Research Challenges in Wireless Networks: A Technical Overview

Sanjay Shakkottai and Theodore S. Rappaport

Abstract— **This technical overview outlines some key network research issues, and differentiates the features of wireless networks and the services that they will support. As the Internet becomes ubiquitously delivered on mobile platforms, new approaches to wireless networking that build upon the advances of wireline networks and cellular system design will be required.**

Keywords— **Multi-user scheduling, wireless LAN, wireless network deployment, wireless data.**

I. Introduction

There has been rapid growth of wireless technology in the past few years, and when coupled with the explosive growth of the Internet, it is clear that there will be increasing demand for wireless data services [33]. Traffic on future wireless networks is expected to be a mix of real-time traffic such as multimedia teleconferencing and voice, and data-traffic such as WWW browsing and file transfers, with users desiring diverse quality of service (QoS) guarantees for different types of traffic. Recently, various mechanisms have been proposed and deployed to support data traffic over wireless media. This has ranged from Wireless Local Area Networks (W-LANs) based on the IEEE 802.11b standard to wide area data networks over wireless telephony services (2.5G and 3G wireless systems).

In this paper, we provide an overview of data services over wireline networks (Section II). In Section III, we describe some fundamental differences between wireline and wireless networks. We then proceed to highlight the salient features of wireless local area networks (W-LAN) and cellular/PCS data networks (Section IV). Next, in Section V, we discuss how various data services over the wireline internet can be supported over wireless networks. Finally, in Section VI, we pose various problems for future research.

II. Services over Wireline Networks

The **wireline network** today consists of thousands of access networks (such as campus wide area networks or small Internet Service Providers) where each supports tens of thousands of users, and these access networks are interconnected by core networks (such as AT&T or Sprint backbone networks) which support hundreds of millions of users. These networks primarily offer two types of services: **guaranteed service** and **best effort service**.

In guaranteed service, the network gives some sort of service guarantee to individual users or groups of users. These guarantees could be in the form of ensuring that the throughput for a group of users is greater than some minimum value or that the delay experienced is smaller than some threshold.

In best effort service, the network makes no promises. This service is typically used by *elastic traffic*. Elastic traffic consists of traffic where users do not necessarily have any minimum requirements, but would like to get as much data through to their respective destinations as quickly as possible. Individual user data flows react to congestion in the network and *adapt* their transmission rate with the aim of minimizing congestion.

In **Guaranteed Services**, some users demand that their traffic receive preferential treatment. For example, a company may decide that it would like dedicated bandwidth between its offices (for example, to implement a Virtual Private Network (VPN)). Then, the network is instructed to reserve bandwidth for this "user" by implementing simple mechanisms in the network infrastructure, such as priority scheduling instructions to each router on the network, to assure that the user is given bandwidth priority. The VPN is an example of traffic that requires *guaranteed service* (in this case, bandwidth availability that exceeds some minimum specification).

Another example of traffic which may require guaranteed service is real time streaming data, generated by applications such as packet video or Internet telephony. Here, the application sends streaming data into the network and often requires Quality of Service (QoS) guarantees in the form of small end-to-end delay and loss. Protocols such as RTP [35] provide support for such traffic including timing recovery, security and content identification. Guaranteed services over the Internet can be implemented over UDP, which is a connectionless data transport protocol that by itself is unreliable due to the fact that UDP simply dumps packetized data into the network and has no correction or recovery mechanism. Higher layer protocols must then implement error recovery or QoS support over UDP depending on the application's requirements. Network support in the form of *differentiated services*, where different users or service classes are treated differently or prioritized by the network infrastructure, may be needed to support guaranteed services [1].

Best Effort Service The prevalent protocol for best effort service for elastic traffic over the Internet is TCP [19]. TCP is a connection-oriented, end-to-end data transfer protocol. It has two objectives, (i) reliable end-to-end transmission of data, achieved by error or loss detection and retransmission, and (ii) congestion control over the Internet. Routers in the network indicate congestion by dropping packets, which in turn causes the source to adaptively decrease its sending rate. In the next generation TCP proposals, packets are marked [13] when congestion occurs. That is to say, in today's TCP, the source adapts its transmission rate based on its failure to receive an acknowledgment (ACK) before some time-out. Next generation TCP has a mechanism called Explicit Congestion Notification (ECN) [13] which

The authors are faculty in the Wireless Networking and Communications Group (WNCG), Department of Electrical and Computer Engineering, The University of Texas at Austin, {shakkott, wireless}@ece.utexas.edu. Invited paper to appear in the Proceedings of the Fifth International Symposium on Wireless Personal Multimedia Communications, Honolulu, HI, 2002.

