij}\gamma_{ij}} - \sum_{\{m,l\}\neq\{i,j\}} p_{ml}h_{mj} - \eta_0 W. \tag{16}$$

As can be seen from (16), δ_{ij} is a function of the SIR allocated for $\{i, j\}$ as well as the current interference levels at j. Moreover, the MSI advertised by a particular session determines the maximum transmission power of other terminals. Hence, the MSI reflects the fact that both transmission power and interference levels within the network affect the chances that a session setup will succeed.

To demonstrate the dynamics between MSI and network performance we propose a hypothetical network with a large number of terminals. We assume that session requests arrive at each of the terminals at a certain rate. Moreover, sessions that are successfully setup never leave the system. For illustration purposes let $\{i, j\}$ be the first session to be setup. This session advertises a nonzero value for δ_{ij} and since there are no other sessions, the network is essentially limited by the AWGN in the network. As other sessions are setup, the MAI at j (the second term on the right-hand side of (16) increases until

$$\frac{W p_{ij} h_{ij}}{r_{ij}\gamma_{ij}} = \sum_{\{m,l\}\neq\{i,j\}} p_{ml}h_{mj} - \eta_0 W. \tag{17}$$

At this point, the MSI advertised by $\{i, j\}$ will be zero. Since no sessions ever leave the network, all other sessions arriving later will be blocked. Although unrealistic, the above scenario does illustrate the strong correlation between the MSI parameter and global network performance. We now turn our attention to the relationship between MSI and resource allocation policies. In doing so we hope to gain insight into the effect that various resource allocation schemes have on DRNP performance.

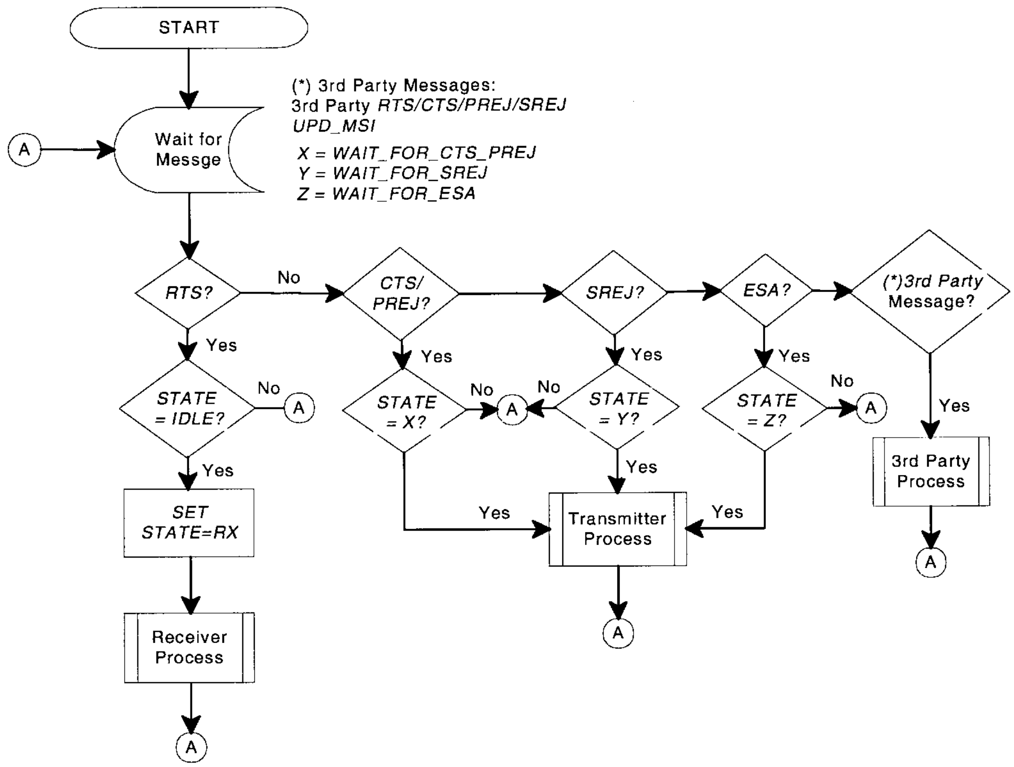

Fig. 5. Flow diagram for dispatch process algorithm.

For illustration purposes, consider a four terminal network (Fig. 13). The following is assumed.

- All of the sessions in the four-terminal network are of the same type. They all have a minimum SIR requirement of γ and minimum data rate requirement of ψ.
- Moreover, $K^t = \{i\}, \{l\}$, and $K^r = \{j\}, \{m\}$. $\{i, j\}$ and $\{l, m\}$ are the only source-destination (S-D) pairs in the network.
- The system bandwidth is W and the background AWGN has a one-sided power spectral density of η_0.

B. Static Network

Consider a static network which is a snapshot of the system where

- $\{i, j\}$ is an ongoing session and has been allocated a transmission power of p_{ij} and data rate r_{ij} by a resource allocation scheme;
- as part of the session setup process, terminal j has advertised a maximum tolerable interference level of ij. This too is resource allocation policy dependent. As a result, terminals $l, m \in K_{ij}$, have set their maximum transmission powers to l and m, respectively;
- terminal m wants to establish a session with terminal l.

1) Minimum Power Allocation: The minimum power allocation scheme is one of the most obvious, and widely utilized resource allocation schemes (especially in centralized networks [8]). This scheme emphasizes conservation of transmission power and only the minimum transmission power required to achieve γ_{ij} and ψ_{ij} is allocated and is, in general, given by

$$p_{ij}^{\min} = \frac{\gamma_{ij}\psi_{ij}\left(\sum_{\{m,l\}\neq\{i,j\}} p_{ml}h_{mj} + \eta_0 W\right)}{Wh_{ij}}. \tag{18}$$

Since only the minimum required SIR is achieved at the receiver, MSI advertised by $\{i, j\}$, δ_{ij}, will be zero. Thus, $\pi_m = 0$. As a result $\{m, l\}$ will not be successful for the duration of $\{i, j\}$. Except for the explicit QoS guarantee, the minimum power allocation scheme functions exactly like the RTS/CTS bandwidth reservation mechanism in 802.11. It essentially reserves network resources for the $\{i, j\}$ session and prevents any other sessions from being established.

Moreover, since only the minimum data rate is allocated, for a given payload size, the transmission time for $\{i, j\}$ is maximized. This implies that not only will the incoming $\{m, ll\}$ session be blocked for the duration of $\{i, j\}$, but that the blocking duration will also be maximized. The above discussion implies that performance may be improved by either:

- decreasing the time duration of the $\{i, j\}$ session;
- increasing the value for MSI advertised by $\{i, j\}$. This will enable other terminals to communicate while still satisfying the MSI constraint imposed by $\{i, j\}$.

2) Maximum Rate Allocation: The maximum rate allocation scheme allocates, if possible, higher data rates than the minimum required at the expense of additional MAI within the network. The maximum rate (MR) scheme is similar to the minimum power (MP) scheme in the sense that only the minimum required SIR is allocated to any session. Any extra

```
Process Transmitter
Retry Count = 0;
if (data to transmit)
{
 while (Retry Count < β)
 {
  STATE = TX;
  Retry Count = Retry Count + 1;
  if ({x,j} ∈ RAL_i ∀ x∈ K^t)
  {
   Wait for τ_BUSY;
   continue;
  }
  else
  {
   Transmit (RTS_ij);
   STATE = WAIT_FOR_CTS_PREJ;
   Set Timeout (τ_WAIT_FOR_CTS_PREJ);
   if (RAL_i is updated) and (τ_WAIT-FOR-CTS-PREJ not expired)
    continue;
   else if (τ_WAIT-FOR-CTS-PREJ expires)
   continue;
   else
   {
    switch (message.type)
    {
     case (PREJ):
      continue;
     case (CTS):
      Extract p_ij and r_ij.
      STATE = WAIT_FOR_SREJ;
      Wait for τ_DATA slots;
      if (message.type == SREJ) and (τ_DATA not expired)
      {
       Extract m,l,δ_ml
       Estimate path loss h_li
       Update RAL_i
       continue;
      }
      else
      {
       Transmit DATA;
       Transmit ESR_ij
       Wait for ESA_ji;
    }
   }
  }
 }
}
```

Fig. 6. Algorithm for transmitter process as executed by i (Algorithm B).

transmitted power, over and above that required to support the minimum data rate and SIR, is allocated to increase the data rate of the session which is given by

$$r_{ij}^{\max} = \frac{W h_{ij} \pi_i}{\gamma_{ij}\left(\sum_{\{m,l\}\neq\{i,j\}} p_{ml} h_{mj} + \eta_0 W\right)}. \qquad (19)$$

The effect of the $\{i, j\}$ session on the static network as a whole is also similar. Since, δ_{ij} is set to zero, then as in the previous case, the network will allow only one session to be active at a time.

