Fundamentals of Wireless Sensor Networks: Theory and Practice

Waltenegus Dargie
Technical University of Dresden, GERMANY

Christian Poellabauer
University of Notre Dame, USA

Part One

Introduction

1

Motivation for a Network of Wireless Sensor Nodes

1.1 What is the difference between passive sensors and active sensor and can you name a few examples for each category (e.g. using Table 1.2 in the book)?

An active sensor requires external power, while a passive sensor can generate an output signal without the need for a power source such as a battery. Examples for active sensors include thermistors, strain gauges, infrared sensors, and Hall effect sensors. Examples for passive sensors include thermocouples, photodiodes, and microphones.

1.2 Consider a Wheatstone bridge circuit using a resistive temperature sensor R_x as shown in Figure 1.2 in the book. Further assume that $R_1 = 10\ \Omega$ and $R_3 = 20\ \Omega$. Assume that the current temperature is $80\,^\circ\mathrm{F}$ and $R_x(80) = 10\ \Omega$. You wish to calibrate the sensor such that the output voltage V_{OUT} is zero whenever the temperature is $80\,^\circ\mathrm{F}$.

(a) What is the desired value of R_2?

The formula to compute V_{OUT} is $V_{OUT} = V_{CC} * (\frac{R_x}{R_3+R_x} - \frac{R_2}{R_1+R_2})$. Since V_{OUT} is supposed to be zero, the second term of this equation must be zero too. As a consequence, R_2 can be computed as $5\ \Omega$.

(b) What is the output voltage (as a function of the supply voltage) at temperature $90\,^\circ\mathrm{F}$, when this increase in temperature leads to an increase in resistance of 20% for R_x?

Once the temperature increases to $90\,^\circ\mathrm{F}$, the output voltage (using the same formula) will be $0.041\dot{6} * V_{CC}$.

1.3 As described in this chapter, using multiple communication hops instead of a single hop affects the overall energy consumption. Describe other advantages or disadvantages of multi-hop communications, e.g. in terms of performance (latency, throughput), reliability, and security.

Multi-hop communications have numerous effects on performance and node or network management approaches. For example, with respect to latency, multi-hop communications require that the same message is being transmitted and received several times. Each relay node incurs additional delays, e.g. for queuing the message. On the

Fundamentals of Wireless Sensor Networks: Theory and Practice Waltenegus Dargie and Christian Poellabauer

other hand, communications over short distances may often be more reliable, allowing a node to use a larger transmission rate (e.g. 11 Mbps instead of 1 Mbps for IEEE 802.11), reducing the communication time and overhead. Since the same message must be transmitted several times (and the limited channel capacity must be shared among the relays for both receiving and transmitting), the maximum achievable throughput will also suffer. Further, reliability can be affected in several aspects, e.g. shorter communication ranges may decrease the error probability, but having multiple nodes (relays) involved in the communication increases the risk of disruptions due to link or node failures. Finally, using multiple relay nodes provides an attacker with more opportunities to intercept or modify a transmission, thereby posing an increased security risk.

1.4 The relationship between the transmitted and the received power of an RF signal follows the inverse-square law (Equation (1.5)), i.e. power density and distance have a quadratic relationship. This can be used to justify multi-hop communication (instead of single-hop), i.e. energy can be preserved by transmitting packets over multiple hops at lower transmission power. Assume that a packet p must be send from a sender A to a receiver B. The energy necessary to directly transmit the packet can be expressed as the simplified formula $E_{AB} = d(A,B)^2 + c$, where $d(x,y)$ (or simply d in the remainder of this question) is the distance between two nodes x and y and c is a constant energy cost. Assume that you can turn this single-hop scenario into a multi-hop scenario by placing any number of equidistant relay nodes between A and B.

(a) Derive a formula to compute the required energy as a function of d and n, where d is the distance between nodes A and B and n is the number of relay nodes (i.e. $n = 0$ for the single-hop case).

$n = 0 : E = d^2 + c$,
$n = 1 : E = 2[(\frac{d}{2})^2 + c]$,
$n = 2 : E = 3[(\frac{d}{3})^2 + c]$, etc.
therefore: $E = (n+1)[(\frac{d}{n+1})^2 + c]$

(b) What is the optimal number of relay nodes to send p with the minimum amount of energy required and how much energy is consumed in this optimal case for a distance $d(A,B) = 10$ and (i) $c = 10$ and (ii) $c = 5$?

Table **??** shows the computation of the energy costs for (i) $c = 10$ and (ii) $c = 5$ for $0..7$ relays. In the first case, the minimum energy is obtained with $n = 2$ relays ($E = 63.\dot{3}$). In the second case, the energy costs for both $n = 3$ and $n = 4$ are 45, i.e., both scenarios (3 or 4 relays) would lead to the optimal case.

1.5 Name at least four techniques to reduce power consumption in wireless sensor networks.

Power management techniques can be found at various layers of a sensor node and sensor network design. Many processors can be operated at multiple frequencies and supply voltages using the DVS (dynamic voltage scaling) technique. Duty cycling refers to a device's ability to turn off the radio component when no transmissions are taking place or no incoming messages are expected. Further, the transmission power of wireless radios can be reduced, thereby limiting how far a signal can travel. Reduced transmission ranges can also lead to less interference and fewer collisions, therefore requiring fewer message

Table 1.1 Energy computation (Exercise 1.4)

Number of Relays	**Energy ($c = 10$)**	**Energy ($c = 5$)**
0	110	105
1	70	60
2	$63.\dot{3}$	$48.\dot{3}$
3	65	45
4	70	45
5	$76.\dot{6}$	$46.\dot{6}$
6	84.3	49.3
7	92.5	52.5

re-transmissions. The network layer is responsible for finding energy-efficient routes, e.g. when all nodes require the same amount of energy for receiving and transmitting a message, a route with the smallest number of relays (hops) is often the most energy-efficient route. Finally, in-network aggregation and elimination of redundant sensor data can reduce the amount of communication needed by sensor nodes, thereby also reducing the energy overheads.

2

Application

2.1 Most applications in wireless sensor networks extract time and frequency domain features to detect interesting events. Define the following features:

(a) Autocorrelation function

The autocorrelation of a random process describes the correlation between the values of the process at different points in time. The correlation can be expressed as a function of the two times or of the time difference. Let X be some repeatable process and i be some point in time after the start of that process – i may be an integer for a discrete-time process or a real number for a continuous-time process – and x_i is the value measured at time i. Furthermore, assume that the process has a mean value of μ_i and a variance of σ_i^2 for all times i. Then the autocorrelation between any two times j and k is:

$$R(j,k) = \frac{E[(x_j - \mu_i)(x_k) - \mu_i]}{\sigma_j \sigma_k} \tag{2.1}$$

(b) Correlation coefficients The correlation coefficients of two discrete sequences X and Y is a measure of the existence of a linear dependency between the sequences. It takes the quotients of the covariance and variance of the individual sequences into account:

$$CC(XY) = \frac{COV(X,Y)}{\sqrt{VAR(X)VAR(Y)}} \tag{2.2}$$

where:
$COV(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2\left(y_i - \bar{Y}\right)^2$ and,
$VAR(X) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$.

(c) Cross-correlation function

it is an indication of the existence of a correlation between two time series measurements $s_1(t)$ and $s_2(t)$, where s_1 and s_2 may represent either the same type of movement measured at different locations, or a single movement measured at the same location but at different times. In case $s_2(t)$ represents $s_1(t+\tau)$, where τ is a specified time lag, the two variables are usually not statistically independent, and large cross correlations between s_1 and s_2 can result. Mathematically, the cross correlation, XC, is described as follows:

$$XC(\tau) = \sum_{i=0}^{N-1} s_1(i)s_2(i+\tau) \tag{2.3}$$

(d) Auto regression function

The notation $AR(p)$ refers to the autoregressive model of order p. The $AR(p)$ model is defined as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-1}\epsilon_t \tag{2.4}$$

where $\varphi_1, ..., \varphi_p$ are the parameters of the model, c is a constant and ϵ_t is a white noise.

(e) Coherence

The spectral coherence measures the relation between two signals or data sets in the frequency domain. If the signals are statistically ergodic and the system is a linear system, coherence can be used to estimate the causality between the input and the output. The coherence between two signals $x(t)$ and $y(t)$ is a real-valued function that is defined as:

$$C_{xy} = \frac{|G_{xy}|^2}{G_{xx}G_{yy}} \tag{2.5}$$

where G_{xy} is the cross-spectral density between x and y and G_{xx} and G_{yy} are the autospectral density of x and y, respectively. The magnitude or power of the spectral density is denoted as $|G|$.

The values of a coherence function always satisfy $0 \leq c_{xy} \leq 1$. For a constant parameter linear system with a single input $x(t)$ and single output $y(t)$, the coherence will be equal to one. If $x(t)$ and $y(t)$ are completely unrelated, the coherence will be zero. If C_{xy} is less than one but greater than zero it is an indication that either the measurements are corrupted by noise, the system relating $x(t)$ and $y(t)$ is not linear or that $y(t)$ is producing output due to input $x(t)$ as well as other inputs.

The coherence of a linear system can be viewed as the fractional part of the output signal power that is produced by the input at a particular frequency. Alternatively, the quantity $1 - C_{xy}$ can be viewed as an estimate of the fractional power of the output that is not contributed by the input at a particular frequency. This leads to the definition of the coherent output spectrum: $G_{vv} = C_{xy}G_{yy}$. G_{vv} provides a spectral quantification of the output power that is uncorrelated with noise or other inputs.

2.2 Explain the difference between time domain and frequency domain features.

The time-domain representation of a signal described the magnitude (amplitudes) of the signal as a function of time. The time instance can be either discrete or continuous. However, in many cases one needs to know the frequency content of the signal rather than the amplitudes of the individual samples. According to the Fourier Theorem, any waveform in the time domain can be represented by the weighted sum of orthogonal sinusoidal waveforms. The same waveform then can be represented in the frequency domain as a pair of amplitude and phase values at each component frequency.

