CROSS-LAYER SCHEDULING AND ROUTING
FOR UNSTRUCTURED AND QUASI-STRUCTURED WIRELESS NETWORKS

Joseph Thomas
University of Maryland, Baltimore, MD 21250

ABSTRACT

Novel and conceptually simple scheduling and routing algorithms are proposed for unstructured (ad hoc) networks. These algorithms are based on an intelligent interference and power management paradigm that borrows notions from successive decoding and cancellation and transmission power control in order to exploit and thereby (or otherwise) suppress interference (thus departing from the “collision-avoidance” objective and its attendant burdens of eliminating hidden nodes) while supporting differential services based on preassigned traffic priorities. The application of the proposed algorithms to quasi-structured networks – where certain nodes may possess additional capabilities and thus alleviate any processing overheads – is also considered.

I. INTRODUCTION

Algorithms for scheduling, i.e. medium access control (MAC), and routing in unstructured (ad hoc) networks are of immense interest in the development of wireless networking protocols. The absence of coordinating infrastructure as in structured wireless (i.e., cellular or local area) networks necessitates cooperation among nodes. Most existing scheduling protocols for unstructured wireless networks are based on the collision-avoidance approach, introduced in slotted Aloha [2], and extended to address the exposed/hidden node problem in the IEEE 802.11 standards via the Request to Send (RTS) and Clear to Send (CTS) signaling mechanism [1], [5]. The best known reactive routing protocols are the dynamic source routing (DSR) [13] and the ad hoc on demand distance vector (AoDV) routing [21] protocols. Several notable efforts at enhancing these protocols to achieve the objectives of energy minimization and throughput improvement explore transmission power control and reliable interference avoidance as briefly reviewed in Section II.

Research leading to the present work follows a different path [27]. Though often ignored, the incorporation of both large- and small-scale fading effects in a simulation model impacts not just the absolute performance measures of MAC and routing protocols but also their relative rankings vis a vis each other [27], [30]. If incorporating such effects in the simulations of these protocols impacts their evaluation, it is natural to infer that incorporating them in their very design would yield appreciable performance gains. Additionally, the vast majority of research efforts in this area seek to avoid interference among users. This work, in contrast, carefully incorporates cross-layer interactions via the effects of fading and interference and proposes an intelligent interference and power management paradigm for scheduling and routing which controls and possibly even exploits multiaccess interference. It utilizes distributed power assignments and successive-decoding-like schemes based on each node’s knowledge of its neighbors’ states to best effect in regard to throughput maximization in scheduling and routing protocols. The essential procedures such as successive decoding executed by each node in such a framework are well understood in communication theory; see, e.g., [29]. Additionally, the proposed algorithms support service differentiation in scheduling and routing. Sections III and IV develop the principles underlying these algorithms and present simulations of their performance. Since these algorithms entail additional processing and storage, Section V considers their performance in quasi-structured networks where certain nodes are endowed with special capabilities that can be harnessed to benefit global performance.

II. RELATED WORK

Transmission power control and management is the primary focus of recent efforts in developing scheduling and routing protocols to deliver throughput enhancements and/or energy savings. These proposals retain the underlying framework of the 802.11 protocol (i.e., collision-avoidance) and AoDV (or DSR) respectively. They aim to either maximize throughput [18], [9], [17], [19] via some variation of transmission power control or to minimize energy consumption [4], [6], [8], [10], [14], [15], [22], [24] via power management (e.g., inducing sleep states when possible) or again via transmission power control. It is understood [11] that throughput maximization and energy minimization are mutually incompatible, implying that one can seek to achieve either one of these objectives for a fixed value of the other parameter.