However, since a higher data rate is allocated to the $\{i, j\}$ session, for a given session size, the duration of the $\{i, j\}$ transmission will be less than in the MP allocation case. Hence,

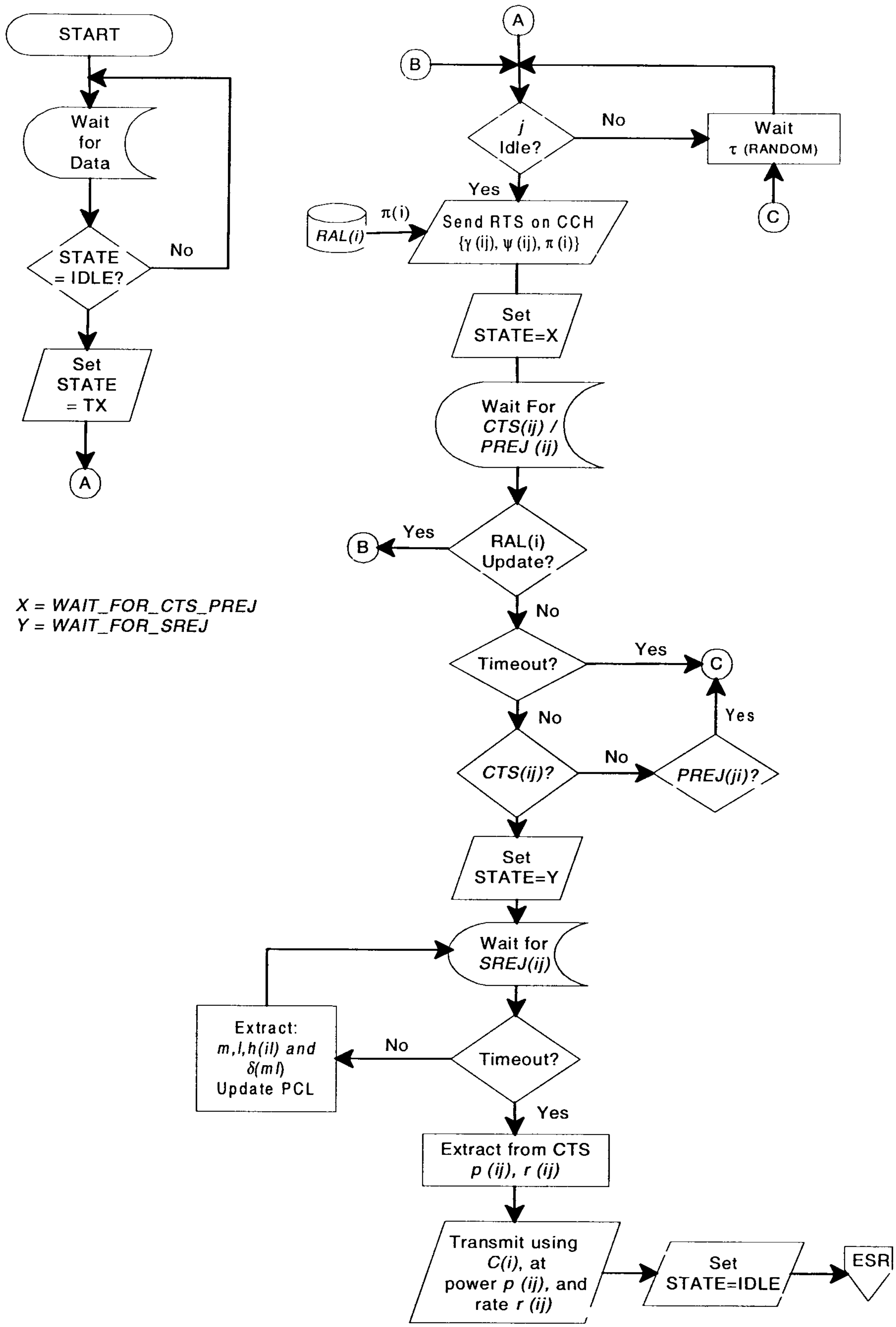

Fig. 7. Flow diagram for transmitter process algorithm as executed by i.

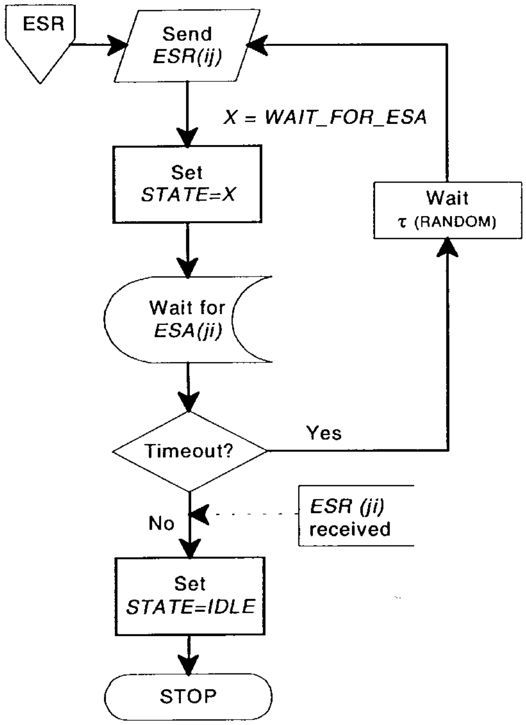

Fig. 8. Flow diagram for transmitter process algorithm as executed by i (Continued).

the fact that this scheme introduces additional MAI into the network is inconsequential as it is the MSI advertised by $\{i, j\}$, δ_{ij}, that is the limiting factor, i.e., no other session will be successful as the transmission power of the other terminals is constrained to zero, irrespective of the MAI.

3) Maximum SIR Allocation: In the maximization of SIR allocation policy, any extra power ration is used to increase the QoS of the session. The SIR at the receiver, ω_{ij}, can be maximized by allocating a transmission power of π_i and data rate ψ_{ij} and is then given by

$$\omega_{ij}^{\max} = \frac{W h_{ij} \pi_i}{\psi_{ij} \left(\sum_{\{m, l\} \neq \{i, j\}} h_{mj} p_{ml} + \eta_0 W \right)}. \tag{20}$$

As opposed to the MP or MR policy δ_{ij} and as a result π_m is may be nonzero. Thus, if terminals l and m are sufficiently separated from terminals i and j,[12] respectively, then $\{m, l\}$ may be successfully setup. The maximum SIR (MaxSir) allocation scheme embodies the essence of CDMA, (i.e., the possibility of multiple simultaneous transmissions), while preserving the QoS guarantees to each active session. Moreover, although ω_{ij} is always guaranteed to be greater than or equal to γ_{ij}, the introduction of $\{m, l\}$ into the network causes ω_{ij} to decrease. The variability in the QoS for currently active sessions is the price to be paid for the increased network throughput.

[12] Note that the success of a session setup is affected by both the maximum transmission power of the transmitter as well as the interference levels at the receiver.

```
Receiver Process
 switch (message.type)
 {
  case (RTS):
   Estimate Path Loss h_ij;
   Estimate MAI ξ_ij;
  if (ω_ij ≥ λ_ij)
   Transmit(CTS_ji)
   STATE = WAIT_FOR_SREJ;
  else
   Transmit(PREJ_ji)
   STATE = IDLE;
  case (ESR):
   Transmit ESA_ji
   STATE = IDLE;
 }
```

Fig. 9. Algorithm for receiver process as executed by j (Algorithm C).

C. Dynamic K-Terminal Network

The discussion presented previously in the context of the four terminal network can be readily extended to a general K-terminal where session requests arrive at a certain rate. Here too, network throughput is strongly correlated to the MSI advertised by each successful setup or equivalently the maximum transmission power allocated to an incoming session. Over the long term, network throughput is dependent on the average MSI advertised by each session. This local parameter has global ramifications as it determines the transmission powers for all sessions arriving within the duration of the session in question. We investigate the effect of average MSI on network throughput in detail through simulations in the following sections.