2.3 A 2D accelerometer sensor measures the movement of a structure to an ambient excitation. The normalized raw data that are collected for 1 second from the x- and y-axis are given below. In each case, the measurement is one-dimensional and should be read from left to right and top to bottom.

$$x = \begin{bmatrix} 0.13 & 0.13 & 0.13 & 0.11 & 0.09 & 0.08 & 0.06 & 0.05 & 0.04 & 0.02 \\ -0.01 & -0.02 & -0.01 & -0.02 & -0.04 & -0.06 & -0.11 & -0.12 & -0.13 & -0,10 \\ 0.12 & 0.00 & -0.06 & -0.03 & 0.00 & 0.02 & 0.02 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.02 & 0.03 & 0.03 & 0.02 & 0.03 & 0.02 & 0.02 \\ 0.03 & 0.02 & 0.02 & 0.03 & 0.03 & 0.02 & 0.01 & 0.05 & 0.05 & 0.03 \\ 0.08 & -0.04 & 0.02 & -0.03 & -0.07 & 0.06 & 0.18 & 0.14 & 0.08 & 0.04 \\ 0.03 & 0.03 & 0.02 & 0.00 & -0.03 & -0.07 & -0.13 & -0.21 & -0.31 & -0.31 \\ -0.42 & -0.37 & -0.28 & 0.31 & -0.01 & -0.28 & 0.12 & -0.12 & 0.04 & -0.01 \\ 0.03 & 0.03 & 0.02 & 0.03 & 0.03 & 0.03 & 0.03 & 0.02 & 0.02 & 0.02 \\ 0.03 & 0.02 & 0.03 & 0.03 & 0.03 & 0.03 & 0.02 & 0.02 & 0.03 & 0.12 \end{bmatrix}$$

$$y = \begin{bmatrix} -0.01 & -0.02 & -0.02 & -0.02 & -0.04 & -0.04 & -0.03 & -0.02 & -0.02 & -0.02 \\ -0.03 & -0.03 & 0.01 & 0.02 & 0.02 & 0.03 & 0.02 & 0.03 & 0.05 & 0.13 \\ -0.01 & 0.04 & -0.02 & -0.06 & 0.02 & -0.01 & 0.01 & 0.00 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.02 & 0.02 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & -0.02 & -0.07 \\ 0.03 & -0.09 & -0.05 & -0.06 & -0.14 & -0.18 & -0.03 & 0.05 & 0.01 & -0.05 \\ -0.04 & -0.02 & -0.02 & -0.03 & -0.04 & -0.05 & -0.07 & -0.04 & 0.00 & 0.01 \\ 0.02 & 0.11 & 0.00 & -0.07 & 0.40 & -0.06 & -0.09 & 0.17 & -0.03 & 0.04 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.00 & 0.01 & 0.02 & 0.01 & 0.01 \\ 0.01 & 0.02 & 0.02 & 0.02 & 0.01 & 0.01 & 0.01 & 0.02 & 0.00 & -0.02 \end{bmatrix}$$

(a) Calculate the autocorrelation for both sequences.

The autocorrelation of the sequence from the x-axis accelerometer sensor is:

$$\begin{bmatrix} 0.0156 & 0.0097 & 0.0074 & 0.0056 & 0.0046 & 0.0041 & 0.0036 & 0.0033 & 0.0028 & 0.0025 \\ 0.0019 & 0.0015 & 0.0014 & 0.0012 & 0.0010 & 0.0008 & 0.0003 & 0.0000 & -0.0002 & -0.0001 \\ 0.0008 & 0.0004 & -0.0009 & -0.0003 & -0.0016 & -0.0015 & -0.0000 & -0.0012 & -0.0026 & -0.0045 \\ -0.0055 & -0.0062 & -0.0063 & -0.0058 & -0.0050 & -0.0041 & -0.0035 & -0.0024 & -0.0008 & 0.0000 \\ 0.0008 & 0.0017 & 0.0018 & 0.0037 & 0.0031 & 0.0043 & 0.0063 & 0.0043 & 0.0027 & 0.0022 \\ 0.0023 & 0.0015 & 0.0011 & 0.0006 & 0.0000 & -0.0000 & -0.0001 & -0.0004 & -0.0008 & -0.0011 \\ -0.0010 & -0.0007 & -0.0003 & 0.0001 & -0.0003 & -0.0008 & -0.0008 & -0.0008 & -0.0011 & -0.0009 \\ -0.0012 & -0.0013 & -0.0016 & -0.0002 & -0.0011 & -0.0012 & -0.0009 & -0.0020 & -0.0012 & -0.0014 \\ -0.0018 & -0.0011 & -0.0005 & -0.0013 & -0.0023 & -0.0027 & -0.0027 & -0.0024 & -0.0021 & -0.0016 \\ -0.0008 & -0.0000 & 0.0010 & 0.0012 & 0.0024 & 0.0028 & 0.0045 & 0.0047 & 0.0066 & 0.0115 \\ 0.0066 & 0.0047 & 0.0045 & 0.0028 & 0.0024 & 0.0012 & 0.0010 & -0.0000 & -0.0008 & -0.0016 \\ -0.0021 & -0.0024 & -0.0027 & -0.0027 & -0.0023 & -0.0013 & -0.0005 & -0.0011 & -0.0018 & -0.0014 \\ -0.0012 & -0.0020 & -0.0009 & -0.0012 & -0.0011 & -0.0002 & -0.0016 & -0.0013 & -0.0012 & -0.0009 \\ -0.0011 & -0.0008 & -0.0008 & -0.0008 & -0.0003 & 0.0001 & -0.0003 & -0.0007 & -0.0010 & -0.0011 \\ -0.0008 & -0.0004 & -0.0001 & -0.0000 & 0.0000 & 0.0006 & 0.0011 & 0.0015 & 0.0023 & 0.0022 \\ 0.0027 & 0.0043 & 0.0063 & 0.0043 & 0.0031 & 0.0037 & 0.0018 & 0.0017 & 0.0008 & 0.0000 \\ -0.0008 & -0.0024 & -0.0035 & -0.0041 & -0.0050 & -0.0058 & -0.0063 & -0.0062 & -0.0055 & -0.0045 \\ -0.0026 & -0.0012 & -0.0000 & -0.0015 & -0.0016 & -0.0003 & -0.0009 & 0.0004 & 0.0008 & -0.0001 \\ -0.0002 & 0.0000 & 0.0003 & 0.0008 & 0.0010 & 0.0012 & 0.0014 & 0.0015 & 0.0019 & 0.0025 \\ 0.0028 & 0.0033 & 0.0036 & 0.0041 & 0.0046 & 0.0056 & 0.0074 & 0.0097 & 0.0156 & 0.0000 \end{bmatrix}$$

The autocorrelation of the sequence from the y-axis accelerometer sensor is:

$$\begin{bmatrix}
0.0002 & 0.0002 & 0.0001 & -0.0000 & 0.0000 & -0.0000 & -0.0001 & -0.0002 & -0.0002 & -0.0002 \\
-0.0002 & -0.0003 & -0.0003 & -0.0003 & -0.0003 & -0.0002 & -0.0002 & -0.0002 & -0.0002 & -0.0002 \\
-0.0000 & 0.0000 & -0.0000 & 0.0000 & 0.0000 & -0.0001 & -0.0003 & -0.0002 & -0.0002 & -0.0005 \\
-0.0005 & -0.0003 & -0.0003 & -0.0003 & 0.0000 & -0.0001 & -0.0001 & 0.0004 & 0.0004 & 0.0004 \\
0.0005 & 0.0008 & 0.0001 & 0.0007 & 0.0013 & -0.0002 & 0.0009 & 0.0003 & -0.0001 & 0.0007 \\
0.0001 & 0.0001 & 0.0000 & 0.0001 & 0.0002 & 0.0003 & 0.0001 & -0.0001 & -0.0002 & -0.0002 \\
-0.0002 & -0.0002 & -0.0004 & -0.0006 & -0.0005 & -0.0005 & -0.0002 & -0.0001 & -0.0001 & -0.0003 \\
-0.0001 & -0.0003 & 0.0000 & -0.0004 & -0.0007 & 0.0001 & -0.0007 & -0.0005 & -0.0001 & -0.0006 \\
-0.0009 & -0.0002 & 0.0000 & -0.0002 & -0.0003 & -0.0001 & 0.0001 & 0.0001 & -0.0000 & 0.0000 \\
-0.0001 & -0.0002 & 0.0001 & 0.0004 & 0.0004 & 0.0000 & 0.0016 & 0.0001 & 0.0001 & 0.0036 \\
0.0001 & 0.0001 & 0.0016 & 0.0000 & 0.0004 & 0.0004 & 0.0001 & -0.0002 & -0.0001 & 0.0000 \\
0.0000 & 0.0001 & 0.0001 & -0.0001 & -0.0003 & -0.0002 & 0.0000 & -0.0002 & -0.0009 & -0.0006 \\
-0.0001 & -0.0005 & -0.0007 & 0.0001 & -0.0007 & -0.0004 & 0.0000 & -0.0003 & -0.0001 & -0.0003 \\
-0.0001 & -0.0001 & -0.0002 & -0.0005 & -0.0005 & -0.0006 & -0.0004 & -0.0002 & -0.0002 & -0.0002 \\
-0.0002 & -0.0001 & 0.0001 & 0.0003 & 0.0002 & 0.0001 & 0.0000 & 0.0001 & 0.0001 & 0.0007 \\
-0.0001 & 0.0003 & 0.0009 & -0.0002 & 0.0013 & 0.0007 & 0.0001 & 0.0008 & 0.0005 & 0.0004 \\
0.0004 & 0.0004 & -0.0001 & -0.0001 & 0.0000 & -0.0003 & -0.0003 & -0.0003 & -0.0005 & -0.0005 \\
-0.0002 & -0.0002 & -0.0003 & -0.0001 & 0.0000 & 0.0000 & -0.0000 & 0.0000 & -0.0000 & -0.0002 \\
-0.0002 & -0.0002 & -0.0002 & -0.0002 & -0.0003 & -0.0003 & -0.0003 & -0.0003 & -0.0002 & -0.0002 \\
-0.0002 & -0.0002 & -0.0001 & -0.0000 & 0.0000 & -0.0000 & 0.0001 & 0.0002 & 0.0002 & 0.0000
\end{bmatrix}$$

(b) Calculate the correlation coefficients of the sequences.