The present work focuses on throughput maximization and therefore it is of interest to next briefly review cur-
rently known throughput maximization enhancements. In [17], power control is employed in a manner that ensures compatibility with existing scheduling and routing algorithms, the preservation of modularity in system design being the primary motivation. In [23], power control is used as a tool to control the topology of the network as seen by individual nodes; however, geographical distances are an important consideration in this work, and this is problematic for at least two reasons, namely the physical configuration of an ad hoc network varies with time and moreover, geographical proximity of one node to another is a reliable metric in this context only if the propagation medium is strictly free space. The algorithms in [18], [3], [19] all seek to limit interference in some sense. In [18] however, the MAC protocol must choose the destination of the node that seeks medium access. This is somewhat unusual, as noted also in [16], since this choice should be made by the network layer and passed down to the MAC protocol whereas no such synergy is sought between the network and link layer functions in [18]. The algorithm in [9] relies on centralized coordination while that in [3] relies again on the geographical locations of nodes thus limiting its use in fading environments where relative geographical locations are not as relevant as radio channel properties as noted above. In [19], multiple concurrent transmissions are allowed but again the focus is consistently on limiting interference rather than exploiting it, as explored in the present work. The paradigm introduced in the sequel (a detailed account of which constitutes part of [27]) uses the notion of admission control which is also employed in [25], [26] and more recently in a special case [7] of another algorithm proposed in [27].

III. A CROSS-LAYER SCHEDULING PARADIGM

A well-understood requirement [29] used in structured networks for the successful detection of a desired signal from a received mixture (of multiple interfering signals and noise) is that the signal to interference plus noise ratio (SINR) exceed a certain threshold. Together with the maximum transmission power of each node dictated by its dynamic range, this can yield an iterative procedure to obtain an optimum transmission power. This concept has been used by the author to develop a distributed cross-layer scheduling and routing algorithm for unstructured networks in [27] where a node is allowed access provided it can meet the SINR threshold for its intended destination and additionally its admission does not lead to a violation of SINR criteria for any existing sessions of higher priority; routing decisions use the hop-sequence that attains the extremum of an objective function of the sample mean and sample variance of the hop-SINRs for each candidate hop-sequence in that framework.

The present paper is more ambitious. It makes the case for a paradigm shift from the conventional collision-avoidance approach to the scheduling problem in random-access networks. The principal idea underlying the proposed approach is that of intelligent interference and power management encompassing concepts from the domain of interference cancellation, suppression, or even exploitation, in addition to transmission power control. The essential objective in the present framework is to regulate admission of traffic to the network in order to meet the two important requirements, namely limit outages in existing sessions while facilitating differential services. For instance, a packet that is currently being served may be dropped to serve another packet only if the latter has a sufficiently higher priority than the former, as is precisely and formally detailed below.

In what follows, the setting is essentially that of a source node attempting to transmit a packet to a destination node in a single hop, across a physical channel that suffers both large-scale (i.e., path-loss) and small-scale (e.g., Rayleigh) fading in addition to multiaccess interference. Further, as is customary, time is described as a sequence of discrete slots; additionally, all nodes have omnidirectional antennas. The coherence time of the channel is assumed to exceed the entire duration of the packet’s transmission including the protocol overhead steps such as route discovery, multiaccess admission, and return receipt acknowledgement. Though multihop routing is inevitable in an unstructured network, the sequence of hops in a route is determined by the choice of the route which is the subject of Section IV. None of these decisions involve any centralized control; they are all made using locally available information alone.

Service differentiation is sought by associating priorities with packets. Each traffic class is assigned a global priority factor that lies in the unit interval (0, 1]. Packets belonging to the same traffic class are mutually discrimination by the occupancy level of the link buffer associated with their intended destination, as detailed in [27]. For simplicity, the sequel will assume that each packet belongs to a distinct traffic class and is thus preassigned a global priority factor, denoted by $q_{i}^{(n)}$ for packet $i$ at node $n$, with $0 < q_{i}^{(n)} \leq 1$.