D. Message Loss in DRNP

In distributed message-based systems, message loss is the most common error condition. In the case of wireless systems, message loss is most often caused by collisions on the shared channel. In our case this is the CCCH. We have shown through exhaustive reachability analysis that DRNP can recover from all error conditions. However, message loss does lead to performance degradation and in some cases loss of QoS guarantees. In this section, we qualitatively analyze the effect of control message loss on the performance of the network. Although this is highly resource allocation policy dependent, it is instructive to at least isolate the most troublesome scenarios. In DRNP message loss effects the third-party terminals as well as the S-D pair for a particular session. As such, we investigate the effect of message loss in each phase of the protocol on the transmitter, receiver, and third-party terminals.

1) RTS/CTS (Session Setup): The effect of losing a RTS or CTS message on network performance is the same for both the MP and MR schemes. The loss of either a RTS or CTS message is detected at the transmitter through a timeout mechanism. Since an exponential backoff scheme is used, each loss of either message leads to exponentially increasing delays in session setup. As far as the third-party terminals are concerned, even if they fail to receive the RTS message sent out by the transmitter, as long as they successfully receive

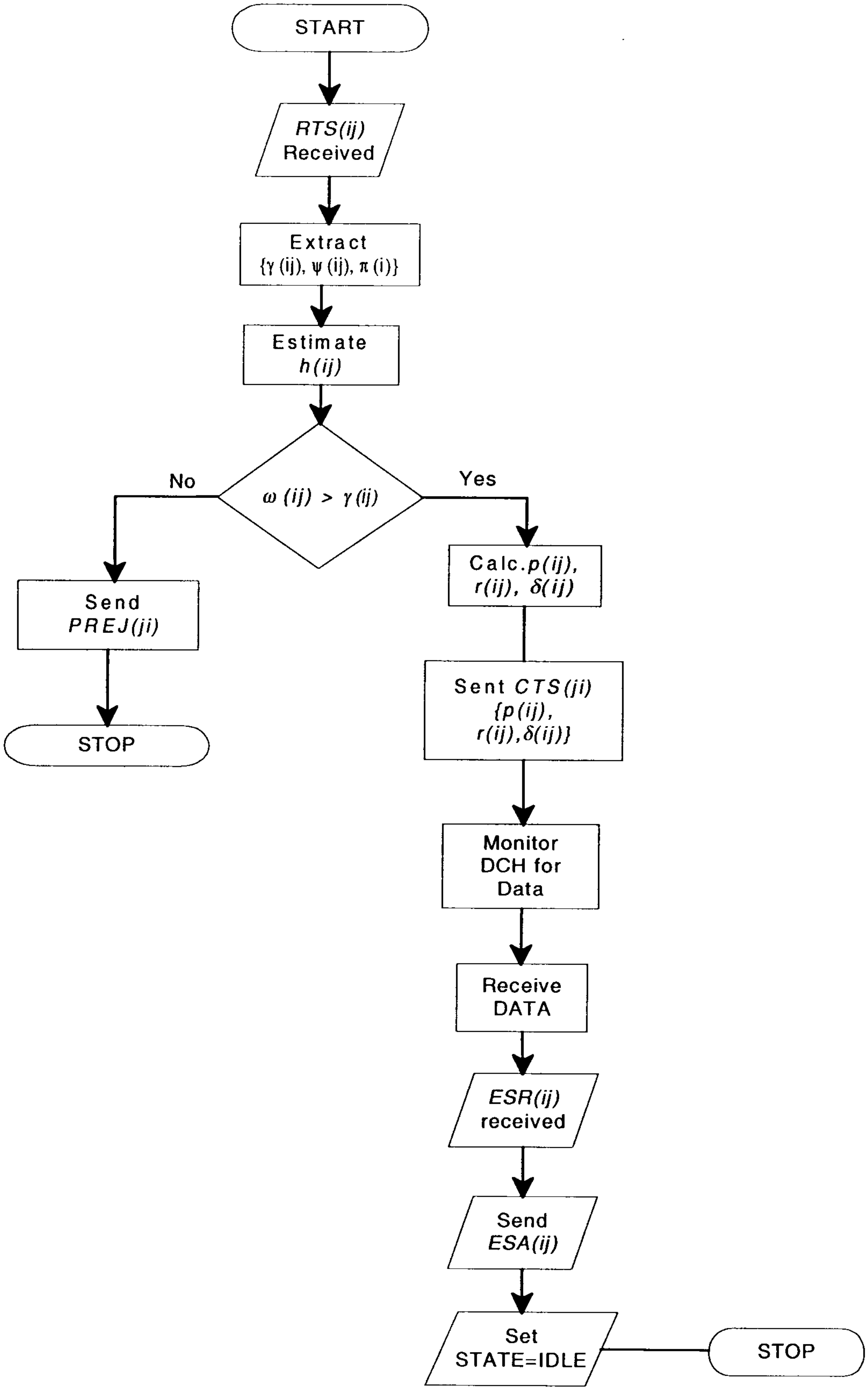

Fig. 10. Flow diagram for receiver process as executed by j.

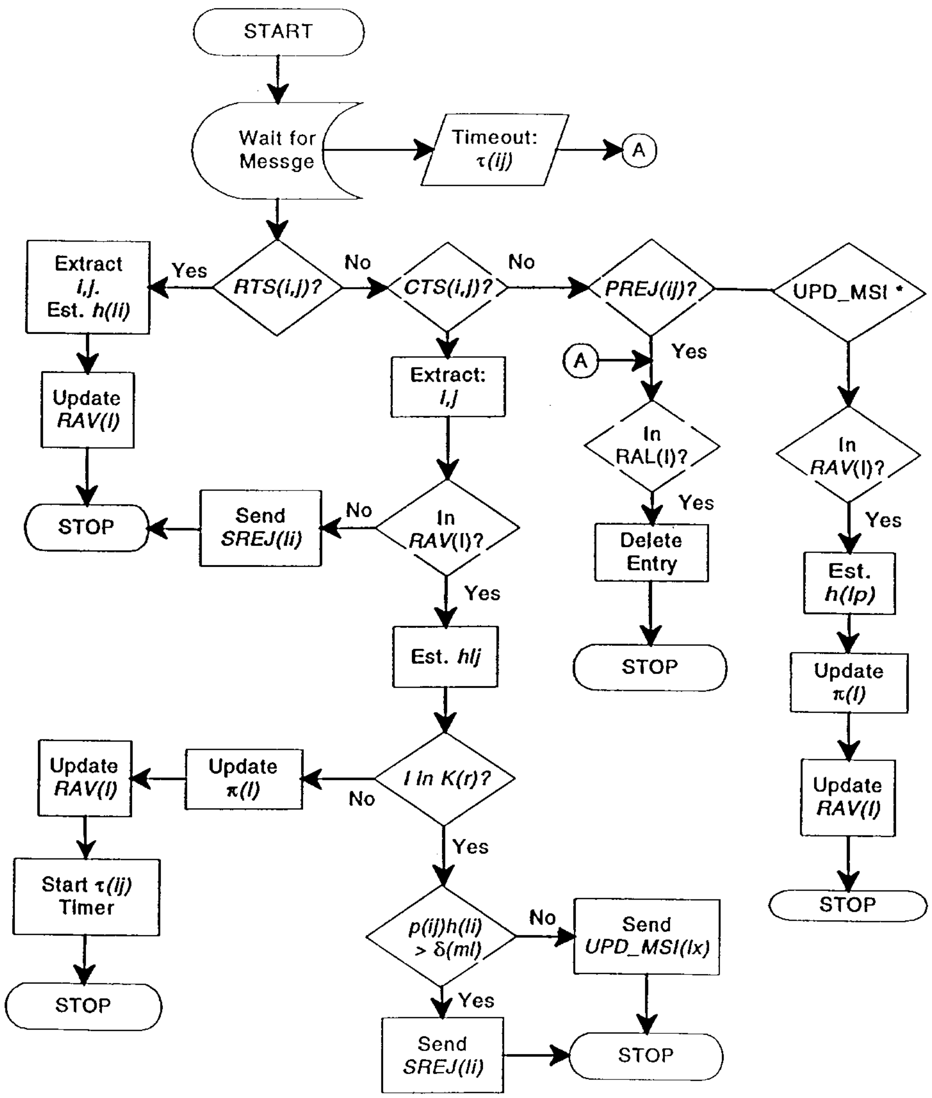

Fig. 11. Algorithm for the third-party process as executed by l (Algorithm D).

the CTS message, they will be able to update their respective maximum transmission powers.

