5

Upper Layer Techniques

5.1 Introduction

The upper layers refer to the layers in the communications model, as shown in Figure 4.1, that are immediately above the physical layer. In the open systems interconnection (OSI) model, these are the link layer that provides link control and the network layer that provides message routing. In the transmission control protocol/internet protocol (TCP/IP) model, they are the upper part of the network interface layer that provides link control and the internet layer that provides message routing. In both models, these layers are the layers that are the closest to the physical layer and therefore they have significant impact on the performances of transmission, detection and estimation in the physical layer. Thus, they will be considered in this chapter for energy harvesting wireless communications.

In the upper layers, the link layer provides control functions for the successful operation of the physical layer and sometimes, it is called the media access control (MAC) layer. These control functions or the MAC protocols coordinate the transmission and reception in the physical layer for each node by performing resource allocation and scheduling. For example, which user should be activated in multi-user scheduling, which channel should be assigned in access control, and how long and how much power each user should use in duty cycle and power management. These functions are vital for the smooth operation of the physical layer. In energy harvesting wireless communications, such resource allocation and scheduling have additional constraints from the energy availability. For example, some user may be activated and assigned a channel but may not have enough energy to perform data transmission. Thus, the MAC protocol designs should take the energy availability into account to maximize the overall efficiency in energy harvesting wireless communications.

The network layer in the OSI model or the internet layer in the TCP/IP model provides routing functions. In many communications systems, the source and the destination may be too far away from each other or may belong to different networks so that a proper route between them needs to be set up to allow their information exchange. The choice of the route will certainly affect the performance of the physical layer, as different routes may have different channel qualities as well as different end-to-end delay or latency due to different traffic. For energy harvesting wireless communications, new routing protocols should be designed, as the nodes on the route chosen may not have enough energy to forward the message so that the energy availability should be accounted for in the routing metric.

Energy Harvesting Communications: Principles and Theories, First Edition. Yunfei Chen.

We will focus on the MAC protocols and the routing protocols for energy harvesting wireless communications in this chapter. There are two main changes brought up by energy harvesting. Firstly, unlike battery power or mains connection, the energy harvester could provide unlimited power. Secondly, unlike battery power or mains connection, the harvested power is not stable or has great uncertainty. The new MAC protocols and routing protocols for energy harvesting wireless communications need to take these two changes into account in their designs. We will discuss the MAC protocols for energy harvesting communications first. Then, we will discuss the routing protocols for energy harvesting communications. Finally, we will discuss other important issues for energy harvesting communications, such as scheduling and effective capacity.

5.2 Media Access Control Protocols

Since the MAC layer is the layer closest to the physical layer, it has the largest impact on the physical layer transmission and reception. Also, a wireless sensor network (WSN) is an important application of energy harvesting wireless communications. Thus, we will focus on the MAC layers for a WSN in this section. For a WSN, due to its low-power operation and large-scale deployment, it is very difficult to recharge or replace the batteries in most cases. Hence, one primary objective of a WSN is to extend its network lifetime. Many methods have been proposed to extend the lifetime of a WSN. For example, power management at the nodes can be introduced to use the energy more efficiently, but this only delays the drainage of the battery. Incremental deployment can be used to replace the old nodes with drained battery by new nodes but this is not environmentally sustainable. Energy harvesting is a promising solution to this issue.

There are many MAC protocols proposed for conventional WSNs without energy harvesting, and they can be categorized into synchronous and asynchronous protocols. The synchronous protocols include S-MAC (Ye et al. 2002), T-MAC (Dam and Langendoen 2003) and the beacon mode of IEEE 802.15.4. They synchronize the transmitting node and the receiving node so that the nodes have their active or sleep states aligned in the time domain. This reduces the listening time but incurs extra overhead on maintaining the synchronization. The asynchronous protocols include B-MAC (Polastre et al. 2004), X-MAC (Buettner et al. 2006) and RI-MAC (Sun et al. 2008), where different nodes sleep and activate at different times independently. This reduces the hardware and overhead requirements, but such an approach requires a long listening time at the transmitter to wait for the receiver to wake up in some cases. These are all based on battery power.

Energy harvesting can extend the network lifetime but also bring great uncertainty. In energy harvesting WSNs, the amount of energy harvested at different nodes may vary. Hence, the synchronous protocols may not work, as some nodes may become unusable due to insufficient energy during the communications. On the other hand, the asynchronous protocols allow nodes to operate independently and hence can be adapted to energy harvesting WSNs. The main concern in energy harvesting WSNs is the uncertainty of the energy supply, similar to the physical layer. Thus, most MAC protocol designs focus on the energy availability.

5.2.1 Duty Cycling

As mentioned before, the network lifetime is the primary objective in WSNs. Thus, many methods have been proposed to save energy so that the network lifetime can be prolonged. For example, transmission power control can be implemented at the sensing node so that a lower transmission power can be used whenever possible (Ramanathan and Hain 2000). Also, dynamic voltage scaling can be used to improve the sensing circuit for higher energy efficiency. These methods require hardware modifications to the commonly used nodes. Alternatively, one can keep the power constant and use the common circuit but control the transmission time to save energy. This leads to the duty cycling method. In the duty cycling method, there is at least one operational mode when the node consumes no energy or negligible energy, such as a sleep mode. The node goes to sleep immediately after it finishes the communications task to save energy. The smaller the duty cycle is or the longer the sleep time is, the more energy the node can save but the less tasks the node can perform. One needs to strike a balance between the sleep mode and the operation mode. This boils down to the design of the duty cycle in the MAC protocols so that the sensing node can accomplish the task while saving as much energy as possible.

In energy harvesting wireless communications, the uncertainty in energy supply brings extra challenges to the design. Specifically, the sensing node may want to transmit data but due to insufficient energy harvested, the transmission is either not possible or given up in the middle, causing an energy outage. Thus, the duty cycle should be adapted to the residual energy or the energy arrival process at the node. This is the main difference between duty cycle designs for energy harvesting systems and those for conventional systems. To this end, several duty cycle designs have been studied. They are slightly different depending on the source of energy harvested. For wireless power transfer, since this transfer is intentional, there is less uncertainty in the energy supply but the charging sequence for a mobile charger (Peng et al. 2010) or the fairness for a fixed charger (Kim et al. 2011) should be accounted for. For ambient energy harvesting, these energy sources are less controllable and hence there is more uncertainty in the energy supply so that duty cycle adjustment is more important (Hsu et al. 2006; Yoo et al. 2012). We start with the MAC protocol using wireless power transfer (Kim et al. 2011).

5.2.1.1 Wireless Power Transfer

In Kim et al. (2011), an experimental system was built and simulated. Later in Kim and Lee (2011), the performance of this system was analyzed. This protocol was called energy adaptive MAC (EA-MAC). Specifically, consider a WSN using radio frequency (RF) power transfer, which assumes a star topology with one master node and I slave nodes. At each node, there is a set of RF front dedicated to data transmission and a separate set of RF front dedicated to power transfer so that energy harvesting and data transmission can be performed at the same time without interfering with each other. The master node operates with fixed power connection and broadcasts RF power to the slave nodes. It is always active and ready to receive data from the slave nodes. The slave nodes keep harvesting the RF energy. For data transmission, they operate in two states: the sleep state when they switch off; and the active state when they transmit data. A contention-based carrier sensing multiple access with collision avoidance (CSMA/CA) is used so that in the active state the slave node only transmits the data when it acquires the channel. Otherwise, it goes back to sleep. The whole process is described by Figure 5.1.

The master node has a fixed location. Hence, the energy harvested at each slave node is different, as according to the Friis formula, the received power is proportional to the inverse of the squared distance so that slave nodes closer to the master node can

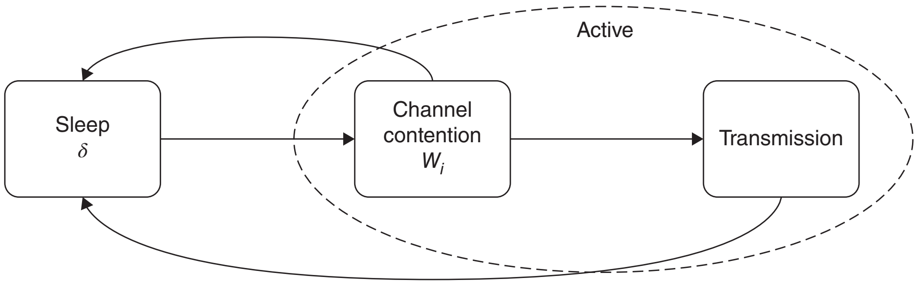

Figure 5.1 The state transition process of each node.

harvest more energy and vice versa. Consequently, the amount of energy available at different slave nodes will be different. This causes unfairness, as for the same amount of data, some slave nodes may not be able to complete the transmission due to insufficient energy. Also, due to insufficient energy, some slave nodes may have less chance to access the channel in CSMA/CA.

