Sādhanā Vol. 32, Part 6, December 2007, pp. 645–681. © Printed in India

A tutorial survey of topics in wireless networking: Part II

ANURAG KUMAR[1] and D MANJUNATH[2]

[1]Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012
[2]Department of Electrical Engineering, Indian Institute of Technology, Bombay, Mumbai 400 076
e-mail: anurag@ece.iisc.ernet.in;dmanju@ee.iitb.ac.in

MS received 24 October 2006; revised 13 April 2007; accepted 12 June 2007

Abstract. This is the second part of the survey of recent and emerging topics in wireless networking. We provide an overview of the area of wireless networking as that of dealing with problems of resource allocation so that the various connections that utilise the network achieve their desired performance objectives.

In Part I we provided a taxonomy of wireless networks as they have been deployed. We then provided a quick survey of the main issues in the wireless 'physical' layer. We then discussed some resource allocation formulations in CDMA (code division multiple access) cellular networks and OFDMA (orthogonal frequency division multiple access) networks.

In this part we begin with a discussion of random access wireless networks. We first provide an overview of the evolution of random access networks from Aloha to the currently popular 802·11 (Wi-Fi) networks. We then analyse the performance of the 802·11 random access protocol. We briefly discuss the problem of optimal association of nodes to Wi-Fi access points. Next, we consider topics in ad hoc multihop wireless networks. We first discuss topology and cross layer control. For the latter, we describe the important maximum weight link scheduling algorithm. The connectivity and capacity of randomly deployed networks are then analysed. Finally, we provide an overview of the technical issues in the emerging area of wireless sensor networks.

Keywords. Wireless networks; random access networks; 802·11 protocol; Wi-Fi networks; sensor networks.

This is the second part of the two-part survey of wireless networks. The separation of the survey into two parts is rather adhoc and it is important that both the parts be read as one unit and not independently. The section, figure and equation numbers are continued from Part I. This facilitates easy cross referencing. However, the references in each part are self contained. We begin the second part with a discussion of random access networks.

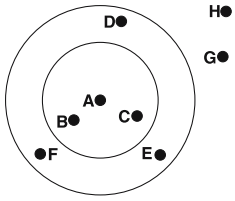

Figure 3. Decode and interference regions for a wireless transmission from A. The inner circle corresponds to the 'decode' region and the space between the two circles corresponds to the 'interference' region. When no other nodes are transmitting, the transmissions from A can be decoded at B and C. D and E can neither decode this transmission nor can they decode another transmission because the the transmission from A will cause significant interference. There will be insignificant interference at G and hence, while A is transmitting G can possibly decode another transmission, e.g. from H. Note that in practice, the shapes of these regions will not be regular.

6. Access networks: Distributed packet scheduling

In much of the discussion in the previous section, we see that there is a centralised mechanism of allocating the channel, (or spectrum or, more generally, a resource) to the nodes that want to transmit. We also saw that in cellular networks, this central control is located at the base station. In this section, we consider an alternative mechanism where the nodes coordinate the transmissions, i.e. the use of the resource, through distributed protocols. Such mechanisms were briefly discussed in § 1 while discussing figure 1.

A transmission on a wireless channel can be decoded by nodes that can receive the signal with a sufficient SINR. Thus, nodes geographically close to the transmitter will be able to simultaneously decode the received signal and at nodes that are not close to the transmitter, the transmission will cause interference. For nodes that are very far from the transmitter, the interference can be insignificant and it may be possible for that node to decode the signal transmitted by a node that is closer (figure 3).

Our primary interest in this section is in wireless networks in which the node locations are such that all nodes can decode the transmissions from all the other nodes in the network. Such networks are called single hop networks because all the nodes are one-hop away from each other. They are also called broadcast networks because every transmission can be decoded by every other node in the network. Colocated networks is yet another name for such networks.

A *multiple access* network is one where there is a single channel and many transmitters have to *share* the use of this channel in a suitable manner. The protocol that helps in such a sharing is called a multiple access protocol. Clearly, broadcast networks need to use a multiple access protocol. It is possible that more than one node of the network may want to use the channel at the same time. The protocol should have a mechanism to perform an arbitration among the nodes contending to use the channel and allocate the channel to only one of the contending nodes. This arbitration is typically a distributed algorithm and takes the form of prescribing transmission attempts and forced silences on the channel. Thus some of the capacity of the wireless channel is lost in the arbitration for channel access and the rest is available to the nodes for transmission of data packets. The fraction of the capacity lost in arbitration is a measure of the efficiency of the protocol. Simple protocols, even of low efficiency, are of interest if the available capacity is significant compared to the capacity required by the nodes in the network. In the following we will survey the more popular distributed multiple access protocols and their performance.

6.1 *Random access: From Aloha to CSMA*

Recall from § 1 that distributed scheduling or multiplexing algorithms can be of the 'random access,' 'polled access' or 'reservation access' types. In random access, when a node has a packet to transmit, it will transmit according to some randomisation. In polled access, a

distributed polling protocol, typically some variation of a token passing scheme, is used to determine the access rights to the channel. In reservation access, either random or polled access protocols are used to reserve the channel for use by a specific node at specific times. Our interest here is in random access protocols.

Random access protocols can be motivated by the following simple example. Consider a 10 Mbps channel that is shared by 100 users each requiring an average of 1 Kbps. The requirement of a user may occur in 'bursts,' e.g. each node having to send 1000 bit packets, on an average once every second. On a 10 Mbps channel it takes 100 μs to transmit the packet. Since the throughput requirement is small and bursty, TDM and polling schemes cause excessive delays. Further, for both these schemes to work correctly they need to know the number of nodes in the network. In most wireless networks, the number of nodes that may want to access the multiple access channel is typically a random number.

To motivate simple designs for the multiple access protocols, assume that a satellite channel is used in the above example. It is easy to see that if a receiver is simultaneously receiving transmissions from more than one node, then it cannot decode any of the transmissions because the signals interfere with each other. A collision is said to have occurred when there are such overlapping transmissions. In a satellite channel, the propagation delay is about 250 ms between all node-pairs. This means that the signal that is being received by a node at time t was transmitted at $(t - 250\,\text{ms})$. Thus if a transmitter, say node A, receives a signal while it is transmitting, say from node B, the signal from node B was transmitted 250 ms ago and all other nodes are receiving this signal at the same time. Further, since the packet transmission time is only 100 μs, the transmission from B is not going to interfere with the reception of the transmission from A. Thus the best option here seems to be to transmit the packet and 'hope for the best,' i.e. hope that none of the other nodes' transmissions overlap with it. Indeed, that is what the Aloha protocol does—whenever a node has a packet to transmit, it just transmits it. It can be shown that for Poisson packet arrivals, a large number of stations and fixed packet lengths, the maximum throughput that can be achieved, also called the capacity, is 0·18 packets per mean packet transmission time. We also see that for the example network that we consider above, this protocol suffices. The 'pure' Aloha scheme described above can be improved by dividing time into slots and restricting all packet transmissions to be confined to the slots. For such a protocol, it can be shown that 36 % of the slots can contain successful packets.

In a local area network where the propagation delays are small compared to the packet transmission times, the channel status can be sensed nearly instantaneously. (Typically, the maximum delay is bounded by limiting the geographic span of the network.) Hence, in such networks, nodes can sense the channel for the presence of a carrier and begin transmission only if the channel is free. Such networks are called *carrier sense multiple access* (CSMA) networks. Carrier sensing does not guarantee that there will be no collisions on the channel. This is because the propagation delay is non-zero and a second node can begin transmission before the signal from the first node to transmit on to an idle channel can be sensed by it. In pure CSMA networks, the cost to capacity of a collision is approximately equal to the longest of the colliding packets. The collision cost can be further minimised by either collision detection or collision avoidance.

In a wireline network, it is typically possible to continue sensing the channel even after a node begins transmission and this ability is exploited by *carrier sense multiple access with collision detect* (CSMA/CD) protocol. Ethernet uses such a protocol. Typically, in wireless networks, a node cannot receive a signal while it is transmitting in the same spectral band. This is because a significant fraction of the transmit power will be coupled back to the receiver. To detect a transmission from another node while it is transmitting, a node needs to distinguish

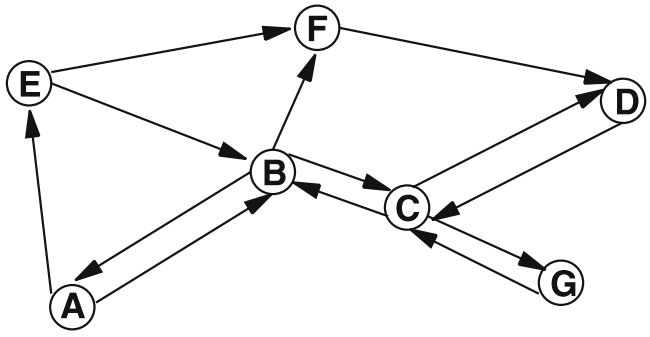

Figure 4. Hidden nodes in wireless multiaccess networks and the CSMA/CA protocol.

the presence of a weak signal in the presence of strong interference, i.e. the SINR will be very low and typically below the threshold for detection. This in turn implies that it is nearly impossible to detect a collision while transmitting. Here *collision avoidance* (CA) techniques can be used alongside CSMA thus yielding the class of CSMA/CA protocols.

To help understand the design of CA techniques, consider a multihop, multiaccess wireless network. Consider the node distribution shown in figure 4 and the corresponding network graph. Here a directed edge from node x to node y indicates that the transmission from x can be decoded at node y. (For a more general model, we can say that the transmission from x will interfere with a reception at y.) We call the nodes whose transmissions can be received at node x as the 'in-neighbours' of x. Likewise, the 'out-neighbours' of x will be those nodes that can receive the transmissions from x. In the network of figure 4, consider the following situation. If B is transmitting to C, then D cannot sense the carrier because it is not an out-neighbour of B. Hence, with just carrier sensing, node D can begin transmission and cause a collision at C. We say that D is *hidden* with respect to BC transmissions. Similarly, in figure 4, A is hidden with respect to CB transmissions. Note that CB transmissions could correspond to either data packets from C to B or acknowledgement packets in that direction in response to data packets from B. Thus BC transmissions could also be affected if A were to transmit while B is receiving an acknowledgement from C. For the example of a BC transmission, a collision avoidance protocol would essentially prevent the neighbours of C from causing a collision at C during a packet reception. It would also prevent the neighbours of B from causing a collision when it would be receiving an acknowledgement for the earlier packet. The basic idea of a collision avoidance protocol is to signal the start and end of data exchange to the neighbours of the transmitter–receiver pair so that these neighbours do not transmit during the data exchange.

6.2 *IEEE 802·11 (Wi-Fi) Networks*

A CSMA/CA protocol is used in the IEEE 802·11 ('Wi-Fi') networks. According to this standard, the network configuration could be either of the following.

1. Independent or 'ad hoc network' mode where the nodes communicate directly with each other. In this mode the nodes form an independent multihop wireless network. An appropriate routing protocol and a corresponding routing algorithm will be used to select the paths for the packet flows. The nodes in this configuration are called a basic services set (BSS).

2. 'Infrastructure' mode where wireless communication is always between a node (STA), and an access point (AP). The AP is connected to the wired network and provides a service similar to the base station of a cellular network. In this mode, the STAs need to 'associate' with an AP using an 'association protocol'. An AP and the STAs associated with it form a BSS and a set of BSSs is called an 'extended service set' (ESS).

We first discuss some physical layer (PHY) issues and then describe the medium access control layer (MAC) for these networks.

There are many PHY standards corresponding to different frequency bands in which the network operates and also the data rates that can be used by the nodes of the network. The initial 802·11 standard had three PHY standards—infrared, 1 and 2 Mbps frequency hopping spread spectrum (FHSS) and 1 and 2 Mbps direct sequence spread spectrum (DSSS). The next version defined two standards—(i) 802·11a in the 5·2 GHz band and (ii) 802·11b in the 2·4 GHz band. 802·11a uses OFDM with a data rate of 54 Mbps that could drop down to 6 Mbps depending on the channel conditions. The 802·11b uses a single carrier with a data rate of 11 Mbps that could drop down to 1 Mbps. Since they use different frequency bands and different modulation techniques, 802·11a and 802·11b are not interoperable and the latter is more popular and more widely deployed. The 802·11g is an extension of the 802·11b and uses DSSS, OFDM or both to support data rates in the range of 5·5 to 54 Mbps.

In addition to the data rates, that can be changed according to the channel conditions, multiple channels are defined for each PHY standard. In 802·11b and 802·11g, the centre frequencies of the channels are separated by 5 MHz and the 30 dB-bandwidth is mandated to be 22 MHz (i.e. the ratio of the peak energy to the energy 22 MHz away from the centre frequency is to be more than 30 dB) while the 50 dB bandwidth is to be 44 MHz. This allows some of the channels to become functionally non-overlapping, i.e. nodes communicating on these channels can co-exist in the same geographical area without causing significant interference to each other. Typically, it is assumed that channels 1, 6, 11 and 14 are non-overlapping. In 802·11a, 12 non-overlapping channels are defined. Table 1 summarises the above discussion.

We now describe the MAC protocols of the 802·11 networks. Two basic protocols are defined—a polling based protocol called the point coordination function (PCF) and a distributed access protocol called the distributed coordination function (DCF). Typically, PCF and DCF can co-exist in a network. The PCF is initiated by the point coordinator, usually the AP, by transmitting a beacon frame. The eligible nodes in the BSS are then polled and the data that needs to be transmitted to them are transmitted along with the polling message. If the polled node has data to transmit it will also transmit a packet in response to the poll.

Table 1. Summary of the 802·11 PHY standards.

Standard	Band	Data rates	Number of channels
IEEE 802·11 (dated)	2·4 GHz	2 &1 Mbps	1
IEEE 802·11b	2·4 GHz	11, 5·5, 2 and 1 Mbps	14
IEEE 802·11a	5·0 GHz	54, 48, 36, 24, 18, 12, 9 &6 Mbps	12
IEEE 802·11g	2·4 GHz	1–54 Mbps	14
IEEE 802·11n		work in progress, data rate up to 540 Mbps	

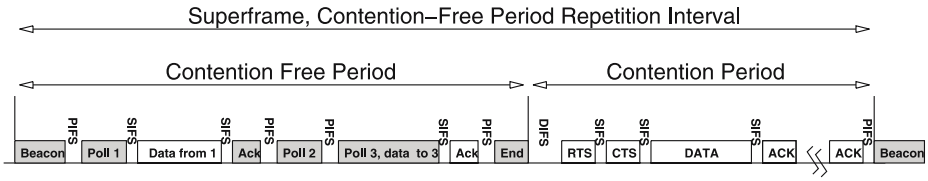

Figure 5. Beacon frame, PCF and DCF periods in an IEEE 802·11 network. The maximum duration for which the contention free period will last is called the *CFP_Max_Duration*. RTS is transmitted by sender and CTS by the intended receiver of the packet.