However, if a third-party terminal is a third-party receiver (only in the MaxSIR allocation case) and fails to receive the RTS message, it cannot determine the path loss from itself to the transmitter. This entails that the third-party terminal cannot determine if the session being setup will cause its SIR to drop below the negotiated minimum. In this case the affected third-party receivers will opt to issue a SREJ to delay session setup and update the offending transmitter's RAL. However, the tradeoffs regarding SREJ discussed previously apply.

2) Update MSI (UPD_MSI): UPD_MSI messages are critical in maintaining QoS guarantees in DRNP, but are only necessary when the MaxSIR policy is used. In the MP and MR case, the MSI advertised by the CTS message during session setup is zero and is never updated.

3) SREJ: As with UPD_MSI messages, SREJ messages are expected to be issued by third party receivers with higher frequency when the MaxSIR allocation scheme is used. When MP and MR schemes are used, the only case where a SREJ will be issued is if one or more third party terminals fail to receive the RTS/CTS exchange during session setup of a particular session. Even in this case, only a single SREJ needs to be issued (no UPD_MSI messages) and thus the possibility of it colliding with another message are extremely unlikely. Moreover, since the only likely candidates for collision with the SREJ message are session setup/tear down messages, QoS guarantees are still maintained as the incoming session setup is interrupted by the collision.

4) ESR/ESA (Session End): Contrary to the discussion above, loss of session tear down encompasses one scenario where the MP/MR schemes may actually be worse than the MaxSIR scheme. Since the former advertise a MSI of zero, loss of the ESR/ESA messages by the third party terminals essentially blocks off any incoming sessions, until the respective time outs occur. Although the above scenario is also possible with the MaxSIR case, it is certainly less frequent.

It is clear from the above discussion that there is a tradeoff involved in using the MaxSIR scheme. Although it does allow

Process 3rd Party
switch (*message.type*)
{
 case (RTS_{ij}):
  Extract i, j;
  Estimate path loss h_{il};
  $RAL_l = RAL_l \cup \{i,j\}$;
 case ($PREJ_{ji}$):
  $RAL_l = RAL_l - \{i,j\}$;
 case (CTS_{ji}):
  Extract p_{ij}, r_{ij}, δ_{ij} and τ_{ij};
  Estimate path loss h_{lj};
  Start τ_{ij} timer;
  if ($l \notin K^r$)
  {
   Calculate μ_{lj};
   Update π_l;
  }
  else if ($l \in K^r$) and ($p_{ij}h_{il} \leq \delta_{ml}$)
  {
   $\delta_{ml} = \delta_{ml} - p_{ij}h_{il}$;
   Transmit UPD_MSI_{lx};
  }
  else if ($l \in K^r$) and ($p_{ij}h_{il} \leq \delta_{ml}$)
  {
   Pick slot 1 to τ_{DATA}; Transmit $SREJ_{li}$;
  }
 case (ESR_{ij}):
  Extract *i, j*;
 case (ESA_{ji}):
  $RAL_l = RAL_l - \{i,j\}$;
}
if (τ_{ij} has expired) **and** (ESRij or ESAji) is not received
 $RAL_l = RAL_l - \{i,j\}$;

Fig. 12. Flow diagram for the third-party process as executed by l.

multiple simultaneous sessions, the SIR allocated to each session can fluctuate as sessions arrive and leave the network. Moreover, due to the presence of multiple sessions, there is a higher risk of losing control messages due to collisions on the CCCH. In the case of MaxSIR resource allocation, this can lead to loss of QoS guarantees. The extent of this phenomenon is investigated in the next section through simulations.

V. Performance Analysis of DRNP

Session level simulations are used to: 1) verify the feasibility of DRNP in a realistic WLAN environment and 2) study the effect of message loss on the performance of DRNP.

A. Simulation Model

Modeling of wireless networks essentially falls into two categories: 1) simulating the physical radio channel which in our case also includes spread spectrum modeling and 2) network level modeling which involves modeling network topology as well as traffic patterns and traffic characteristics. For simulation purposes, the path loss (in decibels) is given by [21]

$$g(d) = L_o + 10 \cdot \alpha_0 \cdot \log\left(\frac{\lambda}{4\pi d_{\mathrm{ref}}}\right) + 10 \cdot \alpha \cdot \log\left(\frac{d}{d_{\mathrm{ref}}}\right) + X_\sigma \tag{21}$$

where

- L_0 antenna gain/loss;
- λ wavelength;
- d_{ref} reference distance from transmitter;
- d transmitter–receiver distance;
- a_0 path loss exponent below d_{ref}, usually within line of sight (LOS);
- a path loss exponent above d_{ref}; usually OBS (obstructed).
- X_σ represents the shadow fading component and is a Gaussian random variable with a standard deviation of σ.

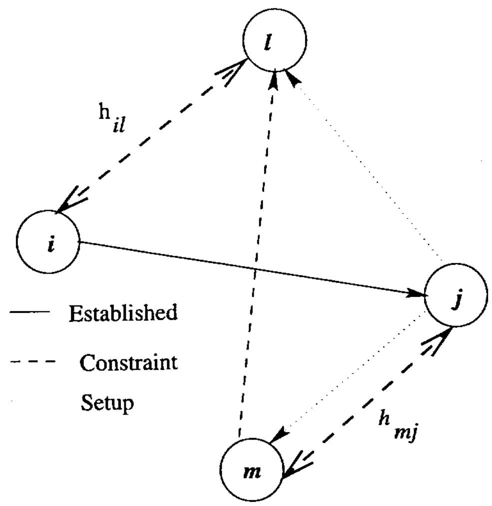

Fig. 13. A four-terminal WLAN.

TABLE II
PATH LOSS PARAMETERS USED IN SIMULATIONS

Parameter	Value
L_0	-2 dB
f_c	2.4 GHz, λ = 125 mm
d_{ref}	10 m
α_0	2
α	3
X_σ	4 dB

The values of the above parameters used in the simulation are presented in Table II. A carrier frequency of 2.4 GHz is chosen as most current WLAN's operate in this unlicensed ISM band. Moreover, it is assumed that LOS is present for T-R separations that are less than 10 m. As such, a path loss exponent of two is chosen for T-R separations within 10 m. For T-R separations greater than 10 m, an obstructed channel is assumed with a path loss exponent of three. In addition, the shadow fading component is represented by a Gaussian random variable with a standard deviation of four.

1) Spread-Spectrum Modeling: The spread spectrum parameters used in the simulations are shown in Table III. The system bandwidth W is set to 50 MHz. Since the maximum possible data rate is 1 Mbit/s the minimum processing gain is therefore 50. Under these circumstances the standard Gaussian approximation is quite adequate for simulation purposes [22]. Moreover, it is important to note that DRNP is essentially independent of the model chosen. In an actual WLAN, DRNP will rely on empirical estimates of the MAI. However, if more precise modeling of DRNP performance is required (especially when estimating performance in actual WLAN environments), other approximations such as the improved Gaussian approximation may be more appropriate.[13] Also note that the physical constraint on the maximum transmission power is set to an extremely high value. This is to ensure that all terminals are within range and any blocking that occurs is caused by DRNP power constraints.

TABLE III
SPREAD SPECTRUM PARAMETERS USED IN SIMULATIONS

Parameter	Value
W	50 MHz
Max Data Rate	1 Mbits/sec
Max Tx Power (Θ)	1000.0

TABLE IV
QOS PARAMETERS FOR VOICE (V) AND DATA (D) SESSIONS USED IN SIMULATIONS

Parameter	Minimum Value	Max. Allowed
SIR (γ) (dB)	7.0 (V) 10.0 (D)	10.0 (V) 16.0 (D)
Data Rate (Kbits/sec)	64.0 (V) 64.0 (D)	256 (V) 256 (D)

Terminals are assumed to be distributed uniformly in a two dimensional 50 m $\times$ 50 m grid.

