Bandwidth Allocation in Wireless Ad Hoc Networks: Challenges and Prospects

Xueyuan Su, Yale University
Sammy Chan, City University of Hong Kong
Jonathan H. Manton, The University of Melbourne

ABSTRACT
Ensuring each user experiences a satisfactory quality of service is an important challenge for network designers, especially in wireless networks, where resources are relatively scarce and interference is relatively high. Accordingly, there has been recent interest in bandwidth allocation in wireless ad hoc networks, the focus of this article. After highlighting the main challenges, we survey recently proposed solutions, which address the problem at the network or MAC layer, individually or jointly. We also classify these solutions according to some major design criteria, and suggest the directions of future work on bandwidth allocation.

INTRODUCTION
Wireless ad hoc networks are decentralized wireless networks; every node is willing to forward packets for other nodes, and routing decisions are made dynamically based on network connectivity. Originally developed for military applications, advances in transmission technologies and portable computing devices have led to a growing interest in deploying wireless ad hoc networks in commercial applications, such as cooperative mobile data exchanges, virtual classrooms, and home networking. With flows competing for the shared wireless medium, it is important to have an effective and efficient bandwidth allocation mechanism to allow fair sharing of bandwidth. Currently, the most widely used medium access control (MAC) protocol in ad hoc networks is IEEE 802.11. Unfortunately, it is based on random access and thus inherently lacks the ability to manage bandwidth allocation.

If, despite many years of research, there is still no quality of service (QoS) in the Internet, is QoS important in wireless ad hoc networks? We believe there is an urgent need in wireless ad hoc networks for QoS in general and for bandwidth allocation in particular; the Internet has survived without QoS essentially by increasing the capacity of its links to meet demand. A wireless ad hoc network does not have this luxury; power (especially for battery-powered equipment) and bandwidth are precious, and interference is harder to deal with.

The decentralized and dynamic nature of ad hoc networks means that information must be passed from node to node about the network topology and end-to-end flow rates. Furthermore, distributed bandwidth allocation algorithms are generally favored over centralized ones. Two key criteria of any such distributed algorithm are the amount of data passed from node to node and the convergence time of the algorithm (how long it takes to reach a steady state after a perturbation). A third criterion is the fairness of the algorithm.

The bandwidth allocation problem considered here is the following. A network of nodes and (wireless) links is specified. At any instance in time, there are a number of end-to-end flows. Each flow has its own required bandwidth. In the special case where a node is greedy and wants to use as much bandwidth as possible, its desired bandwidth can be set to infinity in the following algorithms. Each node must determine the bandwidth to allocate for each flow passing through it in a fair and efficient way.

A bandwidth allocation algorithm must operate across the MAC and network layers of the protocol stack. The network layer is responsible for end-to-end flow rate calculations, while the MAC layer must schedule single-hop transmissions so as to ensure each flow receives adequate bandwidth over each link.

The bandwidth allocation problem is discussed in detail in the next section. It is followed by a section summarizing the recent results in this area and their prospects.

CHALLENGES
Figure 1 illustrates the well-known fairness problem in wireless ad hoc networks. There are four nodes arranged in the chain topology. Nodes B and C are within transmission range of each other. Therefore, flows $f_1$ and $f_2$, initiating from nodes B and C, respectively, interfere with each other and contend for the bandwidth. If the
IEEE 802.11 MAC protocol is being used — carrier sense multiple access with collision avoidance (CSMA/CA) with binary exponential backoff (BEB) — and node $B$ gets the medium first, its backoff window will be reset to the minimum value. When node $C$ tries to transmit, it is likely node $B$ is already transmitting a packet; hence, node $C$’s backoff window is likely to keep increasing, leading to $f_1$ monopolizing the channel while $f_2$ is starved with very limited throughput.

Different from the wireline case, flows $f_1$ and $f_2$ in Fig. I do not use the same link, yet interfere with each other. Therefore, even though node $B$ only wishes to talk with node $A$, and node $C$ only with node $D$, if fairness is to be achieved, it is necessary for nodes $B$ and $C$ to exchange information about their traffic flows and come to an agreement on how to schedule their transmissions so that both flows achieve sufficient throughput. As the network layer is responsible for traffic flows and the MAC layer for scheduling transmissions, a bandwidth allocation scheme must decide how these two layers interact.

The following challenges can thus be identified.

**Interference model:** The local broadcasting and spatial interference characteristics of wireless networks make the bandwidth allocation problem very different from that in wireline networks (Fig. 1). Each node must build up a picture of how transmissions from other nodes will compete with its own transmission requirements. Some bandwidth allocation schemes may choose to work with a simplified interference model, potentially sacrificing bandwidth utilization for a saving in algorithmic complexity.