The EA-MAC protocol aims to tackle this fairness issue by introducing two additional changes to the MAC protocol. First, the duty cycle is adaptive to the harvested energy level at each slave node. More specifically, when the slave node wakes up from the sleep mode, it will check if its harvested energy reaches a threshold δ. This threshold is the amount of energy required to transmit one data packet. If it does, it moves to the active mode and starts to contend for the channel. If it acquires the channel, it will finish the transmission and then go back to the sleep mode. If the channel is busy or its energy is below δ, it will go back to the sleep mode and will keep checking and harvesting energy until it reaches δ before contending for the channel again. Secondly, the CSMA/CA algorithm is adaptive to the harvested energy level too. The maximum number of backoff slots will be set to $w_i * 2^{B_i} - 1$, where B_i is the backoff exponent in the usual CSMA/CA but w_i is a weighting factor as the ratio of the harvested energy level at the ith node to the average harvested energy among all slave nodes. This is compared with the conventional CSMA/CA algorithm with a maximum number of backoff slots of $2^{B_i} - 1$ at the ith slave node.

The use of δ ensures that all slave nodes transmit at the same energy level to avoid interrupted or impossible transmission due to insufficient energy. It changes the duty cycle to level up the harvested energy at different slave nodes, or it compensates the spatial difference among nodes with different duty cycles. The use of w_i ensures that nodes with less harvested energy can have a smaller backoff time so that they will not spend too much energy on contending the channel. It protects slave nodes with less harvested energy from being disadvantaged in channel access, as nodes with less energy can check the channel status less frequently than nodes with more energy so that eventually their energy levels will reach an equilibrium.

Adding a few more assumptions, the performance of EA-MAC can be analyzed. In this case, assume that all nodes can hear from each other well so that there is no hidden terminal problem. The data packet at each node has the same size and each slave node has only one data packet to transmit at a time. Each slave node has a deterministic energy harvesting rate determined by the distance only, due to the constant power broadcast from the master node. Define the time interval between two contentions as a round.

During the jth round, the ith slave node spends a time period of $T_{s,i,j}$ in the sleep mode, $T_{c,i,j}$ on the channel contention, and $T_{t,i,j}$ on packet transmission. Thus, the throughput of the ith slave node is given by (Kim and Lee 2011)

$$S_i = \frac{d \sum_{j=1}^{J} T_{p,i,j}}{\sum_{j=1}^{J} T_{c,i,j} + \sum_{j=1}^{J} T_{t,i,j} + \sum_{j=1}^{J} T_{s,i,j}} \tag{5.1}$$

where J is the total number of rounds, $T_{p,i,j}$ is the time interval of successful data transmission during the jth round, and d is the data rate.

Further, in the EA-MAC protocol, the slave node only wakes up to check the channel status and performs data transmission when its energy reaches the threshold δ determined by the amount of energy used to transmit one data packet. Thus, one has at the ith slave node

$$(P_i - P_s)T_{s,i,j} = (P_c - P_i)T_{c,i,j} + (P_t - P_i)T_{t,i,j} = \delta \tag{5.2}$$

where $P_i = \eta P_{tx} G_t G_r \left(\frac{\lambda}{4\pi d_i}\right)^2$ is the received energy at the ith slave node, η is the conversion efficiency of the energy harvester, P_{tx} is the master node transmission power, G_t and G_r are the antenna gains at the master and slave nodes, respectively, λ is the wavelength for power transfer, d_i is the distance between the master node and the ith slave node, P_s is the power consumption during the sleep mode, P_c is the power consumption for channel contention, P_t is the power consumption for data transmission, and other symbols are defined as before. Using (5.2) in (5.1), one has

$$S_i = \frac{d \sum_{j=1}^{J} T_{p,i,j}}{(1+\alpha_i) \sum_{j=1}^{J} T_{c,i,j} + (1+\beta_i) \sum_{j=1}^{J} T_{t,i,j}} \tag{5.3}$$

where $\alpha_i = \frac{P_c - P_i}{P_i - P_s}$ and $\beta_i = \frac{P_t - P_i}{P_i - P_s}$. Assuming that $\bar{T}_{p,i}$, $\bar{T}_{c,i}$, and $\bar{T}_{t,i}$ are the averages of $T_{p,i,j}$, $T_{c,i,j}$, $T_{t,i,j}$ across J rounds so that $\bar{T}_{p,i}J = \sum_{j=1}^{J} T_{p,i,j}$, $\bar{T}_{c,i}J = \sum_{j=1}^{J} T_{c,i,j}$, and $\bar{T}_{t,i}J = \sum_{j=1}^{J} T_{t,i,j}$. One further has

$$S_i = \frac{d\bar{T}_{p,i}}{(1+\alpha_i)\bar{T}_{c,i} + (1+\beta_i)\bar{T}_{t,i}}. \tag{5.4}$$

Thus, one needs the time averages to calculate the throughput.

Using a 2-D discrete-time Markov chain to describe the backoff process, the average contention time can be calculated as (Kim and Lee 2011)

$$\begin{aligned}\bar{T}_{c,i} = {} & \frac{W}{2} \sum_{v=0}^{m-1} q_i^v (1 - q_i)[w_i 2^{B_{min}} (2^{v+1} - 1) - v - 1] \\ & + \frac{W}{2} q_i^m [w_i 2^{B_{min}} (2^{m+1} - 1) - m - 1]\end{aligned} \tag{5.5}$$

where W is the time duration of each backoff slot, w_i is the weighting factor, m is the maximum number of backoff slots minus 1, q_i is the probability of a busy channel during contention at the ith slave node, and $B_{min} = \min\{B_i\}$. The details of the calculation can be found in Kim and Lee (2011). Similarly, one has

$$\bar{T}_{t,i} = T(1 - q_i^{m+1}) \tag{5.6}$$

and

$$\bar{T}_{p,i} = q_{s,i}T(1 - q_i^{m+1}) \tag{5.7}$$

where T is the time duration of one data packet and is a constant, q_i is still the probability of a busy channel during contention at the ith slave node, and $q_{s,i}$ is the probability of successful transmission without collision at the ith slave node. These two probabilities can be calculated as

$$q_i = 1 - \prod_{j\in\Phi, j\neq i} (1 - q_{t,j}) \tag{5.8}$$

with

$$q_{t,i} = \frac{\bar{T}_{t,i}}{(1+\alpha_i)\bar{T}_{c,i} + (1+\beta_i)\bar{T}_{t,i}} \tag{5.9}$$

and

$$q_{s,i} = \prod_{j\in\Phi, j\neq i} (1 - \tau_j q_{c,j}) \tag{5.10}$$

with

$$q_{c,i} = \frac{\bar{T}_{c,i}}{(1+\alpha_i)\bar{T}_{c,i} + (1+\beta_i)\bar{T}_{t,i}} \tag{5.11}$$

and

$$\tau_i = \frac{2(1-2q_i)(1-q_i^{m+1})}{w_i 2^{B_{min}}(1-(2q_i)^{m+1})(1-q_i) + (1-q_i^{m+1})(1-2q_i)} \tag{5.12}$$

where Φ represents the set of slave nodes that contend for the channel with the ith node.