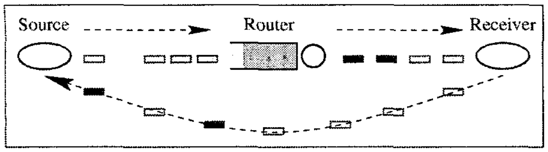

Fig. 1. Marking and rate adaptation in TCP. Marked packets are shown as filled rectangles, and the source is seen to decrease its transmission rate when marks are received.

is used to notify the receiver whenever congestion occurs in the network. This information is eventually sent back to source which then adapts its transmission rate. We have illustrated this behavior in Figure 1, where the source adapts its transmission rate based on network feedback in the form of marked packets. Typical best effort service applications that fall in this category include file transfers and web access.

III. DIFFERENCES BETWEEN WIRELINE AND WIRELESS NETWORKS

Before studying mechanisms by which data services can be supported over wireless networks, we briefly describe the characteristics of wireless systems. Based on channel considerations, wireless networks have two peculiarities which distinguish them from conventional wireline networks.

(i) The wireless channel is time varying, has memory due to multipath and is subject to errors. This causes *bursts of errors* to occur during which packets cannot be successfully transmitted on the link.

(ii) Channel state varies randomly in time on both slow and fast time scales [20], [32]. **Fast channel variations** due to fading are such that states of different channels can asynchronously switch from "good" to "bad" within a few milliseconds and vice-versa. **Slow channel variation** means that the average channel state condition depends on user location and interference level. Therefore, some users may inherently demand more channel access time than others based on their location or mobile velocity, even if their data rate requirement is the same as other users.

Another important difference that exists between wireless and wireline networks has to do with the effect of mobility on the IP routing protocol (network layer). In a cellular network, mobility implies that users will move between multiple cells. This means that network protocols must adapt packet routing paths so that a user continues to receive data as it moves from cell to cell. In this paper, we do not deal with multi-cell mobility issues, but refer the reader to [22], [31] for more details. More recently, attention has been given to ad-hoc networks, where all nodes of a wireless network may be mobile. Unlike wired networks, ad-hoc (or sensor) networks have new challenges due to node failures and unreliable channels in a power-limited environment [21], [30], [10].

IV. WIRELESS NETWORK ARCHITECTURES

Based on scale, we classify today's wireless networks into two classes, (i) wireless local area networks (W-LANs) and (ii) cellular/PCS wireless wide area networks.

A. Wireless LANs

A wireless LAN is a data transmission system designed to work over short distances, typically within buildings. The most popular Wireless LAN implementation use the IEEE 802.11b standards [2]. This standard describes the physical layer as well as the data link layer for media access, and any higher layer protocols such as TCP/IP or RTP/UDP/IP operate on top of these layers. The 802.11b standard allows the network to operate in one of two modes: the *infrastructure mode* and the *ad-hoc mode*. The architecture in the infrastructure mode consists of base-stations (access points) connected to the wireline network, and the mobile users connecting to these access points. On the other hand, in the ad-hoc mode, wireless mobile users communicate directly among themselves in a peer-to-peer fashion.

The unique feature of a 802.11b based wireless LAN is the Media Access Control (MAC) sub-layer of the data link layer. The MAC uses the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) mechanism. In this scheme, when a mobile user has data to transmit, the user first senses the carrier. If the user detects no other transmission, the user waits for an additional random amount of time, and transmits if the carrier is still available. If the data packet is received successfully by the base-station, it sends an acknowledgment (ACK) back to the mobile. If the ACK is not received by the mobile (either due to the data packet or the ACK being corrupted), the mobile retransmits the data using the same procedure as before. This MAC is invisible to the higher layers (e.g. TCP or UDP), but presents itself as a time-varying, but reliable channel. This has serious implications of TCP performance due to the fact that this time-varying *channel behavior* is interpreted by TCP as *congestion in the network*, causing TCP to decrease its transmission rate, whereas what should ideally happen is that TCP should increase its time-out to match the MAC's attempt to successfully connect to the channel.

B. Cellular Wireless Networks

Another way to architect a wireless network is to overlay data networks over wireless telephony services. This approach has been in used in the cellular/PCS industry. For brevity, we just consider the forward-link problem, where the direction of the data flow is from the base-station to the mobile user. In most of the cases we study here, the extension to the reverse-link problem where the data flows from the mobile users to the base station is straightforward.

Consider a cellular system with a fixed base-station and a number of mobile users. Data flows (packets) arrive from the wired Internet to the cell base station and are destined to the mobile users, with the packets for each user being queued temporarily at the base station (a separate queue is maintained for each user). The objective of the base station is to schedule these packets to various mobiles in a timely manner.

The earliest method used to send data to a mobile user on a cellular network consisted of sending data over a voice channel by means of a modem over a FDMA/TDMA network. As data traffic is typically bursty, this method has the disadvantage that when a user is not transmitting data, the bandwidth on the voice channel is wasted.