The end of the PCF mode of medium access is signalled using the 'End' frame. After the contention free period (CFP), a contention period (CP) using the DCF-based MAC begins which continues till the end of the superframe period in which the CFP and the CP alternate. Figure 5 shows the alternating sequence and also a sample of the types of packets that are transmitted in the network. The DCF uses a CSMA/CA protocol and we describe it next.

The basic ideas of the CSMA/CA protocol for IEEE 802·11 DCF were first described as the MACA (Karn 1990) and MACAW (Bhargavan *et al* 1994) protocols. The design of this CSMA/CA protocol assumes that the network is symmetric—the in-neighbours of a node are also its out-neighbours. The actual transmission of a packet is preceded by a handshake protocol in which the sender sends a *request to send* (RTS) packet. If the receiver receives the RTS successfully, then it transmits a *clear to send* (CTS) packet. In addition to alerting the receiver about an impending packet destined to it, the RTS allows the neighbours of the sender to defer transmissions till the sender receives an acknowledgement. Similarly, the CTS from the receiver enables the neighbours of the receiver to desist from transmitting while it is receiving a packet. This sequence is shown in figure 5.

An important aspect of the DCF protocol is the mandated silence periods between transmissions. The minimum silence that should be sensed by a node before beginning the transmission defines the priority for that transmission. Some silence between transmissions is necessary for the modems to turn around from transmitting to receiving, and vice versa, and also to start to synchronise to a new carrier. The excess is used as prioritisation mechanism. The smaller the spacing, the higher the priority. When a node wants to sense the network to initiate an RTS corresponding to a new packet transmission, it has to wait for a period called DCF interframe spacing (DIFS) while the nodes that transmit the CTS, the data packet following a CTS and the corresponding ACK, wait for a duration called the short interframe spacing (SIFS). Since SIFS $<$ DIFS, the handshake transmissions corresponding to an ongoing packet transmission have priority over new transmissions. The point coordinator waits for PCF interframe spacing (PIFS) after the finish of an ongoing transmission before transmitting a beacon or the polling packet. SIFS $<$ PIFS $<$ DIFS implying that the initiation of the CFP has priority over new transmission initiations. We will see later that multivalued DIFS can be used to prioritise among different traffic classes.

Since propagation delays in the network are non-zero, transmissions, especially RTS, could still be involved in collisions. Like in Ethernet and other random access protocols, colliding transmissions re-try after a random backoff delay. If a packet is involved in a collision, it chooses a random backoff interval to wait before making another transmission attempt. The number of slots (the slot duration is specified by the DCF standard) in a backoff period, is a random number uniformly distributed in $[0, CW_k - 1]$ where k denotes the number of collisions experienced by the packet. Typically $CW_k = \min\{2^k \times$

$CW_{\min}, CW_{\max}\}$. An important difference between the DCF and Ethernet is that in DCF counting down of the backoff timer occurs only when the channel is idle. This reduces the probability of a collision due to the accumulation of backoffs that expire during an ongoing transmission.

6.3 *HIPERLAN*

While 802·11 is an IEEE standard for wireless LAN, ETSI has defined a similar standard called the 'high performance radio LAN' (HIPERLAN). HIPERLAN is essentially like the IEEE 802·11 at the physical layer but has significant differences with it in the channel access method, which is called channel access control (CAC). Packets in HIPERLAN are assigned a priority. A brief description of HIPERLAN CAC follows.

The CAC of HIPERLAN uses an 'elimination-yield, non-pre-emptive multiple access' (EY-NPMA) mechanism. When a node has a packet to transmit, if the channel is sensed to be idle for a period called the 'channel-free-interval' then the node begins transmission of its packet. If the channel is sensed busy, then a contention resolution mechanism called the 'synchronised channel access' starts at the end of the current transmission. This access protocol has three phases—prioritisation, elimination and yield phases. The prioritisation phase reserves the channel for the highest priority packets that are contending for the channel. There are H slots in this phase and nodes that need to transmit packets of priority p transmit a radio pulse in slot p if the preceding prioritisation slots, i.e. slots $1, \ldots, (p-1)$, are idle. Of course, priority 1 is the highest priority. A radio pulse in slot p^* effectively reserves the channel for packets of priority p^* and prevents lower priority packets from competing for access. Let p^* be the highest priority packets that are contending for the channel during an instance of 'synchronised channel access'. Only nodes with priority p^* packets contend in the next two phases.

The second phase of the 'synchronised channel access' is called the elimination phase in which nodes transmit for a random duration from the beginning of the phase and those that do not transmit for the longest duration are eliminated. Specifically, each node with priority p^* packets transmits the carrier for a random number of slots, say E_i by node i, and then senses the channel immediately after that. If the channel is idle in the subsequent slot then it has survived the elimination phase and is eligible for the yield phase. The third phase is called the yield phase in which, the nodes that transmit the earliest win that phase. Specifically, each eligible node i.e. those surviving the elimination phase, listens to the channel for a random number of slots, say Y_i by node i, and then transmits the data packet. E_i and Y_i are chosen according to a truncated geometric distribution by each node. We reiterate that the nodes that choose the largest E_i win in the elimination phase while the node(s) that choose the smallest Y_i wins the yield phase. More than one node may transmit simultaneously in the yield phase causing a collision. A collision resolution mechanism is then invoked. Figure 6 shows a sample of the channel activities. See (Vukovic 1998; Anastasi *et al* 1998) for a throughput-delay analysis of this protocol.

A second version of HIPERLAN called HIPERLAN-II has been defined and is essentially a TDM scheme. Since the usage of this protocol is limited, we will not pursue it any more.

6.4 *Saturation throughput of a Colocated IEEE 802·11-DCF network*

As can be seen from the description of the 802·11-DCF MAC earlier, an analytical model of the DCF protocol can be quite complex and possibly intractable. However, by making a few reasonable approximations, a tractable model that can provide insights into the performance of this protocol can be obtained. We discuss one such model next.

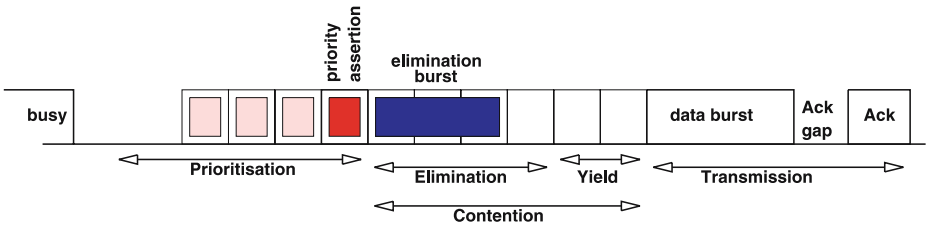

Figure 6. The prioritisation phase lasts four slots with packets of priority four surviving this phase. Nodes with priority 4 packets transmit in the elimination phase. All of these nodes transmit for three of less slots in this phase. Those nodes that transmitted for three slots in the elimination phase contend in the yield phase and the minimum yield was two slots. Note that the timing of the events are not to scale.

Saturation throughput analysis is an important development in understanding the performance of the CSMA/CA protocol in 802·11. Here we assume that all nodes will always have packets to transmit, i.e. a successfully transmitted packet is immediately replaced by another packet that needs to be transmitted. The throughput of the network under this saturation assumption is called the *saturation throughput*. Saturation throughput analysis has been used in systems before, notably in the study of switching systems. It is important to note that in general, the saturation throughput is not the same as the capacity of the network. It is, however, a good indicator of the capacity and for some special systems. For a class of input-queued switches (Jacob & Kumar 2001) and for a class of Aloha networks (Kumar & Patil 1997) it has been shown that the queues will be stable (stationary queue lengths will be almost surely bounded) if the arrival rate is less than the saturation throughput.

The saturation throughput model of (Bianchi 2000) is a popular one and is widely used. In the following, we describe a fixed-point analysis from (Kumar *et al* 2005) of a generalisation of the model of (Bianchi 2000). We consider a 'single cell' network of N saturated nodes in which the transmission of any node can interfere with the reception at all the other nodes. Thus for successful reception, only one node should be transmitting at any time. This model is applicable in the 'infrastructure mode' where all the nodes are associated with the same access point, or in the 'ad hoc' mode when the geographical spread of the networks is such that each node is within the interference range of all the other nodes. We further assume homogeneous nodes, i.e. the parameters of the backoff process and the state machine that implements it is identical at all the nodes. Different timeout parameters are used in IEEE 802·11e to provide service differentiation and will be discussed later.

The basic method of obtaining the saturation throughput is the renewal reward argument. When the channel is idle and the nodes are counting down, time is assumed divided into slots of unit length. In any slot, the channel could become busy or remain idle. A busy period begins from the time that there is a transmission and could correspond to a successful transmission, of

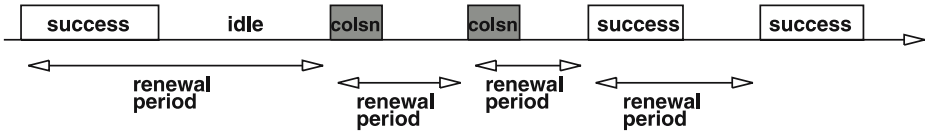

Figure 7. The channel alternates between busy and idle periods. The busy periods could correspond to collisions or to successful transmissions.

average duration T_s units of time (or T_s slots) , or to a collision, of average duration T_c units of time. Let T_i be the average number of idle slots between busy periods. Figure 7 illustrates this alternation between the busy and idle periods. Let P_s denote the probability that an attempt is successful. For the analysis model, we can show that the busy periods are i.i.d. and so are the idle periods. Further, the busy and idle periods can also be shown to be independent of each other. Then, using the renewal reward argument, the saturation throughput in an N-node network, $S(N)$, is given by

$$S(N) = \frac{P_s T_s}{P_s T_s + (1 - P_s)T_c + T_i}. \tag{13}$$

$S(N)$ is the fraction of time that the channel carries successful transmissions. In Eqn. 13, T_s and T_c are constants and are determined by the protocol parameters and the packet lengths while T_i and P_s depend on the randomness in the protocol execution. The saturation assumption is invoked here and we can say that during the idle periods all nodes are in a backoff state. To obtain the T_i and P_s, consider a tagged node in the network. Let β be the time average of the attempts per slot of each of the nodes during the countdown period. We assume that the aggregate attempt process of the other $(N - 1)$ nodes is independent of the attempt process of the tagged node. This is the key 'decoupling approximation'. We further assume that in each idle slot, each of the N nodes transmit with probability β. Thus if β were known, the number of slots in an idle period would follow a geometric distribution. Further, an attempt is successful if and only if exactly one node attempts a transmission in a slot conditioned on one or more nodes attempting transmission in the slot. From this the expressions for the mean idle period and the probability of success follow easily and can be seen to be

$$T_i = \frac{1}{1 - (1 - \beta)^N},$$

$$P_s = \frac{N\beta(1 - \beta)^{N-1}}{1 - (1 - \beta)^N}. \tag{14}$$

In the following we derive a fixed point equation for β.

Let K be the maximum number of attempts that a node will make to transmit a packet, i.e. a packet is discarded after K unsuccessful attempts. Let b_k be the average backoff period (in countdown slots) before the k-th attempt to transmit a packet. The constants K and b_k can be obtained from the protocol parameters. Let R_j be the number of transmission attempts and X_j be the total backoff duration (in countdown slots) for packet j from the tagged node. Clearly, both $\{R_j\}$ and $\{X_j\}$ are sequences of i.i.d. random variables. Let γ be the conditional probability that a packet transmission attempt results in a collision. Then R_j has a truncated (at K) geometric distribution and X_j is the sum of R_j random backoff periods. We see that

$$\mathsf{E}(R) = 1 + \gamma + \gamma^2 + \cdots + \gamma^{K-1},$$

$$\mathsf{E}(X) = b_0 + b_1\gamma + b_2\gamma^2 + \cdots + b_{K-1}\gamma^{K-1}. \tag{15}$$

Using renewal reward arguments, we can say that on an average, during a period $\mathsf{E}(X)$, $\mathsf{E}(R)$ transmission attempts are made. We can thus write the attempt rate, $\mathsf{E}(R)\,/\mathsf{E}(X)$, as a function of γ, say $G(\gamma)$ and we have

$$G(\gamma) := \beta = \frac{1 + \gamma + \gamma^2 + \cdots + \gamma^{K-1}}{b_0 + b_1\gamma + b_2\gamma^2 + \cdots + b_{K-1}\gamma^{K-1}}. \tag{16}$$

A transmission attempt by the tagged node will be successful if no other node attempts a transmission in the same slot. Since all the nodes are statistically identical and assumed to be independent, the probability that a transmission attempt by the tagged node will result in a collision is given by

$$\Gamma(\beta) = 1 - (1 - \beta)^{N-1}.$$

We thus have the following fixed point equation to obtain γ, the probability that a packet transmission attempt results in a collision.

$$\gamma = \Gamma\left(G(\gamma)\right) \tag{17}$$

Solving this equation will give us γ which can be used to obtain the attempt probability $\beta = G(\gamma)$ from Eqn. 16. This in turn can be used in Eqn. 14 to obtain P_s, the probability that a transmission attempt will be successful. A relaxation method, can be used to iteratively obtain the fixed point solution to Equation 17.

Given that there is a fixed point equation that models the system performance, an immediate question is the uniqueness of the solution. A non-unique solution could imply that the system has multiple 'operating points' each one corresponding to a different solution. This could affect the throughput of the protocol. In (Kumar *et al* 2005) it has been shown that Eqn. 17 has a unique fixed point if b_k is a non-decreasing sequence for $k \geq 0$.

Figure 8 plots the saturation throughput as a function of N for different protocol parameters. An interesting observation is that the saturation throughput does not vary much with the number of nodes, thus suggesting that the back-off mechanism correctly adapts the attempt probabilities to keep the throughput roughly constant.