From the above, the throughput of the ith slave node can be calculated by following an iterative procedure. First, initial values for q_i and $q_{s,i}$ are chosen. Using these initial values, the time averages of $\bar{T}_{c,i}$, $\bar{T}_{p,i}$, and $\bar{T}_{t,i}$ can be calculated using (5.5)–(5.7). Then, using the time averages of $\bar{T}_{c,i}$, $\bar{T}_{p,i}$, and $\bar{T}_{t,i}$, one can calculate S_i, $q_{t,i}$, $q_{c,i}$, and τ_i using (5.4), (5.9), (5.11), and (5.12), respectively. Using $q_{t,i}$, $q_{c,i}$, and τ_i, the values of q_i and $q_{s,i}$ can be updated by (5.8) and (5.10), respectively. The whole process iterates until S_i converges. It was shown in Kim and Lee (2011) that this value converges to a unique solution after around 5 iterations. It was also shown there that the analytical calculation above matches well with the simulation result and that the throughput increases with the transmission power and decreases with the distance from the master node, as expected.

Further, the Jain's fairness index of the protocol is defined as $F = \frac{(\sum_{i=1}^{I} S_i)^2}{I\sum_{i=1}^{I} S_i^2}$ and can be examined, which is a value between 0 and 1 (Jain et al. 1999). It was reported in Kim et al. (2011) that the EA-MAC protocol with adaptive contention using w_i can achieve a fairness index of around 0.9, while the EA-MAC protocol without adaptive contention when $w_i = 1$ can achieve a fairness index of around 0.4. Thus, the EA-MAC protocol improves fairness by adapting the protocol to the harvested energy.

5.2.1.2 Ambient Energy Harvesting

The work in the above subsection uses the RF power transfer. This reduces the uncertainty in power supply but leads to unfairness due to different user locations. On the other hand, if ambient energy harvesting is used, where the sensing nodes are powered by the Sun or wind, etc., there is more uncertainty in power supply but less unfairness among users. For example, the sensors in the same area may receive the same amount of solar power (except in some extreme cases where some sensors are located in shaded areas) so that the geographical locations will not affect the harvested energy much, but the Sun comes and goes causing uncertainty. In this subsection, we study the duty cycle problem in WSNs where the sensing nodes harvest the ambient energy. The fundamental problem here is very similar to what we studied in Section 4.2.5, where we have to ensure that the total energy consumed is smaller than the total energy harvested. The objective is to maximize the duty cycle, which determines the performance of the WSN, by taking the random energy arrival into account. We will consider the more complicated protocol in Hsu et al. (2006) first, followed by the less complicated protocol in Yoo et al. (2012).

To enable adaptive duty cycle using ambient energy harvesting, one needs the energy arrival model, the energy consumption model and the energy storage model when power management is used, so that the duty cycle of the sensing node can be adapted. In this case, denote $P_s(t)$ as the harvested energy at time t and $P_c(t)$ as the consumed energy at time t. The harvester has a conversion efficiency of η. Also, assume that the energy storage has an initial energy of G_0. The duty cycle is assumed to have a linear relationship with the utility or the performance of the WSN. For example, the amount of transmitted data is proportional to the transmission time. For sensing nodes that detect intrusion, the detection probability increases linearly with the duty cycle too. Also, the linear relationship is a good approximation to the non-linear relationships in some cases. Thus, to maximize the utility, we need to maximize the duty cycle whenever possible.

The duty cycle needs to be adjusted at different times. Assume that each sensor operates with a time slot of T seconds and that the adaptation of the duty cycle is performed for I time slots. Assume that P_i is the average amount of energy harvested during the ith time slot, $i = 1, 2, \cdots, I$. Also, P_c is the power consumption assumed constant during each time slot, D_i is the duty cycle in the ith time slot to be adapted, and G_i is the amount of energy in the energy storage at the beginning of the ith time slot.

There are two possible cases on the energy use in each time slot. If the power consumption P_c is larger than P_i in the ith time slot, the consumed energy will be taken from both the energy harvester and the energy storage. On the other hand, if the power consumption P_c is smaller than P_i, the consumed energy will only be taken from the energy harvester.

Thus, the energy change in the storage during the ith time slot follows

$$G_i - G_{i+1} = TD_i[P_c - P_i]^+ - \eta T(1 - D_i)P_i - \eta TD_i[P_i - P_c]^+ \tag{5.13}$$

where $[x]^+ = x$ when $x > 0$ and $[x]^+ = 0$ when $x < 0$. In (5.13), TD_i is the active time of the node and $T(1 - D_i)$ is the sleep time of the node. Also, the first term represents the energy taken from the storage when $P_c > P_i$, the second term represents the energy added to the storage when the node sleeps, and the third term represents the energy added to the storage when the node is active with $P_c < P_i$.

Using (5.13), the duty cycle adaptation problem can be formulated as (Hsu et al. 2006)

$$\max_{D_1, D_2, \cdots, D_I} \left\{ \sum_{i=1}^{I} D_i \right\} \tag{5.14}$$

$$G_i - G_{i+1} = TD_i[P_c - P_i]^+ - \eta T(1 - D_i)P_i - \eta TD_i[P_i - P_c]^+ \tag{5.15}$$

$$G_1 = G_0 \tag{5.16}$$

$$G_{I+1} \leq G_0 \tag{5.17}$$

$$D_{max} \leq D_i \leq D_{min}, i = 1, 2, \cdots, I \tag{5.18}$$

where the total duty cycle is maximized over all D_i, with a constraint on the energy change in (5.15), an initial energy of G_0 in (5.16), a constraint on the final energy that must be larger than the initial energy in (5.17), and practical limits on the duty cycle in (5.18). The constraint in (5.17) is also called the energy neutrality condition. Several comments can be made. First, the optimization above does not consider the storage capacity. In reality, the battery capacity is limited with a finite size B. This can be applied in (5.15) and (5.17). Secondly, although the optimization contains the non-linear functions $[x]^+$, these functions depend on constants only and do not depend on the variables to be optimized. Thus, standard linear programming methods can be used to solve the optimization problem. The optimum duty cycles will depend on P_i. Thus, one must have full knowledge of all P_i for $i = 1, 2, \cdots, I$. This is the deterministic model of energy arrival process discussed in Section 3.4.

Hence, the optimization algorithm can be implemented in three steps: the first step acquires knowledge of the past and future energy availability for P_i; the second step solves the optimization problem using linear programming; and the last step dynamically adapts the duty cycle based on the optimum solutions. Next, we simplify the algorithm in (5.14) further. First, we define two sets as

$$S = \{i | P_i - P_c \geq 0\} \tag{5.19a}$$

$$W = \{i | P_c - P_i > 0\}. \tag{5.19b}$$

Using (5.19a,b), (5.15) can be summed up over all I time slots to give

$$\sum_{i=1}^{I}(G_i - G_{i+1}) = \sum_{i \in W} TD_i[P_c - P_i] - \sum_{i=1}^{I} \eta TP_i + \sum_{i=1}^{I} \eta TP_iD_i - \sum_{i \in S} \eta TD_i[P_i - P_c] \tag{5.20}$$

where the term on the left-hand side of the equation is the overall energy change in the storage and the term on the right-hand side of the equation is the total energy used over I time slots. For energy neutrality operation, we would like to set $\sum_{i=1}^{I}(G_i - G_{i+1}) = 0$ so that all harvested energy is used over the I time slots to maintain the energy level in the storage. This gives

$$\sum_{i=1}^{I} P_i = \sum_{i \in W} D_i \left[\frac{P_c}{\eta} + P_i \left(1 - \frac{1}{\eta} \right) \right] + \sum_{i \in S} P_cD_i \tag{5.21}$$

from (5.20). The left-hand side of the equation represents the total energy harvested, while the right-hand side of the equation represents the energy used during W and S time slots, respectively. Using (5.21), the optimization problem is simplified as (Hsu et al. 2006)

$$\max_{D_1, D_2, \cdots, D_I} \left\{ \sum_{i=1}^{I} D_i \right\} \tag{5.22}$$

$$\begin{aligned} &\sum_{i=1}^{I} P_i = \sum_{i \in W} D_i \left[\frac{P_c}{\eta} + P_i \left(1 - \frac{1}{\eta} \right) \right] + \sum_{i \in S} P_c D_i \\ &D_{max} \leq D_i \leq D_{min}, i = 1, 2, \cdots, I \end{aligned} \tag{5.23}$$

where the constraints on the energy change in each time slot are replaced by a constraint on the overall energy change over all I time slots. Hsu et al. (2006) proposed a simple iterative solution to (5.22) by using only simple arithmetic operations and sorting, which are suitable for embedded computation. Further, the error between the predicted harvested energy and the actual harvested energy was accounted for to improve the performance of the algorithm. The performances of the optimal algorithm in (5.14), the adaptive algorithm in (5.22), and the simple algorithm without duty cycle adaptation were compared using solar energy harvesting. It was shown that the performances of the optimal algorithm and the adaptive algorithm are graphically indistinguishable from each other, both of which are better than the simple algorithm without duty cycle adaptation. The performance gain varies from around 50 to 0%, depending on the conversion efficiency η. Some other variants of this problem can also be studied. For example, an energy storage with finite size can be considered. In this case, the energy change is limited by the storage size. Also, $\sum_{i=1}^{I}(G_i - G_{i+1})$ does not have to be zero and can be relaxed.