2) Traffic Modeling: Two classes of traffic characterized by their disparate QoS requirements were simulated. Nonreal time data traffic was modeled using TCP/RPC model proposed by Anderlind [23]. This model effectively encompasses various TCP characteristics such as slow start and variable packet sizes. The T4 parameter, which defines the average time between TCP bursts, was used to vary network wide load. Voice sessions are assumed to be Poisson arrivals. A standard two state ON-OFF model with a voice activity factor of 0.4 and average burst length of 500 ms was used to model voice traffic. Furthermore, voice sessions are limited to 20% of the overall network load.

The QoS parameters for the voice and data sessions are summarized in Table IV. To ensure that simulation runs are comparable across resource allocation policies, the maximum possible data rate and SIR that can be allocated to a session are limited to a multiple (in this case four) of the minimum data rate and SIR.

The blocking rate is essentially a throughput measure and represents those sessions that have exceeded their backoff count and are deemed lost due to the receiver not being available. The loss of QoS (QoSLoss) parameter only applies to real mode simulations and denotes the percentage of sessions that are successfully set up but lose their QoS guarantees (their SIR falls below the negotiated minimum).

Using the above models, a custom network simulator was used to analyze the performance of DRNP under a variety of scenarios. The simulations were performed in two modes, genie and real. The genie mode simulations consider an idealized DRNP network where no message loss occurs. It represents the best case scenario and is a useful gauge of performance degradation when message loss is considered.

[13] The computational complexity of using the improved Gaussian approximation, especially in long simulations proved to be prohibitive.

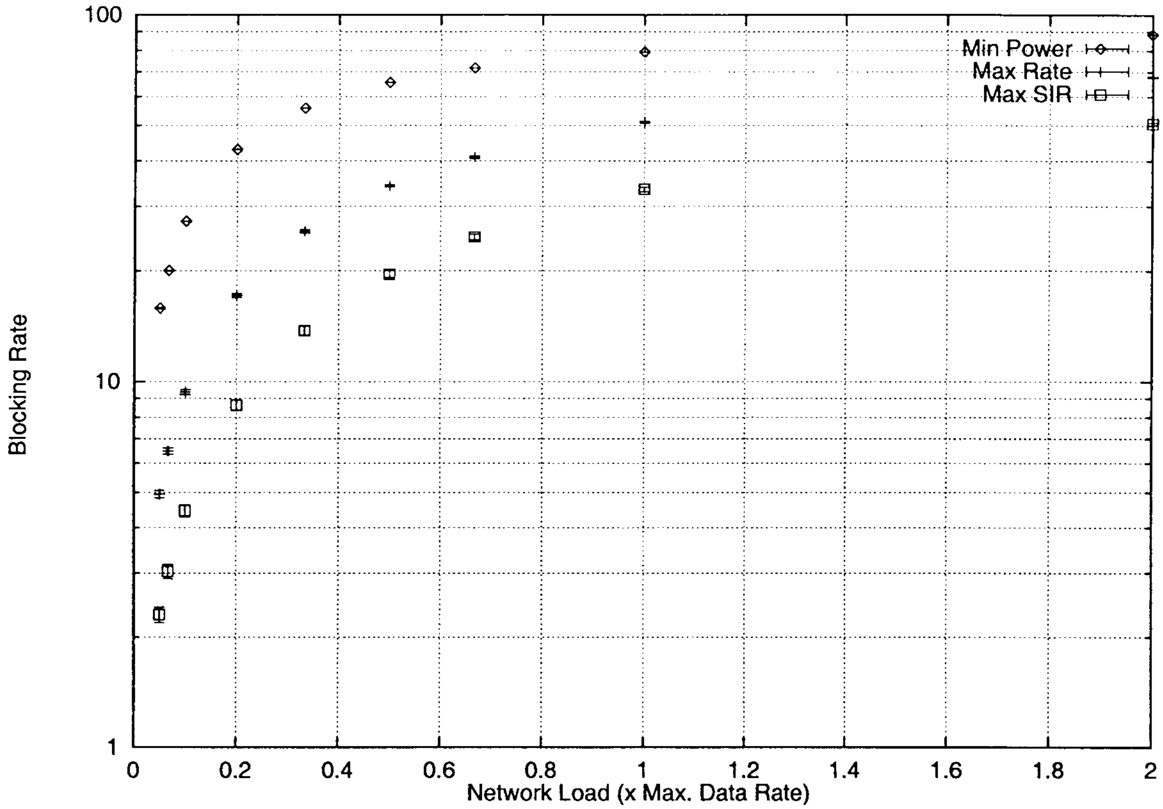

Fig. 14. Blocking rates versus network load ($\beta = 0$).

Real mode simulations illustrate realistic scenarios that involve loss of control messages. It is important to note that even in real mode simulations we assume that all terminals are within range of each other. A terminal that is not transmitting will receive all control messages except when there is a collision on the CCCH. Specifically, the percentage of sessions that lose their QoS guarantees is investigated for the following scenarios.

- The deaf transmitter scenario where a transmitter is incapable of receiving any control messages. In this case, it has no knowledge of any new sessions while it is transmitting.
- Loss of UPD_MSI message through collisions. This scenario takes into account the deaf transmitter phenomenon as well as the fact that update MSI messages are prone to collisions.
- The SREJ mode of simulation demonstrates the efficiency of the SREJ mechanism toward limiting loss of QoS.

Each simulation run consists of a randomly generated network. The session requests are generated using the traffic model outlined above. The transmitter is assigned first from the pool of terminals in the network. Next, the receivers are chosen randomly from the set of remaining terminals, i.e., no loopback sessions are permitted. A run is deemed complete after 150 000 session arrivals. Each run is repeated 30 times with independently generated random networks and traffic patterns. The required metrics along with their 95% confidence intervals were then computed and plotted.

B. Genie Mode Simulation Results

We now compare the blocking probabilities for the three resource allocation policies discussed previously, i.e., MP, MR, and MaxSIR. The effect of the backoff count parameter on global blocking probabilities is also investigated.

1) Blocking Rates: Fig. 14 shows the blocking rates for the MP, MR, and MaxSIR allocation schemes for a network where the backoff count is set to zero, i.e., an incoming session that cannot obtain resources is blocked and forced to leave the system. Contrary to results obtained in cellular multimedia CDMA systems [8]–[10], where capacity is maximized by using the MP scheme, and as predicted by the qualitative discussion relating to the four terminal network, the MaxSIR scheme seems to be the best overall performer. This can be attributed to the incremental nature of DRNP as opposed to the global RM schemes proposed for cellular architectures. The performance gap is especially evident in light to moderate load conditions. For instance, with a network load of 0.1 times channel capacity, blocking rate for the MP scheme is an order of magnitude greater than that for the MaxSIR scheme (almost 12 times as great). Although, the gap does narrow in extremely high load conditions, the MaxSIR scheme still allows 50% of the traffic through (almost full link capacity),[14] while the MP scheme causes almost 90% of the sessions to be blocked. The above discussion demonstrates that although the same resource allocation policies are applicable in centralized and

[14] This is achieved by having multiple active sessions each having been allocated the minimum required data rate; i.e., 64 Kbits/s.