The correlation matrix of the sequences X and Y is given as:

$$\begin{bmatrix}
1.0000 & -0.1675 \\
-0.1675 & 1.0000
\end{bmatrix}$$

Therefore the correlation coefficient, r_{xy}, of the sequences is -0.1675.

(c) Calculate the FFT of both sequences.

The FFT of the sequence from the x-axis accelerometer is given as:

$$\begin{bmatrix}
0.0100 & 0.8025-1.4785i & 3.0096+1.8724i & -0.4807+1.0864i & 0.4183-2.9520i \\
1.9381-0.2695i & -1.1829+0.8001i & -0.1823-1.2541i & 1.0273+0.2045i & 0.1763+0.3967i \\
-0.4573-0.4685i & 0.6988-1.3108i & 0.3343+1.0120i & -1.0753-0.0329i & 0.7558-0.2726i \\
0.7838-0.0541i & 0.2727+0.2382i & -1.0284-0.4318i & 0.3909-0.5318i & 0.5362+0.9460i \\
-0.2044+0.1761i & -0.0719-0.5883i & 0.2495-0.3662i & 0.2535+0.9499i & -0.7291+0.3037i \\
0.5000-0.7900i & 0.6070-0.0010i & -0.3308+0.5336i & -0.8158+0.0329i & 0.4814-0.5377i \\
0.8173+0.5392i & -0.6272+0.2862i & -0.4873-0.3354i & 0.2527-0.3275i & 0.6537+0.8278i \\
-0.5433+0.2637i & -0.1924-0.6864i & 0.2368-0.3595i & 0.3944+0.7553i & -0.7052+0.5154i \\
-0.0256-0.5615i & 0.5330-0.4780i & 0.2767+0.6045i & -0.8132+0.3938i & -0.1372-0.3806i \\
0.4714-0.2308i & 0.2453+0.3640i & -0.5775+0.0268i & -0.0907-0.1701i & 0.0211+0.0109i \\
0.2300 & 0.0211-0.0109i & -0.0907+0.1701i & -0.5775-0.0268i & 0.2453-0.3640i \\
0.4714+0.2308i & -0.1372+0.3806i & -0.8132-0.3938i & 0.2767-0.6045i & 0.5330+0.4780i \\
-0.0256+0.5615i & -0.7052-0.5154i & 0.3944-0.7553i & 0.2368+0.3595i & -0.1924+0.6864i \\
-0.5433-0.2637i & 0.6537-0.8278i & 0.2527+0.3275i & -0.4873+0.3354i & -0.6272-0.2862i \\
0.8173-0.5392i & 0.4814+0.5377i & -0.8158-0.0329i & -0.3308-0.5336i & 0.6070+0.0010i \\
0.5000+0.7900i & -0.7291-0.3037i & 0.2535-0.9499i & 0.2495+0.3662i & -0.0719+0.5883i \\
-0.2044-0.1761i & 0.5362-0.9460i & 0.3909+0.5318i & -1.0284+0.4318i & 0.2727-0.2382i \\
0.7838+0.0541i & 0.7558+0.2726i & -1.0753+0.0329i & 0.3343-1.0120i & 0.6988+1.3108i \\
-0.4573+0.4685i & 0.1763-0.3967i & 1.0273-0.2045i & -0.1823+1.2541i & -1.1829-0.8001i \\
1.9381+0.2695i & 0.4183+2.9520i & -0.4807-1.0864i & 3.0096-1.8724i & 0.8025+1.4785i
\end{bmatrix}$$

The FFT of the sequence from the y-axis accelerometer is given as:

$$\begin{bmatrix}
-0.0100 & 0.6053-0.0123i & -1.2426+0.7345i & -0.0903-0.6244i & -0.0039+0.8793i \\
0.1917+0.5692i & -0.2959+0.1878i & 0.4854-0.7606i & -0.1048+0.4401i & -0.4671-0.1115i \\
0.2410+0.2155i & 0.0431-0.0941i & 0.2038-0.0333i & -0.4258-0.0002i & 0.1479+0.0042i \\
0.0113+0.2104i & 0.3898-0.2785i & -0.2337+0.2153i & -0.4067-0.6490i & 0.2652+0.2948i \\
-0.2636+0.1783i & -0.0368+0.0343i & -0.2171-0.3750i & 0.4998-0.1641i & -0.0630+0.6133i \\
-0.4700-0.2600i & 0.1786-0.2433i & 0.3631-0.0934i & 0.1616+0.6267i & -0.8868-0.2938i \\
0.3640-0.6721i & 0.5043+0.4546i & 0.0248+0.7559i & -0.7083-0.4020i & 0.5606-0.9172i \\
0.3703+0.4462i & -0.3310+0.7305i & -0.5515-0.3396i & 0.6866-0.7003i & 0.3494+0.2780i \\
-0.4314+0.5834i & -0.5176-0.1664i & 0.6376-0.4034i & 0.3602+0.0731i & -0.2618+0.2294i \\
-0.5033-0.1525i & 0.3760-0.0214i & 0.1130+0.1323i & 0.0595+0.0464i & -0.2708-0.1779i \\
0.1900 & -0.2708+0.1779i & 0.0595-0.0464i & 0.1130-0.1323i & 0.3760+0.0214i \\
-0.5033+0.1525i & -0.2618-0.2294i & 0.3602-0.0731i & 0.6376+0.4034i & -0.5176+0.1664i \\
-0.4314-0.5834i & 0.3494-0.2780i & 0.6866+0.7003i & -0.5515+0.3396i & -0.3310-0.7305i \\
0.3703-0.4462i & 0.5606+0.9172i & -0.7083+0.4020i & 0.0248-0.7559i & 0.5043-0.4546i \\
0.3640+0.6721i & -0.8868+0.2938i & 0.1616-0.6267i & 0.3631+0.0934i & 0.1786+0.2433i \\
-0.4700+0.2600i & -0.0630-0.6133i & 0.4998+0.1641i & -0.2171+0.3750i & -0.0368-0.0343i \\
-0.2636-0.1783i & 0.2652-0.2948i & -0.4067+0.6490i & -0.2337-0.2153i & 0.3898+0.2785i \\
0.0113-0.2104i & 0.1479-0.0042i & -0.4258+0.0002i & 0.2038+0.0333i & 0.0431+0.0941i \\
0.2410-0.2155i & -0.4671+0.1115i & -0.1048-0.4401i & 0.4854+0.7606i & -0.2959-0.1878i \\
0.1917-0.5692i & -0.0039-0.8793i & -0.0903+0.6244i & -1.2426-0.7345i & 0.6053+0.0123i
\end{bmatrix}$$

2.4 To improve the expressiveness of frequency domain features, it is preferred to compute the Short Time Fourier Transform (STFT) of a time series sequence instead of the FFT of the entire frame.

(a) Divide the 1-second frame into 10 subframes such that there is an overlap of 50% between each subframe except the first and the last ones.

(b) Calculate the STFT for each window.

(c) Now reduce the overlap to 25% and compute the STFT and compare the results with the results obtained from the 50% overlap subframes.

2.5 How can oversampling of sensor data overcome the effect of noise?

The basic idea in oversampling is to increase the sampling rate of the signal to the point where a low-resolution quantizer suffices. By oversampling, it is possible to reduce the dynamic range of the signal values between successive samples and thus reduce the resolution requirements on the quantizer.

2.6 One of the time domain features used to recognize interesting events is the zero-crossing rate, which can be expressed as:

$$ZCR(s) = \frac{1}{T}\sum_{i=1}^{T} F(s(i) \cdot s(i-1) < 0)$$

where s is a discrete, time-series sequence; $s(i)$ and $s(i-1)$ are two consecutive samples. $F = 1$ if the evaluation is true, $F = 0$ otherwise.

(a) Compute the zero crossing rates for the two time series measurements given above.

ZCR(X-axix) =

$$\begin{aligned}
&F(0.02 \times (-0.01) < 0) + F((-0.10) \times 0.12 < 0) + F(< 0) + F(0.08 \times (-0.04) < 0) + \\
&F((-0.04) \times 0.02 < 0) + F(0.02 \times (-0.03) < 0) + F((-0.07) \times 0.06 < 0) + \\
&F((-0.28) \times 0.31 < 0) + F(0.31 \times (-0.01) < 0) + F((-0.28) \times (-0.12) < 0) + \\
&F(0.12 \times (-0.12) < 0) + F((-0.12) \times 0.04 < 0) + F(0.04 \times (-0.01) < 0) + \\
&F((-0.01) \times 0.03 < 0)
\end{aligned}$$

= 13

ZCR(Y-axis) =

$$\begin{aligned}
&F((-0.03).01 < 0) + F(0.13 \times (-0.01) < 0) + F((-0.01) \times (0.04) < 0) + \\
&F(0.04 \times (-0.02) < 0) + F((-0.06) \times 0.02 < 0) + F(0.02 \times (-0.01) < 0) + \\
&F((-0.01) \times 0.01 < 0) + F(0.01 \times (-0.02) < 0) + F((-0.07) \times 0.03 < 0) + \\
&F(0.03 \times (-0.09) < 0) + F((-0.03) \times 0.05 < 0) + F(0.01 \times (-0.05) < 0) + \\
&F((-0.07) \times 0.40 < 0) + F(0.40 \times (-0.06) < 0) + F((-0.09) \times 0.17 < 0) + \\
&F(0.17 \times (-0.03) < 0) + F((-0.03) \times 0.04 < 0)
\end{aligned}$$

= 17

(b) What conclusion can be drawn from the zero-crossing rate?

The zero-crossing rate reveals how often a signal (measurement) crosses the zero-reference line. It is a direct indication of the fundamental frequency of the signal. For an accelerometer sensor, if the calibration position is known, the zero-crossing rate can be used to estimate the orientation of the sensor. For example, if the sensor is calibrated by standing it up (say, along the $y - axis$), then it will produce an acceleration of $1g$ if it is laid flat with a displacement of 90^0 either in the $z - axis$ or in the $x - axis$.

2.7 Another interesting feature is the spectral centroid, a frequency domain feature which represents the balancing point of the spectral power distribution:

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] \cdot n}{\sum_{n=1}^{N} M_t[n]}$$

where $M_t[n]$ is the magnitude value of the spectrum at position (frequency) n.