In Yoo et al. (2012), the authors reported two new MAC protocols called duty-cycle scheduling based on residual energy, DSR-MAC, and duty-cycle scheduling based on prospective increase in residual energy, DSP-MAC, to reduce the end-to-end delay and also to increase fairness among sensing nodes.

Specifically, in the DSR-MAC protocol, the duty cycle is set as (Yoo et al. 2012)

$$\begin{aligned} &D_i = D_{max} - D_{max} \frac{G_i - G_t}{G_{max} - G_t}, \text{when } G_i > G_t \\ &D_i = D_{max}, \qquad\qquad\qquad\quad \text{when } G_i \leq G_t \end{aligned} \tag{5.24}$$

where G_i is the residual energy at the ith node, G_t is an adjustable threshold to meet the minimum requirement of the concerned application, G_{max} is the maximum possible E_i (not necessarily the battery capacity), and D_{max} is the maximum duty cycle depending on the application. Thus, in this protocol, the authors proposed to reduce the duty cycle when the residual energy increases and vice versa.

In the DSP-MAC protocol, the prediction of future energy increase is used to adjust the duty cycle more aggressively. Assume that the average harvested power over a time duration of T is P_i and that the average power consumption over this duration is P_c. Thus, the residual energy level will be increased from G_i to $G_i + (P_i - P_c)T$ at the end of time duration T, when $P_i > P_c$ and $G_i > B_t$. Thus, if the values of P_i, P_c, and T are available and when $P_i > P_c$ and $G_i > B_t$, the duty cycle of the ith node will be set as (Yoo et al. 2012)

$$D_i = D_{max} - D_{max}\frac{G_i + (P_i - P_c)T - G_t}{G_{max} - G_t} \tag{5.25}$$

which effectively replaces G_i with $G_i + (P_i - P_c)T$ in (5.24). When $P_i < P_c$ or $G_i < B_t$, the DSP-MAC switches to the DSR-MAC mode. It is clear that the effectiveness of this protocol depends heavily on the accuracy of the estimated values of P_i, P_c, and T. A large estimation error will lead to a significant performance degradation. In Yoo et al. (2012), the authors used

$$T = \frac{G_{max} - G_i}{P_i - P_c} \tag{5.26}$$

to calculate the value of T when $P_i > P_c$. Using the RI-MAC in Sun et al. (2008) as a benchmark, the authors showed in Yoo et al. (2012) that both the end-to-end delays and the packet delivery rates of DSR-MAC and DSP-MAC are better than the RI-MAC, and the performance gains increase when the number of nodes increases. They also showed that the DSR-MAC and DSP-MAC have similar Jain's fairness index, both of which are higher than that of the RI-MAC.

5.2.2 Other Issues in MAC Protocols

In addition to the study of duty cycle in the MAC protocols, other issues in the MAC protocols have also been investigated in the literature.

For example, in Iannello et al. (2012), the performances of two existing MAC protocols have been evaluated in the case when ambient energy harvesting is used. Two classical protocols were considered: time division multiple access (TDMA) where each node is allocated a fixed time slot and no other nodes can use it even when the concerned node does not have any data or enough energy to transmit, and the ALOHA structure where different nodes contend to access the channel. The main effect of ambient energy harvesting is the availability of energy when it needs to transmit data. The energy uncertainty was considered to study the delivery probability and the time efficiency of the two MAC protocols. Numerical results showed that TDMA always has a larger delivery probability than ALOHA, as TDMA has fixed channel access. For the time efficiency, ALOHA is not necessarily better than TDMA either. Unfortunately, the authors did not provide any comparison between energy harvesting MAC protocols and traditional MAC protocols to examine the effect of energy harvesting, which is perhaps more interesting than the comparison between TDMA and ALOHA.

In Ha et al. (2018), the authors proposed a harvest-then-transmit MAC protocol (HE-MAC) as an extension of the enhanced distributed coordination function used in current IEEE 802.11 standards considering wireless power. In this case, the sensor nodes receive data as well as wireless power from the hybrid access point. To coordinate the transmission of data and the transfer of power from the access point, the distributed coordination function needs to be re-designed. They used the Markov chain model to analyze the steady-state rate and based on this analysis, the energy harvesting rate was maximized subject to constraints on data performances. In Naderi et al. (2014), the authors optimized the wireless charging method for wireless powered WSNs.

Specifically, the position, the frequency and the number of wireless power transmitters have been investigated in terms of the sensor charging time. Based on this investigation, the wireless power transfer efficiency has been optimized, with minimum disruption to the data communication at the sensor nodes. The authors have also given some guidelines on how to choose the maximum energy harvesting threshold, the power transmitters, how to request charging, and different priorities of charging and data transmission. They showed a network throughput improvement of 300% compared with classical methods.

In Peng et al. (2010), Kim and Lee (2011), Kim et al. (2011), Naderi et al. (2014), and Ha et al. (2018), wireless power transfer is used. In this case, there is less uncertainty in the energy availability due to the intentional power transfer, and the design objective is more on the coordination between the data transmission time and the power transfer time or the adaptation to the transferred power. On the other hand, in Hsu et al. (2006), Iannello et al. (2012), and Yoo et al. (2012), ambient energy harvesting is used. In this case, there is more uncertainty in the energy availability due to the unreliable energy source so that the design objective is more on the scheduling of transmission time with energy constraints. Since the information processor and the energy harvester are normally separated in this case, there is no need for the coordination between data transmission time and power transfer time. Thus, the energy source renders a fundamental difference in the design objective.

There are other studies on MAC protocols for energy harvesting communications. For example, in Eu et al. (2011), the performances of existing CSMA and polling-based MAC protocols using ambient energy harvesting were evaluated and compared in terms of network throughput, fairness and end-to-end delay, similar to Iannello et al. (2012). The energy arrival models and the energy harvesters were studied in detail to be used for performance evaluation. In Fafoutis and Dragoni (2011), a new energy on-demand MAC protocol was proposed for WSNs using ambient energy harvesting. The idea is similar to Hsu et al. (2006) by maximizing the sensor performance using a consumed energy as close to the harvested energy as possible. This is again a design of duty cycle or transmission time based on the energy availability. A similar problem was also studied in Nguyen et al. (2014). Finally, in Yang et al. (2012) and Liu et al. (2015b), the charging time was considered with contention time and transmission time for wireless power transfer, which was also considered in Ha et al. (2018).

In summary, the energy source determines the design objective in the MAC protocols. If the energy source is the ambient environment, the data transceiver and the energy harvester operate independently at the sensor node such that the energy uncertainty is more important than the coordination between data transmission and energy harvesting. If the energy source is an intentional power transmitter that operates at the same band as data, the energy supply is more controllable with less uncertainty but the coordination between data transmission and energy harvesting due to limited channel access time becomes more important.

5.3 Routing Protocols

Routing is another main function of the upper layers. For WSNs, this function is particularly important, as most WSNs are low-power and low-data-rate applications such that the access point is normally out of the direct transmission ranges of the sensor nodes. Hence, multi-hop communications have to be used to forward the data packet from the sensing node to the access point. The physical layer performs the actual transmission and reception of the data packet, the MAC protocol coordinates the transmission and reception of the data packet at each node, while the routing protocol coordinates the transmission and reception of the data packet across the network. Specifically, the routing protocol needs to identify the best route from the sensor to the access point for multi-hop communications.