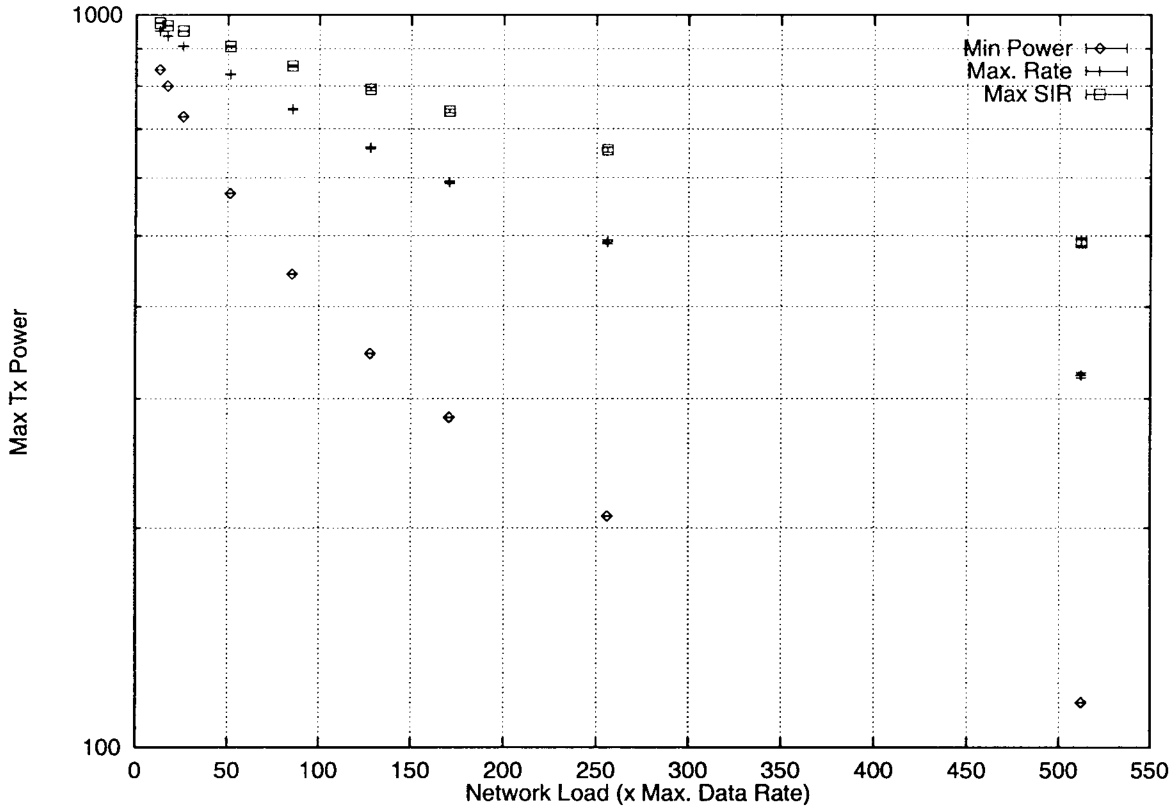

Fig. 15. Average maximum transmission power versus network load ($\beta = 0$).

distributed networks, their performance within the context of these networks is quite different.

The qualitative arguments relating to the MSI, and hence the maximum transmission power (MaxTx), being the limiting factor for network performance are confirmed in Fig. 15 which shows the respective values for average Max for the three resource management schemes. The MaxTx values for all three schemes are quite similar for very light traffic conditions. This is to be expected as the network is idle most of the time and an incoming session will essentially experience no power constraints. But as the traffic load is increased, the MaxTx values for both the MP and MR schemes deteriorate rapidly, i.e., a large proportion of the incoming sessions find the network busy, resulting in their transmission powers being constrained to zero. The MaxSIR scheme seems to be most robust with MaxTx's values close to the maximum limit for low and moderate traffic loads. The preceding discussion confirms the strong correlation between network performance and the MSI parameter.

The backoff count parameter dictates the maximum number of times a session can request resources from network before being discarded. Higher values for this parameter can improve blocking rates at the expense of relatively larger buffer usage and delay. Fig. 16 illustrates the blocking rates for the various RA policies with the maximum backoff count β set to three. As can be seen by comparing Figs. 16 and 14, the MaxSIR scheme seems to derive the most benefit from the nonzero backoff count. For instance, at a network load of two, the blocking rate was reduced from ~50% to slightly over 20%. The MP scheme is still an order of magnitude worse than the MR scheme. The performance of these schemes seems invariant to low values of the backoff count, especially for high levels of traffic. Fig. 17 shows that the MaxTx parameter is essentially independent of the backoff/retry policy used in the network. The improvements seen in the blocking rates are positively correlated with the comparatively higher values for MaxTx.

C. Real Mode Simulation Results

We now turn our attention to investigating the effect of message loss on DRNP's ability to maintain QoS guarantees. Unless stated otherwise, τ_{DATA} is set to ten slots. Note that all simulations in real mode were performed with the MaxSIR allocation scheme and the backoff count set to zero.

The deaf transmitter mode can be thought of as the best case real mode scenario. Here, the only messages lost are due to a transmitter's inability to receive control messages. This implies that terminals, while transmitting, have no knowledge of sessions that arrive or leave. Thus when these transmitters issue RTS messages for future session setups, they advertise an out of date value for the MaxTx. Without the SREJ mechanism, this causes loss of QoS for already active sessions.

As Fig. 18 shows, under light load conditions the percentage of sessions that violate their QoS is limited to approximately 0.05%, but under heavy load conditions (twice channel capacity), in spite of the heavy blocking rates (~50%), 5% of sessions end up losing their QoS guarantees. This is quite substantial. Ideally, we would like to limit QoS loss rates to 1–2%.

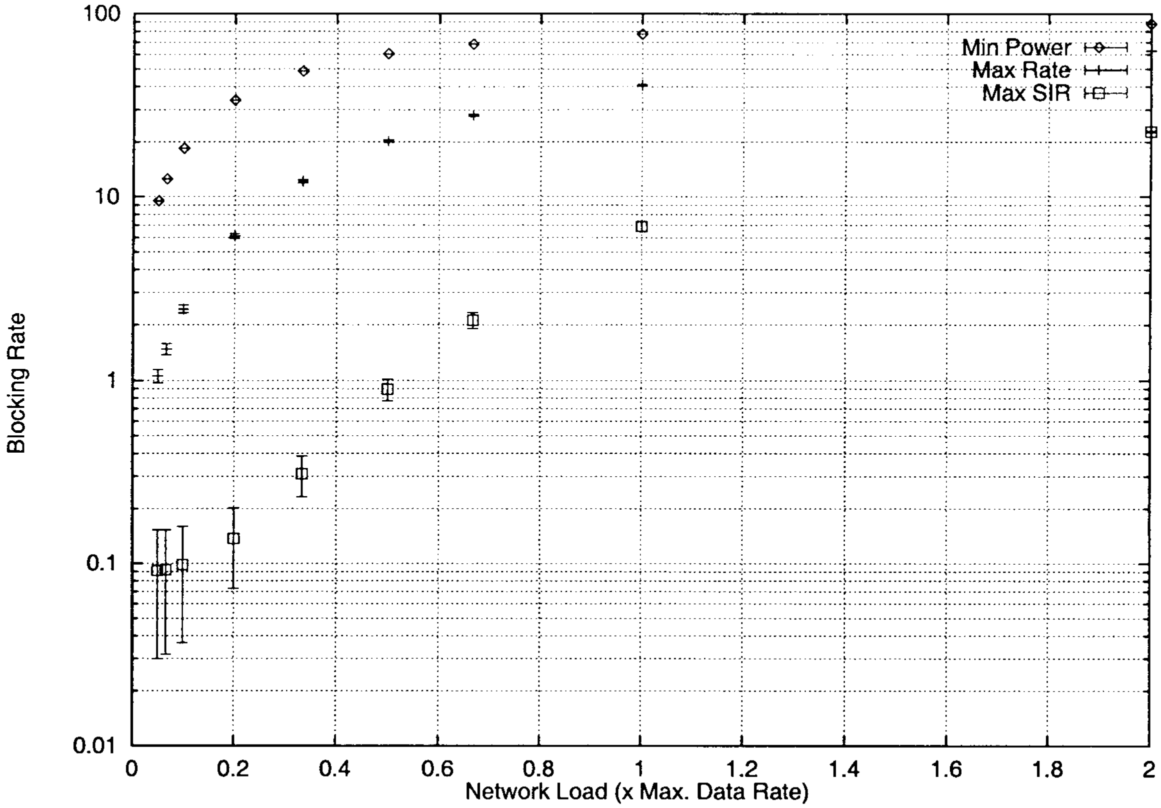

Fig. 16. Blocking rates versus network load ($\beta = 3$).