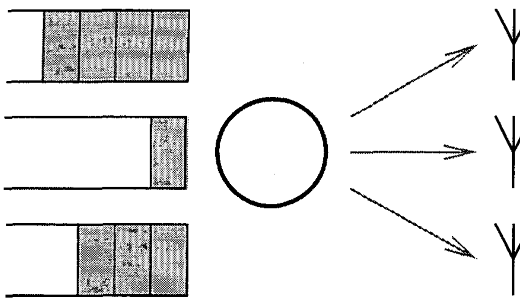

Fig. 2. Time-division scheduling: Queues at the base-station.

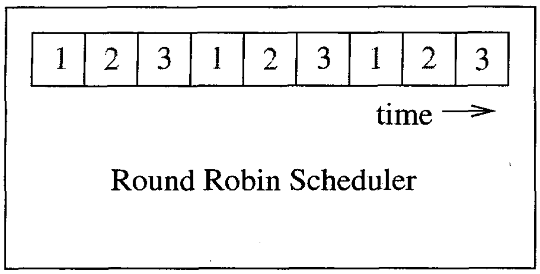

Fig. 3. Time-division scheduling based on the round-robin strategy.

A more efficient way is to dedicate some bandwidth for data transport and allocate this bandwidth in a dynamic manner among various data users. For instance, time could be divided into fixed size time-slots and users could be allocated time-slots in a dynamic manner. The simplest scheduling algorithm is illustrated in Figure 3, where users are periodically allocated slots irrespective of whether there is data to be sent to a particular mobile user. Note that this simple TDM scheme leads to wasted bandwidth just as in the previous case. Further, it suffers from a more subtle problem, namely, this approach is independent of the channel state (it is a channel state independent scheduling mechanism).

The 3G wireless systems have a mechanism built in where the forward-link channel state information for each user is fed back to the base-station. For example, in HDR used in cdma2000, the channel state is fed back once every 1.667 msec. This slot size is short enough so that each user's channel quality stays approximately constant within one time-slot. In each time-slot, one user is scheduled for transmission. Each user constantly reports to the base station its "instantaneous" channel capacity, i.e., the rate at which data can be transmitted if this user is scheduled for transmission. In an HDR system (and in the generic variable channel model, as well) a "good" scheduling algorithm should take advantage of channel variations by giving some form of priority to users with instantaneously better channels.

Examples of such systems include HDR over CDMA [6] and 1xEV-DV over CDMA proposed by Qualcomm Inc., and EDGE over GSM [32]. These proposals are essentially time-division overlays over the underlying physical-layer protocol to enable data traffic, and have the flexibility of allowing smart scheduling mechanisms to improve efficiency based on instantaneous user demand and channel states for each user.

Example IV.1: We illustrate the significant gains due to channel state dependent scheduling algorithms by means of a simple example. Consider a wireless system consisting of three users as in Figure 2. For this example, we consider access time to be slotted, and the channels to be constant over a time-slot. We assume that the channels are either ON or OFF, equally likely, and the channels being independent of each user. Thus, in this system, there are eight possible (instantaneous) channel states ranging from (ON,ON,ON) to (OFF,OFF,OFF).

When a user's channel is ON, one packet can be transmitted successfully to the mobile user during the time-slot. The system is assumed to be a TDM system; thus, the base-station can transmit to only one user each time-slot. The associated scheduling problem is to decide which user is allowed access to the channel during each time-slot.

A naive scheduling rule would be to employ the round-robin mechanism as shown in Figure 3. In such a scheme, the users periodically get access to the channel, with each user getting $1/3$ of the slots. As the channel of each user equally likely to be ON or OFF in each time-slot, it follows that over time, on an average each user will get a data rate of $1/6$ packets/slot.

On the other hand, suppose that the base-station had knowledge of the instantaneous channel state. Then, a simple policy would be to schedule a user whose channel is in the ON state. If more that one user's channel is in the ON state, then, pick a user randomly (equally likely) among those users whose channel is ON. This simple example assumes that all the users have identical traffic demands. For the case where some users have greater needs than others, such users would be assigned access with greater likelihood.

We observe that this policy ensures that *no data is sent by the base-station if and only if all users' channels are OFF* (which occurs on an average, $1/8$ of the time). Thus, the *total data rate* achieved by this state dependent rule is $1 - 1/8 = 7/8$ packets/slot. As this rule is symmetric across users, it follows that on an average the *data rate per user is* $7/24$ packets/slot, which is almost twice the throughput as the round-robin scheme which provided $1/6$ packets/slot. In addition, the base station does not radiate power during bad channel conditions, thereby decreasing interference levels in the wireless system.

This gain achieved due to channel-state dependent scheduling is called the *multi-user diversity* gain [23], [44].