Calculate the spectral centroid for the two time series sequences given above

$$C_t(x) = \frac{[1 : 100] \times x}{ones(1, 100) \times x} = -1259.33$$

where X is a column vector

$$C_t(y) = \frac{[1 : 100] \times y}{ones(1, 100) \times y} = -1405$$

where y is a column vector

2.8 In structural health monitoring, inspection techniques are classified into global and local inspection techniques. Explain the difference between these techniques. For which of these techniques are wireless sensor networks suitable?

Local monitoring:

- Detect tiny incipient cracks or defects in structures
- Sophisticated ultrasound, thermal, X-ray, magnetic, or optical imaging techniques.
- Such imaging equipments are expensive, power consuming, and bulky

Global monitoring

- Discover damage large enough to influence the properties of the entire structure or large sections of it.
- Significant damage to an entire cable on a bridge or an entire column of a building.
- Infer damage from changes in the modes of structural response due to external excitations, either ambient or forced.

Global monitoring is more suitable for wireless sensor networks.

2.9 Explain how the property of a pipeline changes at a location where gas and oil leakages occur.

Fluid pipelines generate a hot-spot at the location of the leak, whereas gas pipelines generate a cold-spot due to the gas pressure relaxation.

2.10 Explain how an acoustic sensor can be used to monitor the content of a pipeline.

If the speed of a sound through a fluid medium is known, a time-of-flight approach can be used to determine the type of material that passes through a pipeline.

Alternatively, the Swept Frequency Acoustic Interferometry (SFAI) can be used to characterize fluids (gases, mixtures, suspensions, emulsions, liquids, etc.). The technique is based on setting up standing waves in fluid-filled cavity and determining very accurately the sound speed and sound absorption as a function of frequency.

2.11 Explain the principle of piezoelectric sensor to measure movement?

When a mechanical stress is applied to a piezoelectric material, it generates an electric charge which is proportional to the applied stress. This property can be used to measure movement.

2.12 How can a magnetic sensor be employed to measure the movement of vehicles?

Because of their metallic parts, vehicles disturb the magnetic field distribution of the earth as they drive. The magnitude of disturbance is proportional to the size of the metallic objects in the vehicles and the speed of drive. This can be measured by a magnetic sensor in which a voltage is induced as the magnetic field of the earth is disturbed. By considering the induced voltage and its polarization, one is able to determine the type of car and the speed of drive as well as the direction of the drive.

2.13 What is an electromyography and for what application can it be used?

Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by skeletal muscles. EMG is performed using an instrument called an electromyography, to produce a record called an electromyogram. An electromyography detects the electrical potential generated by muscle cells when these cells are electrically or neurologically activated. The signals can be analyzed to detect medical abnormalities, activation level, and recruitment order or to analyze the biomechanics of human or animal movement

2.14 Describe the three phases of a Parkinson's disease.

Persons who are being given an external stimulation to remnant cells in the substantia nigra to produce more dopamine can be found in one of the three phases:

(a) The exhibition of typical symptoms in the form of tremor and slow movement when the motivation has worn off. This is known as the *off state*.

(b) Normal movements free of tremor when the medication is balanced. This is known as the *on state*; and

(c) An exaggerated involuntary movements when the medication is at a highest concentration. This is known as *dyskinesia*.

2.15 What is a heat unit?

Often a heat unit is defined as the amount of heat required to raise the temperature of one gram of water 1^oC. The heat unit is also alternatively referred to as the British Thermal Unit (BTU), which is the amount of heat required to raise the temperature of one pound of water through 1^oF ($58.5^oF - 59.5^oF$) at sea level (30 inches of mercury).

3

Node Architecture

3.1 A vibration sensor outputs an analog signal with a peak-to-pick voltage of 5V at a frequency of 100Hz.

(a) What should be the minimum sampling frequency, so that no information should be lost during the digitization process?

The minimum sampling rate should satisfy the Nyqusit rate, which is twice greater than the frequency of the analog signal, i.e. it has to be sampled at 200Hz.

(b) Suppose a resolution of 0.025V is required to detect an interesting event. What should be the resolution of the ADC in terms of bits to convert the analog signal into a digital signal?

According to equation (3.1):

$$0.025 = \frac{10V}{2^M} - - > 2^M = 400 - - > m = lb(400) = 8.6$$

Therefore an ADC of 9 bits resolution is required.

3.2 What is the drawback of using a multi-channel ADC?

At high sampling frequency, the crosstalk and uncorrelated noise in multi-channel ADCs is high and undermine the signal-to-noise ratio on the individual channels. Moreover, signal coupling creates spurs that can reduce the spurious free dynamic range (SFDR) and total harmonic distortion (THD).

3.3 What is aliasing?

Aliasing occurs when an analog signal is sampled at a rate below the Nyquisit rate.

Suppose the analog signal $s(t)$ is a band limited signal with bandwidth, $B = 2f_b$. The sampling process is expressed as:

$s(t) \times \sum_{-\infty}^{\infty} \delta(t - nT)$.

Recall that multiplication in time domain is equivalent to convolution in Frequency domain. Hence, the sampling process is expressed as:

$s(f) * \sum_{-\infty}^{\infty} \delta(f - nf_s) = \sum_{-\infty}^{\infty} s(f - nf_s$. Where f_s is the sampling frequency. This shows that the spectrum of the baseband signal repeats itself at a period f_s. If $f_s > 2f_b$,

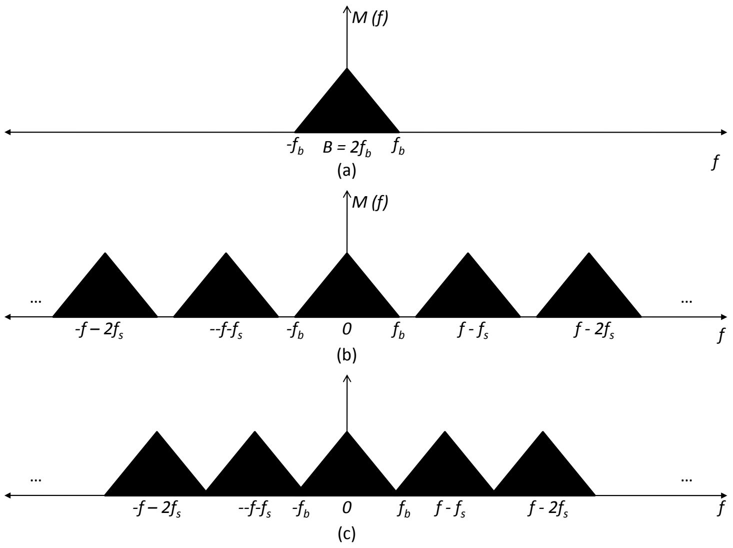


Figure 3.1 Aliasing

there will be no overlap between the adjacent replications of the baseband spectrum (see Figure **??** (b)). If, however, $f_s < 2f_b$, there will be an overlap between the adjacent spectrums, and this is called aliasing (see Figure **??** (c)). An aliasing effect cannot be removed by a lowpass filter.

3.4 Define each of the following terms as applied to discrete time signal processing systems:

(a) Linearity

In a linear system, there is a linear relationship between the input and the output of the system. More generally, a linear system is characterized by its impulse response:

$$\begin{aligned} y[n] &= T\left\{x[n]\right\} = T\left\{\sum_{\infty}^{k=\infty} x[k]\delta[n-k]\right\} \\ &= \sum_{\infty}^{k=\infty} x[k]T\left\{\delta[n-k]\right\} = \sum_{\infty}^{k=\infty} x[k]h_k[n] \end{aligned}$$

Hence, if we have to cascaded linear system with impulse responses $h_1[n]$ and $h_2[n]$ respectively:

$$x[n] * (h_1[n] + h_2[n]) = x[n] * h_1[n] + x[n] * h_2[n]$$

(b) Time-invariance

A system is time-invariant if the impulse response of the system satisfies: $h[k] = h[n-k]$.

(c) Causality

The impulse response of a causal system satisfies: h[n] = o, n < 0

(d) Stability A system is stable if:

$$S = \sum_{-\infty}^{\infty} |h[k]| < \infty$$

3.5 While they are not most energy-efficient, microcontrollers are the predominant processors in wireless sensor networks. Explain some of the reasons?

Microcontrollers offer greater programming flexibility compared with the other types of small-scale processors; hence, they are useful for many applications.

3.6 Explain the reason why using the von Neumann architecture is not efficient for a wireless sensor node.

In the Von Neumann architecture, there is a single memory space for data as well as program code and a single bus interfaces the memory unit with the processor. This means each data and instruction transfer requires a separate cycle, as a result of which computation is slow in low-scale processing subsystems.

3.7 Why are parallel busses not desirable in a wireless sensor node?

Parallel busses require large space. This is difficult to accommodate in small sensor nodes.

3.8 What is the side-effect of using a serial bus that supports full-duplex communication?

A full-duplex serial bus requires four data lines instead of two data lines that are required for a half-duplex bus. Besides the space required for housing the extra data lines, it also mean more pines on the IC packages.

3.9 Explain the following terms in the context of the serial bus, SPI

(a) Serial Data Out

SDO pin carried data out of the device

(b) Serial Data In

SDI pin carries data into the device

(c) Serial Clock

SD is responsible for controlling the sending and receiving of data.

3.10 How can a Master component communicate with multiple slaves in:

(a) I^2C

I^2C uses an addressing mechanism which enables a master device to communicate with more than one slave device. Each device is connected with the SDA and SCL wires as shown in Figure **??** (a). Figure **??** (b) shows an example configuration. The advantage of this configuration is that a new device can be added or an existing one can be removed without affecting the configuration.