The criterion for the best route depends on the application requirements. For some applications, the network throughput is the most important metric. Thus, the best route can be chosen to maximize the network throughput. For other applications, the end-to-end delay or latency is more important. In this case, the best route can be chosen to minimize the delivery time with the quickest route. For WSNs, especially for low-power and lower-data-rate WSNs, throughput and latency are often not the main concern. The primary objective for these networks is to design a network that can transmit as many data packets as possible with a lifetime as long as possible.

The new energy harvesting feature at the sensors provides a promising solution to the network lifetime issue but also creates new problems for routing in energy harvesting systems. The main problem is that the energy supply at the sensor becomes random so that there might be energy outage when the sensor has an empty queue to forward data but it does not have enough energy to do so. Thus, the energy at the sensors must be utilized efficiently. This has been studied in the previous section on the duty cycling of the MAC protocol. For routing, in order to forward the data packet efficiently from the sensor to the access point via several hops, the helpers or the relays must be chosen carefully to avoid any energy outage.

To this end, for efficient routing, the dynamic nature of the energy availability at each sensor must be taken into account in the choice of the best route. Traditional fixed routing algorithms use predefined and fixed paths for packet forwarding. This requires the topology of the whole network as well as fixed nodes. However, the fixed routing algorithm cannot adapt to any changes in the operating environment, such as energy. Thus, they are not suitable for energy harvesting communications systems. On the other hand, opportunistic routing does not require the network topology and explores the broadcast nature of wireless to find helpers or relays en route to the access point in real-time. Thus, opportunistic routing is more suitable for energy harvesting communications systems.

To account for the dynamic nature of the energy availability in routing, some studies calculate the routing metric based on the residual energy at the nodes as current energy availability, some calculate the routing metric based on the harvesting rate as future energy availability, and some use both. In the following, we will discuss the use of the ambient energy first, followed by the use of wireless power for the routing protocols in energy harvesting communications systems.

5.3.1 Ambient Energy Harvesting

Similar to the MAC protocols, when the energy is harvested from the ambient sources, such as wind and vibration, there is great uncertainty in the energy availability. This uncertainty requires the redesign of routing protocols for energy harvesting networks.

In Shafieirad et al. (2018), the authors proposed a new energy-aware opportunistic routing protocol for large-scale WSNs. In this protocol, the sensor node sends its data packet to the access point or the fusion center via multiple hops. The candidate relays used to forward the data packet are selected to maximize the amount of delivered data packets, instead of maximizing the network throughput or minimizing the latency. This was called the "Max-SNR" routing protocol in Shafieirad et al. (2018).

Specifically, consider a network of N sensor nodes uniformly distributed within a circle of radius R that has one access point located at the center of the circle. Each node harvests energy from the environment and stores the harvested energy in a battery with infinite capacity. The whole process is divided into multiple time slots in a finite time horizon. Each node also has a data buffer with a finite size to store the data packets to be transmitted. The data packets come from the sensor node's own sensed data as well as from its neighbors that require forwarding. However, it only accepts data packets from the neighbors when its data buffer has room and when the neighbor is within its reception range of R_r. Assume that R_r is much smaller than R so that most sensor nodes need multi-hop communications to send the data packet to the access point, except those nodes within R_r of the access point. The design problem of the routing protocol is to choose the best neighbor in each hop to forward the data packet to the access point that maximizes the number of delivered packets.

The key point here is to choose the best neighbor. This choice must take the amount of available energy at each potential relay into account to avoid energy outage and it also needs to minimize the energy consumption or the number of hops so that the data packet can arrive at the access point with minimum energy. Assume that the ith node has an amount of energy E_i available at the time of selection and that the distance between this node and the access point is d_i. In Shafieirad et al. (2018), it proposed to use the selection criterion of

$$C_i = \frac{E_i}{d_i^m} \tag{5.27}$$

where m is the path loss exponent. Thus, the more energy available at the ith node, the more likely that the ith node will be chosen. Also, the shorter the distance between the ith node and the access point, the more likely that it will be chosen, as the energy consumption is inversely proportional to d_i^m. This is effectively the signal-to-noise ratio (SNR) that will be received at the access point if the ith node is chosen. Then, a timer is set at each node for a waiting time of

$$T_i = \frac{1}{C_i} \tag{5.28}$$

before the sensor node in the previous hop is allowed to forward the data packet. One can see that the larger the received SNR at the access point is, the smaller the waiting time will be and hence the more likely that it will be chosen. This is why it is called the "Max-SNR" protocol.

Using the maximum SNR criterion, the number of delivered packets has been analyzed. Denote $P_{i\to k}$ as the probability that the ith node's data packets will be forwarded by the kth node, $i, k = 1, 2, \cdots, N$. Denote $N(i)$ as the set of the ith node's neighbors that are within the transmission range R_r of the ith node with enough data buffer to store the forwarded data packets. Denote k as the node that has the maximum SNR. It was derived in Shafieirad et al. (2018) that the probability $P_{i\to k}$ is given by

$$P_{i\to k} = \int_0^\infty [1 - F_{E_k(d_k^m t)}] \sum_{n\in N(i), n\neq k} d_n^m f_{E_n}(d_n^m t) \left[\prod_{l\in N(i), l\neq n,k} F_{E_l}(d_l^m t) \right] dt \tag{5.29}$$

where $f_{E_i}(\cdot)$ and $F_{E_i}(\cdot)$ are the probability density function and the cumulative distribution function of the available energy E_i, respectively. Since the energy is harvested, these functions are determined by the energy arrival process as well as the energy conversion process at the harvester, as discussed in Section 3.4.

Using $P_{i\to k}$ in (5.29), the average number of data packets transmitted from the ith node to the kth node can be derived as

$$M_{ik} = \left(\sum_{l\in N(i)} M_{li} + M_i \right) P_{i\to k} \tag{5.30}$$

where M_i is the average number of data packets from the ith node's own sensed data, and M_{li} is the average number of data packets received from the lth neighbor that need to be forwarded to the access point. Thus, one has

$$M_{ik} - \sum_{l\in N(i)} M_{li}P_{i\to k} = M_iP_{i\to k} \tag{5.31}$$

where $M_iP_{i\to k}$ is the average number of delivered packets for the ith node's own data by the kth node, and $\sum_{l\in N(i)} M_{li}P_{i\to k}$ is the average number of data packets for the ith node's received data (to be forwarded by the ith node for its neighbors) by the kth node. Rewriting (5.31) in matrix form for all nodes, one has

$$\mathbf{AM} = \mathbf{P} \tag{5.32}$$

where $\mathbf{A}$ and $\mathbf{P}$ are the matrices determined by $P_{i\to k}$ and $\mathbf{M}$ is the vector containing all M_{ik}. The matrix equation in (5.32) can be solved for $\mathbf{M}$. Further, if the energy constraint is considered, it becomes an optimization problem of

$$\min_{\mathbf{M}} ||\mathbf{AM} - \mathbf{P}||^2 \tag{5.33}$$

$$M_{ik} \geq 0 \tag{5.34}$$

$$\sum_{l\in N(i)} M_{il}E \leq E_i \tag{5.35}$$

where the last constraint is the energy causality that the consumed energy must be smaller than the available energy, and E is the amount of energy required to transmit each data packet. The total average number of data packets transmitted by the ith node is then given by

$$M_i^t = \sum_{l\in N(i)} M_{il} \tag{5.36}$$

using the solutions from (5.33). More details of the derivation can be found in Shafieirad et al. (2018). It was shown that the "Max-SNR" protocol can achieve a delivery ratio of almost 100% for an exponential energy with parameter 1 and $N = 200$. Its performance is significantly better than the previous protocols that do not account for the energy availability, under the same conditions.

Note that, although the above results are derived for energy harvesting systems, it can be used for conventional systems too, as the derivation does not specify the randomness of the available energy E_i. Hence, the randomness could be caused by energy harvesting, or by energy consumption. Note also that the above derivation assumes independent energy availability at each node. In some energy harvesting systems, the nodes may harvest the same source and hence have correlated energies.

In Cao et al. (2016), both the energy currently available at the node and the energy to be harvested for future use at the node were used in the routing metric calculation. Specifically, in Cao et al. (2016), each node is assumed to have a table about its own energy status determined by three parameters: residual energy; energy harvesting rate; and energy harvesting density. These three parameters also determine the energy statuses for all its neighbors.