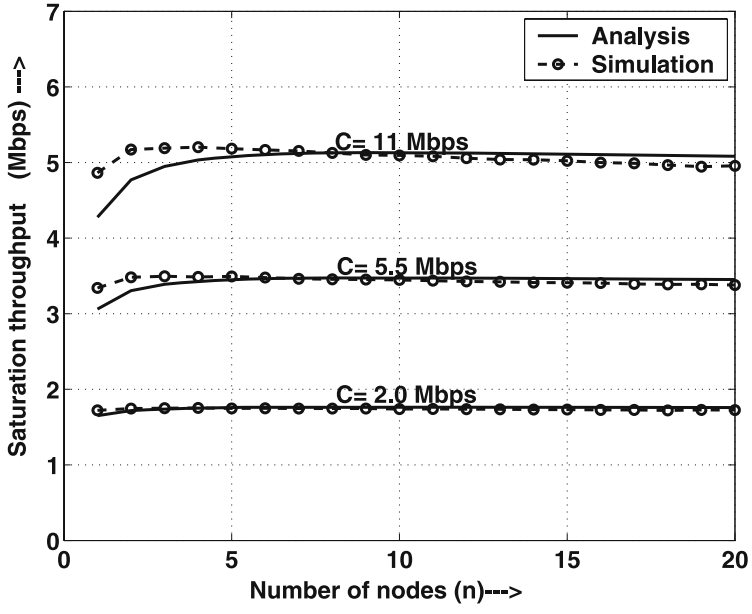

Figure 8. Saturation throughput as a function of the number of nodes in the network, for various PHY rates and packet size of 1500 bytes.

The decoupling approximation made above significantly simplifies the analysis. It is clearly a strong approximation. Nevertheless, it is by far the most popular of the 802·11 model because it provides good and easy insight into the behaviour of the 802·11 protocol. It is not a bad approximation because in (Bordenave *et al* 2005) it has been shown that as $N \to \infty$, this decoupling approximation is asymptotically exact. This implies that for large N, the decoupling approximation is a reasonable one. The goodness of the approximation is further evidenced by the analytical results from the model being close to those from simulations that model the protocol operation more accurately. Another reason for the popularity is that it lends itself to easy extensions that can be used to develop additional models some of which are cited below.

The above method can be extended to obtain a fixed point analysis for interfering WLANs. In (Panda *et al* 2005), the analysis is extended to a deployment that has two adjacent WLANs that completely interfere with each other but cannot decode each others' transmissions. A fixed point for 802·11 with capture is available in (Ramaiyan *et al* 2005b).

There have been many other analytical studies on the saturation capacity of the 802·11 DCF protocol. The study (Chhaya & Gupta 1997) accounts for the hidden terminals and capture of a signal by a receiver in the event of a collision. A capacity analysis of the basic handshake protocol is available in (Tay & Chua 2001). See (Cali *et al* 1998; Carvalho & Garcia-Luna-Aceves, 2003) for more approximate models for the throughput of the protocol. Numerical results from a Markovian model without the decoupling assumption are given in (Kumar *et al* 2005). More recently, (Sharma *et al* 2006b) has developed a Markovian model of the backoff process without the decoupling assumption. Here the system state at time t is described by $\mathbf{X}(t) = [X_i(t)]$ where $X_i(t)$ is the number of nodes in backoff stage i and $\sum_i X_i(t) = N$. It is shown that for a large number of nodes in the network, the system state stays close to a 'typical state'. Renewal reward arguments are used to obtain the throughput in this typical state.

There have also been many papers that go beyond the saturation throughput analysis. Tickoo & Sikdar (2004) present a delay analysis. The many parameters in the protocol, e.g. superframe duration, ratio of CFP and CP and backoff intervals, adapting them to the traffic conditions can improve the performance of the protocol. An example of such an adaptation is described in (Dong *et al* 2003) where past throughputs under the CFP and CP access methods are used as a feedback to adapt the PCF and DCF frame sizes to maximise the network throughput.

6.5 *Service differentiation and IEEE 802·11e WLANs*

QoS at the MAC layer can be provided by either of two mechanisms—service differentiation and per-flow 'time' reservation with admission control. The transmission rate on the channel depends on the channel characteristics for the sender–receiver pair and the throughput (almost always referred to as bandwidth in the literature) depends on the time and the transmission rate. We consider only service differentiation. Service differentiation mechanisms provide 'better than best effort' service and do not provide any absolute guarantees. Thus this kind of QoS is more suitable to elastic traffic. Service differentiation is provided by dividing traffic (or nodes) into different classes. Traffic of the same class compete with each other and receive 'best-effort-within-class' service while the classes receive different grades of service in the aggregate. A simple method would be to assign absolute priorities to the classes and provide a non pre-emptive priority service. (Recall that this is achieved by the prioritisation phase of HIPERLAN.) A second method would be to reserve capacity for the different classes. This can be accomplished in many ways.

- The CFP of the 802·11 can be used to serve different classes of traffic in different ratios by varying the polling rate and the duration for which the node is allowed to transmit when given the opportunity.
- The 802·11 DCF protocol parameters used by the different classes could be such as to give different success probabilities to the different classes.

The service provided to the different traffic classes, especially to the traffic classes defined for elastic traffic, could also be determined to achieve some fairness criterion, e.g. max–min or proportional fairness. Of course, any scheme that is proposed should be amenable to a distributed implementation. Recall that scheduling for fairness in centralised access wes discussed in § 5.2.

Essentially the first method above is used in the HCF Controlled Channel Access (HCCA) mechanism to provide service differentiation in the 802·11 protocol. This is a polling mechanism in which the AP, which in this case will also be called a hybrid coordination controller (HCF), starts a CFP at regular intervals. For each node, a service interval (SI) is calculated and the node is polled by the HCF with a period equal to SI. Every time it is polled, a node is allowed to transmit for a maximum duration of TxOP. The HCF, being a polling access method, standard analysis methods of polling systems will apply.

The second method listed above is used in the Enhanced DCF (EDCF), an extension to the IEEE 802·11 DCF, to classify and prioritise medium access among the traffic classes, called access classes (AC). Service differentiation is provided by mandating different interframe spacings and different backoff windows and the backoff window growth in the event of a collision for the different classes. Clearly, this provides service differentiation by changing the probability of obtaining channel access. The following parameters are defined for each AC.

- Arbitration interframe space (AIFS) that specifies the minimum number of slots that the AC should sense the channel to be free before attempting a transmission. Higher priority nodes will start backoff countdown earlier than the lower priority nodes and will, hence, have a higher success probability.
- Different minimum and maximum contention windows, $CW_{\min}$ and $CW_{\max}$, are specified for each class. Clearly, having a lower $CW_{\min}$ will increase the probability of success at the first and subsequent early attempts. Similarly, a lower $CW_{\max}$ will increase the success probability if the packet experiences many collisions.
- Different multipliers for increasing the value of CW after each collision are also specified.
- The transmission opportunity (TxOP) limit specifies the maximum time for which a node can transmit after acquiring the channel. Allowing high priority nodes a larger TxOP implies that their contention cost per bit (or packet) can be reduced.

Almost identical ideas were described in (Aad & Castelluccia 2001) to provide service differentiation within the framework of the 802·11 MAC.

Each node maintains a separate queue for each AC. An internal contention mechanism chooses the AC that will be transmitting a packet. This packet will be transmitted after the backoff counter has counted down and after the channel has been idle for the period specified by the AIFS of the AC. This allows the higher priority nodes to get a chance to transmit earlier than the lower priority nodes. Figure 9 describes this. Further, the higher priority may be allowed to transmit for longer durations based on their TxOP.

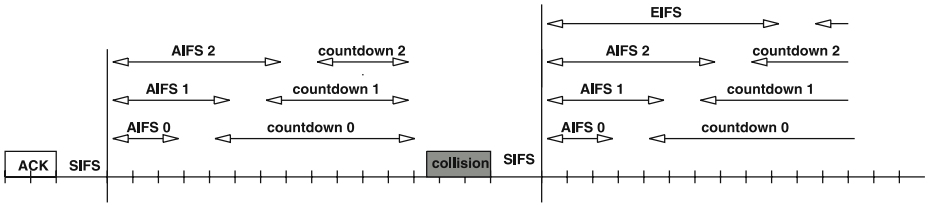

Figure 9. Three ACs (0, 1 and 2) are shown. The AIFS for the three classes and the countdown period is shown. The collision could be from either two nodes of the same AC or nodes from different ACs having their backoff counters reach 0 in the same slot. After a collision, an EIFS is defined during which the non-colliding nodes disable their countdown.

To analyse the EDCF access mechanism in the same framework as that of the single class network from the previous section, we can just extend the fixed point method analysis described in § 6.4 for multiple classes (Ramaiyan *et al* 2005a). Recall that the parameter in the saturation throughput model is the attempt rate. In the analysis of the EDCF, we assume that the effect of the different MAC parameters described above essentially translate to an attempt probability, β_i for class i and hence a different conditional collision probability γ_i. Thus we can generalise Equations 15 and 16 as follows.

$$\mathsf{E}(R_i) = 1 + \gamma_i + \gamma_i^2 + \cdots + \gamma_i^{K_i - 1},$$

$$\mathsf{E}(X_i) = b_{i,0} + b_{i,1}\gamma_i + b_{i,2}\gamma_i^2 + \cdots + b_K \gamma_i^{K_i - 1},$$

$$G_i(\gamma_i) := \beta_i = \frac{1 + \gamma_i + \gamma_i^2 + \cdots + \gamma_i^{K_i - 1}}{b_{i,0} + b_{i,1}\gamma_i + b_{i,2}\gamma_i^2 + \cdots + b_{i,K_i - 1}\gamma_i^{K_i - 1}},$$

$$\gamma_i = \Gamma(\beta_1, \ldots, \beta_n) = 1 - \prod_{j=1, j \neq i}^{n} (1 - \beta_j),$$

$$= \Gamma(G_1(\gamma_1), \ldots, G_n(\gamma_n)).$$

The above generalisation accounts for the different $CW_{\min}$, $CW_{\max}$, and the multiplier for increasing the value of CW after a collision, but does not take into account the different AIFS and TxOP for the different traffic classes. In the above, $b_{i,j}$, is the average backoff duration for AC i for the j-th attempt and K_i is the maximum number of attempts that a packet of AC i will make. The last equation above can be written compactly as

$$\boldsymbol{\gamma} = \Gamma(\mathbf{G}(\boldsymbol{\gamma})). \tag{18}$$

This is a vector fixed point equation and it can be shown that the fixed point exists. Depending on the $\{b_{i,j}\}$ and $\{K_i\}$ there will be many interesting properties for the fixed point and we enumerate some of these below. First, we make the following definition. If a solution to Eqn. 18 is such that $\gamma_i = \gamma_j$ for all $1 \leq i, j \leq n$, then we say that such a fixed point is balanced else we will say that it is unbalanced. Balanced fixed point essentially means that all the classes receive the same throughput.

- For the homogeneous case, if an unbalanced fixed point exists, then the solution is not unique because any permutation of a solution is also a solution of Eqn. 18.
- If all backoff and attempt parameters are the same for all the ACs, then by symmetry we can look for a balanced fixed point with $\gamma_i = \gamma$ for $1 \leq i \leq n$. If such a balanced fixed point exists, then it will be the unique balanced fixed point.
- However, it is possible that even when all the MAC parameters are the same, there exists unbalanced fixed points. This will correspond to multistability and short term unfairness in the system where some nodes get to use the channel for extended durations while locking out the other nodes. It is interesting to note that in the context of Ethernet, this has been investigated as the *capture* phenomenon (Molle *et al* 1987).

Using similar ideas, the model of (Sharma *et al* 2006b) can also analyse multiclass networks without the decoupling assumption made above. Li & Battiti (2003) analyse a network with service differentiation by a suitable adaptation of the model of Bianchi (2000). In the analysis of (Li & Battiti 2003), flows of different classes are assumed to use different DIFS. Each flow is modelled as a three-dimensional Markov chain $(s_i(t), b_i(t), d_i(t))$ where $s_i(t)$ is the state of the backoff counter, $b_i(t)$ captures the number of collisions experienced by the packet and $d_i(t)$ captures the number of slots that node i has to wait.

Adaptive algorithms have also been proposed to provide service differentiation. Romdhani *et al* (2003) describe an adaptive service differentiation scheme in which after a successful transmission the nodes update their CW adaptively by taking into account the estimated collision rate. Analysis of the 802·11 DCF with VoIP and TCP traffic is available in (Harsha *et al* 2006a).

6.6 *Association in IEEE 802·11 WLANs*

Recall that in the infrastructure mode of the 802·11 network, the wireless nodes (STAs) need to connect through an Access Point (AP), i.e. it should *bind* to an AP. Further, since an 802·11 network is allowed to operate over multiple channels some of which are non-overlapping, from the same physical location, an STA may have a choice of APs to which it can bind. Since the radio path between the STA and the APs could be different, it is possible that the transmission rates at which these associations can be made are different. Thus we have to solve a problem of association of the STAs to APs.

An obvious rule is for the STA to associate with the AP whose 'received signal strength indicator' (RSSI) is the highest. This would seem reasonable because the RSSI is an indicator of the proximity of the AP to the STA and also the transmission rate at which the AP-STA binding can be made. However, this rule does not take into account the carried load on the AP which would impact the QoS that the STA can obtain through that AP. An association scheme that takes into account the QoS that can be obtained by the STA through the APs needs to be devised.

A simple scheme is for the STA to associate with the AP that has the highest 'potential throughput,' i.e. the maximum rate at which data can be transferred if the STA is associated with that AP. Vasudevan *et al* (2005) describe the following algorithm to calculate the potential throughput using the delay in the beacon transmission time measured by the STA. Recall that associated with the beacon frame is a 'target beacon transmission time' (TBTT) which is the time at which the 'periodic' beacon frame needs to be transmitted. If the load on the AP is high, then the beacon frame will be delayed. Let this delay be denoted by T_B. Note that T_B is the delay from the time that the beacon frame was ready to be transmitted and the time at which it was successfully transmitted and can be called the access delay. We will assume that

T_B will also be the expected access delay that will be experienced by a data packet, we can calculate the expected time between successful transmissions of packets on the network, T, to be

$$T = T_B + T_{\text{RTS}} + (T_{\text{SIFS}} + T_{\text{CTS}}) + \left(T_{\text{SIFS}} + \frac{DATA}{R}\right) + (T_{\text{SIFS}} + T_{\text{ACK}}).$$

where T_{RTS}, T_{CTS} and T_{ACK} are the transmission times of RTS, CTS and ACK respectively, T_{SIFS} is the duration of SIFS, *DATA* is the packet length and R is the transmission rate under which the association is sought. The potential throughput is calculated as $\frac{DATA}{T}$. The STA can calculate the potential throughput for all the APs and join the one that offers the highest potential throughput.