Each node in the network is assumed to be capable of decoding a desired packet from a received mixture-sequence that is a sum of this desired signal (possibly faded) and other interfering signals and noise and additionally to be capable also of concurrently decoding multiple packets (transmitted to it by different sources) from its received mixture sequence. The latter could be achieved for instance by successive decoding and subtractive inter-
ference cancellation [29]. In sum, this allows recovering even a weak desired signal in the presence of a strong interferer: the latter is first decoded and its (thus accurately reconstructed) contribution is subtracted from the received mixture, thereby allowing subsequent successful detection of the originally sought weak signal. Since this step of first decoding such a strong interferer (noting that this would not be possible if the interferer was weak or even at the same received power as the desired signal) contributes to its near-perfect cancellation, this concept is referred to as interference-exploitation or alternatively, as an exploitation of a “near-far” interference-limited environment, i.e. one in which a strong interfering signal from a source that appears to be near the destination overwhelms a weak and apparently distant source’s desired transmission. When interferers are significantly weaker than the desired signal then conventional interference suppression strategies may be employed. On the other hand, when they are at approximately the same power as the desired signal in the received mixture then the request for access may be denied depending on the relative priorities of the requesting packet’s traffic and that of the packet that will be dropped as a result of such interference if the former were to be admitted. These principles are easily extended to the case of multiple desired packets in the received mixture at each node.¹

A node initiates a request to transmit a packet to a destination of its choice by broadcasting a Request to Send (RTS) signal at its maximum possible transmission power. This RTS packet includes, among other details, information on the intended destination’s address and the priority (say $q$) and length of the actual packet that the source intends to transmit. The intended destination computes the optimum power that the transmitter should use in order to allow the receiver to decode the packet in question. Each node is assumed, additionally, to have information about its neighbors’ states: specifically, in the present case, the intended destination can compute the probability $p_i^{(n)}$ that packet $i$ at neighbor node $n$ (with priority $q_i^{(n)}$) will have to be dropped as a predicted consequence of potentially allowing the currently sought transmission at the above-mentioned optimum power. Based on such computations for every packet at each such neighbor and the individual priorities of these packets and that of new packet seeking transmission, the intended destination must assess the benefit of allowing this transmission. The simple rule proposed below allows for such a decision with little additional processing.

Allow the transmission only if

$$q = \max_{n,i} \{q_i^{(n)}(1 - p_i^{(n)})\}$$

In other words, the transmission is allowed only if its priority is high enough to justify the potential dropping of other packets (from other sessions) that are currently being processed. If the transmission is allowed then the intended destination sends a Clear to Send (CTS) signal (at its maximum possible transmission power) back to the requesting source. This CTS packet contains, in addition to the source and destination addresses, the optimum transmission power to be used by the source, and the intended packet’s length and priority. It must be noted here that this information is used by the destination’s neighbors to update their records of its state.

If the transmission is not allowed, the source does not receive a CTS and backs-off for a window of duration that is uniform in the interval $[0, W_t]$. Here, $W_t$ depends on the number of past unsuccessful attempts by this source node at accessing the channel and the priority of its current packet, as in [12]. Similarly, this source’s probability of subsequently attempting to access the channel (i.e., in each timeslot past the back-off window) is also set as a function of these two quantities [12]. The proposed scheduling algorithm’s performance is studied in Section IV in conjunction with the routing scheme proposed there.

IV. A CROSS-LAYER ROUTING ALGORITHM

Route discovery in an unstructured network leads to a choice among several sequences of multiple hops from the source to the destination. It is clearly possible to run any of the well-known reactive (or even proactive) routing algorithms for ad hoc networks, e.g. AoDV or DSR, over the proposed scheduling/MAC protocol. However, it is equally clear that this would be a poor choice given the possibilities that are revealed from the preceding discussions. For example, it would make little sense to lose potential throughput improvements such as those afforded by interference cancellation and prioritization in the above MAC protocol by returning the first discovered route (as in AoDV) or flooding the network with route-replies to the source with all possible routes to the destination (as in DSR). The task of developing a compatible routing algorithm to exploit the scheduling scheme in Section III is nontrivial because straightforward topology control based on geographical proximity is deficient for the reasons detailed earlier in Section II; in addition, the SINR-based cross-layer routing alternative [27] does not allow for in-

¹Successive decoding and cancellation assume that each node is capable of SINR estimation and multiuser channel estimation as in [20].
interference exploitation via successive decoding and cancellation per the present framework.