Denote E_i^r as the residual energy or the energy currently available at the ith node. For the use in this protocol, it was discretized as

$$L_i = k + 1 \tag{5.37}$$

where $k = 1, 2, \cdots, K-1$, K is the maximum energy level allowed at the node and is predetermined, k satisfies $\frac{k}{K} < \frac{E_i^r}{E_m} \leq \frac{k+1}{K}$ with E_m being the largest amount of energy or the battery capacity.

The energy consumption at the ith node is calculated as (Cao et al. 2016)

$$E_i^c = LE_0 + L\alpha d_i^m \tag{5.38}$$

where L is the number of bits in each data packet, E_0 is the energy consumption used to transmit each data bit, α is a constant representing energy loss, d is the distance between two nodes, and m is the path loss exponent.

The energy density at the ith node is calculated as (Cao et al. 2016)

$$D_i = \frac{\sum_{j \in N(i), j \neq i} (R_j T - E_j^c)}{|N(i)| + 1} \tag{5.39}$$

where $N(i)$ is the set of all neighbors of the ith node within its transmission range, $|N(i)|$ is the number of these neighbors, R_j is the energy harvesting rate, and T is the harvesting time. From (5.39), this metric takes both the harvested energy and the consumed energy into account. It will be the net energy added to the storage after the transmission. Comparing (5.27) with (5.39), one sees that (5.27) uses the ratio of current energy to the energy consumption, while (5.39) uses the difference between future energy and the energy consumption.

These values are initialized at the beginning of transmission using (5.37) and (5.39). During each transmission, the node with the largest energy density of D_i determined by (5.39) is chosen for the next hop. After each transmission, the three parameters in the tables at the jth node are updated to prepare for the next transmission. This process repeats until the data packet arrives at the access point. Simulation results have shown that this routing protocol always has a higher average residual energy compared with the conventional protocol. A similar protocol was also proposed in Kawashima and Sato (2013) by incorporating the power generation pattern, which is related to the energy harvesting rate, in the routing metric calculation.

In Martinez et al. (2014), in addition to the residual energy and the harvested energy, the energy wastage was also used in the routing metric calculation. Kawashima and Sato (2013) and Shafieirad et al. (2018) have assumed an infinite capacity for the energy storage at the sensor so that energy can be harvested as much as possible. In practice, the storage capacity is finite and hence too much energy harvesting may cause energy wastage in the network. This is really the other end of the problem, where instead of worrying about insufficient energy for data transmission, Martinez et al. (2014) worried about too much energy for data transmission.

In this case, assume a routing path of σ_n. If the ith node is on this path, it will cause an energy wastage of

$$E_i^{WN} = \max\{0, E_i + E_i^h - E_i^c - B\} \tag{5.40}$$

where E_i is the residual energy or the current energy in the battery as defined before, E_i^h is the energy to be harvested in the next time slot T, E_i^c is the energy to be consumed in the next time slot T, and B is the storage capacity. The equation in (5.40) essentially calculates the amount of energy that would exceed the storage capacity in the next time slot and hence would be wasted, if any. If the ith node is not on this path, it will cause an energy wastage of

$$E_i^{WF} = \max\{0, E_i + E_i^h - B\} \tag{5.41}$$

as it will not consume any energy in the next time slot. Hence, if the path σ_n is chosen for routing, the total energy consumption including that used for data transmission and wasted is

$$C(\sigma_n) = \sum_{i\in\sigma_n}(E_i^c + E_i^{WN}) + \sum_{i\notin\sigma_n} E_i^{WF} \tag{5.42}$$

and the total energy available for routing is

$$E(\sigma_n) = \sum_{i\in\sigma_n}(E_i^c + E_i^h). \tag{5.43}$$

Thus, the routing path is selected so that the total remaining energy after the routing is maximized as

$$\sigma^* = \arg\max_{\sigma_n\in\Omega}[E(\sigma_n) - C(\sigma_n)] \tag{5.44}$$

where Ω is the set of all possible routes and σ^* is the optimum route. This routing metric calculation takes the residual energy E_i, the harvested energy E_i^h and the wasted energy E_i^{WF} and E_i^{WN} into account. Simulation results were provided in Martinez et al. (2014) to show the performance of this protocol using solar energy. As expected, higher residual energy levels can be achieved using the proposed routing protocol.

The above studies only consider routing. In other studies, routing is also considered jointly with other functions in the upper layer. To this end, Avallone and Banchs (2016) considered a joint channel assignment and routing algorithm for mesh networks with extra constraints on the energy availability. This is applicable to systems with multiple channels so that both routing and channel assignment can increase the chance of success for multi-hop communications. In Hasenfratz et al. (2010), different routing protocols were investigated and compared in conjunction with the MAC protocols by maximizing the delivery ratio or the packet loss. Specifically, three routing algorithms of randomized Max-Flow, energy opportunistic weighted minimum energy and randomized minimum path recovery time have been studied.

5.3.2 Wireless Power Transfer

The studies in the previous subsection used ambient energy harvesting. Because of this, there is great uncertainty in the energy availability. Consequently, it is important to perform routing by taking the residual energy, the harvested energy and the wasted energy into account for maximum delivery ratio. In this subsection, wireless power transfer will be considered. As discussed before, this generates less uncertainty in the energy supply. However, since power transfer and data transmission may be performed in the same frequency band, coordination is required. Thus, studies on routing protocols in wireless powered systems mainly consider this coordination.

In Doose et al. (2010), wireless-charging-aware routing protocols were studied. Due to different distances and heights of the sensors to the power transmitter, the amount of energy received at different sensors is different. Hence, to reach a certain level of energy, the charging time for different sensors needs to be different too.

Denote t_i as the average charging time of the ith node and ϵ_i as the standard deviation of the charging time at the ith node. These values can be obtained experimentally. Also, denote $T_{max}(\sigma_n)$ and $\epsilon_{max}(\sigma_n)$ as the maximum charging time and the maximum standard deviation among all nodes on the path σ_n, respectively. These two maximum values will be included in the routing request packet and this packet will be forwarded to each node on the path and will be updated every time a sensor receives it by comparing its own charging time and standard deviation with them.

When the sensor receives the routing request packet, it waits for a delay of $t_i + \epsilon_i$ before forwarding it, so that the nodes with shorter charging times can forward it earlier. When the destination receives all these forwarded request packets from different nodes, it will choose the path with the minimum $T_{max}(\sigma_n)$ so that

$$\sigma^* = \min_{\sigma_n \in \Omega}\{T_{max}(\sigma_n)\}. \tag{5.45}$$

After the destination chooses the path with the minimum maximum charging time, it sends back the route reply packet to inform the nodes on the chosen route or path. In this case, since charging and transmission use the same frequency band, the destination has to optimize the time allocation for charging and transmission with a fixed total of $T_c + T_x = T$, where T_c is the charging time, T_x is the transmission time, and T is the total packet time. The optimization problem proposed in Doose et al. (2010) hence becomes

$$\max_{T_c}\left\{\frac{T_x R}{T}\right\} \tag{5.46}$$

$$(P_r - P_d)T_c - P_t T_x \geq 0 \tag{5.47}$$

$$N\left(T_c + \frac{P+H}{R}\right) \leq L_{max} \tag{5.48}$$

$$\frac{1}{S_0}[1 - kte^{\frac{-4700}{M+273}}] > \frac{1}{S_{max}} \tag{5.49}$$

$$T = T_c + T_x \tag{5.50}$$

where the throughput is maximized with respect to T_c, R is the data rate, P_r is the power harvested from the wireless charger during T_c, P_d is the idle power consumed during charging, P_t is the transmission power for data, N is the total number of nodes on the chosen path, P is the size of data in the packet, H is the size of overhead in the packet, L_{max} is the maximum end-to-end delay or latency allowed, S_0 is the initial equivalent series resistance of the supercapacitor, S_{max} is the maximum equivalent series resistance of the supercapacitor allowed, beyond which the capacitor will not work, k is a design constant, t is the operation time, and M is the absolute temperature. The first constraint makes sure that the consumed energy will be smaller than the harvested energy so that the node can stay alive after transmission. The second constraint limits the total latency caused by routing. The third constraint makes sure that the energy storage will work properly.