In the above scheme the AP is passive in that it does not influence the decision of the STA. The APs could be made to play an active role by periodically broadcasting their load in the beacon messages. Each STA can then decide on the AP that it wants to associate with by considering both the load on the AP and the RSSI to the AP. One such scheme is described in Kasbekar *et al* (2006). In this scheme, a new STA will first calculate a reward function for each of the APs that it can associate with, say C_i for AP i. It will associate with the AP for which the reward function is the maximum. An example reward function is $C_i = \alpha T_i + \beta r_i$ where T_i is the throughput that the STA can obtain if it associates with AP i and r_i is the transmission rate with which it can associate with AP i. Of course T_i is a function of the transmission rates of the STAs associated with AP i. In Kasbekar *et al* (2006), stability analysis is used to determine the load profiles that can be handled with this reward function by the basic policy and its variations.

The above strategies can be considered to be 'online' strategies where an 'arriving' STA measures the grade of service that it can get from each of the APs that it can associate with and associates with the best AP. In contrast, (Tan & Guttag 2004; Bejerano *et al* 2004; Kumar & Kumar 2005) describe 'offline' strategies. Here each STA specifies its throughput demand and the transmission rates with which it can associate with each of the APs in the network. A centralised algorithm decides the optimal association. We will now explain these schemes briefly.

Tan & Guttag (2004) describe the following greedy algorithm to accommodate as many of the demands as possible. Let b_i be the throughput that STA i wants to achieve and $r_{i,j}$ the transmission rate with which STA i can associate with AP j. The 'best fit decreasing' greedy bin packing algorithm is adapted as follows. Let $u_{i,j} := \frac{b_i}{r_{i,j}}$. Note that $u_{i,j}$ is the fraction of time that STA i would be transmitting data if it were to associate with AP j. Sort the STAs in increasing order of b_i and associate the STAs from the beginning of the list. Associate the k-th STA in the sorted list to AP j that has the lowest utilisation and break ties randomly. An improvement of this is also suggested (Tan & Guttag 2004).

An algorithm for max–min fair allocation of the STAs to the APs is described (Bejerano *et al* 2004). STA i is assumed to have a weight w_i. Let $r_{i,j}$ be the transmission rate that STA i can obtain with AP j. A generalised version is first considered in which STA i gets throughput $b_{i,j}$ from AP j with $\sum_j b_{i,j} = b_i$. Let $\rho_i := \frac{b_i}{w_i}$ denote the normalised bandwidth allocated to STA i and $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_n]$. An association $[b_{i,j}]$ is considered feasible if the utilisation of all the APs is less than 1, i.e. if

$$\sum_i \frac{b_{i,j}}{r_{i,j}} < 1.$$

Let $x_{i,j}$ denote the fraction of bandwidth of STA i that AP j provides. Clearly $\sum_j x_{i,j} = 1{\cdot}0$. Define the load on AP j as $\sum_i \frac{x_{i,j} w_i}{r_{i,j}}$.

A weighted max–min fair allocation of a bandwidth to the STA is an allocation in which the normalised bandwidth of one STA cannot be increased without decreasing that of another STA with the same or less normalised bandwidth. It has been shown that under a max–min fair allocation, all the STAs that are associated with an AP have the same normalised bandwidth allocation.

A weighted min–max load balanced association is one in which the load on an AP cannot be decreased without increasing the load on another AP that has the same or higher load. Under min–max load balancing, it can be shown that each STA is associated with APs with the same load. Further, the bandwidth allocated to STA i, $b_i = \frac{w_i}{y_j}$ where y_j is the load on any of the APs to which i is associated.

It has also been shown that a min–max load balancing association and the max–min fair allocation are equivalent.

Using the above results, a fractional association algorithm that is min–max load balanced has been derived. A rounding method is then used to provide the integral associations.

While a max–min fair algorithm seems a reasonable objective, a more detailed analysis shows that this may actually result in significantly decreased throughput. Kumar & Kumar (2005) provide some examples. The primary reason for this is that the STAs that associate at lower rates occupy more time on the channel every time they get access while those that associate at higher rates occupy less time and the probability of access does not depend on the transmission rate. An association that maximises the throughput can be shown to be very unfair. Hence the following is suggested.

Let $\mathbf{A} = [A_{i,j}]$ be the association matrix with $A_{i,j} = 1$ indicating that STA i is associated to AP j. Let $\theta_i(\mathbf{A})$ be the throughput received by STA i when the association matrix is $\mathbf{Z}$. Let the 'utility' of STA i from an association matrix $\mathbf{Z}$ be given by $\log(\theta_i(\mathbf{A}))$. An optimal association would then maximise

$$\sum_i \log(\theta_i(\mathbf{A})).$$

subject to the capacity constraints at each of the APs and also the unique association constraint given by

$$\sum_i A_{i,j} = 1 \text{ for all } j.$$

WiFi hotspots are being widely deployed in public spaces. An important issue in such deployments is the pricing of the association. Mussachio & Walrand (2006) studied the situation where the STA pays directly to the AP is considered and a game theoretic analysis of the interaction between the AP and the STA is carried out. In the analysis, only one AP and STA are assumed. The game progresses in time slots. The client will have a utility function $F(T, \tau)$ where T is the duration for which the STA connects to the AP and τ the duration for which it wanted to connect. The AP quotes a price p_t at the beginning of slot t and the STA can either accept and connect at that price or reject and conclude the game, i.e. go away. Strategies that would achieve equilibrium are derived. STAs are allowed to associate with multiple APs and split their traffic amongst the APs that they associate with Shakkottai *et al* (2006). It is shown that if an appropriate pricing scheme is used, the network throughput is maximised.

6.7 *Other applications of 802·11 networks*

Much of the analysis and, to a large extent, deployment of the 802·11 networks have been as single hop networks for elastic data transfer applications. However, there has been significant interest in using WiFi networks for streaming applications. We briefly review the issues and studies in servicing streaming applications over 802·11 networks.

We illustrate the issues via the VoIP application. The most important expectation of any streaming application is that a specified fraction of the packets from the application be delivered with bounded delay. For example, we could demand that Pr (pkt delay $> D_{\max}$) $< L_p$. For example, $D_{\max} = 50$ ms and $L_p = 0{\cdot}02$ would demand that more than 98 % of the packets be delivered with a delay of less than 50 ms. $D_{\max}$ is called the MAC-delay budget. L_p is determined by the codec and the acceptable quality of speech at the receivers. The quality of speech is measured through the 'mean opinion score' (MOS) metric. MOS perceptive metric and typically requires extensive 'live' tests. There are, however, procedures available to convert delay distributions and loss patterns to MOS estimates. We now examine the various other parameters that determine the delay budget and the VoIP load that can be serviced by a 802·11 network.

In addition to determining L_p, the codec also determines the load on the network. Voice codecs are standardised in the G.7xx series by the ITU with the G.711 (PCM encoding at 64 Kbps), G.721 (ADPCM encoding at 32 Kbps) and G.729 (CS-ACELP encoding at 8 Kbps) being the more widely used. Most codecs have a 'frame' period, i.e. they output a speech sample every T_f seconds. Typically T_f is 10 or 20 ms. The codec may have voice activity detector (VAD) in which case no samples are generated when there is no speech activity. Thus the use of VAD results in variable bit rate (VBR) speech.

The speech samples have to packetised. Every speech packet contains a significant amount of header information, corresponding to the different protocols, e.g. PHY, MAC, UDP and RTP. To amortise this overhead, more than one speech sample may be included in one speech packet. Thus there is the 'packetisation delay'. The packetisation delay also determines $D_{\max}$. Packet delay variations are smoothed out by using a 'playout buffer' the depth of which also determines $D_{\max}$. The error correction and loss concealment mechanism that is used also determines $D_{\max}$.

We now see how voice calls can be supported in an 802·11 network. It is reasonable to assume that the two nodes in conversation are not associated with the same AP. This means that corresponding to each call, there will be packet flows in both directions—from the AP to the STA and vice versa. This means that, on an average half the packets will be flowing from the AP. At high loads, we can assume that the collision window is almost always $CW_{\min}$, and we can show that a maximum of

$$\lceil (R(2(T_{\text{voice}} + T_{\text{SIFS}} + T_{\text{ack}} + T_{\text{DIFS}}) + (T_{\text{slot}} \times CW_{\min})))^{-1} \rceil$$

calls can be simultaneously supported (Hole & Tobagi 2004). Here R is the packet rate from the voice source, T_{voice} is the packet transmission time, T_{ack} is the ack transmission time and T_{slot} is the slot duration.

Note that the above analysis is independent of the QoS requirement for the calls in terms of the delay distribution and hence the quality of the speech in terms of the MOS metric. If a target MOS is to be achieved, then a significantly lower number of voice calls can be supported. Of course, a higher number of voice calls with a target MOS can also be supported by using the QoS provisioning mechanisms of the 802·11 protocol.

Recall that there are two ways in which QoS can be provided to the voice calls. The first is the use of PCF, where we could emulate a circuit multiplexing like scheme. Here, the important parameters are the superframe duration and the duration of the contention free period within the superframe. The number of calls that can be supported is essentially the number of voice packets that can be accommodated in a superframe. Veeraraghavan *et al* (2001) provide some simple calculations and simulation results to support voice using PCF.

The second method provide QoS to give priority to voice packets via the QoS mechanism of 802·11e. Here, the voice packets will see a higher success probability but bandwidth is not reserved for them like when PCF is used. See (Harsha *et al* 2006a; Harsha *et al* 2006b) for an analysis of this approach.

With the 802·11 hardware becoming a commodity item with widespread availability, there has been significant interest in using this technology for innovative applications. The Digital Gangetic Plain (DGP) project in Uttar Pradesh and the Ashwini project in Andhra Pradesh (Raman & Chebrolu 2007) provide Internet connectivity to a large rural areas which do not have any significant communication infrastructure. A hierarchical architecture is used in which each village has an access point and the nodes in the village connect to the access point. A mesh of point-to-point links interconnect the access points. The the point-to-point use an 802·11 protocol to exploit the cost advantage of its hardware. Of course, powerful, specially designed antennas are needed to create these point-to-point links.

We mention here that there is also significant interest in the use of 802·11 in multihop networks. There are many performance issues. See (Xu & Saadawi 2002) for an analysis of problems associated with the basic MAC protocol. The performance of TCP transfers over a multihop 802·11 network is analysed in (Kherani & Shorey 2004). Some suggestions on improving the spatial re-use available in (Ye *et al* 2003). These are just a sample of some of the literature in the area and not an exhaustive list.

7. Ad hoc Internets

We now consider wireless ad hoc internets (WANETs). In contrast to AP-based networks that we discussed in the previous section, there is no association between the nodes and fixed infrastructure. Information transport services are built over a set of mobile, arbitrarily located nodes with every node behaving both like a mobile host and also a wireless router. Hence we call them internets. Obvious applications for such networks include providing communication services in emergency situations, e.g. areas affected by storms, floods and earthquakes. An ad hoc Internet can also provide connectivity to fleets of vehicles operating in areas with no networking infrastructure. Of course, there are also many obvious military applications. In all of these applications, as has been discussed in § 2, we can identify a set of 'point-to-point flow' requirements between nodes in the network with each of the flows having their own QoS requirements, e.g. a minimum throughput requirement.

Recall, from our discussion in § 6, that the wireless channel is a broadcast medium and that a transmission from one node reaches all the nodes in the network. The received power will of course depend on the distance and the characteristics of the radio channel between the transmitter and the receiver. In figure 3, we took a simple view and, depending on the strength of the received signal, we divided the region around the transmitter into decode and interference regions. In this model, no other transmission can be decoded by an intended receiver in the decode and interference regions of a transmitter. However, if another transmitter–receiver pair is such that the receiver is outside of the decode or interference

region of the transmitter then this transmission can co-exist with the first transmission. This suggests the important feature of wireless networks—spatial re-use, the same frequency band or channel will be simultaneously used in geographically separated areas of the network.

Spatial re-use increases the capacity of the network and is achieved by controlling the transmission power such that the received power at nodes 'far' away from the transmitter is low enough to cause minimal interference to its receiving from nearer neighbours. In other words, the transmission powers are maintained at levels to allow multiple simultaneous transmissions without interfering with each other. This means that the transmission range of a node should not cover the entire operational area and a WANET will be a multihop network with the intermediate nodes co-operating in the communication process by relaying the packets. In fact, the transmission range should be as low as possible to increase the re-use factor. One might argue that the multiple hops can actually reduce capacity because a packet has to be transmitted multiple times. Informally, observe that although increasing the transmission range decreases the average number of hops linearly, spatial re-use factor is reduced in proportion to the square of the range. Thus, the capacity increases as the transmission range decreases. Hence, as low a transmission range as possible should be used. However, the transmission range cannot be made too low because then the network becomes disconnected, i.e. there may not be a path between every pair of nodes in the network. Thus keeping the network connected becomes an important issue. The capacity of the network and the connectedness are thus important issues and we will discuss them in different settings. In addition, note that WANET nodes are expected to use battery energy. Hence optimising energy consumption is important.

Wireless networks are typically analysed by assuming either of the following two models. Let x_i be the location of node i in the network. In the *protocol model* it is assumed that node i can receive transmissions from node j if and only if the Euclidean distance between them is less than a pre-specified constant $r_{i,j}$, i.e. if $|x_i - x_j| < r_{i,j}$. $r_{i,j}$ is called the 'cutoff' for i-j transmissions. Typically, it is assumed that $r_{i,j} = r$ for all i, j. It may be further assumed that while node i is transmitting to node j, and node k is transmitting simultaneously, then the following inequality is satisfied: $\|x_k - x_j\| > \gamma(1 + \delta)$. In the *physical model* it is assumed that transmissions from node i will be decoded at node j if the SINR constraint

$$\frac{Ph_{i,j}}{N_j + \sum_{k \in \mathcal{T}, k \neq i} Ph_{k,j}}$$

is satisfied. Here P is the transmission power, $h_{i,j}$ is the i-j channel gain, N_j is the receiver noise at j and $\mathcal{T}$ is the set of nodes transmitting simultaneously with i.