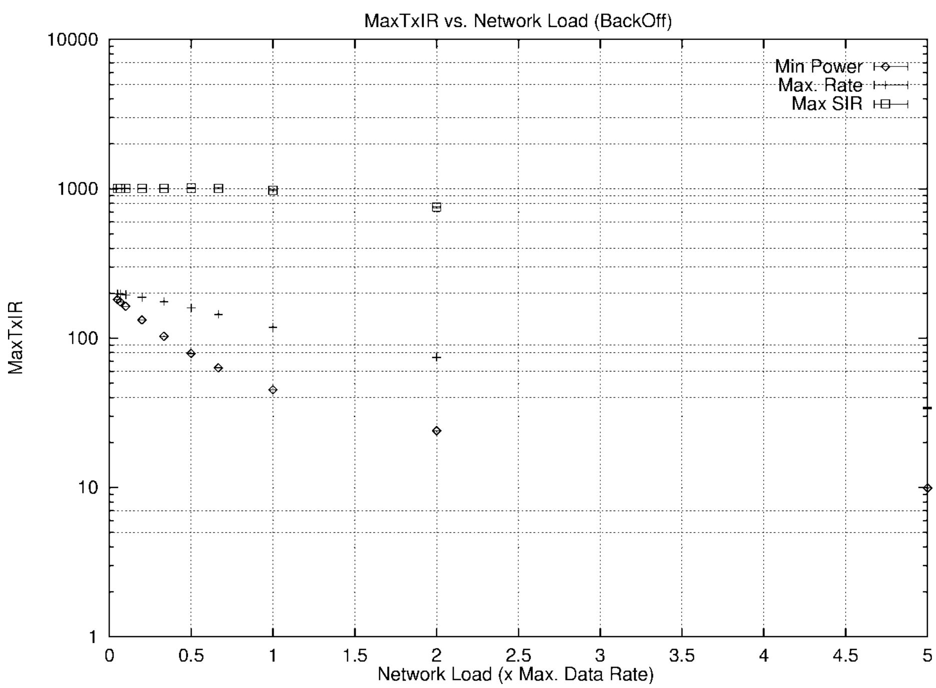

Fig. 17. Average maximum transmission power versus network load ($\beta = 3$).

Including the effect of UPD_MSI message loss only makes matters worse. The SIR for slightly over 8.4% of the sessions fall below their negotiated minimum. However, this result does demonstrate DRNP's robustness, even under extreme traffic load conditions. This corroborates our results from the Promela validation model, i.e., there is no cascade effect in loss of QoS guarantees. DRNP is self-healing and can successfully recover from an incorrect state within a finite time frame.

The situation changes substantially with the introduction of the SREJ mechanism. Even though SREJ messages are themselves lost due to collisions the loss of QoS is limited to ∼1.9% which is almost a four fold improvement.

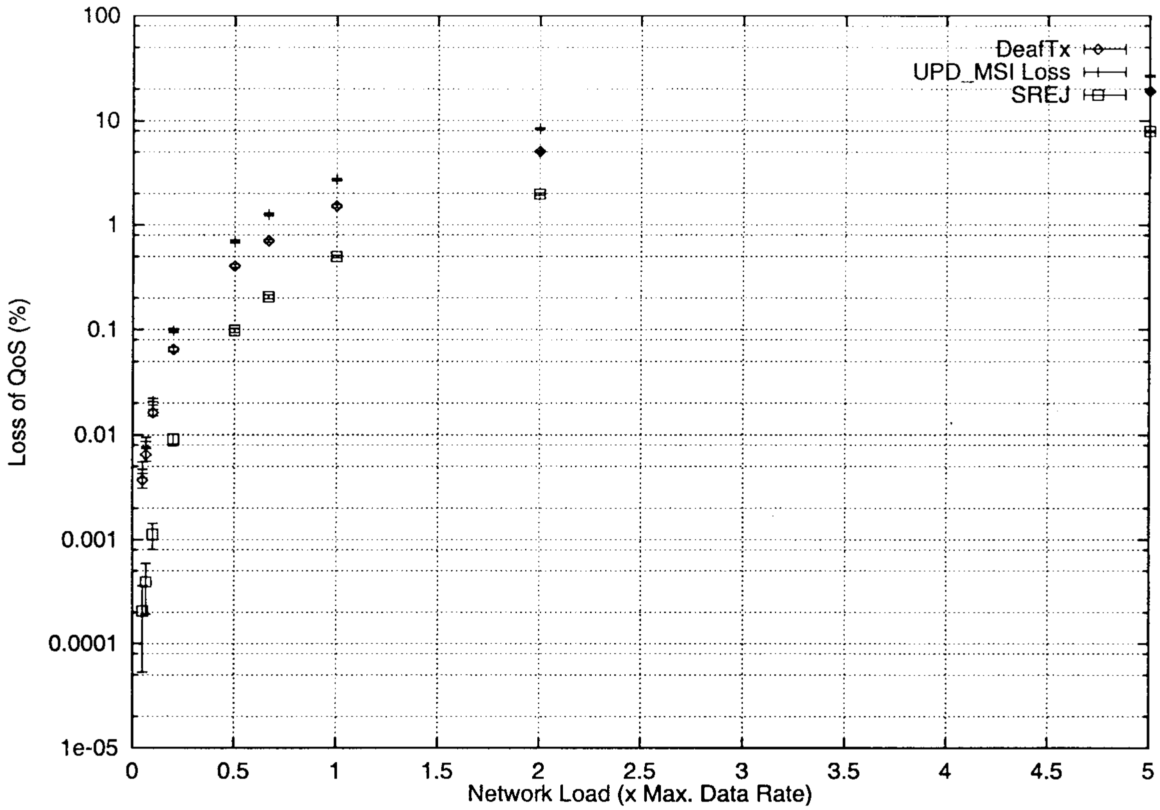

Fig. 18. QoSLoss rates for deaf transmitter, loss of UPD_MSI and SREJ scenarios.

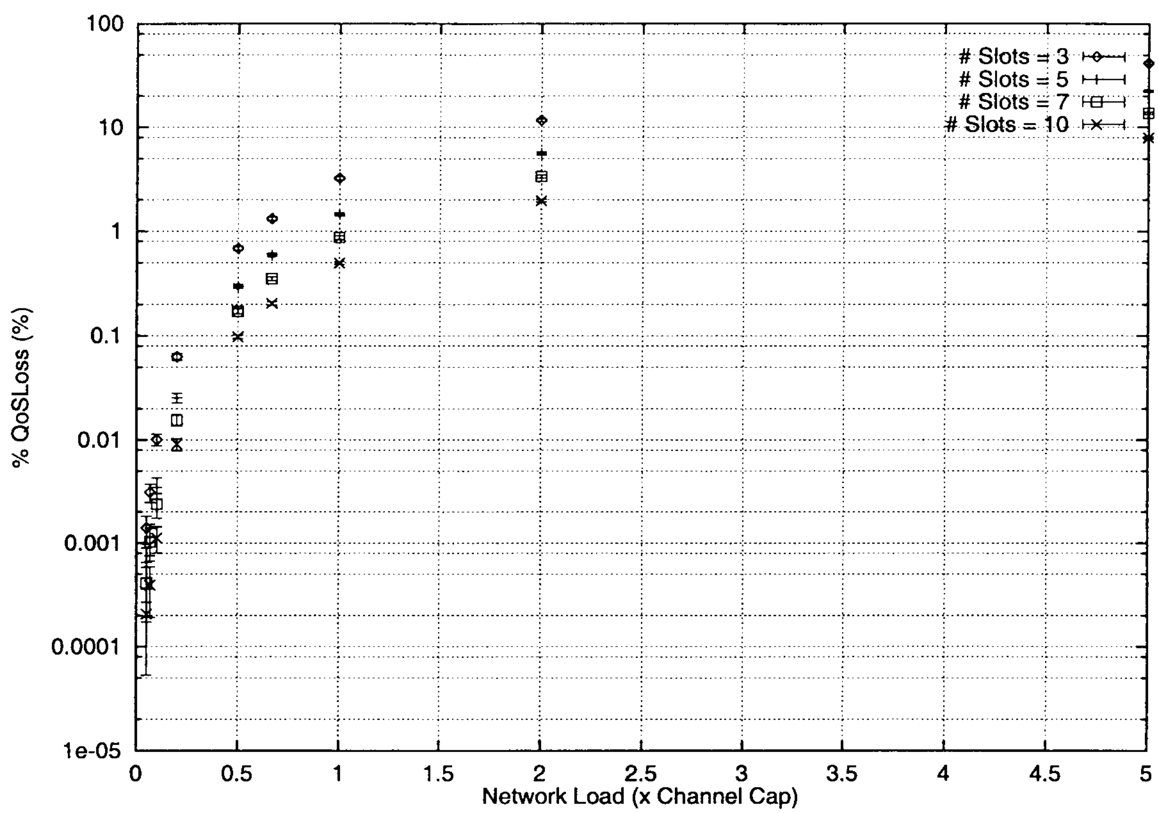

Fig. 19. QoSLoss rates versus τ_{DATA} size.

Fig. 19 illustrates the QoSLoss rates for various sizes of the τ_{DATA} window. As expected, the performance improves as the window is increased. Even with very small values of τ_{DATA}, QoSLoss rates are limited to 10% for very high network load conditions. Although unacceptably high, these results again demonstrate the effectiveness of the SREJ mechanism. As stated earlier, with a moderate value of $\tau_{\text{DATA}} = 10$ slots, QoSLoss rates are well within acceptable levels for nominal traffic loads.