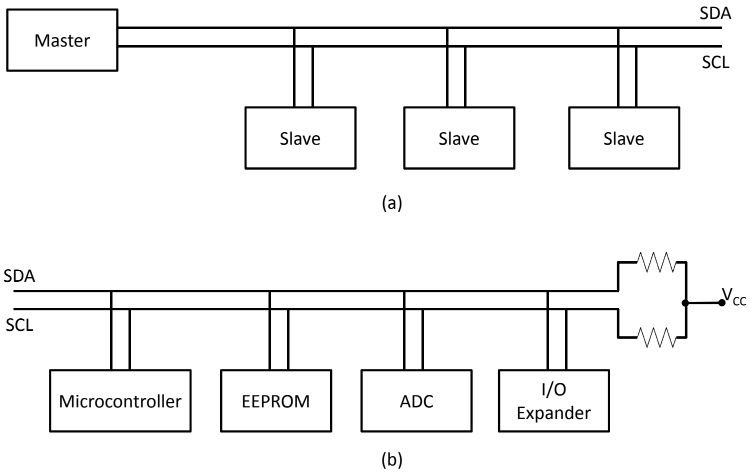


Figure 3.2 Connecting devices with the I^2C serial bus. (a) A schematic diagram. (b) An example configuration.

(b) SPI

SPI enables a single master to connect with multiple slaves through the Slave Select, Chip Select ($\overline{SS}(\overline{CS})$) signal. When SS is low, a slave is enabled. This is displayed in Figure **??**. In (a) a single master is connected with a single device while in (b) a single mater is connected with multiple devices. The problem with this configuration is that an addition or a removal of a device cannot take place without affecting the configuration.

3.11 Explain with the help of diagrams how the data transfer protocol of the I^2C bus functions.

The I^2C protocol is shown below:

In I^2C, the master initiates communication and controls the clock signal. The data transfer procedure is summarized as follows:

(a) Master sends start condition (S) and controls the clock signal

(b) Master sends a unique 7-bit slave device address

(c) Master sends read/write bit (R/W) - 0 - slave receives, 1 - slave transmits

(d) Receiver sends acknowledge bit (ACK)

(e) Transmitter (slave or master) transmits 1 byte of data

(f) Receiver sends an ACK bit for the byte received

(g) If there are more bytes to transmit, step 5 and 6 are repeated.

(h) For a write transaction (master transmitting), master issues stop condition (P) after the last byte of data. For a read transaction (master receiving), master does not acknowledge final byte, just issues stop condition (P) to inform the slave the transmission is completed.

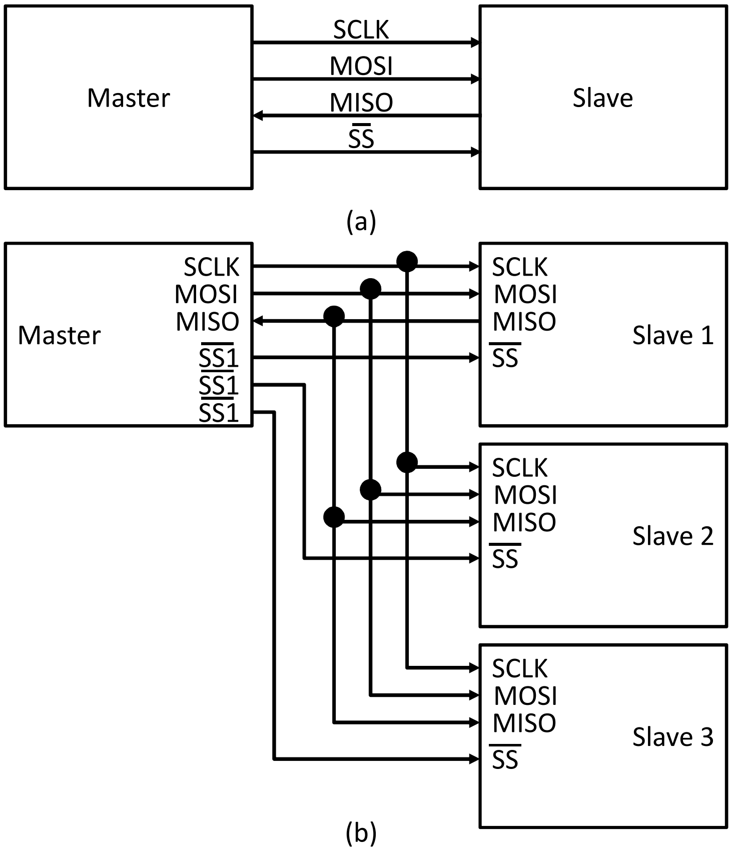


Figure 3.3 Connecting devices with the SPI serial bus. (a) A single master connecting with a single slave. (b) A single master connecting wtih multiple slaves.

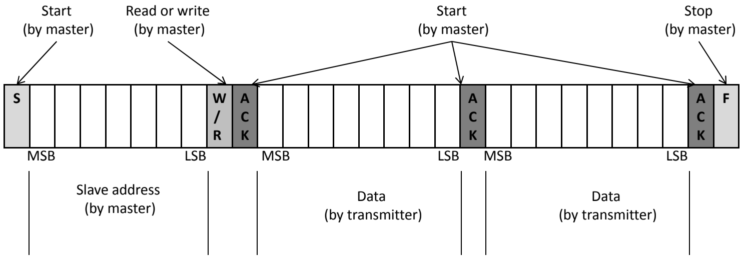


Figure 3.4 The I^2C protocol

Communication in I^2C takes place through the serial data line (SDA) and the coordination of the serial clock line (SCL). Start condition (S) is flagged only when there is SDA transition from 1 two 0. Likewise, stop condition (P) is flagged only when there is SDA transition from 0 to 1. During a repeated start (Sr), a start signal is flagged instead of a stop. One clock pulse is required for each data bit transfer. Data exchange takes place during low clocks. The figure below summarizes data, clock and control signals.

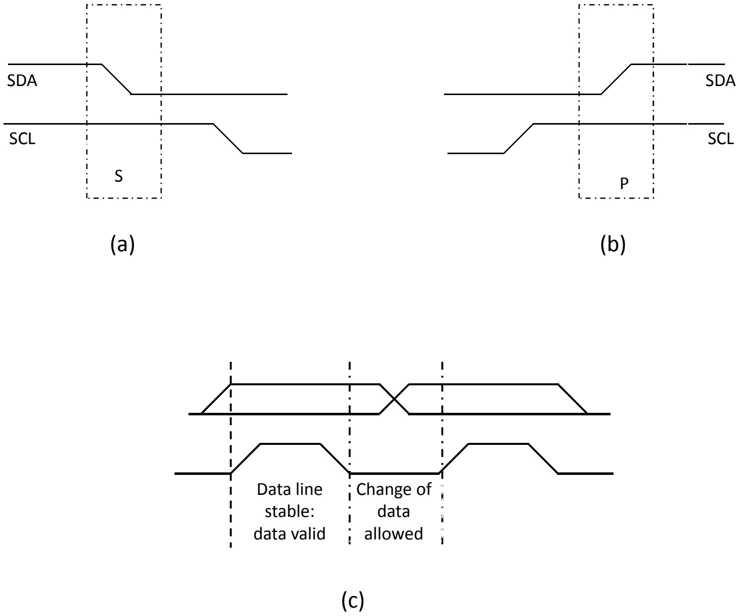


Figure 3.5 Signaling in I^2C protocol

3.12 Explain the basic similarities and differences between a FPGA and an ASIC.

Table 3.1 Comparison between FPGA and ASIC

FPGA	ASIC
Faster time-to-market	Slow time-to-market
No NRE (Non Recurring Expenses)	High NRE
Simple design cycle	Complex design cycle
Remote reprogramming possible	Remote reprogramming not possible
Good for low speed and low volume designs	Good for high speed, big volume designs
High power consumption	Low power consumption
Cheap design tools	Expensive design tools
Mixed signal design is not possible	Mixed signal design possible

Difference between ASICs and FPGAs mainly depends on costs, tool availability, performance and design flexibility. Table **??** summarized some of their differences.

3.13 Explain some of the distinct features of the Super-Harvard architecture.

(a) It provides an internal instruction cache to store frequently needed instructions.

(b) Program memory can be used to store data.

(c) Direct data streaming from an external hardware into the data memory through an I/O controller is possible.

3.14 A large number of commercially available wireless sensor nodes integrate three types of memory architectures: EEPROM (flash memory), RAM and ROM. Explain the purpose of each of them.

The RAM memory is used to temporarily store data and program instructions that are ready to execute. ROM is used to store program instructions and EEPROM is used for data logging.

3.15 The communication subsystem of a wireless sensor node is usually interfaced with the processor subsystem through a SPI bus instead of I^2C bus. Why is it?

The SPI bus provides a high speed data transfer.

3.16 While dynamic memory management is very useful, it cannot be supported in wireless sensor networks, why?

A dynamic memory management strategy requires a memory allocator whose responsibility is to keeps track how much memory is allocated and how much is free. Based on this knowledge, it provides applications with the right amount of memory and reuses a memory block that is no longer in use. In wireless sensor networks, it is usually known at design time how much memory is required by the application (in most cases, there is only a single application), and therefore, the overhead introduced by the memory management service is undesirable.

3.17 What is a virtual memory?

In a system that supports virtual memory, physically fragmented blocks of an active memory can be presented to applications as if they were contiguous. Alternatively, the address space can be extended to include some portion of the disc storage.

3.18 In most communication systems, the last stage in the reception process requires digital-to-analog converters (DAC). But in this book, the DAC is not discussed. Why do you think is the reason?

The data extracted from a wireless sensor network is further processed at the application level using digital signal processing algorithms. Therefore, there is no need to convert the digital signal back to an analog signal.

3.19 Explain two different ways of interfacing an analog temperature sensor with a processor subsystem.

(a) Most processing subsystems integrate internal ADCs which can be directly connected with the temperature sensor.

(b) Alternatively, an external ADC can be used to interface the temperature sensor with the processing subsystem.

3.20 How can two hardware components having different speed communicate with each other through a serial bus?

If the serial bus is SPI, the master adjusts its speed with the speed of the slower device. if I^2C is used, the minimum speed that can be supported is already set, so a device has to satisfy this requirement to access the bus.

4

Operating Systems

4.1 What is a process in the context of operating systems?

A process is an instance of an executable program. In a multi-threaded system, a process may consist of one or more threads.