Example IV.2: We next illustrate the idea of the **throughput region** of a multi-user wireless system. Consider a wireless system with two users. Suppose that packets arrive from the wireline network to the base-station, are temporarily buffered, and then need to be forwarded to the mobile users. Let us denote the rate at which data arrives to mobile users 1 and 2 (from the wireline network) by λ_1 and λ_2 packets/slot, respectively. A natural question that one can ask is what are the pairs of (λ_1, λ_2) that the wireless system can support without the queues at the base-station overflowing.

In this example, we illustrate the solution to the above question for a simple wireless system containing two channel states (the previous example consisted of eight channel states). The channel model is the following: In the first channel state, if user 1 is scheduled, the achieved rate is 'a' packets/slot, while if user 2 is scheduled, the achieved rate is 'b' packets/slot. Similarly, in the second channel state, if user 1 is scheduled, the achieved

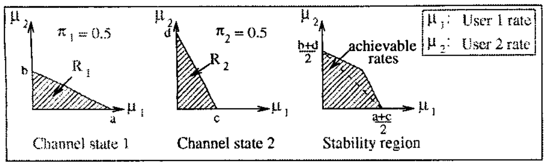

Fig. 4. Throughput region for a simple channel model.

rate is 'c' packets/slot, while if user 2 is scheduled, the achieved rate is 'd' packets/slot.

In Figure 4, we have illustrated this system. The left-most panel illustrates the first channel state. The x-axis indicates the rate allocated to user 1 in channel state 1, and the y-axis indicates the rate allocated to user 2. Suppose that whenever the wireless channel is in state 1, the base-station scheduler decides to transmit user 1 half the time and the other half of the time, user 2 is scheduled. Then, the average rate allocated to user 1 in state 1 would be $a/2$ packets/slot, and the average rate allocated to user 2 would be $b/2$ packets/slot. Similarly, we can choose various splitting rules for channel state 1 (for example, $1/5$ of the time to user 1, $3/5$ of the time to user 2 and do nothing for $1/5$ of the time). The shaded region in the left-most panel in Figure 4 is the collection of average rates to users 1 and 2 over all such splitting rules.

Similarly, the middle panel illustrates the corresponding achievable region in channel state 2.

Now, choose any $v_i \in R_i, i = 1, 2$, where R_1, R_2 are the achievable regions in states 1 and 2, respectively. Then, the **throughput region** is defined as $\mathcal{V} = \{v : v = \pi_1 v_1 + \pi_2 v_2\}$, where π_1, π_2 are the steady-state probabilities of being in channel-states 1 and 2, respectively. Observe that $\mathcal{V}$ is merely the convex combination of the achievable regions in each state. This is illustrated in the right panel in Figure 4. Thus, if the packet arrival rates (λ_1, λ_2) lie within the throughput region, we can be guaranteed that the queues will not overflow.

It can be shown that if the channel state *was not used for packet scheduling*, the throughput region would be that to the left of the dotted line in the right panel of Figure 4. In Example IV.1, we saw that the gain due to channel-state-dependent scheduling was almost twice that of channel-state-independent scheduling. In fact, we can argue that the larger the variation is in the channel, the larger is the gain due to state-dependent scheduling. For Raleigh fading wireless channels, state-dependent scheduling can lead to a doubling of the network throughput as compared to channel-state-independent scheduling [45].

The above examples illustrate the significance of *multi-user diversity* gain in network scheduling. However, a central question is how to design on-line algorithms which achieve this gain, while also supporting diverse quality of service requirements for various users. While this problem is not yet completely answered, there has been extensive research on various aspects of this problem [42], [44], [4], [38], [39], [12], [26], [28].

C. *Interoperability between W-LANs and Cellular/PCS*

As discussed earlier, wireless networks can be implemented as wireless LANs or wide area cellular/PCS networks. Interoperability between these distinct networks (one predominantly indoor and fixed/portable, and the other outdoor and mobile) is crucial for wide scale deployment of products and services over wireless networks. At the time of this writing, dozens of specialized applications, such as web phones and smart-cards are being designed for dual use on IEEE 802.11a/b and cellular/PCS networks [33].

V. CURRENT APPROACH TO DATA SERVICES OVER CELLULAR NETWORKS

In Section II, we described data services that are available over the Internet. In this section, we discuss means by which those services can be provided over a wireless network. Both best effort (through TCP) as well as guaranteed services are treated.

A. *Best Effort Services over Wireless Links*

It is well known that the TCP protocol is very sensitive to packet loss in wireless networks. Briefly, this is because TCP assumes that all packet losses occur due to congestion in the network, and reacts to this by aggressively decreasing the data transmission rate. In this section, we discuss various schemes that have been proposed to alleviate the effects of channel impairments, interference and other non-congestion-related losses on TCP performance over networks with wireless links [5].