VI. Summary

DRNP was designed specifically to allow resources to be allocated in a multimedia environment. The protocol also fulfills the role of a call admission control (CAC) mechanism, blocking those users that violate the QoS constraints of the currently active sessions. It is completely distributed and adaptive to changing network conditions. DRNP has been tested and validated using automated state space search techniques and found to be robust and fault-tolerant. In fact, for reasonable values of τ_{DATA}, the loss of QoS guarantees is limited to around 2%, even under high load conditions.

We investigated the performance of three resource allocation policies (minimum resource, MR, MaxSIR) within the context of DRNP, in terms of network wide metrics such as blocking and QoSLoss rates. Surprisingly, the minimization of resources scheme which emphasizes conservation of transmission powers and yields the highest capacities in cellular (more generally centralized) networks, turned out to be the worst overall performer. This can be attributed to the fact that DRNP is an incremental resource protocol. Since resources are allocated on a per session basis, a session that has been allocated minimum SIR cannot sustain any additional interference without losing its QoS guarantees. This leads to a single session blocking all other sessions. The best overall performer was the MaxSIR policy which tries to maximize the SIR allocated to each session. Although this does increase the MAI within the network, it does allow other spatially dispersed terminals to setup sessions successfully, as each session is allocated a high QoS (SIR) that it needs. Thus we note that although the same resource allocation policies are applicable in centralized and distributed networks, their performance within the context of these networks is quite different.

VII. Future Work

Future research can follow several avenues.

- The first involves retaining the DRNP framework and investigating the performance of other resource allocation policies. The resource allocation schemes presented in this paper are ad hoc, but an adaptive scheme combined with other performance metrics may yield more efficient allocation of resources. For instance, under extremely low traffic load conditions, the MaxSIR policy's performance is not much better than the minimum power scheme. Thus in this case, the latter may suffice, especially if power conservation is paramount.
- DRNP may be modified/improved. This may involve reduction of overall control channel traffic or better metrics in deciding when SREJ messages should be issued. Specifically, since issuing a SREJ requires that the reception of a current session be interrupted, a soft decision metric that takes the overall damage caused may yield better overall performance at the expense of minor QoS guarantee losses. A long term goal involves the extension of DRNP to multihop networks. Scalability issues arising from control channel traffic are especially important here.

References

[1] K. Pahlavan, "Trends in local wireless data networks," in *Proc. IEEE VTS 46th Vehicular Technology Conf.*, vol. 1, pp. 21–25, 1996.

[2] S. Keshav, *An Engineering Approach to Computer Networking: ATM Networks, the Internet & The Telephone Networks.* Reading, MA: Addison-Wesley, May 1997.

[3] A. Elwalid, D. Mitra, and R. H. Wentworth, "A new approach for allocating buffers and bandwidth to heterogenous, regulated traffic in an ATM node," *IEEE J. Select Areas Commun.*, vol. 12, pp. 544–551, Apr. 1994.

[4] S. Floyd and V. Jacobson, "Link sharing and resource management models for packet networks," *IEEE Trans. Networking*, vol. 3, Aug. 1995.

[5] R. Guerin, H. Ahmadi, and M. Nagshineh, "Equivalent capacity and its application," R. Guerin and L. Gun, Eds. "A unified approach to bandwidth allocation and access control in fast packet switched networks," in *Proc. IEEE INFOCOM'92.*

[6] R. W. Nettleton and H. Alavi, "Power control for a spread spectrum cellular mobile radio system," in *Proc. VTC'83*, 1983, vol. 1, pp. 242–246.

[7] S. A. Grandhi, R. Vijayan, D. J. Goodman, and J. Zander, "Centralized power control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. VT-42, pp. 466–68, Nov. 1993.

[8] A. Sampath, P. S. Kumar, and J. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in *Proc. IEEE PIMRC'95*, vol. 1, pp. 21–25.

[9] W. W. S. Wong, "Power assignment strategies and traffic control for wireless multimedia DS-CDMA systems," M.S. thesis, Univ. Toronto, Toronto, Ont., Canada, 1996.

[10] K. W. Shum, "Fuzzy distributed power control in cellular radio networks," in *Proc. PIMRC'95*, 1995, vol. 1, pp. 51–55.

[11] A. Baier, "Multi-rate DS-CDMA: A promising access technique for third generation mobile radio systems," in *Proc. PIMRC'93*, pp. 114–118.

[12] G. J. Holzmann, *Design and Validation of Computer Protocols*. Englewood Cliffs, NJ: Prentice-Hall Softw. Series, 1991.

[13] J. Mikkonen and J. Kruys, "The magic WAND: A wireless ATM access system," in *Proc. ACTS Mobile Summit'96*, pp. 535–542.

[14] G. J. M. Janssen, P. A. Stigter, and R. Prasad, "Wideband indoor channel measurements and BER analysis of frequency selective multipath channels at 2.4, 4.75 and 11.5 GHz," *IEEE Trans. Commun.*, vol. 44, Oct. 1996.

[15] E. Sousa, "A spreading code protocol utilizing an echo from the receiver," in *Proc. Milcom'88*, vol. 1, pp. 357–359.

[16] ——, "Spreading code protocols for distributed spread-spectrum packet radio networks," *IEEE Trans. Commun.*, vol. 36, pp. 272–281, Mar. 1988.

[17] M. Georgiopoulos, "Packet error probabilities in direct sequence spread-spectrum packet radio networks," *IEEE Trans. Commun.*, vol. 38, pp. 1599–1606, Sept. 1990.

[18] J. S. Lehnert and M. B. Pursley, "Error probabilities for binary direct-sequence spread spectrum communications with random signature sequences," *IEEE Trans. Commun.*, vol. COM-35, pp. 87–98, Jan. 1987.

[19] E. Sousa, "Interference modeling in a direct-sequence spread spectrum packet radio network," *IEEE Trans. Commun.*, vol. 38, pp. 1475–1482, Sept. 1990.

[20] J. Whitehead, "Distributed packet dynamic resource allocation (DRA) for wireless networks," in *Proc. IEEE VTS 46th Vehicular Technology Conference*, vol. 1, pp. 111–115, 1996.

[21] K. Pahlavan and A. Levesque, *Wireless Information Networks.* New York: Wiley, 1995.

[22] E. Sousa, "Interference modeling in a direct-sequence spread spectrum packet radio network," *IEEE Trans. Commun.*, vol. 38, pp. 1475–1482, Sept. 1990.

[23] E. Anderlind, "A traffic model for non-real time data users in a wireless radio network," *IEEE Commun. Lett.*, vol. 1, no. 2, pp. 37–39, Mar. 1997.

Sanjay Lal (M'98) received the B.A.Sc. degree in computer engineering and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, Ont., Canada, in 1995 and 1997, respectively.

He is currently a Member of the Technical Staff in the Advanced Interworking Systems Group, Lucent Technologies, Holmdel, NJ. He has worked as a Research Engineer at Nokia Research Center, Irving, TX. He has also conducted research in the area of ATM switching at Motorola's Research and Advanced Development group, Mansfield, MA. His current research interests include resource allocation and performance analysis of wireless information networks.

Elvino S. Sousa (S'79–M'80–SM'96) was born in Graciosa, Acores, Portugal, on December 28, 1956. He received the B.A.Sc. degree in engineering science and the M.A.Sc. in electrical engineering from the University of Toronto, Toronto, Ont., Canada, in 1980 and 1982, respectively, ant the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1985.

Since 1986, he has been with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ont., Canada. From 1986 to 1996, he held the ranks of Assistant and Associate Professor and was also a Natural Sciences and Engineering Research Council (NSERC) University Research Fellow. In 1997, he was promoted to the rank of Full Professor. He has performed research in spread spectrum systems since 1983. His current interests are in the areas of spread spectrum systems, mobile communications, and indoor wireless communications with emphasis on the high level performance of the network and signal processing issues affecting such a performance. At the University of Toronto he has founded the Wireless Communications Lab which is involved in a large number of projects in wireless communications with emphasis on CDMA.

Dr. Sousa has been invited to give lectures and short courses on spread spectrum and wireless communications in a number of different countries throughtout the world. He was the Technical Program Chairman for PIMRC'95 and is the Technical Vice-Chair for Globecom'99.