**Rate allocation in the network layer:** The network layer must calculate the rate each flow is allowed to have. This calculation depends on the active flows in the network, the interference model used, and the definition of fairness adopted. Performance factors include the computational complexity of the algorithm, the amount of additional information that must be transferred between nodes, the convergence time, and the optimality of the solution.

**Bandwidth allocation in the MAC layer:** Once the rate each flow is allowed to have has been calculated, a method is required for scheduling the transmissions at each node to ensure the rates are achieved. Coordinating and scheduling transmissions in a distributed manner is nontrivial; the popular MAC protocols cannot perform distributed scheduling because they are based on random access or intuitive fairness.

**Interaction between the two layers:** The MAC and network layers must interact; the MAC layer must help gather information about the network topology and the active flows in order for the network layer to allocate rates, and the network layer must tell the MAC layer at what rate to transmit each flow. Figure 2 shows the two possible approaches. The first is to construct a single algorithm dealing simultaneously with rate and bandwidth allocations in both layers. The second is a modularized approach, whereby limited information is exchanged between the layers.

---

**SOLUTIONS AND PROSPECTS**

Recent work on bandwidth allocation in wireless ad hoc networks is summarized and critiqued.

**INTERFERENCE MODEL**

In wireless networks, nodes are usually regarded as half-duplex; they cannot transmit and receive data simultaneously. Furthermore, due to the dynamic nature of ad hoc networks — nodes come and go — a common assumption is that all nodes transmit omnidirectionally and at the same frequency. There are essentially two classes of interference models used to determine if two or more nodes can transmit at the same time without interfering with each other: physical models and protocol models [1]. Physical models are based on determining if the signal-to-interference ratio (SIR) at the receiver will exceed $\beta$, the minimum SIR required for successful reception. For example, if $\{X_k; k \in T\}$ is the subset of nodes transmitting simultaneously, the transmission from node $X_i$ to $X_j$ is deemed successful if

$$\frac{P_i}{N + \sum_{k \in T, k \neq i} P_k} \frac{X_i - X_j}{|X_i - X_j|^\alpha} \geq \beta,$$

where $P_i$ is the transmission power of node $X_i$, $N$ is the ambient noise power level, and $\alpha$ is the path loss exponent, which is usually 2 or 4. Using such physical models in the bandwidth allocation problem tends to complicate matters [2], and simpler models known as protocol models have been found to be more effective to date [3]. One of the most popular models is the clique con-
Figure 3. An example of the clique constraint approach.

An example of the clique constraint approach. The clique constraint is not the only protocol model considered in the literature. If it were assumed that nodes within range of each other have different frequencies, such as in [8], the only requirement is that no node simultaneously transmits and receives.

**RATE ALLOCATION SCHEMES**

Rate allocation schemes vary in objectives, approaches, and performance. Max-min fairness is a popular objective; a rate allocation is max-min fair if increasing the rate of one flow would cause another flow, already having an equal or lower rate, to decrease further. This implies that flows which contend with each other will each receive the same amount of bandwidth if possible.

References [4, 5, 7] all aim to allocate max-min fair rates. The difference is that [4] considers only single-hop flows; multihop flows are broken into multiple independent single-hop subflows. The disadvantages of doing so are seen from the example in Fig. 3. When max-min fair allocation for single-hop subflows is calculated, the rate for subflow \( f_{1,1} \) is larger than that of \( f_{1,2} \). However, since subflows \( f_{1,1} \) and \( f_{1,2} \) are both part of flow \( f_1 \), the rate of \( f_{1,1} \) is larger than \( f_{1,2} \). Packets from \( f_{1,1} \) will accumulate and cause congestion at node \( B \). Alternatively, if a reliable transport protocol such as TCP is adopted, the rates of \( f_{1,1} \) and \( f_{1,2} \) will stabilize to the same value; that is, \( f_{1,1} \) would not be able to achieve its allocated rate, and in other network topologies, this may cause the bandwidth to be wasted that could otherwise have been used by another flow.

As the end-to-end throughput determines the QoS perceived by users, we believe it is more reasonable to allocate rates to end-to-end flows, as done in [5, 7].