The WANET nodes are expected to be mobile. Hence it is reasonable to analyse networks by assuming that the location of the nodes is arbitrary. We first consider the case when the node locations are known but arbitrary. Two problems are addressed—(1) maximising the capacity of the network and (2) controlling the transmission ranges and hence the topology of the network. We then assume that the node locations are random and assign a distribution to their locations. We analyse the transmission range required to make the network connected. We also analyse the capacity, after a suitable definition, of such a network.

7.1 *Fixed networks*

We take the view that once the network is deployed and the node locations are relatively stable, the nodes will first discover their neighbours and adjust their transmission powers

to achieve a global or local topology objective. This is essentially an association phase in which each node determines the wireless links on which it will transmit data. In the second phase, routing of the data flows, the scheduling of the transmissions by the nodes and the transmission power and data rate are determined to achieve some performance objective.

7.1a *Topology control:* One method of achieving energy efficiency is to construct a network topology that is 'energy efficient'. The objective here is to retain only those links that correspond to an energy optimal topology while keeping the network connected. The more interesting of these algorithms first determine the transmission power for each node to achieve a local topological property which is guaranteed to make the network connected.

Rodoplu & Meng (1999) have described a distributed algorithm. This algorithm first obtains the geographical area in which each node should search for 'neighbours' to which the node should have a link. For node i and a possible neighbouring node j, the 'relay region' is defined as the set of points k for which the transmission power required from i to achieve a specific SNR at k is more than the sum of the powers required for maintaining the same SNR from i to j and from j to k. The set of nodes which are not in the relay region of other nodes will be the neighbours of i. The links and the transmission power for the activation of the links is determined in this manner for all the nodes. It can be shown that the network so obtained is connected. This algorithm needs location information of the nodes in its neighbourhood.

Gomez *et al* (2001) describe the power aware routing optimisation (PARO) algorithm. In this algorithm a node asks its neighbours to reduce their transmission power if it sees them transmitting to nodes that are further than itself in the same direction, since such nodes can be reached through it. This idea is similar in principle to that of Rodoplu & Meng (1999) in that the power is decreased if there are nearer neighbours who can relay. Thus the scheme of Gomez *et al* (2001) is one with a 'relay node' initiated power reduction mechanism as opposed to the transmitter calculated mechanism of Rodoplu & Meng (1999).

In the scheme described by Wattenhofer *et al* (2001) each node increases its power till it finds a neighbour in every cone of degree θ. It has been shown that choosing $\theta = 5\pi/6$ results in a connected network (Li *et al* 2001). After obtaining a connected network in this manner, the topology is further optimised by eliminating links (i, k) for which the sum of the transmission powers for link (i, j) and (j, k) is less than that for (i, k).

Power efficiency is achieved in a very different manner by the algorithm described by Narayanaswamy *et al* (2002). It is argued that it is best to have all nodes transmit with a common power. The COMPOW of Narayanaswamy *et al* (2002) protocol prescribes that each node obtain a routing table for each possible power level and the lowest power that makes the network connected is chosen as the operating power.

7.1b *Capacity and cross layer control:* Given the physical locations of the node and connectivity model, we can construct the graph G induced by the WANET that represents the topology of the network. Define a vertex in G corresponding to every node in the WANET. If the transmission from node i can be received and decoded at node j we say that $l := (i, j)$ is a link in the network with $t(l) = i$ and $r(l) = j$. Let E be the set of links in the network. The network topology is represented by $G = (V, E)$ with V representing the nodes and E representing the links. In the following we assume that G is given.

We will consider a time-slotted network with M nodes and L links. We will assume that the transmission rate on all the links is the same and that the packets fit into one time slot. The set of links on which transmissions can occur simultaneously will be called an 'activation set'.

Let $c = [c_1, \ldots, c_L]$ be a 0-1 vector that represents an activation set. An activation set can be determined from either the protocol or the physical model described earlier. Let S denote the set of all allowable activation sets.

Consider a network with a flow requirement $F = [f_{i,j}]$ where $f_{i,j}$ represents the long term rate at which packets need to be sent from source node i to destination node j, $1 \leq i, j \leq M$. The first issue that we address is the set of F for which a routing and scheduling algorithm exists for which the packet queues at all the nodes will be stable. Queue stability means that the discrete time queue occupancy process has a stationary distribution. To characterise this set of F, observe that an activation vector can be represented by a corner of the L-dimensional unit hypercube. Note that not all corners of the hypercube are possible activation sets. It can be shown that a requirement F can be supported if and only if F can be routed such that the link flows correspond to an interior point of the convex hull of S. Thus, knowledge of F can possibly help statically design the routes the flows and also the corresponding transmission schedule. Note that this result does not provide us with a method to determine the routing or to determine if a given F can be supported by the network. In fact, determining such a routing or even determining if an F can be supported is a hard problem.

Observe that there are two problems to be solved simultaneously—(1) a routing problem to determine the paths for the packets from i to j and (2) a scheduling problem that determines the sequence of activation sets. Remarkably, a centralised, dynamic scheduling and routing algorithm that supports any F that can be supported without actually knowing F has been described in Tassiulas & Ephremides (1992)! The scheduling policy is essentially a maximum weight matching policy and is described below.

Packets with destination j will be said to belong to class j. Let $Q_{i,j}(\tau)$ denote the number of packets of class j at node i at the beginning of slot τ. For link l define link weight $w_l(\tau)$ as follows

$$w_{j,l}(\tau) = \begin{cases} Q_{t(l),j}(\tau) - Q_{r(l),j}(\tau) & \text{if } r(l) \neq j \\ Q_{t(l),j}(\tau) & \text{if } r(l) = j \end{cases}$$

$$w_l(\tau) = \max_j \; w_{j,l}(\tau) \tag{19}$$

Let $w(\tau) = [w_l(\tau)]$. For slot τ, select the activation set $c^*(\tau)$ that maximises $w(\tau)c^T(\tau)$ where $c^T(\tau)$ is the transpose of $c(\tau)$. For an activated link $l \in c^*(\tau)$ transmit the class for which $w_{j,l}(\tau)$ is the maximum over j. Time averages of the channel error rates can also be used in choosing the link schedules by appropriately weighting $w_{j,l}(\tau)$.

Observe that in the above algorithm, the source of the packets is not important and that the routing is done automatically. Further, the only objective is to support as large a set of F as possible without considering the energy expended. We now consider minimising the energy expended while supporting a given F. One way to achieve this is as follows and is described in Cruz & Santhanam (2003). We first assume that the the paths for the flows are determined, i.e. the routes are determined. This will give us the link flow, say x_l for link l. We now need to do a scheduling that will specify the links that should be active in a slot and the transmission power that will be used on the link, i.e. $p_l(\tau)$ for each link l and slot τ with $p_l(\tau) = 0$ indicating that link l is not active in slot τ. Thus the activation vector for slot τ has the form $p(\tau) = [p_1(\tau), p_2(\tau), \ldots, p_L(\tau)]$. Corresponding to a $p(\tau)$, the SINR at each receiver and the corresponding transmission rate $r_l(\tau)$ can be determined. The long term average rate on link l will be $r_l = \liminf_{\tau\to\infty} \frac{1}{\tau} \sum_{\hat{\tau}=1}^{l} r_l(\hat{\tau})$. The average power consumed

by link l is $p_l = \limsup_{\tau\to\infty} \frac{1}{\tau} \sum_{\hat{\tau}=1}^{t} p_l(\hat{\tau})$. We thus need to obtain a sequence $p(\tau)$ for $\tau = 1, 2, \ldots$ for which the link capacity constraints $r_l \leq x_l$ and peak power constraints are satisfied while minimising the average power consumed in the network $\sum_l p_l$. An iterative ascent algorithm is described in Cruz & Santhanam (2003) to determine the optimum sequence $p(\tau)$ which will essentially be a periodic transmission schedule. To determine the optimum routes, the optimum routing scheme is used in which a feasible set of routes is first assigned and then the flows are incrementally adjusted towards the shortest paths in each step of the iteration.

There are two restrictive features of the above method. Firstly, it is a centralised algorithm. Secondly, the time average of the flow requirements and also of the channel characteristics need to be known *a priori*. The latter is required to obtain the SINR at the receiver. In an important recent paper, (Neely 2005) both these restrictions have been removed. We describe this next.

We assume that the link state is from a discrete set and that corresponding to link state s, the transmission rate on link l is a function of the transmission power vector p, denoted by $r_l(s, p)$ and $r = [r_1, \ldots, r_L]$. The power-optimal dynamic policy is an extension of the throughput optimal policy using Equation 19 and is as follows. For slot τ, calculate the weight of link l as follows.

$$j_l(\tau) = \arg\max_j \{Q_{t(l),j}(\tau) - Q_{r(l),j}(\tau)\}$$

$$w_l(\tau) = \max(Q_{t(l),j_l}(\tau) - Q_{r(l),j_l}(\tau), 0)$$

Choose the power vector $p(\tau)$ that maximises

$$\sum_{i=1}^{M} \left(\sum_{l\in O_i} 2r_l(s(\tau), p(\tau))w_l(\tau) - g_i \left(\sum_{l\in O_i} p_l(\tau) \right) \right)$$

and transmit the packet from class j_l selected above on link l. Here, $O_i = \{l : t(l) = i\}$, i.e. the set of links for which node i is the transmitter and $g_i(\cdot)$ is a convex increasing function of the power expended at node i. Also note that $Q_{j,j} = 0$ because node j will remove the packets destined for itself and pass them to the applications and not queue them. A distributed implementation of this algorithm has also been described by Neely (2005).

Notice that the algorithms that we described above all jointly perform routing, scheduling power control thus encompassing the network, the MAC and the physical layer. Thus the above can be considered to be a cross layer design problem for networks. Other work in this area are described in Singh *et al* (1998); Chang & Tassiulas (2000); Xu *et al* (2000); Kar *et al* (2003); Neely *et al* (2003); Lin & Shroff (2004); Elbatt & Ephrimedes (2004); Lin & Shroff (2006). An excellent survey is now available in Georgiadis *et al* (2006).

7.2 *Random networks*

We now consider random networks specified by the spatial distribution of the nodes in the operational area and a transmission radius or 'cutoff'. The distribution of the node locations could correspond to the ensemble of randomly deployed nodes that are static or to the stationary distribution of time evolving mobile networks. We continue to assume a multihop network. We first explore the statistical properties of the topology of these networks and also some of the tools used in their study. To do this we construct a random graph model for wireless networks when the nodes are placed randomly in their operational area. It is important to

realise that the graph obtained from the random deployment of the nodes is not the same as the traditional Erdos–Renyi random graph, i.e. random graphs where the edges are independent of each other. Erdos–Renyi random graphs have been studied in detail by Bollobas (1985).

The connectivity of random wireless networks are best represented by a 'random geometric graph' (RGG) that are constructed as follows. The network graph $G_w = (V, E)$ for a realisation of the node locations is constructed as follows. Let x_i represent both the node and its location and let $\{x_i\}_{i=1,\cdots,n}$ be a realisation of the node locations. Each node in the network corresponds to a vertex in V. Edge $e_i \in E$, $e_i := (i_1, i_2)$, if the transmission from x_{i_1} can be decoded by x_{i_2}. This in turn can be reduced to the requirement that $\|x_{i_1} - x_{i_2}\| < r_{i_1,i_2}$ where $\|\cdot\|$ denotes a suitable norm, the Euclidean l_2 norm being the most obvious one. r_{i_1,i_2} would depend on the transmission power of i_1 and the receiver sensitivity of i_2 and, as has been described earlier, is called the cutoff for i_1-i_2 transmission. It is easy to see that the randomness of the coordinates of the node in the distribution area introduces randomness in the graph. Also note that in an RGG the edges are not independent. Observe that the RGG essentially corresponds to the protocol model described earlier. In this section our interest will primarily be in RGGs constructed from the protocol model.

The topological properties of random network as a function of the cutoff is of interest. Here we will assume that the cutoffs are the same for every node pair in the network, i.e. $r_{i,j} = r$ for $1 \leq i, j \leq M$. Further, if the nodes are deployed over a finite area, the cutoff could be made a function of n, the number of nodes deployed in the area. Let $r(n)$ be the cutoff for all the nodes when there are n nodes in the network. Two important quantities are of interest—(i) The topological properties of the network as a function of $r(n)$ with the most important property being k-connectedness. (ii) The communication properties of the network—the set of *communication requirements* that the network can support. A generic communication requirement is specified by $\rho_{i,j}$ where $\rho_{i,j}$ represents the throughput requirement from node i to node j. If $\rho_{i,j} = 0$ for $i \neq 1$ we say that the requirement is a broadcast from node 1 to all the nodes in the network. Similarly, If $\rho_{i,j} = 0$ for $j \neq 1$ corresponds to a 'collection' network with data being collected at node 1.

7.2a *Topological properties:* We study the topological properties as the graph properties of the random geometric graph induced by the wireless network. Barring a few exceptions, only asymptotic topological properties are obtained for the case of nodes distributed in a finite area with $n \to \infty$.

Let us first consider the connectivity of the network. In Desai & Manjunath (2002) the probability that a network of n nodes distributed uniformly in the one-dimensional set $[0, z]$ and having a cutoff of r, $P_c(n, z, r)$ is shown to be

$$P_c(n, z, r) = \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{(z - kr)^n}{z^n} u(z - kr) \tag{20}$$

where $u(x)$ is the unit step function. Eqn. 20 can be derived as a special case of the probability of there being k components in the network which is derived (Godehardt & Jaworski 1996) to be

$$\sum_{k=m-1}^{n-1} \binom{n-1}{k} \binom{k}{m-1} (-1)^{k+m-1} (1 - kr)^n.$$

Many more results for the one-dimensional uniform network have been obtained in Godehardt & Jaworski (1996). In Gupta *et al* (2006); Karamchandani *et al* (2005) the node locations are

assumed to be at distances that are i.i.d. exponential from the origin and many topological properties are derived. The exponential distribution of the node locations implies that the node spacings are also independent and have an exponential distribution, i.e. if $X_{(i)}$ is the location of the i-th node from the origin then, $Y_{(i)} := X_{(i+1)} - X_{(i)}$ are independent and exponentially distributed. Thus the probability of the network being connected is just the probability that $Y_i < r$ for $i = 1, \ldots, n-1$. An interesting property of the exponential network is that if $\lambda r < \infty$ then it can be shown that $\lim_{n\to\infty} P_c(n) < 1$, where $P_c(n)$ is the probability that the n-node network is connected.