4.2 What is an intra-process communication and how does it differ from inter-process communication?

A typical inter-process communication is communication between multiple threads of the same process. These threads share the same memory space, so they can communicate with each other by writing into and reading data from the memory. Inter-process communication is communication between two or more processes which do not share a memory space. These processes require a mediator – the operating system – to exchange messages. Alternatively, they can use messages and message listeners to directly communicate with each other.

4.3 Explain the difference between a system program and an application program?

In wireless sensor networks, a system program belongs to the operating system. It is mainly responsible for managing hardware devices (such as a watchdog timer, the radio, a sensor, etc.). An application program belongs to the application or a higher-level service (such as a routing protocol) and accesses hardware devices through one or more system programs.

4.4 What are system calls?

System calls are a set of functions provided by the operating systems to applications and higher-level services so that they can make requests for low-level (mostly hardware related) services without the need to know how the system calls are implemented or the services are actually provided.

4.5 Explain the following terms and what are some of the mechanisms to avoid them?

(a) Race condition A race condition occurs when there is a timing problem with the arbitration mechanism of a shared resource in an operating system that supports concurrent programming. A typical example is when two programs “collide” as they attempt to modify one and the same file, which leads to data corruption.

(b) Deadlock

A deadlock occurs when two processes in a multi-threaded environment simultaneously wait for the other to release a shared resource. According to Coffman et al. [1] deadlock occurs because of one of the following reasons:

i. Tasks claim exclusive control of the resources they require. This is called *mutual exclusion condition*.
ii. Tasks hold resources already allocated to them while waiting for additional resources. This is called *wait for condition*.
iii. Resources cannot be forcibly removed from the tasks holding them until the resources are used to completion. This is called *no preemption condition*.
iv. A circular chain of tasks exists, such that each task holds one or more resources that are being requested by the next task in the chain. This is called *circular wait condition*.

The first condition can be prevented by removing the mutual exclusion condition, i.e. no process may have exclusive access to a resource. For preventing the remaining three conditions, the Coffman et al. adopt the suggestion made by Havender [2]:

i. Each task must request all its required resources at once and cannot proceed until all have been granted.
ii. If a task holding certain resources is denied a further request, that task must release its original resources and, if necessary, request them again together with the additional resources.
iii. The imposition of a linear ordering of resource types on all tasks. In other words, if a task has been allocated resources of type r_i it may subsequently request only those resources of types following r_i in the ordering.

(c) Starvation Resource starvation occurs when a processor is perpetually denied to access the resources it requires. This is particularly the case when the operating system supports priority-based execution – a low priority process may never be scheduled because of a blocking high priority process. Round-robin-based scheduling mechanism avoids resource starvation by dividing the execution time into slots and making sure that each task (process) receives a portion of this time.

4.6 Compare the following scheduling mechanisms:

(a) FIFO scheduling

In a FIFO scheduling mechanism a task is processed based on the first-in-first-out principle.

(b) Sorted queue

In a sorted queue scheduling mechanism, tasks in a queue are first sorted according to a set criterion, for example, based on their length.

[1] Coffman, E. G., Elphick, M., and Shoshani, A. 1971. System Deadlocks. ACM Comput. Surv. 3, 2 (Jun. 1971), 67-78.

[2] Havender. J. W. 1968. Avoiding deadlock in nnllti-tasking systems. IBM Systems Journal. 2 (1968), 74-84.

(c) Round-robin

In round-robin scheduling mechanism, time multiplexing is used to divide time fairly among competing tasks. The execution time is divided into several slots and each task is executed in a slot. When the slot is due, the task is set on hold and the next task is executed. This way, all tasks progress together towards their completion.

4.7 What are interrupts and interrupt handlers?

An interrupt is an asynchronous signal (event) and is generated by a hardware or software component which requires immediate handling. When the processing subsystem receives an interrupt event, it transfer control to an interrupt handler, also called an interrupt service routine (ISR). An interrupt handler is a callback subroutine which is called by the operating system when an interrupt event is received. Interrupt handlers should register for the interrupt types they are interested in.

4.8 Why do most operating systems in wireless sensor networks define a kernel? A monolithic kernel with a small finger print enables dynamic module update and dynamic reprogramming.

4.9 What is a preemptive process? Provide an example. A preemptive process is a privileged process which temporarily preempt (interrupt) a task which is currently being executed and takes over control.

4.10 How is concurrency supported in TinyOS?

In TinyOS tasks are executed up to completion, which means only one task accesses a resource at a time.

4.11 What is a split-phase programming and how is it useful in wireless sensor networks?

A split phase program divides a function call into a call (which will be immediately acknowledged) and return (which will be notified as an asynchronous event when the called function completes execution).

4.12 Explain the difference between configuration components and modules in TinyOS.

A configuration component describes how different modules are interconnected to build an executable service or application, whereas a module is an implementation of an interface.

4.13 Why do threads require their own separate stacks and what is the problem with this approach in wireless sensor networks?

Threads require separate stacks in order to store their own context. This requires a memory space larger than a single-stack based execution. Since memory is a scarce resource in wireless sensor networks, multi-threading is expensive.

4.14 Give three reasons for supporting dynamic reprogramming in wireless sensor networks.

(a) The sensing task may change overtime;

(b) The application code may need debugging and correction; and

(c) Policies related to the environment in which the wireless sensor network operates may change and therefore, the network may need to adapt to this change.

4.15 Explain the difference between event-based and thread-based operating systems. Discuss some of the advantages and disadvantages of the two approaches in the context of wireless sensor networks.

In event-based programming, interaction between processes is based on events and event handlers. Tasks are executed to completion, unless they are interrupted by events. This way concurrency is supported and execution is efficient. Since only one task is executed at a time, long-duration tasks may block short-duration tasks, but this problem can be overcome by using a sorted queue scheduling.

In multi-threaded programming, multiple threads run concurrently. Threads can be suspended ensuring non-blocking operation. However, thread management introduces resource overhead on the operating system.

4.16 Explain the difference between static and dynamic memory allocation.

In a static memory allocation, memory is allocated to a piece of program at compilation time. If the memory requirement of the program is known at compilation time and this requirement remains unchanged, static memory allocation is efficient. But sometimes it is difficult to foresee the memory requirements of a piece of program, in which case static memory allocation is inflexible.

In dynamic memory allocation, the memory requirement of a piece of program is decided at runtime, and memory is allocated accordingly. While it is flexible, programs are usually allocated memory a little more than they require, which can be inefficient in resource-constrained devices.

4.17 How is separation of concern supported in the following operating systems:

(a) Contiki

Contiki distinguishes between core services (which are essential to the OS and remain unchanged once the system is running) and dynamic reloadable services which can be reprogrammed at runtime.

(b) SOS

SOS provides a small monolithic kernel and reloadable (reprogrammable) modules. SOS saves the context of a module outside of the module so that context transfer during dynamic reprogramming is possible.

(c) LiteOS

In LiteOS, applications are not a part of the OS, so each can be independently developed.

4.18 Explain the following concepts in TinyOS:

(a) Commands

Commands are non-blocking requests for service.

(b) Tasks

Tasks are monolithic processes that should be executed to completion

(c) Events

An event is an occurrence of interest outside of a process and prompts the process to act (or handle the event).

4.19 What is the difference between a TinyOS command and a SOS message?

The two are the same in that both are executed asynchronously and both are scheduled before they are processed. Whereas tasks are executable processes, messages require message handlers to process.

4.20 Why is the state of a module stored in a separate memory space (outside of the module) in SOS? So that dynamic module update or reprogramming is possible. If a module's state is stored outside of it, it can easily be replaced or modified.

4.21 Explain how SOS supports dynamic reprogramming.

When a new module is available, a code distribution protocol advertises it in the network. The local distribution protocol evaluates the advertisement in terms of relevance and resource requirements and if all is fine, proceeds with downloading. Once the downloading is successfully completed, module insertion takes place. During module insertion, the kernel creates metadata to store the absolute address of the handler, a pointer to the dynamic memory holding the module state and the identity of the module. Then the SOS kernel invokes the handler of the module by scheduling an init message for the module.

4.22 How is multi-threading supported in a Contiki environment?

Contiki is by default an event-based operating system, but offers multi-threading an alternative library service, which can be dynamically loaded and integrated as a part of the application code.

4.23 What is the function of a program loader in Contiki and why is it important?

The program loader is responsible for dynamically loading a module. If the module is available locally, it loads it from the program memory into the active memory, but if the module is not available locally, then it employs the communication module to fetch the binary image.

4.24 How is module replacement supported in Contiki?

Contiki supports dynamic reprogramming by separating replaceable modules from the core services (which make up the program loader, the communication service and the kernel). The former can be replaced by employing the core services.

4.25 What is the advantage of considering a wireless sensor network as distributed file system in LiteOS?

Users can easily navigate through and program the sensor nodes. For both aspects LiteOS provides intuitive interfaces, particularly, for Linux users.

4.26 What is differential patching in LiteOS?

A differential-patching estimates the location of a program in the active memory and carry out module update based on this knowledge.

4.27 Explain the functions of the following message handlers in SOS:

(a) *init-handler* The *init-handler* is called by the scheduler when first a module is initialized. During dynamic module replacement, it is useful to pass over the context of the previous module.

(b) *final-handler*

The *final-handler* is called before a module is removed from the active memory so that it can gracefully release all the resources it owns.

4.28 Which type of scheduling strategy do the following operating systems employ:

(a) TinyOS: FIFO

(b) SOS: FIFO

(c) Contiki: FIFO, Priority scheduling (for poll handlers)

(d) LiteOS: Priority scheduling with an optional round-robin

4.29 How does TinyOS deal with dynamic reprogramming?

TinyOS requires a separate module (outside of the OS) to support dynamic programming

4.30 Why is separation of concern in TinyOS not a priority?

Because of code efficiency.

Part Two

Basic Architectural Framework

5

Physical Layer

5.1 How can a single ADC be employed by multiple sensors to convert their analog output into corresponding discrete sequences?

An analog multiplexer can be used to serialize the input of two or more sensors.

5.2 Suppose a discrete memory-less channel (DMS) source emits symbols from the ternary alphabet $A = \{-1, 0, 1\}$ with a probability, $P(-1) = 0.5, P(0) = P(1) = 0.25$. But the source can also be configured such that instead of emitting one symbol at a time, it can emit two symbols (A^2) with a probability that is the multiplication of the probabilities of the individual symbols. Show that the entropy of the second configuration is twice greater than the entropy of the first configuration.