From the above optimization, the performance of this routing protocol was studied in Doose et al. (2010) in terms of throughput, latency, network lifetime, and residual energy. It was reported there that the latency decreases with the charging rate and increases with the size of data in the packet. Also, the maximum throughput decreases with the optimal charging time but the network lifetime increases with the optimal charging time.

In Tong et al. (2010) and Li et al. (2011), similar problems have been studied. Specifically, in Li et al. (2011), a joint charging and routing algorithm was proposed, where the sensor node is charged when needed to prolong the network lifetime, while in Tong et al. (2010), the sensor deployment was jointly designed with routing to make the best use of wireless charging.

5.4 Other Issues in the Upper Layers

The previous two sections have mainly examined the MAC protocols and the routing protocols for energy harvesting wireless communications. They are the two most important tasks of the upper layer. Next, we examine some other issues in the upper layer: scheduling; and effective capacity.

5.4.1 Scheduling

In transmission scheduling, the main problem is to adjust the transmission time and the transmission power for each time slot so that all data packets can be delivered by the minimum deadline, assuming randomly arriving data packets. Thus, one aims to minimize the delay with respect to transmission time and transmission power, under different traffic models. In energy harvesting wireless communications, this problem is further complicated by the fact that the energy arrives randomly too so that the transmission time and the transmission power must be adapted to the energy availability. Thus, the transmission scheduling problem in energy harvesting wireless communications requires the minimization of the delivery time with respect to the transmission time and the transmission power, for both randomly arriving data packets and randomly arriving energy. Figure 5.2 shows the transmission scheduling problem considered. This problem was studied in Yang and Ulukus (2012) for a single user case.

Assume that the initial amount of data packet is B_0 bits and the initial amount of energy is E_0 at the transmitter. During the packet delivery, the energy is harvested with

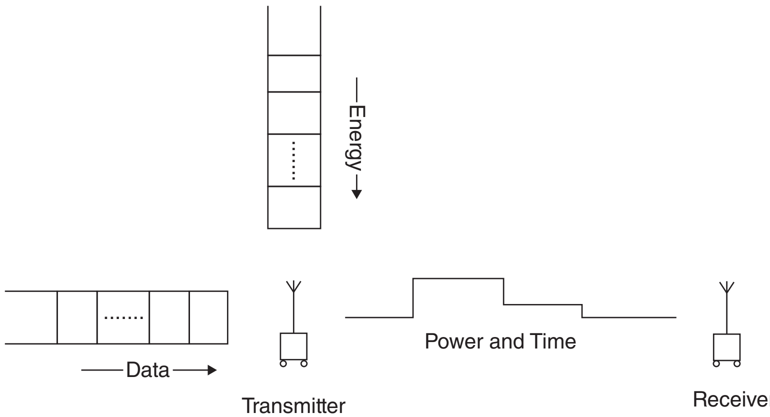

Figure 5.2 The transmission scheduling problem.

an amount of $E_1, E_2, \cdots, E_N$ at time instants of $t_1, t_2, \cdots, t_N$, respectively, where t_N is the last time instant that energy arrives within a total delivery time of T_d. This represents the energy arrival process with known amount and known arrival time, the deterministic model in Section 3.4. Note that this model has simplified the energy harvesting process and the energy storage process by assuming perfect energy conversion and perfect energy storage. In practice, there could be conversion loss or leakage. The data packet arrives with an amount of $B_1, B_2, \cdots, B_M$ at time instants of $T_1, T_2, \cdots, T_M$, respectively, where T_M is the last time instant that data packets arrive within the total delivery time of T_d. Again, the packet arrival time and amount are also known. Finally, within the total delivery time, assume that a sequence of transmission power $P_1, P_2, \cdots, P_K$ with transmission duration of $d_1, d_2, \cdots, d_K$ is adopted before the transmitter finishes the transmission. The transmission power and transmission duration are adapted to the energy arrival and packet arrival to minimize the delivery time. The energy arrival is independent of the packet arrival.

Using these assumptions, two scenarios were considered in Yang and Ulukus (2012). In the first scenario, a simpler case is considered, where the amount of data packet is fixed at B_0 and there is no random packet arrival during T_d. In this case, the total energy consumption and the total amount of bits transmitted before any time instant t are given by

$$E(t) = \sum_{k=1}^{\bar{K}} P_k d_k + P_{\bar{K}+1}\left(t - \sum_{k=1}^{\bar{K}} d_k\right) \tag{5.51}$$

$$B(t) = \sum_{k=1}^{\bar{K}} C(P_k) d_k + C(P_{\bar{K}+1})\left(t - \sum_{k=1}^{\bar{K}} d_k\right) \tag{5.52}$$

where $\bar{K} = \max\{k, \sum_{j=1}^{k} d_j \le t\}$ is the largest time index that the transmission duration is changed before t, and $C(P_k)$ gives the transmission rate as a function of the transmission power. For example, in a static additive white Gaussian noise, $C(P_k) = \ln(1 + P_k)$ follows a logarithmic relationship. Based on the above equation, the transmission scheduling problem becomes

$$\min_{P_1,\cdots,P_K,d_1,\cdots,d_K} \{T_d\} \tag{5.53}$$

$$E(t) \leq \sum_{n:t_n<t} E_n, \quad 0 \leq t \leq T_d \tag{5.54}$$

$$B(T_d) = B_0. \tag{5.55}$$

Here (5.53) shows the minimization of the total delivery time T_d with respect to K transmission power $P_1, \cdots, P_K$ and transmission time $d_1, \cdots, d_K$. Then (5.54) shows a constraint on the energy that the consumed energy must be smaller than the total harvested energy before time instant t, where the right-hand side of the inequality gives the total energy harvested before t. This energy constraint is similar to the one we used before in Section 4.2.5 and will be used in later chapters too. Finally, (5.55) shows a constraint on the data transmission that all the B_0 data bits must be delivered at the end of T_d.

It was reported in Yang and Ulukus (2012) that the optimum solutions to the above problem satisfy the following conditions as

$$P_k = \frac{\sum_{j=n_{k-1}}^{n_k-1} E_j}{t_{n_k} - t_{n_{k-1}}} \tag{5.56}$$

$$d_k = t_{n_k} - t_{n_{k-1}} \tag{5.57}$$

for $k = 1, 2, \cdots, K$, with $n_k = \arg\min_{n:t_n \leq T_d, t_n > t_{n_{k-1}}} \left\{ \frac{\sum_{j=n_{k-1}}^{n-1} E_j}{t_n - t_{n_{k-1}}} \right\}$ and also $\sum_{k=1}^{K} C(P_k)d_k = B_0$ such that n_k is the time index of the energy arrival when the transmission power P_k switches to P_{k+1} or at time $t = t_{n_k}$ the power P_k switches to P_{k+1}. Details of the derivation can be found in Yang and Ulukus (2012).

In the second scenario, the data packet arrives randomly during the delivery time of T_d too but the time and amount of packet arrival are known. In this case, the optimization problem for packet scheduling becomes

$$\min_{P_1,\cdots,P_K,d_1,\cdots,d_K} \{T_d\} \tag{5.58}$$

$$E(t) \leq \sum_{n:t_n<t} E_n, \quad 0 \leq t \leq T_d \tag{5.59}$$

$$B(t) \leq \sum_{m:T_m<t} B_m, \quad 0 \leq t \leq T_d \tag{5.60}$$

$$B(T_d) = \sum_{m=0}^{M} B_m. \tag{5.61}$$

Thus, the two constraints are that the consumed energy must be smaller than the total energy harvested and that the delivered packet must be smaller than the total packet arriving, before any time instant t for $0 \leq t \leq T_d$. A procedure used to calculate the optimum solutions of P_k and d_k was also provided in Yang and Ulukus (2012).

This study provides some useful insights into the transmission scheduling problem with energy harvesting but to make them more practical, more studies need to be performed. For example, the power management needs to be considered. Also, the current study assumes perfect knowledge of the arrival time and amount of energy and packet but in reality such knowledge is not available or not perfectly available.