An active node in an unstructured network is assumed to reactively determine its set of neighbors. It does this by beaconing its presence (at maximum power) and adds nodes from which it receives positive acknowledgements to its set of neighbors; each of these acknowledgements is sent by these nodes to the requesting node at the minimum power necessary for its correct reception, and includes information about the optimum transmission power that should be used by the requesting node for future transmissions. This list is updated reactively. If a physical neighbor determines that it cannot successfully decode the requesting nodes transmissions for any given received power, it does not return a positive acknowledgement. Clearly, this process of neighbor discovery differs both from the obviously-flawed criterion of geographical proximity and from the minimum SINR threshold requirement [27] mentioned earlier.

Given the above construction of neighbor sets for each node, it is now possible to use an AoDV-like or a DSR-like route discovery algorithm. This would however still fail to exploit the essential wealth of possibilities suggested by the MAC protocol proposed in Section III. Consider all possible sequences of multiple hops from the source to the destination as returned in a DSR-like framework based on the above neighbor sets. The best of these sequences must be identified and chosen by defining a suitable means of locally evaluating the various options with respect to the impact of this choice on the network as a whole, keeping in view the need to provide differential quality of service according to the priority ascribed to the packet. In any such given hop-sequence, let $p_i^{(n)}$ denote the probability that some other packet $i$ that is currently being processed by node $n$ along this hop would be dropped if the route corresponding to this hop-sequence were chosen. If such outage probabilities corresponding to all the other packets that are being concurrently processed along this route is small, then this might appear to be a good route. Incorporating the priorities $q_i^{(n)}$ of these other packets, this may restated as seeking a route with a high sample mean $\tilde{\mu}$ of the product $q_i^{(n)}(1 - p_i^{(n)})$ over all values of $i$ and $n$ on this route. The hop-sequence corresponding to the maximum $\tilde{\mu}$ is determined at the source and chosen via the usual mechanism of route replies. Route-caches are maintained and updated. These and several related issues such as bidirectionality of route-reply paths, frequency of neighbor set updates, and network connectivity are discussed in [27] for this protocol.

The performance of the proposed scheduling and rout-
two mutually perpendicular diametric chords on the original disc (of radius 1000 m). The remaining eight nodes in the fifth group were isolated from these clusters and were located at uniformly random positions on the remaining surface of the original disc. The traffic patterns, channel characteristics, and other details were the same as before; however, the mobility of each node within a cluster was limited to its local disc in each case, whereas the isolated nodes could move along the remaining surface of the original disc. The significant difference between this scenario and the preceding one is that of enormous received power disparities between different sessions. The 802.11/DSR and 802.11/AoDV schemes are unable to accommodate for such near-far limited interference and therefore experience an increasing number of failures in attempting to access the channel and an increasing depletion of available routes. The present scheme, in contrast, is able to exploit the near-far interference-limited nature of these channels in its successive cancellation approach and, as Fig. 2 shows, its packet delivery ratio improves significantly relative to the preceding case.

Fig. 3 shows a plot of the normalized delay (i.e., normalizing the smallest delay in the data used in this figure to unity) against the actually realized throughput for the highest priority class in this clustered scenario. Again, not surprisingly, the improvement afforded by the proposed scheme over the other two schemes is enormous. Other related results such as energy consumption comparisons, protocol overheads etc are detailed in [27].