An end-to-end perspective is also taken in [6]; however, an objective different from max-min fairness is used; the objective is to maximize the sum of all flow rates (referred to in the literature as maximizing the spatial reuse of the transmission spectrum) subject to the constraint that each flow is guaranteed a basic fair share of the spectrum. The problem is formulated as a constrained optimization problem for the whole network and solved by an algorithm in a centralized manner. An alternate approximate optimal solution is proposed for the distributed manner of operation. From the scheme in [4] to the one in [6], the end-to-end perspective is obvious progress, although the distributed algorithm in [6] can only provide an approximate solution.

Some researchers are approaching the rate allocation problem from an economics perspective and seeking a market-based solution. Pricing [5, 7] is a good example of such schemes, and can be thought of in terms of market supply and demand. Actually, pricing has shown its efficiency long ago when applied to wireline networks [9]. In order to adopt the price-based approach, utility functions are used to characterize the resource requirements and the degree of satisfaction of individual users. Users are charged according to the overall price of the resources they are using. Different pricing structures and utility functions can lead to different algorithms, some better suited to distributed operations than...
others. In [5] users are charged by the sum of the prices of all the resources they use. Careful choices of utility functions allow proportional fairness or max-min fairness to be achieved. However, the proposed iterative algorithm has a relatively high computational complexity and a relatively long convergence time, which might limit its applicability.

A price-based rate allocation scheme is proposed in [7]. It charges each user the maximum price rather than the sum of all prices. It turns out that the fair rates are determined implicitly via a two-stage mechanism, which in essence, starts by transmitting packets in a round-robin fashion so that all flows receive the same rate, and then reduces the transmission rates for those flows experiencing congestion. What allows this scheme to work is the simplifying assumption that different nodes transmit at different frequencies; that is, only transmissions having a common node interfere with each other. The two-stage mechanism is as follows. Time is slotted. Consider a node and all the flows passing through or originating at this node. List these flows so that they can be examined in a round-robin fashion. At the start of each time slot, the node examines the next flow on the list. If the number of pending packets for this flow at the upstream node or downstream node significantly exceeds the number of pending packets for this flow at the current node, this flow is skipped and the next flow on the list considered. Otherwise, the node releases a packet for this flow; the packet is now pending transmission. (Technically, if the flow is being relayed by this node, the next time slot is not used; twice the bandwidth is required for flows being relayed compared with flows originating at the node.) Then the second stage, implemented in the MAC layer, is responsible for scheduling the packet transmission.

Another direction on rate allocation research is from the perspective of congestion control. The objective is to maximize the instantaneous flow rates, while ensuring stability of buffers and fair resource sharing. The rate control mechanism proposed in [10] is a dual congestion controller since it can be interpreted as a gradient algorithm for the dual of an optimization problem. While most rate allocation schemes require the flow rates to be changed instantaneously in response to feedback information from the network, a primal-dual congestion control mechanism is proposed in [11], which updates the flow rates gradually to mimic the response of TCP to congestion feedback.

Future work on rate allocation schemes in the network layer should focus on low computational complexity and fast convergence speed. Current research in this area combines theories from economics and mathematics, involving areas such as game theory, optimization theory, control theory, and graph theory.

**MAC Protocols**

Once the network layer has calculated the data rate to which each end-to-end flow is entitled, a mechanism is required in the MAC layer for controlling the access of single-hop flows to the medium in such a way as to ensure each flow receives its entitled data rate. Such mechanisms usually belong to one of two types: contention-based or cooperative.

Contention-based algorithms essentially have only two control mechanisms at their disposal, persistence probabilities and backoff timers, by which is meant the following. Each node having a packet ready for transmission attempts to transmit that packet with a probability called the persistence probability. (Setting the persistence probability to one disables this mechanism.) Should a collision occur (another node attempts to transmit a packet at the same time), each node involved in the collision will wait for a certain (possibly random) amount of time, called the backoff time, before attempting retransmission.

Cooperative algorithms pass small but extra amounts of information between nodes, allowing the nodes to coordinate their transmissions so as to avoid collisions. An example of a cooperative algorithm is given presently.

The algorithm in the relatively early [12] tries to achieve MAC layer fairness by adjusting the persistence probabilities. Unlike the other papers cited in this section, the algorithm it proposes does not interact with the network layer. That means it attempts to achieve a fair bandwidth allocation based solely on the levels of congestion each node sees; this limitation means it is not possible for end-to-end fairness to be achieved in general.

As discussed earlier, to achieve end-to-end fairness, the network layer must inform the MAC layer of the desired rate for each subflow. The MAC layer must therefore handle two types of scheduling, referred to in [6] as intra-node coordination and inter-node coordination. Intra-node coordination ensures each node transmits packets from different subflows in a fair order, so no subflow is unduly delayed or starved of bandwidth. Inter-node coordination ensures each node has a sufficient opportunity to transmit data, so no flow is unduly delayed or starved of bandwidth.