The entropy of the first configuration, $H(A)$, is:

$$\begin{aligned} &= -p(-1)log(p(-1)) - p(0)log(p(0)) - p(1)log(p(1)) \\ &= -0.5log(0.5) - 0.25log(0.25) - 0.25log(0.25) \\ &= -1.5log(0.5) \end{aligned}$$

The entropy of the second configuration, $H(a^2)$, is:

$$\begin{aligned} = \quad & p(-1,-1)log(p(-1,-1)) - p(-1,0)log((-1,0)) - p(-1,1)log(p(-1,1)) - \\ & p(0,-1)log((0,-1)) - p(0,0)log((0,0)) - p(0,1)log((0,1)) - \\ & p(1,-1)log((1,-1)) - p(1,0)log((1,0)) - p(1,1)log((1,1)) \\ = \quad & -3log(0.5) \end{aligned}$$

From the two values, it can be concluded that $H(A^2)$ = $2H(A)$.

5.3 The following codes are given:

$C_1 = \{1, 10, 01\}$
$C_2 = \{0, 00001\}$
$C_3 = \{0, 10, 11\}$
$C_4 = \{01, 11\}$
$C_5 = \{0, 00, 000\}$

(a) Which of the given codes are uniquely decodable?

C_2,C_3, and C_4 are uniquely decodable whereas C_1 and C_5 are not. For example, if we decode the sequence 11011 with $(C_1 : a1 = 1, a2 = 10, ac = 01)$, it can be decoded as a1, a2, a1,a1 or as a1,a1,a3,a1. So it's not uniquely decodable. For $(C_5 : a1 = 0, a2 = 00, a3 = 000)$, if we decode the sequence 000000, it can be decoded as a1, a2, a3 or a1, a3, a2 or a2, a1, a3 or a2, a3, a1 or a3, a1, a2 or a3, a2, a1.

(b) Which of the given codes are instantaneously decodable?

C_3 and C_4, since C_3 and C_4 are prefix-free codes, which indicates that they are also the instantaneously decodable codes. Instantaneously decodable codes must be the uniquely decodable codes. So the non-uniquely decodable codes must be non-instantaneously decodable codes. Consequently, C_1 and C_5 cannot be instantaneously decodable. For C_2, if the sequences 000001 need to be decoded, one can decode only after the bit 1 is received. Thus, there is a delay in decoding.

(c) Which of the given codes can be an optimal prefix-free code for some probability assignment?

C_3 and C_4.

5.4 Suppose an information source emits symbols from an alphabet $X = \{x_1, ..., x_8\}$ with corresponding probabilities $\{0.2, 0.35, 0.15, 0.1, 0.09, 0.06, 0.04, 0.01\}$.

(a) Calculate an upper bound on the average codeword length achievable with a binary Shannon or Huffman code if single symbols are encoded at a time.

It should be remembered that $H(x) \leq \bar{R} < H(x) + 1$.

$$\begin{aligned} H(x) &= \sum_{i=1}^{K} p(x_i) log_2(p(x_i)) \\ &= -0.2log_2(0.2) - 0.3log_2(0.3) - 0.15log_2(0.15) - 0.1log_2(0.1) \\ &\quad -0.09log_2(0.09) - 0.06log_2(0.06) - 0.04log_2(0.04) - 0.01log_2(0.01) \\ &= 2.54 \end{aligned}$$

Thus, the upper limit is equal to $2.54 + 1 = 3.54$.

(b) Construct a binary Huffman code for the given source.

The figure below displays the Huffman coding.

From the figure, the following table can be generated.

$Symbol$	$Probability$	$Code$
x_1	0.35	0
x_2	0.2	100
x_3	0.15	101
x_4	0.1	1100
x_5	0.09	1101
x_6	0.06	1110
x_7	0.04	11110
x_8	0.01	11111

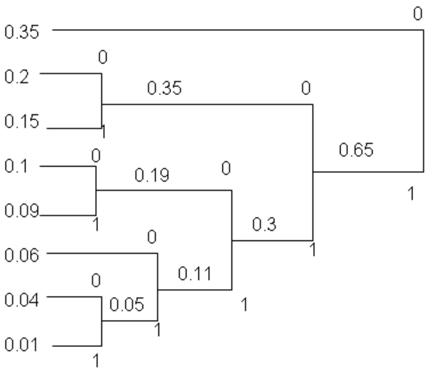


Figure 5.1 The Huffman coding

(c) Calculate the average codeword length achieved by the Huffman code and compare it with the calculated bounds.

The average codeword length, $\bar{R}$:

$$\begin{aligned}\bar{R} &= 0.35 \times 1 + 0.2 \times 3 + 0.1 \times 4 + 0.09 \times 4 + 0.06 \times 4 + 0.04 \times 5 + 0.01 \times 5 \\ &= 2.65\end{aligned}$$

$2.54 < 2.65 < 3.54$, and hence, the codeword length of the Huffman code is smaller to that of the upper bound.

5.5 Refer to the analog signal shown in Figure **??**.

[h]

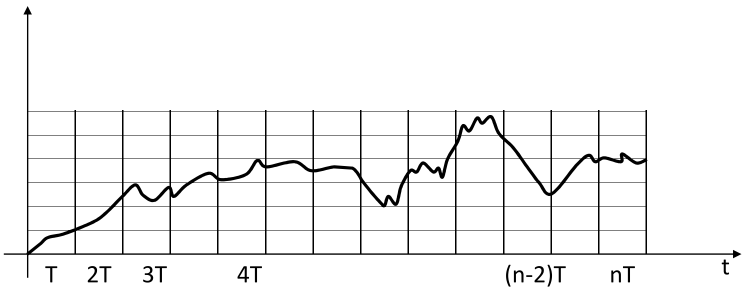


Figure 5.2 Source coding an analog signal

(a) How can the signal be encoded with a 3 bit PCM?

The figure below illustrates how the analog signal can be converted to a discrete sequence and encoded with a 3 bit PCM. As can be seen, approximation has been

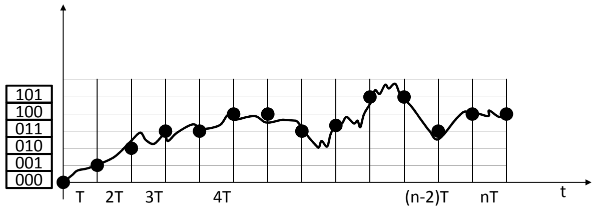


Figure 5.3 A 3 bit PCM encoder

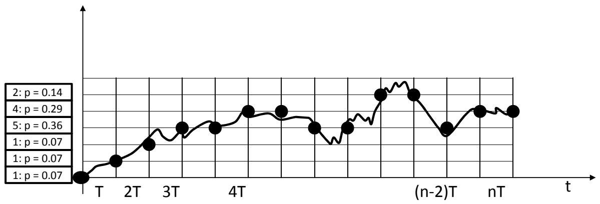


Figure 5.4 A 3 bit PCM encoder

made by the ADC when the analog signal is either below or above the specified discrete levels. This approximation is the source of quantization error. In the figure, each discrete value is encoded by a 3 bit codeword regardless of the fact that some values occur more frequently than others.

(b) How can the signal be encoded with a delta encoder?

(c) Illustrate how a PCM encoder with a codebook of different symbol length can be used to efficiently encode the signal.

The figure below illustrates that the analog signal has some values that are more predictable than others. Out of the 14 discrete samples that can be obtained between $[0, nT]$, the third (considering 0 as the first level) discrete level occurs five time and the fourth discrete level occurs four times. If these occurrences were representative of the future samples, the Huffman coding or a code similar to it can be used to efficiently encode the analog signal.

(d) Manchester coding is a line encoding technique that is useful for minimizing the effect of DC voltage during data transmission and for dynamic clock recovery. Illustrate how the PCM bit stream can be encoded with a Manchester coding.

The figure below illustrates how Manchester coding can be employed to encode the PCM output. In (b), the PCM representation of the analog signal is displayed. As can be seen, the figure contains too many successive zeros that can produce a significant DC. The Manchester code shown in (c) reduces this effect by causing a symbol to have a transition from low to high or vice versa in each clock cycle.

(e) Now instead of Manchester coding, encode the PCM stream with a differential Manchester coding.

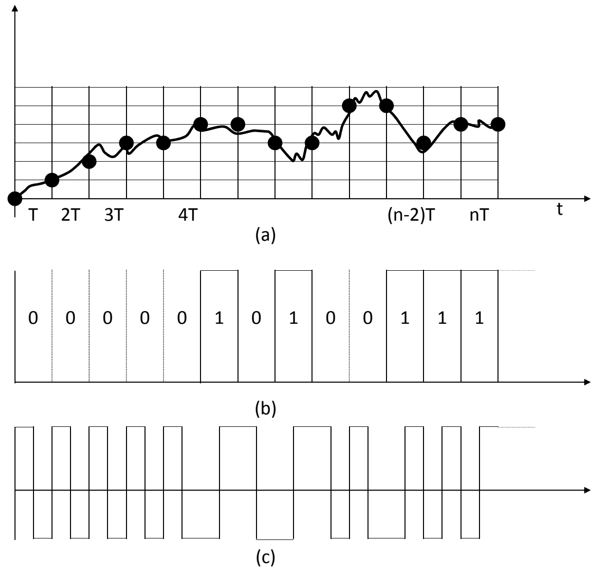


Figure 5.5 Manchester coding

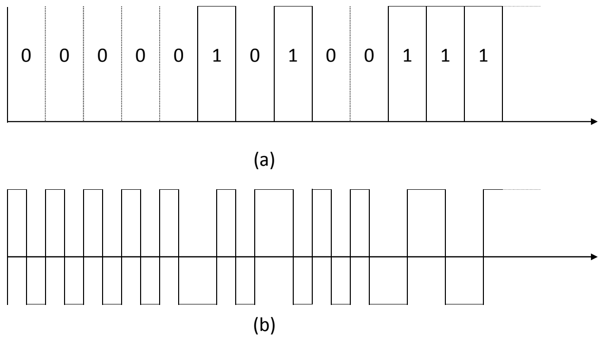


Figure 5.6 Differential Manchester coding

The figure below displays the differential Manchester coding where the symbol “0” is represented by the presence of a transition in the first half cycle either from high to low or low to high and “1” is represented by the absence of a transition in the first half cycle.