V. APPLICATION TO QUASI-STRUCTURED NETWORKS

Many of the unstructured networks found in practice consist of nodes that have some variations in their capabilities. In certain cases, these variations are sufficiently pronounced for these networks to be classified as quasi-structured rather than as ad hoc networks. For instance, clusterheads equipped with additional processing or memory capabilities in a cluster of nodes may devote their resources to aiding in or to some extent managing the cooperative protocols that are being executed by an active group of source/destination nodes in their respective clusters. A particularly relevant example is that of a crisis management scenario where individual workers equipped with mobile communication transceivers communicate with each other and are aided by more powerful transceivers located in a stationary fire-truck-like vehicle in the vicinity. Another example is that of a wireless local area network that is adaptively reconfigurable [28]. Such reconfigurability is often desired even in infrastructure-based networks since channel-access demands vary with time. It is assumed to have a powerful central base-station/access-router (represented by $C$ in Fig. 4) to which are connected several small stations ($B_0$ and $B_1$ in Fig. 4) via very high bandwidth wired links, e.g. fiber. These local-stations have limited processing power, are inexpen-
sive, and are readily introduced from an economic perspective. Their primary role is to receive and transmit signals between the central-station and the mobile nodes (represented by $M_0$ and $M_1$ in Fig. 4) that they serve. In practice, in such a configuration, a mobile node would be served by a set of local-stations which mutually exchange information via the central-station. This set of local-stations assigned to a given mobile node is dynamically updated depending on the quality, e.g. SINR, of the channel between the mobile node and each potentially available local-station.

How are quasi-structured networks relevant to the proposed scheduling/MAC and routing schemes? Assuming that nodes are equipped with successive-decoding/interference-cancellation capabilities for physical layer processing, the primary additional overhead in the scheduling and routing algorithms proposed in this work arises from the burden of storing additional information such as neighbor configurations, and additional computations of, for example, projected outage probabilities. It is clear that the presence in a network of nodes with special capabilities such as additional computing power, storage memory, or high capacity wired links to the wired infrastructure can contribute significantly to alleviate the above overhead. Three ways in which such assistance can be derived are outlined below:

(i) If the special nodes have no additional processing/memory capability but have a high capacity link to the wired infrastructure as in Fig. 4, then the wired infrastructure could replace a sub-sequence of radio-link hops in the source-to-destination route.

(ii) In the absence of any link to the wired infrastructure, multi-hop routes that include special nodes (i.e., ones with increased computing power) are obviously preferred to routes that do not, since the former have more effective interference suppression capabilities.

(iii) The most effective way in which such special nodes can assist is by taking on some of the computational burdens of other neighboring nodes in scheduling or route discovery decisions.

The clustered configuration with isolated nodes used in Section IV above was used to model a quasi-structured network. One of the eight nodes in each of the four clusters was assumed to have special processing capabilities emulating the notion of a clusterhead for each cluster. Of the remaining eight isolated nodes, two were assumed to be fixed and connected to the wired infrastructure via high-bandwidth links. All other simulation details such as channel characteristics, mobility models, traffic patterns etc were assumed to be the same as for the simulations depicted in Fig. 2. The packet delivery fractions plotted in

---

**Fig. 5. Packet delivery ratios in a clustered hybrid quasi-structured network architecture**

**Fig. 4. An example of a quasi-structured network architecture**

---

**VI. CONCLUSIONS**

A novel distributed scheduling and routing paradigm that seeks to maximize throughput while respecting priorities has been presented. The underlying philosophy advanced in this paper revolves around the principle of intelligent interference and power management encompassing concepts from the domain of interference cancellation, suppression, or even exploitation, in addition to transmission power control, thus eliminating concerns about effectively addressing the hidden-node problem which is inherent in the customary collision-avoidance approach. This
framework regulates admission of traffic to the network in the scheduling and medium access control phase such that two important requirements are met: namely outages in existing sessions are sought to be limited and differential services are facilitated. Similarly, the routing algorithm chooses the best possible route with respect to these two criteria. This paradigm allows for multiple concurrent transmissions (assuming, quite reasonably that receivers are equipped with successive decoding capabilities) at optimum powers such the resulting session outages are limited and high priority sessions are differentially preferred. Notwithstanding concerns that these benefits come only with the additional computation and storage cost entailed by the proposed algorithms (relative to existing ones) it must be noted that, in practice, this approach would merely exploit physical/link layer capabilities that are already available at each node. Additionally, per the development in [27], the present algorithms are readily extended to include a judicious choice of frequency band from those available in a dynamic spectrum allocation setting.

REFERENCES

[1] IEEE Standard 802.11: Wireless LAN medium access control (MAC) and physical (PHY) layer specifications, 1999