There are two different approaches to improving TCP performance in wireless systems. The first is to hide any non-congestion related losses from the TCP sender. This requires no changes to the existing sender implementations. Protocols that adapt this approach attempt to make the lossy wireless link appear as a higher quality link with a reduced effective bandwidth. Hence, most of the losses seen by the TCP sender are actually caused by congestion, and not the wireless channel.

The second class of techniques attempts to cause the sender to be aware of the existence of wireless hops and to realize that some packet losses are not due to congestion. The sender can therefore avoid invoking congestion control algorithms when non-congestion-related losses occur. It is also possible to have a "wireless-aware" transport protocol to coexist with link-layer schemes to achieve good performance.

Based on the fundamental philosophy of these schemes, we can classify them into three groups ([5]), namely, (i) end-to-end approaches, (ii) split-connection approach and (iii) link layer based network protocols.

(i) The end-to-end approach: These proposals attempt to make the TCP sender handle losses through the use of two techniques. First, they use some form of selective acknowledgments (SACKs) to allow the sender to recover from multiple packet losses in a window without resorting to a timeout. Second, they attempt to have the sender distinguish between congestion and other forms of losses using an explicit loss notification (ELN) feedback mechanism.

(ii) Split-connection based approach: These approaches take data from a wireline network and terminate the flow at the base station. The base station then initiates a second, separate connection between itself and the mobile user. The data packets are temporarily buffered at the base station, and the second connection can use techniques such as negative or selective acknowledgments, rather than just standard TCP, to provide connectivity over the wireless link. In addition to latency induced by buffering, this approach has the disadvantage that as the mobile is handed off from cell to cell, the buffered packets need to be moved to the next new base station.

(iii) Link layer based network protocols: These try to hide the link-related losses from the TCP sender by using Forward Error Correction (FEC) and/or local retransmissions (ARQ) over the wireless link. The local retransmissions can be done such they are tuned to the characteristics of the wireless link and therefore can provide a significant improvement in throughput performance. However, there is a problem in that the TCP sender may not be fully shielded from wireless link losses and therefore there might be some redundant retransmissions. Today's IEEE 802.11b uses such an approach.

We refer to [5] for more details of the above schemes.

We observe that the schemes described above primarily focus on improving TCP performance over a single link. A question for future research is to study the above schemes in conjunction with multi-user scheduling, where the users could have widely differing bandwidth requirements.

B. Guaranteed Service over 3G Wireless Networks

The other major service class we have discussed is guaranteed service. We list now a few important questions that one can ask about this service in the wireless context.

(i) How can multiple *real-time* data users be supported simultaneously with good quality of service (QoS) for all users, namely, with packet delays not exceeding given thresholds with high probability?

(ii) How can a *mixture* of real-time and non-real-time users be supported simultaneously with real-time users receiving their desired QoS and non-real-time users receiving the maximum possible throughput without compromising the QoS requirements of real-time users?

(iii) How can bandwidth be fairly allocated among various users, especially when some users inherently demand more channel access time than others, even if their data rate requirement is the same.

Though these questions have not been answered fully, there has been extensive research on various aspects of the above problems. Below, we discuss some of the available results.

In [42], optimal scheduling for a wireless system consisting of N queues and a single server is studied. The arrival processes to each of the queues are assumed to be i.i.d. Bernoulli processes. The channel perceived by each queue is also assumed to be an i.i.d. Bernoulli process. The authors show that the policy which minimizes the total number of packets in the system in a stochastic ordering sense is the one which serves the longest queue. In [36], the authors have studied the problem of scheduling multiple real-time streams with deadlines, over a shared channel. The channel is considered to be ON-OFF. It is observed that a channel state dependent version of the earliest deadline date (EDD) policy is not always optimal in the sense of minimizing the number of packets lost due to deadline expiration. Zhang and Wasserman [47] study a dynamic program approach to forward-link scheduling, where for various amounts of channel information, the authors characterize policies which maximize a discounted cost using a dynamic program.

The authors in [7] study the effect of the wireless link on the performance of various scheduling algorithms using a simulation-based approach. They conclude that a channel-state dependent scheduler can lead to significant improvement in channel utilization for typical wireless LAN configurations. This idea is further explored in [14] in the context of fair queueing. Other related work on fair queueing in the context of wireless networks is available in [24], [25].

The authors in [23] were among the first to study wireless scheduling in multi-user context, and explicitly characterize the *multi-user diversity* gain. In [44], the author proposed a scheduling rule called the proportionally fair rule, which explicitly makes use of the channel state information and also provides fair allocation of bandwidth across users. A simulation based approach is available in [39] where the authors study means by which a mixture of real and non-real time users can be supported with users having different QoS requirements. Analytical justification for some of the results observed is provided in [38], [37].

In [45], the authors consider opportunistic beam-forming, and multi-user scheduling based on the proportional fair rule, and show that significant gains can be accrued due to multi-user diversity. Finally, recent work based on a utility based approach is available in [3].