The scheduling algorithm adopted in [6] maintains different queues at each node for different subflows. When a packet reaches the head of a queue, it is tagged with the time it reached
Both intra-node and inter-node coordination mechanisms use in their calculations the start and finish times tagged to packets; these tags are shared between neighboring nodes by piggybacking on the short flow-control packets (RTS, CTS, and ACK packets).

<table>
<thead>
<tr>
<th>Distributed algorithm</th>
<th>[2][5][9][10]</th>
<th>[3][8]</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-to-end perspective</td>
<td>[4]</td>
<td></td>
</tr>
<tr>
<td>Low complexity</td>
<td>[7] + [13]</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Summary and comparisons of different schemes.

The motivation for bandwidth allocation is to ensure each user receives appropriate quality of service. Achieving this in wireless ad hoc networks requires the network and MAC layers to cooperate; the network layer determines (at least implicitly if not explicitly) an appropriate bandwidth allocation entitlement for each flow. Based on this determination, the MAC layer coordinates transmissions so that each flow receives its entitled bandwidth allocation. Factors to look for when comparing bandwidth allocation schemes are: how the network and MAC layers interact; how much extra information is passed around the network to support the algorithm; the time it takes for the flow rates to reach steady state (convergence rate); the interference model used; and the computational and implementation complexity of the scheme. Some previously proposed solutions to this problem are discussed, including one recently proposed by us that we believe demonstrates the feasibility of relatively straightforward and effective bandwidth allocation schemes in wireless ad hoc networks.

ACKNOWLEDGMENT
This work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. CityU 111208). We appreciate the anonymous reviewers whose comments led to improvement of the survey.

CONCLUSION

The motivation for bandwidth allocation is to ensure each user receives appropriate quality of service. Achieving this in wireless ad hoc networks requires the network and MAC layers to cooperate; the network layer determines (at least implicitly if not explicitly) an appropriate bandwidth allocation entitlement for each flow. Therefore, subflows requiring a large fraction of the bandwidth will gain slots more easily, and once they have gained their quota of slots, they will stop contending.

This procedure for gaining extra slots is called the Greedy Self-Contention (GSC) algorithm. An alternative algorithm is also proposed in [13]. This Cooperative Token Forwarding (CTF) algorithm removes the need for subflows in the same clique to contend for bandwidth. Enough information is exchanged between neighboring nodes to allow each of them to calculate the rate of each subflow as dictated by the network layer, and a token forwarding scheme is used to allow each node in turn to assign all its subflows to slots in one go.

In summary, the key performance indicators to consider when developing new MAC protocols for bandwidth allocation in wireless ad hoc networks are the implementation complexity, the convergence rate, and the amount of extra information passed among nodes. In contention-based MAC algorithms, often the limiting factor is the rate of convergence, as dictated by the quality of the contention resolution technique used. On the other hand, cooperative algorithms are limited by how efficiently and effectively they exchange information among neighboring nodes and how they use this information to coordinate transmissions among nodes.

Based on our major design criteria, Fig. 4 classifies and compares the schemes we have discussed in this section.
REFERENCES


BIographies

XUEYUAN SU received his B.E. degree in communication and information engineering from the University of Electronic Science and Technology of China in 2005, and his M.Phil. degree in electronic engineering from the City University of Hong Kong in 2007. Currently, he is pursuing his Ph.D. degree in computer science at Yale University. His research interests mainly lie in the area of parallel and distributed computing, and its various applications.

SAMMY CHAN (eeschan@cityu.edu.hk) is an associate professor in the Department of Electronic Engineering, City University of Hong Kong. He joined the university in 1994, after working with Telecom Australia Research Laboratories for six years. Over the years he has worked on traffic management in ATM networks, ATM switch architectures, handoff techniques for mobile networks, optimal file placement in video on demand systems, and TCP over wireless networks. Recently, his research interest has shifted to resource allocation and routing in ad hoc networks.

JONATHAN MANTON holds a distinguished Chair at the University of Melbourne with the title Future Generation Professor. He is also an adjunct professor in the Mathematical Sciences Institute at the Australian National University. Previous positions include Queen Elizabeth II Fellow, professor at the Australian National University, and executive director of the Australian Research Council. His traditional research interests range from pure mathematics to electrical engineering. Recently, though, led by a desire to participate in the convergence of life sciences and mathematical sciences, he has started to apply his expertise to research in neuroscience.