Asymptotic results for the graph properties are obtained by considering a sequence of graphs G_n with G_n being an RGG of n nodes. $G_n(r_n)$ will be the n-node RGG with cutoff r_n. Let c_n be the smallest cutoff required to make G_n connected. c_n is often called the critical radius and is defined as

$$c_n := \inf\{r > 0 : G_n(r) \text{ is connected with high probability}\},$$

A *monotone increasing* property of a graph is one that continues to hold when edges are added to the graph. For RGGs constructed from node locations X_i, $i = 1, 2, \ldots$, increasing the cutoff r adds new edges and hence connectivity of the graph is a monotone increasing property.

Consider a sequence G_n of random geometric graphs with graph G_n obtained using a cutoff r_n. A sequence t_n of cutoffs is a *weak threshold* function for a monotone property $\mathcal{P}$ iff $r_n/t_n \to \infty$ implies that $\mathsf{Pr}\,(G_n \text{ has property } \mathcal{P}) \to 1$ and $r_n/t_n \to 0$ implies $\mathsf{Pr}\,(G_n \text{ has property } \mathcal{P}) \to 0$. We can also define a *strong threshold* function t_n for property $\mathcal{P}$ as follows: if for all $0 < \epsilon < 1$, $r_n = (1-\epsilon)t_n$ implies $\mathsf{Pr}\,(G_n \in \mathcal{P}) \to 0$ and $r_n = (1+\epsilon)t_n$ implies $\mathsf{Pr}\,(G_n \in \mathcal{P}) \to 1$. See Friedgut & Kalai (1996); Goel *et al* (2004) for alternate notions of thresholds.

For one-dimensional uniformly distributed networks, we can show that $\frac{\log n}{n}$ is a strong threshold for connectivity of the graph G_n. We can make a stronger claim. Let c_n be the minimum cutoff required to make the n-node network connected and let d_n be the 'largest nearest neighbour distance', i.e.

$$d_n = \max_i \min_{j\neq i} |x_i - x_j|.$$

We can show that $\lim_{n\to\infty} \frac{nc_n}{\log n} = 1$ with probability 1 while $\lim_{n\to\infty} \frac{nd_n}{\log n} = \frac{1}{2}$ with probability 1.

The book by Penrose (2003) is an excellent tutorial for RGGs. A recent survey of some of the newer results and some new proofs is available in Iyer & Manjunath (2006). The topological properties of RGGs constructed from the physical model, also called signal-to-interference ratio graphs (STIRGs) have been studied in Dousse *et al* (2002); Dousse & Thiran (2004).

Finite network analysis for networks in higher dimensions is difficult and there are no known results for RGGs constructed using the Euclidean norm. Desai & Manjunath (2005) obtains the probability that a labelled graph occurs under the l_∞ norm.

7.2b *Communication properties:* The most important communication property of interest is the 'capacity' of the network. The notion of capacity has to be more formally defined. Recall that for fixed networks we defined a 'capacity region' as the set of flow rates that could be carried by the network.

We first define the *transport capacity* for a unicast network. It is a measure that captures the degree of spatial re-use and also the fact that information reaches the destination over

multiple hops. It is defined as the sum of the products of the distance travelled by every bits towards its destination in unit time. For example, we will say that the transport capacity of the network is one bit-meter per second if one bit moved one meter towards the destination in one second. The results that we describe below will assume that n nodes are distributed on a square of unit area.

Let us first consider the transport capacity of the network. We will assume a time slotted network with R bits being transmitted in each slot and slot duration of unit length. Recall that for the network to be connected, the transmission range of the nodes needs to be $O\left(\sqrt{\frac{\log n}{n}}\right)$. Consider the protocol model. In this model, when a node is transmitting, since the transmitter needs acknowledgements from the receiver, no other node within its transmission range can be transmitting. We will assume a perfect scheduling scheme to be in operation at all times. This ensures that spatial re-use is maximum at all times. Thus, in a given slot, $O\left(\frac{n}{\log n}\right)$ nodes will be transmitting simultaneously. We can reasonably assume that in routing a packet, the next hop will be the farthest node away from the transmitter towards the destination. This distance is clearly $O\left(\sqrt{\frac{\log n}{n}}\right)$. Thus we can say that the transport capacity of the random network is $O\left(R\sqrt{\frac{n}{\log n}}\right)$ bit-meters per second.

Now consider the throughput per node. Let $\lambda(n)$ be the throughput for a node when there are n nodes in the network. Since the transmission range depends on n, the number of hops to the destination will also depend on n. Let $h(n)$ be the average number of hops to the destination. We can thus write

$$n\lambda(n)h(n) \leq O\left(\frac{n}{\log n}\right)$$

Clearly, $h(n) = O\left(\sqrt{\frac{n}{\log n}}\right)$ and we see that

$$\lambda(n) = O\left(\frac{1}{\sqrt{n \log n}}\right).$$

In fact this can be made more formal and Gupta & Kumar (2000) show that for some constants c_1 and c_2,

$$\lim_{n\to\infty} \mathsf{Pr}\left(\lambda(n) = \frac{c_1 R}{(1+\delta)^2\sqrt{n \log n}} \text{ is feasible}\right) = 1$$

$$\lim_{n\to\infty} \mathsf{Pr}\left(\lambda(n) = \frac{c_2 R}{\delta^2\sqrt{n \log n}} \text{ is feasible}\right) = 0$$

Thus $\lambda(n) = \Theta\left(\frac{1}{\sqrt{n \log n}}\right)$.

For the physical model Gupta & Kumar (2000) show the following

$$\lim_{n\to\infty} \Pr\left(\lambda(n) = \frac{c_1 R}{\sqrt{n \log n}} \text{ is feasible}\right) = 1$$

$$\lim_{n\to\infty} \Pr\left(\lambda(n) = \frac{c_2 R}{\sqrt{n}} \text{ is feasible}\right) = 0$$

It can be shown that with power control, $\lambda(n) = \Theta\left(\frac{1}{\sqrt{n}}\right)$ (Agarwal & Kumar 2004).

The above results assume that relaying is allowed in the network. It can be shown that in the physical model, if there were only direct transmissions without any relaying then $\lambda(n) = O(n^{-1/(1+\alpha/2)})$ (Grossglauser & Tse 2001). This is because a large number of transmissions will be over long ranges hence reducing the spatial re-use.

To increase the capacity, a network can exploit the fact that the nodes are mobile. Grossglauser & Tse (2001) propose a two-phased protocol in which each packet is relayed only once. Each source transmits the packet to its one-hop neighbours to relay to the destination. Since the nodes are assumed to be mobile, either the relay nodes or the source will eventually be 'near' the destination. The packet then makes one more hop to the destination. In each slot $n_s = \theta n$ nodes are randomly chosen to be senders and the rest as receivers. By retaining the links that satisfy the SINR constraint and transmitting to the nearest receiver, it can be shown that, on an average, $O(n)$ nodes can transmit simultaneously. By alternating between senders and relays transmitting, it can be shown that

$$\lim_{n\to\infty} \Pr\left(\lambda(n) = c_3 R \text{ is feasible}\right) = 1$$

for some constant c_3.

The above result is interesting because it shows how to exploit mobility of the nodes in the network to enhance capacity. This does not however consider the delay that will be experienced by the packets. The delay will of course depend on the mobility characteristics of the nodes. Recently, Bansal & Liu (2003); Sharma *et al* (2006a) and others have studied delay capacity trade-offs for different mobility models.

8. Ad hoc wireless sensor networks (WSNs)

Advances in microelectronics technology have made it possible to build inexpensive, low power, miniature sensing devices. Equipped with a microprocessor, memory, radio and battery, such devices can now combine the functions of sensing, computing, and wireless communication into miniature *smart sensor nodes*, also called *motes*. Since smart sensors need not be tethered to any infrastructure because of on-board radio and battery, their main utility lies in being *ad hoc*, in the sense that they can be rapidly deployed by randomly strewing them over a region of interest. For a survey of this technology, see Akyildiz *et al* (2002); a textbook treatment is provided in Zhao & Guibas (2004). Several application of such wireless sensor networks have been proposed, and there have also been several experimental deployments. Examples of applications are:

- Ecological monitoring: wild-life in conservation areas, remote lakes, forest fires (Jiang *et al* 2005); (Yu *et al* 2005)

- Monitoring of large structures: bridges, buildings, ships, and large machinery, such as turbines
- Industrial measurement and control: measurement of various environment and process parameters in large factories, such as continuous process chemical plants (Ameer *et al* 2006)
- Assistance in navigation and guidance through the geographical area where the sensor network is deployed (Li *et al* 2002); (Li & Rus 2005).
- Defence applications: monitoring of intrusion into remote border areas; detection, identification, and tracking of intruding personnel or vehicles (Iannone *et al* 2004); (He *et al* 2004); (Arora *et al* 2004); (Brooks *et al* 2004); (Gui & Mohapatra 2005); (Biswas & Phoha 2004); (Cyairci *et al* 2004).

The 'ad hoc' nature of these wireless sensor networks means that the devices and the wireless links will not be laid out to achieve a planned topology. During the operation, sensors would be difficult or even impossible to access and hence their network needs to operate autonomously. Moreover, with time it is possible that sensors fail (one reason being battery drain) and cannot be replaced. It is, therefore, essential that sensors *learn about each other* and *organise into a network* on their own. Another crucial requirement is that since sensors may often be deployed randomly (simply strewn from an aircraft), in order to be useful the devices need to determine their locations. In the absence of a centralised control, this whole process of self-organisation needs to be carried out in a distributed fashion.

A smart sensor may have only modest computing power, but the ability to communicate allows a group of sensors to collaborate to execute tasks more complex than just sensing and forwarding the information, as in traditional sensor arrays. Hence, they may be involved in online processing of sensed data in a distributed fashion so as to yield partial or even complete results to an observer, thereby facilitating control applications, interactive computing and querying. A distributed computing approach will also be energy efficient as compared to mere data dissemination since it will avoid energy consumption in long haul transport of the measurements to the observer; this is of particular importance since sensors could be used in large numbers due to their low cost yielding high resolutions and large volumes of sensed data. Further, by 'arranging computations' among only the neighbouring sensors the number of transmissions is reduced, thereby, saving transmission energy. A simple class of distributed computing algorithms would require each sensor to periodically exchange the results of local computation with the neighbouring sensors. Thus the design of distributed signal processing and computation algorithms, and the mapping of these algorithms onto a network, is an important aspect of sensor network design.

8.1 *Design elements of WSNs for measurement and inference*

Thus, here we are interested in the *self-organising distributed instrumentation* aspect of ad hoc wireless sensor networks. We can identify the following elements in the design of such systems.

(i) A geographical area, space, or structure needs to be monitored for some environmental parameter, or some activity. The scenario and the measurements and inferences that need to be made are defined. Performance objectives are also defined: e.g. network lifetime, accuracy of inferences being made (such as probability of errors in the inference, and delay in the inference).

(ii) Several sensor nodes are deployed in the area. Each device has one or more sensors, corresponding to the modality that needs to be monitored. For example, in a security situation where vehicular intrusions are expected, vibration and infrared sensors may be deployed. Given the sensitivity and range of the sensors, a certain number of sensors will need to be deployed so that there is adequate coverage of the area to be monitored. This yields a *measurement system*. In some applications the sensor nodes may be deployed randomly over a large area, and thus would not be placed at known locations. In such situations there would need to be an algorithm that localises the nodes. Since the nodes would be very small, with limited energy resources, and may not have a line-of-sight view of the sky, a GPS type approach may not be suitable. Several approaches have been explored based on the use of a few beacon nodes that know their locations; see Doherty *et al* (2001) and Karnik & Kumar (2004b).

(iii) The measurement system yields certain observations, and, given the desired inference that needs to be made, a signal processing algorithm can be obtained. For example, in a security setting if normally the area does not have any activity, then a change detection algorithm needs to be devised (Niu *et al* 2005; Prasanthi & Kumar 2006).

(iv) There are now two approaches to carrying out the signal processing algorithm. All the measurements may be sent to a base station where a central computer performs the calculations. This approach will require a lot of communication, as the same data will be repeatedly transmitted over the network. The alternative is to reduce the computations to simple calculations that can be performed in a distributed manner in the network nodes. Each node takes inputs from neighbours, performs some computation and then forwards its results. Such simple calculations could be: obtaining the average, the mode, the maximum or the minimum of observed values. See Giridhar & Kumar (2005) and (Giridhar & Kumar 2006) for an analysis of general function computation over random networks, and Khude *et al* (2005) for the time and energy complexity of the computation of the maximum function over random networks.

(v) The distributed computation needs to be carried out over an ad hoc wireless network. Such a *computation network* will need to be self-organised from the RF transceivers in the nodes. A topology needs to be defined, and then a packet scheduling algorithm (the MAC (medium access control)) determines how packet transmissions are scheduled in the network. In the survey (Santi 2005) and the book (Zhao & Guibas 2004), the focus is on network topologies that use very little energy in communication. In many situations the networks also need to achieve throughput and delay performance. The work reported in Karnik & Kumar (2004a) provides a self-organisation approach that optimises the computation throughput. Yet another approach for self-organisation with performance objectives is provided by Jia *et al* (2004).

Finally, in the operational network, the measurements are made, the distributed computations are made in the nodes, and the ad hoc wireless network moves the computations between the nodes, thus eventually resulting in the desired inferences being made.

8.2 *An extended example*

In order to illustrate some of the issues, we will discuss some examples primarily from the work of Karnik & Kumar (2004a), and Khude *et al* (2005).