C. Deployment of Network Access Points

In today's cellular/PCS networks, site deployment is conducted through computer-aided modeling of propagation characteristics, as well as the user densities and the corresponding RF interference levels [32]. In today's W-LAN deployments, however, IT professional are not accustomed to considering RF/interference issues. For efficient build-out of W-LANs and 3G networks, site specific methods of determining locations of access points and in-situ throughput performance of wireless modems offer promise [18], [34], [41]. However, with so many different user classes and data rate requirements, future deployment tools will need to "convert" from a wireless *system* design mindset with a *propagation* emphasis, to a wireless *network* design mindset with a user *throughput or QoS* emphasis. Early work shows that there are some first-order effects that relate the propagation parameters (SNR, RSSI) to end-user network parameters (throughput, delay), independent of manufacturer or user classes.

Consider a two user network based on 802.11b. The upshot of Figures 5 and 6 (see [18] for more details) is that, independent of the modem manufacturer and their proprietary designs, the end-user throughput appears to be described by functions of RSSI and SNR.

This holds promise for network deployment CAD software tools that IT professionals can use without knowledge of specific propagation issues while incorporating such vital data in their strategy. SitePlanner and LANfielder are examples of commercial deployment and measurement tools that incorporate a network-centric approach to wireless design [40].

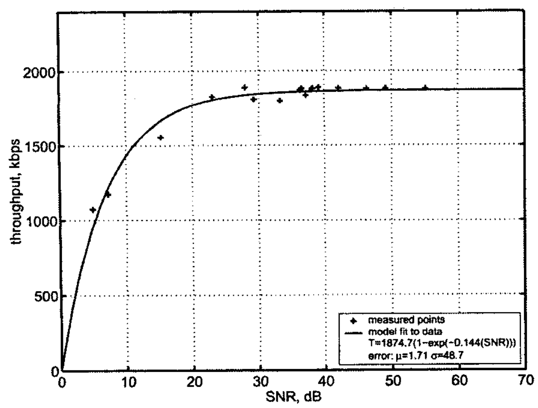

Fig. 5. SNR vs. throughput for WaveLAN WLAN 802.11b modem in an indoor environment. Throughput was measured on a round-trip basis between two nodes, using random messages and unprotected UDP.

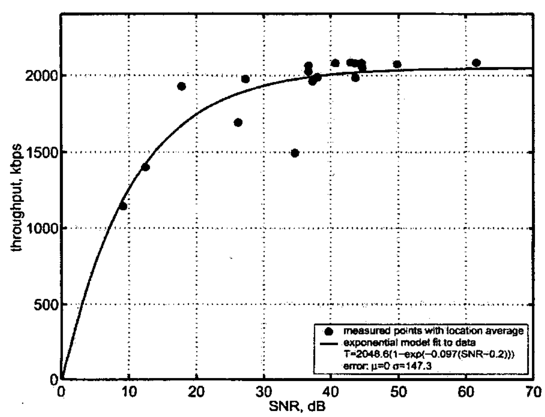

Fig. 6. SNR vs. throughput for 3COM WLAN 802.11b modem in an indoor environment. Throughput was measured on a round-trip basis between two nodes, using random messages and unprotected UDP.

VI. Future Research Problems

In this section, we discuss some problems that appear to have relevance for the future. We emphasize that the problems listed are by no means exhaustive, but provide an overview of technical challenges and emerging research areas.

(i) **Design of scheduling algorithms that interact well with TCP:** Most scheduling algorithms today are designed for flows which do not react to congestion. On the other hand, as well have discussed earlier, most studies of TCP over wireless restrict themselves to single user models. A design and analysis methodology for wireless scheduling algorithms specifically for controlled flows (such as TCP flows) and studying TCP in a multi-user context is an area of interest.

(ii) **Design for multi-carrier systems:** Most down-link scheduling algorithms today (for 3G wireless systems) are for a single-carrier (such as a shared time-division) system. Fourth generation wireless systems are expected to be based on a multi-carrier scheme (OFDM) with hundreds of carriers. In this scheduling context, an interesting area to explore is the design of good algorithms with low complexity.

(iii) **Fundamental properties of multiuser diversity:** It has been shown (see Section IV-B) that multi-user diversity gains can substantially improve wireless network throughput. However, given different types of QoS constraints, how does the diversity gain grow with the number of users? The answer to this question is of great relevance from a practical perspective. The associated design problem would be to come up with good algorithms which make use of multiuser diversity gain. Finally, most of the studies so far are confined to a single-cell scenario. Extending these studies to the multi-cell case would be of interest.

(iv) **Fast deployment of new algorithms:** With networking technology and services changing so quickly, an adaptable infrastructure which can quickly change with evolving technologies will be needed. Active networks [9], [43], [8], [29] is a promising area to enable fast deployment of new protocols and algorithms over existing networks.