In Antepli et al. (2011), this problem was extended to a two-user Gaussian broadcast channel. In this work, assuming a fixed amount of data available at the transmitter that needs to be sent to user 1 and user 2 within a certain period of time, the effect of energy harvesting was examined. This corresponds to the first scenario in Yang and Ulukus (2012) but with B_1 and B_2 for user 1 and user 2, respectively, instead of B_0. The optimum solutions were obtained following numerical procedures.

5.4.2 Effective Capacity

Effective capacity is a concept proposed in Wu and Negi (2003) as a link layer channel model. Most existing channel models in the literature are physical layer channel models that determine the performance metrics of transmission and reception in the physical layer, such as symbol error rate, channel capacity, and outage probability. These models cannot be used to examine the link layer performance metrics, such as delay, delay violation probability, and network throughput, etc., directly. However, in systems that require quality of service (QoS) guarantees, such performance metrics are important to evaluate the link layer connection for admission control and resource allocation. To provide this QoS support, one needs an analysis of the queuing behavior of the connection, which cannot be extracted from the physical layer channel models. Thus, Wu and Negi (2003) aimed to obtain a link layer channel model that examines the link layer performance metrics directly without using the physical layer channel models. Figure 5.3 shows the link layer channel model with queuing considered.

Specifically, this model involves two random processes: the traffic process that describes the arrival of the data packet at the transmitter; and the service process that describes the departure of the data packet in the proposed channel. If the arriving data packet cannot be processed by the channel in time, the data packets will start to queue at the transmitter. Thus, the delay in the link layer is determined by the arriving rate in the traffic process and the departure rate in the service process.

Assume that the traffic process has a constant arriving rate of μ and that the queue has an infinite size. The service process is denoted as $S(t)$. The effective capacity that describes the service process is defined as

$$\alpha(u) = \frac{1}{u} \lim_{t \to \infty} \frac{1}{t} \log E\{e^{-uS(t)}\} \tag{5.62}$$

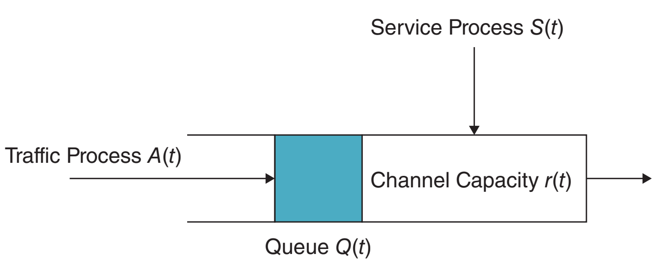

Figure 5.3 The link layer channel model.

where u is the argument of the function, $S(t) = \int_0^t r(\tau)d\tau$ is the service provided by the channel or the partial sum of $r(t)$ up to time instant t, $r(t)$ is the instantaneous rate of the channel, $\log E\{e^{-uS(t)}\}$ is the log-moment-generating-function of $S(t)$, and $u \geq 0$. Using the effective capacity, it can be shown that for a constant traffic arriving rate of μ, one has

$$\sup_t Pr\{D(t) \geq D_{max}\} \approx \gamma(\mu)e^{-\theta(\mu)D_{max}} \tag{5.63}$$

where $D(t)$ is the delay in the channel determined by the difference between the traffic and the service, D_{max} is the maximum delay allowed in the system to guarantee the QoS, $Pr\{D(t) \geq D_{max}\}$ is the probability that the delay is larger than the maximum delay that violates the QoS requirement or the delay-violation probability, $\gamma(\mu) = Pr\{D(t) \geq 0\}$ is the probability that the delay exists, and $\theta(\mu) = \mu\alpha^{-1}(\mu)$ is called the QoS exponent with $\alpha^{-1}(\cdot)$ being the inverse function of the effective capacity $\alpha(\cdot)$.

The values of $\gamma(\mu)$ and $\theta(\mu)$ are the parameters of the proposed link layer channel model that determine the quality of the connection. Thus, for a communications system that allows a maximum delay of D_{max} with delay-violation probability of ϵ, or $Pr\{D > D_{max}\} \leq \epsilon$, the maximum constant traffic rate supported by this channel to fulfill the delay requirement is μ, which can be solved using

$$\epsilon = \gamma(\mu)e^{-\theta(\mu)D_{max}}. \tag{5.64}$$

In this case, the service of the channel satisfies

$$Pr\{S(t) \leq \Phi(t)\} = Pr\{D > D_{max}\} \leq \epsilon \tag{5.65}$$

where $\Phi(t)$ is the upper bound given by $\Phi(t) = \mu[t - D_{max}]^+$ and $[\cdot]^+$ is the non-negative function defined as before. Thus, the link layer channel model using effective capacity links its channel parameters to the link layer performance metrics in a simple and direct relationship.

In Wu and Negi (2003), an example of the Rayleigh fading channel is given. It was shown that for a Rayleigh fading channel, the moment-generating-function of the service process can be approximated as

$$E\{e^{-uS(t)}\} \approx \frac{1}{|u\delta\mathbf{R} + \mathbf{I}|} \tag{5.66}$$

where $\delta = \frac{t}{N}$ is the time separation by dividing t into N slices, $\mathbf{R}$ is the covariance matrix of the Rayleigh channel gains, and $\mathbf{I}$ is the identity matrix.

For the additive white Gaussian noise channel, the instantaneous service rate is given as

$$r(t) = \ln\left[1 + \frac{a^2P(t)}{N_0}\right] \tag{5.67}$$

where a is the static channel gain and $P(t)$ is the transmission power at time t. For fading channels, a becomes a random variable and hence the expectation in the moment-generating function needs to be derived. Thus, the power allocation can be studied to optimize the link layer delay using the effective capacity model. For example, in Yu et al. (2016), the effective capacity was used to optimize the spectral efficiency and the energy efficiency in green communications.

For energy harvesting wireless communications, especially for ambient energy harvesting, the power supply or the transmission power $P(t)$ may be random, as discussed in Chapter 4. Considering this randomness, new effective capacity can be calculated for variable-power transmission, such as in Gong et al. (2014). If fixed-power transmission is used instead, this randomness does not exist any more. Also, if wireless power is used, the harvested power may suffer from fading. Since power transfer and information transmission share the same channel, a and $P(t)$ may be correlated. Thus, the expression of the instantaneous rate will be more complicated due to energy harvesting.

5.5 Summary

In this chapter, the upper layer for energy harvesting wireless communications has been investigated. We have mainly focused on two functions in the upper layer: MAC protocols; and routing protocols. For the MAC protocols, only asynchronous protocols are viable in energy harvesting wireless communications, and for the routing protocols, only opportunistic routing algorithms are possible, due to the dynamic nature of the energy supply. The design objectives of these protocols are highly related to the energy sources used.

If ambient energy harvesting is used, where the sensors harvest energy from the ambient sources, such as the Sun and wind, the uncertainty in the energy supply is the main concern. In this case, the communications process follows a best-effort method, where the sensor transmits as many data as possible, subject to constraints on the energy availability. This energy availability is determined by the energy consumption (the traffic model) and the energy arrival (the harvesting model). In particular, for the MAC protocols, the duty cycle or the transmission power can be adjusted according to the residual energy. For transmission scheduling, both the transmission power and the transmission time are adapted to the traffic model and the harvesting model. For the routing protocols, the routing metric used to select the routing path should include the current energy level, the future energy level and possibly the energy wastage for energy-aware routing.

On the other hand, if wireless power is used, the sensors harvest energy from an intentional power transmitter. In this case, the uncertainty in energy supply has been greatly reduced, as power transfer is controllable. However, in some applications, power transfer and data transmission are performed in the same frequency band so that time multiplexing is necessary, as the sensor node cannot receive power and transmit data at the same time. Thus, coordination between energy harvesting and data transmission is important. For the MAC protocols, the design focuses on the fairness problem to prolong the network lifetime, as different sensors receive different amounts of power due to the different locations. For the routing protocol, the charging time needs to be considered in the routing path selection, as it affects the throughput and the latency. Alternatively, one can consider the effective capacity model that accounts for delay.

Up until now, we have covered the challenges brought by energy harvesting to wireless communications. In the following chapters, we will discuss several state-of-the-art energy harvesting communications systems as applications to cover the opportunities created by energy harvesting. These include wireless powered communications, energy harvesting cognitive radios, and energy harvesting relaying.