Consider a scenario where sensors are randomly deployed in a geographical region to gather statistics of a spatial process; for example, the temperature of an environment, or the level of

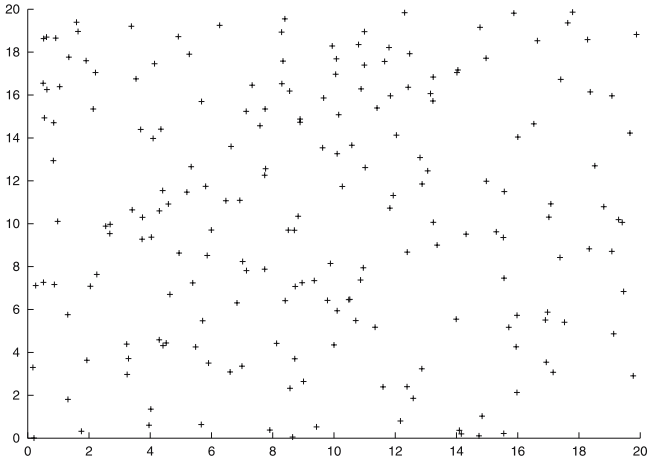

Figure 10. Random deployment of 200 sensors in a region of area 200 m^2. Each has transmission range of 4 m.

some chemical contamination. Suppose that the observer is interested in knowing the maximum value of the quantity being measured. Instead of each sensor sending its measurement to the observer, sensors can compute the maximum in a distributed fashion, and communicate only the result to the observer. A simple distributed algorithm for each sensor can be to collect measurements from its neighbours, compare them with its local value and only forward the maximum of these values. One communication structure suitable for such a computation is a spanning tree which sensors can form in a distributed fashion. Moreover, a sensor needs to transmit the local maximum to its parent only after it receives the corresponding values from its children. The algorithm ensures that the observer ultimately receives the maximum value in the network.

From this example, it is clear that the higher the communication throughput of sensors, the more rapidly is the computation of the maximum proceed, thereby allowing the network to track the variations of the temperature or contamination with time. The example also shows that the network topology and the transmission protocol are critical factors that determine the communication throughput. It is in these two aspects that optimal self-organisation of sensor networks is studied. Instead of limiting to a particular task (maximum computation) a general distributed computation scenario is considered in which sensors are continuously sampling and processing a spatio-temporal process. The problems of building an *optimal topology* and tuning to an *optimal value of channel access rate* are formulated, and distributed algorithms for solving them are obtained. A complete scheme by which sensors can adaptively organise themselves into an efficient distributed information processing instrument is described. The adaptive approach not only allows sensors to tune themselves for given task objectives and constraints but also to adapt to various conditions, such as energy depletion, task reprogramming, etc. The following is the step-by-step summary of this scheme.

8.2a *Deployment:* Sensors are randomly dispersed in a region to monitor a spatio-temporal process, e.g. the temperature of an environment. Figure 10 shows a random deployment of 200 sensor nodes in a 20 m by 20 m field.

8.2b *Discovery:* Once deployed, sensors discover and learn about each other to form a preliminary network. The sensors are spatially distributed, therefore, the discovered links should form a connected network. Since the time required for discovering a connected network

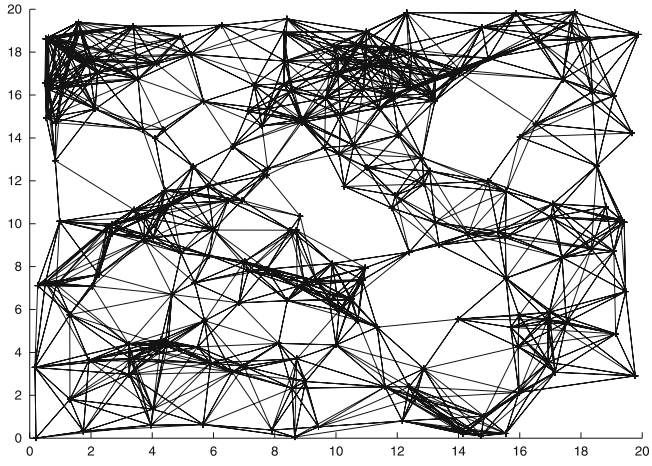

Figure 11. The discovered network after 1000 slots from deployment; sensor i is connected to j if i receives the identity of j successfully at least once in 1000 slots.

depends on the initial transmission attempt rates, these are selected so as to minimise the discovery time. Eventually, all links will be discovered; in practice, sensors stop at some pre-programmed time. Figure 11 shows the topology that was discovered after 1000 slots.

8.2c *Topology formation:* Sensors construct a topology optimal for the given computational task. In figure 12 we show an optimal topology obtained from our algorithm reported by Karnik & Kumar (2004a). In figure 13 we show an alternate topology obtained from a clustering algorithm.

8.2d *Performance tuning:* Sensors optimise the computing rate of the network by tuning certain parameters such as the transmission attempt rates. In this example the objective is to compute the maximum of the measurements, in a pipelined fashion, and deliver the results to an operator station at the origin. Figures 14 and 15 show the computation rate and the

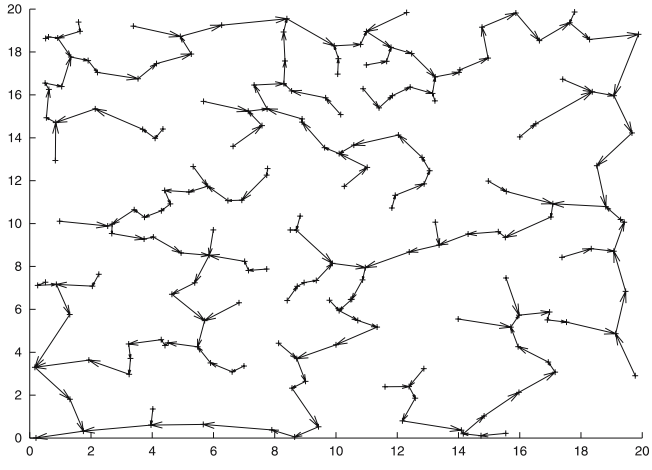

Figure 12. An optimal computing topology (MAWSS).

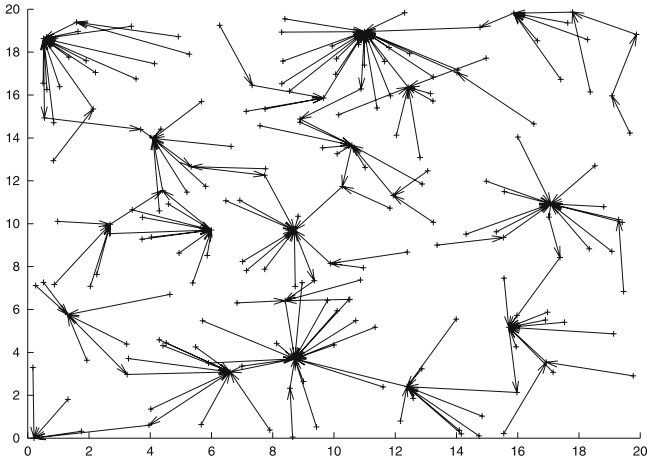

Figure 13. A cluster-like computing topology (TOP).

computation delay. Results are shown with the optimal MAWSS topology, and an alternative cluster-like topology, called TOP, with fixed attempt probabilities (a), and with the MAWSS topology and adaptive attempt rates obtained from the MMTAP algorithm (for self-tuning the attempt probabilities; see Karnik *et al* (2004)). For fixed attempt rates MAWSS performs significantly better than TOP, but for sampling rate greater than 3·5 in a 100 slots, the performance drops sharply owing to network congestion. On the other hand, with MMTAP the attempt probabilities are self-tuned to the local network topology and we see that a much higher throughput is maintained even under overload. An obvious advantage of adaptive attempt probabilities is that there is no need to 'guess' what fixed attempt probability will work well for each network instance, an obviously intractable task!

8.2e *Distributed computation:* It is also of interest to compare the throughput and energy expenditure of various computation algorithms. In the above discussion we have only

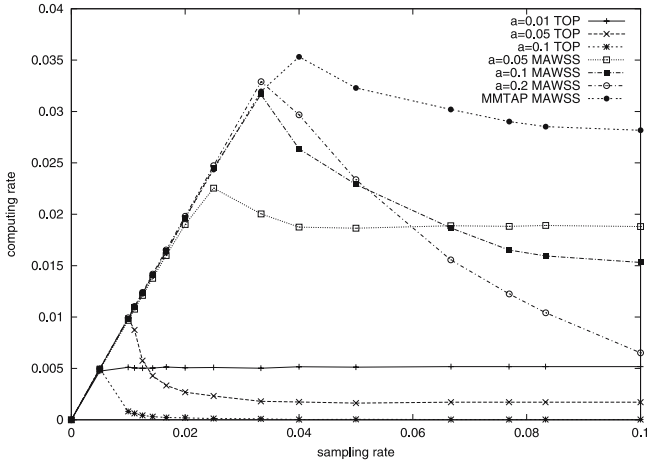

Figure 14. Computing rate (number of maximum calculations produced per slot) vs. sampling rate per slot.

examined maximum computation over a tree. If we had ideal centralised scheduling, what is the best one can do, and how do specific algorithms compare in performance? Our study of this question for the specific problem of distributed maximum computation is provided in Khude *et al* (2005). The results are in the form of 'in probability' scaling laws as the number of nodes in a fixed area increases. The following table summarises the results.

Algorithm	Energy expenditure	Computation time
Optimum algorithms	$\Theta(n)$	$\Theta\left(\sqrt{\frac{n}{\log n}}\right)$
Multi-hop transmission	$\Theta\left(n\sqrt{\frac{n}{\log\ n}}\right)$	$\Theta(n)$
Tree algorithm	$\Theta(n)$	$\Theta(\sqrt{n \log n})$
Ripple algorithm	$\Theta(n\sqrt{n\ \log\ n})$	$\Theta(\sqrt{n\ \log\ n})$

Both the tree algorithm and multihop transmission are organised on a tree rooted at the operator station. The difference is that in multihop transmission all the data is sent to the operator station where the maximum is computed. On the other hand, in the tree algorithm the maximum is recursively computed as the data propagates towards the root. In the ripple algorithm nodes exchange their current estimate of the maximum with their neighbours. Each node's influence propagates like a ripple; when the influences of all the nodes reach the operator station the computation is complete. By optimum we mean that no algorithm can do better than this scaling, and there is an algorithm that achieves this scaling. We conclude that, among the algorithms compared, the tree algorithm is the best, though it is slightly worse than optimum in its energy consumption.

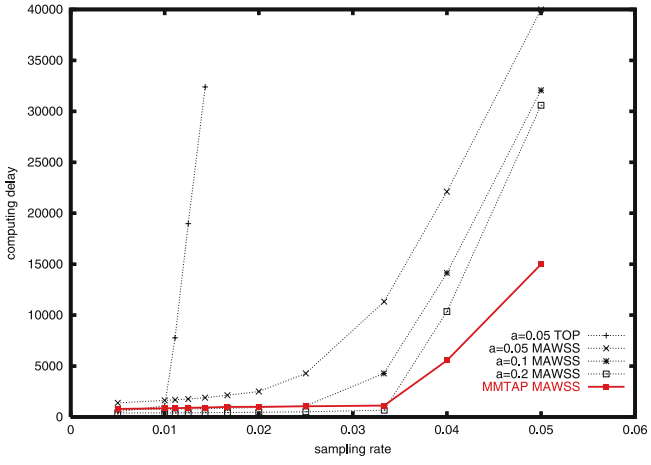

Figure 15. Computing delay (delay between the sample generation and the corresponding maximum computation) vs. sampling rate per slot.

9. Summary

Although wireless communication has been known for over a hundred years, the recent advances in highly integrated low power electronic devices and in wireless physical layer communications has resulted in the rapid development and deployment of a variety of wireless networks. By far the most important commercial application of such technologies is in permitting small mobile devices to access the wired infrastructure. There has also been a resurgence of interest in ad hoc wireless networks, particularly in the new area of wireless sensor networks. In this paper we have first provided a taxonomy of the various wireless network systems, then we have reviewed wireless physical layer techniques, and then we have surveyed a variety of resource allocation and network design problems in wireless networks. With the increasing demand for tetherless communication, and new emerging paradigms for utilising embedded pervasive wireless devices, we expect that this area will continue to be the 'new frontier' in communication networking for many more years to come.

References

Aad I, Castelluccia C 2001 Differentiation mechanisms for IEEE 802·11. *Proc. IEEE Infocom*

Agarwal A, Kumar P R 2004 Capacity bounds for ad-hoc and hybrid wireless networks, ACM SIGCOMM *Comp. Comm. Rev. Spec. Issue on Sci. of Networking Design* 34(3): 71–81

Akyildiz I F, Su W, Sankarasubramaniam Y, Cayirci E 2002 Wireless sensor networks: A survey, *Comp. Networks* 38: 393–422

Ameer P M, Anurag Kumar, Manjunath D 2006 Analysis of network architectures for Zigbee sensor clusters, *Proc. Networks 2006*, New Delhi

Anastasi G, Lenzini L, Mingozzi E 1998 Stability and performance analysis of HIPERLAN. *Proc. IEEE Infocom* 134–141

Arora Anish, Dutta P, Bapat S, Kulathumani V, Zhang H, Naik V, Mittal V, Cao H, Demirbas M, Gouda M, Choi Y, Herman T, Kulkarni S, Arumugam U, Nesterenko M, Vora A, Miyashita M 2004 A line in the sand: a wireless sensor network for target detection, classification and tracking. *J. of Comp. Networks* 46: 603–634

Bansal N, Liu Z 2003 Capacity, delay and mobility in wireless ad-hoc networks. *Proc. IEEE Infocom*

Bejerano Y, Han S-J, (Erran) Li Li 2004 Fairness and load balancing in wireless lans using association control. *Proc. ACM MobiCom.* 315–329

Bharghavan V, Demers A, Shenker S, Zhang L 1994 MACAW: A media access protocol for Wireless LANs. *Proc. ACM SIGCOMM* 212–225

Bianchi G 2000 Performance analysis of the IEEE 802·11 distributed coordination function. *IEEE J. in Selected Areas in Comm.*, SAC-18, (March) 535–547

Biswas Pratik, Phoha Shashi 2004 A sensor network test-bed for an integrated target surveillance experiment. *Proc. 29 Annual IEEE Inter. Conf. on local Computer Networks (LCN'04)*

Bollobas B 1985 *Random Graphs*. Academic Press

Bordenave C, McDonald D, Prouti A 2005 Random multi-access algorithms: A mean field analysis. *Proc. 43rd Allerton Conf. on Comm., Control and Computing*

Brooks Richard, Friedlander David, Koch John, Phoha Shashi 2004 Tracking multiple targets with self-organizing distributed ground sensors. *J. of Parallel and Distributed Computing* 64: 874–884