(v) **Multi-hop networks:** In this paper, we have discussed two types of network architectures: wireless local area networks based on IEEE 802.11b standards, and wide area cellular networks. A fast emerging area of research lies at the intersection of these two architectures, namely, multi-hop wide area cellular networks, wherein the base station uses mobile users as *relays* to improve network capacity. An extreme case of such multi-hop networks are wireless sensor networks, where the number of nodes participating in multi-hop communications can range in the hundreds, if not more. Current research about these networks include capacity and scaling laws for large ad hoc networks [16], [17], [15], and design of distributed algorithms for routing, MAC, coverage and location identification [21], [30], [27], [10], [46], [11]. However, there are still many open problems in both the fundamental nature of these networks (for example, capacity and scaling with reliability issues, time-varying channels, spatial distribution of users, etc.) as well as practical, distributed algorithms for routing, congestion control and secure communication over such networks.

(vi) **Modeling network performance:** Understanding the properties of mixed traffic and service types over wireless networks in practical propagation environments will be crucial for deploying and analyzing the performance characteristics of future scheduling approaches and protocols. Creating models that translate actual channel characteristics, such as path loss and interference, into useful models for use at the network layer and above, while benchmarking such models with actual network field data will enable scheduling and traffic techniques to be developed accurately and expeditiously.

References

[1] J. Alvarez and B. Hajek. On using marks for pricing in multiclass packet networks to provide multidimensional QoS. Presented at *Workshop on the Modeling of Flow and Congestion Control Mechanisms*, Ecole Normale Superieure, September 2000. http://www.comm.csl.uiuc.edu/~hajek

[2] IEEE 802.11b standards. http://grouper.ieee.org/groups/802/11/main.html.

[3] R. Agrawal, A. Badekar, R. La, and V. Subramanian. Class and channel condition based weighted proportional fair scheduler. In *Proc. ITC-17*, Salvador De Bahia, Brazil, December 2001.

[4] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting. CDMA data QoS scheduling on the forward link with variable channel conditions, April 2000. Bell Laboratories Tech. Rep.

[5] H. Balakrishnan, V. Padmanabhan, S. Seshan, and R. H. Katz. A comparison of mechanisms for improving TCP performance over wireless links. *IEEE/ACM Transactions on Networking*, December 1997.

[6] P. Bender, P. Black, M. Grob, R. Padovani, N Sindhushayana, and A. Viterbi. CDMA/HDR: A bandwidth efficient high speed wireless data service for nomadic users. *IEEE Communications Magazine*, pages 70–77, July 2000.

[7] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. Tripathi. Enhancing throughput over wireless LANs using channel state dependent packet scheduling. In *INFOCOM '96*, 1996.

[8] S. Bhattacharjee, K. L. Calvert, and E. W. Zegura. A survey of active network research. In *IEEE Intl. Conf. on Network Protocols*, pages 220–228, Atlanta, GA, October 1997.

[9] M. S. Blumenthal and D. Clark. Rethinking the design of the Internet: The end-to-end argument vs. the brave new world. *ACM Tranacations on Internet Technology*, 1(1):70–109, August 2001.

[10] D. Braginsky and D. Estrin. Rumor routing algorithm for sensor networks, 2001. Available at http://lecs.cs.ucla.edu/~estrin/.

[11] L. Doherty, L. El Ghaoui, and K. S. J. Pister. Convex position estimation in wireless sensor networks. In *Proceedings of IEEE Infocom*, Anchorage, AK, April 2001.

[12] A. Eryilmaz, R. Srikant, and J.R. Perkins. Stable scheduling policies for broadcast channels. In *Proceedings of the IEEE Symposium on Information Theory*, Lausanne, Switzerland, July 2002. To appear. An earlier version appeared in *Modelling and Design of Wireless Networks, Proceedings of SPIE*, E. K. P. Chong (editor), vol. 4531, pages 70–78.

[13] S. Floyd. TCP and explicit congestion notification. *ACM Computer Communication Review*, 24:10–23, October 1994.

[14] C. Fragouli, V. Sivaraman, and M. Srivastava. Controlled multimedia wireless link sharing via enhanced class-based queueing with channel-state-dependent packet scheduling. In *Proc. IEEE Infocom*, pages 572–580, April 1998.

[15] M. Grossglauuser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *Proceedings of IEEE INFOCOM 2001*, Anchorage, Alaska, April 2001.

[16] P. Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming. Edited by W.M. McEneany, G. Yin, and Q. Zhang*, pages 547–566, Boston, 1998. Birkhauser.

[17] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, IT-46(2):388–404, March 2000.

[18] B. Henty. Throughput measurements and empirical prediction models for IEEE 802.11b wireless LAN (WLAN) installations. Masters Thesis, Dept. of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, August 2001, http://scholar.lib.vt.edu/theses/available/etd-08142001-172523/

[19] V. Jacobson. Congestion avoidance and control. *ACM Computer Communication Review*, 18:314–329, August 1988.

[20] W. C. Jakes. *Microwave Mobile Communications*. IEEE Press, 1993.

[21] D.B. Johnson, D.A. Maltz, and J. Broch. DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks. In *Ad Hoc Networking, edited by C. E. Perkins*, pages 139–172. Addison-Wesley, 2001.

[22] David B. Johnson and David A. Maltz. Protocols for Adaptive Wireless and Mobile Networking. In *IEEE Personal Communications*, 3(1):34-42, February 1996.

[23] R. Knopp and P.A. Humblet. Information capacity and power control in single-cell multiuser communications. In *Proceedings of the ICC*, Seattle, WA, June 1995.

[24] S. Lu, V. Bharghavan, and R. Srikant. Fair scheduling in wireless packet networks. In *Proceedings of ACM SIGCOMM*, 1997.

[25] S. Lu, V. Bharghavan, and R. Srikant. Fair scheduling in wireless packet networks. *IEEE/ACM Transactions on Networking*, 7(4):473–489, August 1999.

[26] X. Lu, E. P. K. Chong, and N. Shroff. Transmission scheduling for efficient wireless utilization. In *Proceedings of IEEE INFOCOM 2001*, Anchorage, Alaska, April 2001.

[27] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *Proceedings of IEEE Infocom*, Anchorage, AK, 2001.

[28] M. J. Neely, E. Modiano, and C. E. Rohrs. Power and server allocation in a multi-beam satellite with time varying channels. In *Proceedings of IEEE INFOCOM 2002*, New York, NY, June 2002.

[29] S. M. Nettles, S. K. Song, S. Shannon, and M. Hicks. Evolution in action: Using active networking to evolve network support for mobility, 2002. Submitted to IWAN.

[30] C.E. Perkins. *Ad Hoc Networking*. Addison-Wesley, 2001.

[31] Charles E. Perkins and David B. Johnson. Mobility Support in IPv6. In *Proceedings of the Second Annual International Conference on Mobile Computing and Networking (MobiCom'96)*, November 1996.

[32] T.S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 2002.

[33] T. S. Rappaport, A. Annamalai, R. M. Buehrer, W. H. Tranter Wireless communications: past events and a future perspective. *IEEE Communications Magazine*, Volume: 40 Issue: 5 Part: Anniversary, pages 148–161, May 2002

[34] T. S. Rappaport and R. R. Skidmore. Method and system for automated optimization of antenna positioning in 3-D. U.S. Patent 6,317,599, November 13, 2001.

[35] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A transport protocol for real-time applications, January 1996. RFC 1889.

[36] S. Shakkottai and R. Srikant. Scheduling real-time traffic with deadlines over a wireless channel. *Baltzer/ACM Wireless Networks*, 2002.

[37] S. Shakkottai, R. Srikant and A. L. Stolyar. Pathwise Optimality and State Space Collapse for the Exponential Rule *Proceedings of IEEE Symp. on Info. Theory (ISIT)*, Lausanne, Switzerland, July, 2002.

[38] S. Shakkottai and A. Stolyar. Scheduling for multiple flows sharing a time-varying channel: The exponential rule. *Translations of the AMS, A volume in memory of F. Karpelevich*, 2002. To appear.

[39] S. Shakkottai and A. L. Stolyar. A study of scheduling algorithms for a mixture of real and non-real time data in HDR. In *Proceedings 17th International Teletraffic Congress*, Salvador da Bahia, Brazil, September 2001.

[40] SitePlanner, Wireless Valley Communications Inc., http://www.wirelessvalley.com.

[41] R. R. Skidmore, T.S. Rappaport. System for creating a computer model and measurement database of a wireless communication network. U.S. Patent No. 6,442,507 , August 27, 2002.

[42] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, 39:466–478, March 1993.

[43] D. Tennenhouse, J. Smith, W. Sincoskie, D. Wetherall, and G. Minden. A survey of active network research. *IEEE Communications Magazine*, 35(1):80–86, January 1997.

[44] D. Tse. Forward link multiuser diversity through proportional fair scheduling, August 1999. Presentation at Bell Labs. Also, see Appendix of [45].

[45] P Viswanath, D. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6):1277–1294, June 2002.

[46] W. Ye, J. Heidemann, and Deborah Estrin. An energy-efficient MAC protocol for wireless sensor networks. In *Proceedings of IEEE Infocom*, New York, NY, June 2002.

[47] D. Zhang and K. M. Wasserman. Stochastic transmission policies for time-varying channels with memory under incomplete channel state information. In *Proc. 38th Annual Allerton Conf. Commun., Control, and Comput.*, Urbana, IL, 2000. University of Illinois.