Cali F, Conti M, Gregori E 1998 IEEE 802·11 Wireless LAN: Capacity analysis and protocol enhancement. *Proc. IEEE INFOCOM 1998*

Carvalho M M, Garcia-Luna-Aceves J J 2003 Delay analysis of the IEEE-802·11 in Single-Hop Networks. *Proc. 11th IEEE Int. Conf. on Network Protocols (ICNP'03)*

Chang J H, Tassiulas L 2000 Energy conserving routing in wireless ad hoc networks. *Proc. of IEEE INFOCOM*

Chhaya H S, Gupta S 1997 Performance modelling of asynchronous data transfer methods of IEEE 802·11 MAC Protocol. *Wireless Networks* 3: 217–234

Cruz R L, Santhanam A 2003 Optimal routing, link scheduling and power control in multi-hop wireless networks. *Proc. IEEE Infocom*

Cyairci Erdal, Tezcan Hakan, Dogan Yasar, Coskun Vedat 2004 *Wireless sensor networks for underwater surveillance systems. Ad Hoc Networks* 4: 431–446

Desai M P, Manjunath D 2002 On the connectivity of finite ad hoc networks. *IEEE Comm. Lett.* 10(6): 437–490

Desai M P, Manjunath D 2005 On range matrices and geometric random graphs. *Proc. 3rd Int. Sym. on Modelling and Optimisation in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*

Doherty L, Pister K, Ghaoui El L 2001 Convex position estimation in wireless sensor networks. *Proc. IEEE Infocom*

Dong X J, Ergen M, Varaiya P, Puri A 2003 Improving the aggregate throughput of access points in IEEE 802·11 Wireless LANs. *Proc. 28th Ann. Local Computer Networks Conf.*

Dousse O, Baccelli F, Thiran P 2002 Impact of intereference on connectivity of ad hoc networks. *Proc. of IEEE INFOCOM*

Dousse O, Thiran P 2004 Connectivity *vs* capacity in dense ad hoc networks. *Proc. of IEEE INFOCOM*

Elbatt T, Ephrimedes A 2004 Joint scheduling and power control for wireless ad hoc networks. *IEEE Trans. on Wireless Communications* 3(1): 74–85

Friedgut E, Kalai G 1996 Every monotone graph property has a sharp threshold. *Proc. of the Am. Math. Soc.* 124: 2993–3002

Georgiadis L, Neely M J, Tassiulas L 2006 Resource allocation and cross layer control in wireless networks. *Foundations and Trends in Networking* 1(1): 1–144

Giridhar Arvind, Kumar P R 2005 Computing and communicating functions over sensor networks. *IEEE J. on Selected Areas in Comm.* 23(4): 755–764

Giridhar Arvind, Kumar P R 2006 Towards a theory of in-network computation in wireless sensor networks. *IEEE Comm. Mag.* 44(4): 98–107

Godehardt E, Jaworski J 1996 On the connectivity of a random interval graph. *Random Structures and Algorithms* 9: 137–161

Goel A, Rai S, Krishnamachari B 2004 Sharp thresholds for monotone properties in random geometric graphs. *Proc. ACM Symp. on Theory of Computing (STOC)*

Gomez J, Campbell A T, Naghshineh M, Bisdikian C 2001 Conserving transmission power in wireless ad hoc networks. *Proc. 9th IEEE Conf. on Network Protocols (ICNP) 2001*

Grossglauser M, Tse D 2001 Mobility increases the capacity of wireless ad hoc networks. *Proc. of IEEE INFOCOM* 1360–1369.

Gui Chao, Mohapatra Prasant 2005 Virtual Patrol: A new power conservation design for surveillance using sensor networks. *Proc. IPSN*

Gupta B G, Iyer S K, Manjunath D 2006 On the topological properties of one-dimensional exponential random geometric graphs. *(Revision submitted)*

Gupta P, Kumar P R 2000 The capacity of wireless networks. *IEEE Trans. on Inf. Theory* 46(2) 388–404

Harsha S, Kumar A, Sharma Vinod 2006a An analytical model for the capacity estimation of combined VoIP and TCP file transfers over EDCA in an IEEE 802·11e WLAN. *Proc. 14th IEEE Int. Workshop on Quality of Service (IWQoS)*

Harsha S, Kumar A, Sharma Vinod 2006b An analytical model for the capacity estimation of combined VoIP and TCP file transfers over EDCA in an IEEE 802·11e WLAN. *Proc. 14th IEEE Int. Workshop on Quality of Service (IWQoS).*

He Tian, Krishnamurthy Sudha, Luo Liqian, Yan Ting, Gu Lin, Stoleru Radu, Zhou Gang, Cao Qing, Vicaire Pascal, Stankovic John A, Abdelzaher Tarek F, 2004 Vigilnet: An integrated sensor network system for energy-efficient surveillance. *Proc. ACM Mobisys*

Hole D P, Tobagi F A 2004 Capacity of an IEEE 802·11b Wireless LAN supporting voip. *Proc. IEEE GLOBECOM*

Iannone Luigi, Benbadis FArid, de Amorium Marcelo Dias, Fdida Serge 2004 *Some Applications of Wireless Sensor Networks*

Iyer S K, Manjunath D 2006 Topological properties of random wireless networks. *Sādhanā:*

Jacob L, Kumar A 2001 Establishing the region of stability for an input queuing cell switch. *IEE Proc. – Comm.* 148(6): 343–347

Jia Xiaohua, Li Deying, Du Dingzhu 2004 QoS topology control in ad hoc wireless networks. *Proc. IEEE Infocom.*

Jiang Chunyu, Dong Guozhu, Wang Bin 2005 Detection and tracking of region-based evolving targets in sensor networks. *Proc. Computer Comm. and Networks, ICCCN 2005*

Kar K, Kodialam M, Lakshman T V, Tassiulas L 2003 Routing for network capacity maximisation in energy-constrained ad hoc networks. *Proc. IEEE INFOCOM 2003* 673–681

Karamchandani N, Manjunath D, Iyer S K 2005 On the clustering properties of exponential random networks. *Proc. IEEE Inter. Symp. World of Wireless, Mobile and Multimedia Networks (WoWMoM)*

Karn P 1990 A new channel access method for packet radio. *Proc. Ninth Computer Networking Conference* 134–140

Karnik Aditya, Kumar Anurag 2004a Distributed optimal self-organisation in ad hoc wireless sensor networks. *Proc. IEEE INFOCOM.*

Karnik Aditya, Kumar Anurag 2004b Iterative localisation in ad hoc wireless sensor networks: One-dimensional case. *Proc. Conf. of Signal Processing and Comm. (SPCOM), Bangalore*

Karnik Aditya, Kumar Anurag, Borkar Vivek 2004 Distributed self-tuning of sensor networks. *Proc. WiOpt'04: Modelling and Optimisation in Mobile, ad hoc and Wireless Networks* Cambridge, UK

Kasbekar G, Kuri J, Nuggehalli P 2006 Online association policies in IEEE 802·11 WLANs. *Proc. of the Fourth Int. Sym. on Modelling and Optimisation in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*

Kherani A A, Shorey R 2004 Throughput analysis of TCP in multi-hop wireless networks withIEEE 802.11 MAC. *Proc. IEEE WCNC* 237–242

Khude Nilesh, Kumar Anurag, Karnik Aditya 2005 Time and energy complexity of distributed computation in wireless sensor networks. *Proc. IEEE Infocom. IEEE*

Kumar A, Kumar V 2005 Optimal association of stations and APs in an IEEE 802.11 WLAN. *Proc. of the Nat. Conf. on Comm.*

Kumar A, Patil D 1997 Stability and Throughput analysis of CDMA-ALOHA with Finite number of users and code sharing. *Telecomm. Systems* (*a Baltzer Science Journal*) 8: 257–275

Kumar A, Altman E, Miorandi D, Goyal M 2005 New Insights from a fixed point analysis of single cell IEEE 802.11 WLANs. *Proc. IEEE Infocom*

Li B, Battiti R 2003 *Supporting Service Differentiation with Enhancements of the IEEE 802.11 MAC Protocol: Models and Analysis*. Tech. Report. DIT-03-024, Deptt of Information and Communication Technology, University of Trento

Li L, Halpern J Y, Bahl V, Wang Y M, Wattenhofer R 2001 Analysis of a cone-based distributed topology control algorithm for wireless multihop networks. *Proc. of the ACM Symposium on Principles of Distributed Computing (PODC)*, 264–273

Li Qun, Rus Daniela 2005 Navigation protocols in sensor networks. *ACM-Transactions*, on Sensor Networks 13–35

Li Qun, DeRosa Michael, Rus Daniela 2002 *Distributed algorithms for guiding navigation across a sensor network*. Technical Report. Dartmouth College

Lin X J, Shroff N B 2004 Joint rate control and scheduling in multihop wireless networks. *Proc. IEEE Conference on Decision and Control* 1484–1489

Lin X J, Shroff N B 2006 The impact of imperfect scheduling on cross-layer congestion control in wireless networks. *IEEE Trans. on Networking*, 14(2): 302–315

Molle M L, Sohraby K, Venetsanopoulos A N 1987 Space–time models for asynchronous CSMA protocols for local area networks. *IEEE Selected Areas in Comm.*, 5(6): 956–968

Mussachio J, Walrand J 2006 WiFi access point pricing as a dynamic game. *IEEE Trans. on Networking*, 14(2) 289–302

Narayanaswamy S, Kawadia V, Sreenivas R S, Kumar P R 2002 Power control in ad hoc networks: theory, architecture, algorithm and implementation of the COMPOW protocol. *Proc. of European Wireless 2002*. Next Generation Wireless Networks: Technologies, Protocols, Services and Applications

Neely M J 2005 *Energy Optimal Control of Time Varying Wireless Networks* (Preprint)

Neely M J, Modiano E, Rohrs C E 2003 Dynamic power allocation and routing for time varying wireless networks. *Proc. IEEE INFOCOM*.

Niu Ruixin, Varshney Pramod K, Cheng Qi 2005 Distributed detection in a large wireless sensor network. *EURASIP J. on Wireless Comm. and Networking* 5: 462–472

Panda M, Kumar A, Srinivasan S H 2005 Saturation throughput analysis of a system of interfering IEEE 802.11 WLANs. *Proc. WoWMoM*

Penrose M D 2003 *Random Geometric Graphs* (London: Oxford University Press)

Prasanthi Venkata K, Kumar Anurag 2006 Optimising delay in sequential change detection over ad hoc wireless sensor networks. *Proc. IEEE SECON.*

Ramaiyan V, Kumar A, Altman E 2005a Fixed point analysis of single cell IEEE 802.11e WLANs: Uniqueness, multistability and throughput differentiation. *Proc. ACM Sigmetrics*

Ramaiyan V, Kumar A, Vasudevan N 2005b Fixed point analysis of the saturation throughput of IEEE 802.11 WLANs with capture. (*Preprint*)

Raman B, Chebrolu K 2007 Experiences in using WiFi for rural internet in India. *IEEE Comm. Magazine*, 104–110

Rodoplu V, Meng T H 1999 Minimum energy mobile wireless networks. *IEEE Journal on Selected Areas in Comm.*, 17(8)

Romdhani L, Ni Q, Turletti T 2003 Adaptive EDCF: Enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks. *Proc. of Wireless Comm. and Networking Conf. (WCNC)*

Santi Paolo 2005 Topology control in wireless ad hoc and sensor networks. *ACM Computing Surveys*, 37(2): 164–194

Shakkottai Srinivas, Altman Eitan, Kumar Anurag 2006 The case for non-cooperative multihoming of users to access point in IEEE 802.11 WLANS. *Proc. IEEE Infocom*

Sharma G, Mazumdar R, Shroff N 2006a Delay and capacity trade-offs in mobile ad hoc networks: A global perspective. *Proc. IEEE Infocom*

Sharma G, Ganesh A, Key P 2006b Performance analysis of contention based medium access control protocols. *Proc. IEEE Infocom*

Singh S, Woo M, Raghavendra C S 1998 Power-aware routing in mobile ad hoc networks. *Proc. the 4th IEEE/ACM Conf. on Mobile Computing and Networking (MOBICOM)* 181–190

Tan G, Guttag J 2004 *Capacity Allocation in Wireless LANs*. Technical Report: 973, MIT CSAIL, Cambridge, MA

Tassiulas L, Ephremides A 1992 Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. on Automatic Control*, 37(12): 1936–1948

Tay Y C, Chua K C 2001 A capacity analysis for the IEEE 802.11 MAC protocol. *Wireless Networks*, 7: 159–171

Tickoo O, Sikdar B 2004 Queueing analysis and delay mitigation in IEEE 802.11 Random access MAC based wireless networks. *Proc. IEEE INFOCOM*

Vasudevan S, Papagiannaki K, Diot C, Kurose J, Towsley D 2005 Facilitating access point selection in IEEE 802.11 wireless networks. *Proc. of ACM SIGCOMM/USENIX Internet Measurement Conference*

Veeraraghavan M, Cocker N, Moors T 2001 Support of voice services in IEEE 802.11 wireless LANs. *Proc. IEEE INFOCOM*, 488–497

Vukovic I 1998 HIPERLAN Type 1: Performance analysis of the channel access control protocol. *Proc. IEEE Vehicular Tech. Conf.*

Wattenhofer R, Li L, Bahl P, Wang Y M 2001 Distributed topology for power efficient operation in multihop wireless ad hoc networks. *Proc. IEEE INFOCOM*

Xu S, Saadawi T 2002 Revealing the problems with the 802.11 medium access control protocol in multihop wireless ad hoc networks. *Computer Networks*, 38: 531–548

Xu Y, Heidemann J, Estrin D 2000 *Adaptive energy conserving routing for multihop ad hoc networks*. Research Report 527, USC/Information Sciences Institute

Ye F, Yi S, Sikdar B 2003 Improving spatial re-use of 802.11 based ad hoc networks. *Proc. IEEE Globecom*

Yu Liyang, Wang Neng, Meng Xiaoqiao 2005 Real-time forest fire detection with wireless sensor networks. *Proc. Wireless Communication, Networking and Mobile Computing Conf.*

Zhao Feng, Guibas Leonidas 2004 *Wireless Sensor Networks: An Information Processing Approach*. 1 edn. Morgan Kufmann, Elsevier