(f) Discuss the difference between the bit streams generated by the Manchester and differential Manchester encoding techniques.

In Differential Manchester Coding, the presence or absence of a transition determines whether a “0” or “1” is being transmitted. This has two advantages: (a) Often it is easier to detect transitions rather than measuring the signal's amplitude to determine the symbol. and (b) it is not relevant to decide whether a transition is from high to low or vice versa to determine a symbol. This means a Differential Manchester coding works exactly in the same manner if the signal is inverted.

5.6 The feedback loop shown in Figure **??** is a very useful concept in linear systems. It is the basic principle for designing stable amplifiers and oscillators. Moreover, most receivers employ the feedback loop to set up an automatic gain control (AGC) to ensure that the power of the received message signal remains constant despite changes in the channel's properties. The feedback loop is characterized by the overall loop gain, which is the ratio of the output voltage to the input voltage, $G_{loop} = \frac{V_{out}}{V_{in}}$. Calculate the overall loop gain, G_{loop}.

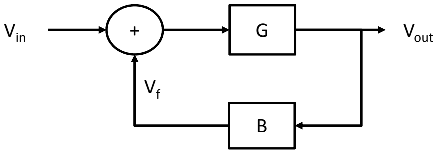


Figure 5.7 A feedback loop

$$\begin{aligned} V_{out} &= (V_{in} + V_f)\,G \\ &= GV_{in} + GBV_{out} - - > GV_{in} = V_{out}(1 - GB) \\ G_{loop} &= \frac{V_{out}}{V_{in}} \\ &= \frac{G}{1-GB} \end{aligned}$$

5.7 The half-wave rectifier shown in Figure **??** is one of the two components of an envelope detector which is responsible to extract the baseband message signal from the carrier. Sketch the output of the rectifier for the corresponding sinusoidal input signal.

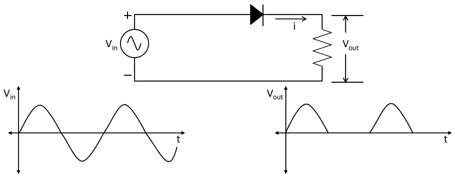


Figure 5.8 A half-wave rectifier

5.8 Now instead of the half-wave rectifier, the full-wave bridge rectifier shown in Figure **??** is used. How does the output wave form look like?

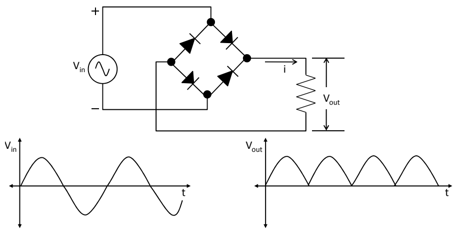


Figure 5.9 A full-wave bridge rectifier

5.9 A lowpass filter is required to separate the baseband signal from the carrier in an amplitude modulated signal. Explain how the lowpass filter shown in Figure **??** can be used for this purpose.

Recall that a capacitor is like a short circuit at a high frequency and an open circuit at high frequency. Therefore, at low frequency, v_{out} is high while at high frequency, the voltage drop across the capacitive element is very small.

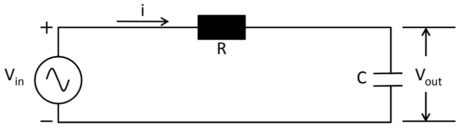


Figure 5.10 An RC lowpass filter

Based on Figure **??**, drive an expression for:

(a) The voltage drops across the resistor and the capacitor.

$$\begin{aligned} V_R &= \frac{R}{r+\frac{1}{j\omega C}} V_{in}. \\ V_C &= \frac{\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}} V_{in} \\ &= \frac{1}{1+j\omega RC} V_{in} \end{aligned}$$

(b) The current that circulates in the filter.

$$\begin{aligned} i &= \frac{V_{in}}{\left(R+\frac{1}{j\omega C}\right)} \\ &= \frac{V_{in}(j\omega C)}{1+j\omega RC} \end{aligned}$$

At a high frequency, $j\omega C >> 1$ and the above equation can be reduced to:

$$i = \frac{V_{in}}{R}$$

This indicates that the capacitive element is effectively a short circuit. As a result, there will be no voltage drop across it.

(c) The transfer function, $H_C(s) = \frac{V_{out}(s)}{V_{in}(s)}$, where $s = j\omega$ is the Laplace operator.

$$\begin{aligned} V_{out}(s) &= \tfrac{1}{1+sRC} V_{in} \\ H_C(s) &= \tfrac{1}{1+sRC} \end{aligned}$$

5.10 A modulating signal, $m(t) = 5cos(2\pi \; 1KHz \, t)$, is used to amplitude modulate a carrier signal, $c(t) = 10cos(2\pi \; 100MHz \, t)$.

(a) Compute the time domain expression of the modulated signal

Recall that $cos(x) \times cos(y) = \frac{1}{2}\left(cos(x+y) + cos(x-y)\right)$. Hence, the modulated signal is:

$$c_{mod}(t) = \tfrac{50}{2}\left(cos((100,000,000 + 1000)t) + cos((100,000,000 - 1000)t\right)$$

(b) Compute the frequency domain expression of the modulated signal

As can be seen from the time domain feature, there are two frequency components, one at $f_c - f_m = (100,000,000 - 1000)Hz$ and another at $f_c + f_m = (100,000,000 + 1000)Hz$. The frequency domain expression is simply the Fourier transformation of two cosine signals at the specified frequencies.

(c) Suppose the message signal is sampled at a period T using the Dirac's delta function as shown in Figure **??**. How does the spectrum of the sampled signal look like?

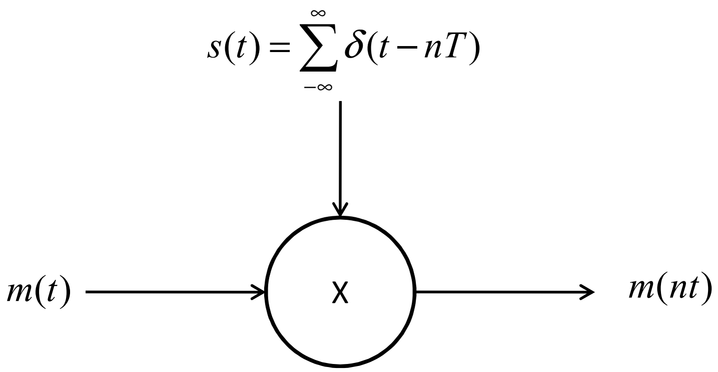


Figure 5.11 Sampling a modulating signal with the Dirac's delta function

$$\begin{aligned} s(t) &= \textstyle\sum_{-\infty}^{\infty} \delta(t - nT) \leftrightarrow S(j\Omega) = \frac{2\pi}{T} \sum_{-\infty}^{\infty} \delta(\Omega - k\Omega_s) \\ m(nT)(m_s[n]) &= m(t)s(t) \leftrightarrow M_s(j\Omega) = \tfrac{1}{2\pi} m(j\Omega) * S(j\Omega) \end{aligned}$$

Thus, $M_s(j\Omega) = \frac{1}{T}\sum_{-\infty}^{\infty} m(t)(j(\Omega - k\Omega_s))$. The Fourier transform of the sampled sequence consists of periodic repetition of the Fourier transform of m(t).

(d) What precondition should be satisfied in order to reconstruct the continuous modulating signal from the sampled sequences?

The sampling frequency has to be at least twice higher than the bandwidth of the message signal.

5.11 The path loss index, γ, describes how an electromagnetic wave is attenuated as it propagates through a space. In free space, where there is no obstacle between the transmitter and the receiver, $\gamma = 2$. That means, the power of a propagating electromagnetic wave falls as the function of the square of a distance – $P_r \approx \frac{P_t}{4\pi\rho^2}$, where ρ is the distance in meter. If, however, there is an obstacle between the transmitter and the receiver, this figure is greater than 2. Figure **??** displays a simple model in which the electromagnetic wave reaches the receiver after being reflected once. Drive an expression of the path loss index for the model. Assume that the reflector is an ideal reflector and there is a line-of-sight link between the transmitter and the reflector and the receiver and the reflector. Assume also $\rho >> h$.

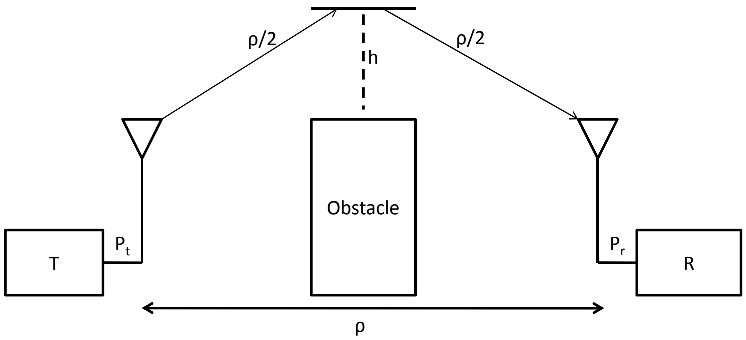


Figure 5.12 Single reflection model for an electromagnetic propagation

If the scattering object is lossless, the power received by it is proportional to: $P_{scat} \propto \frac{P_t}{2\pi\left(\frac{\rho}{2}\right)^2}$. Likewise, the received power is proportional to: $P_r \propto \frac{P_{scat}}{2\pi\left(\frac{\rho}{2}\right)^2}$. Subsequently, the received power as a function of the transmitted power can be expressed as: $P_r \propto \frac{P_t}{2\pi\left(\frac{\rho}{2}\right)^4}$. Thus, the pathloss index for this model is 4.

5.12 Figure **??** displays the block diagram of a part of a receiver. It consists of an omnidirectional antenna, a RF amplifier, a local oscillator, an intermediate-frequency amplifier and a detector (an envelope detector). While it is possible to detect the modulating signal after mixing the received, modulated signal with the local carrier signal, it is useful to have the intermediate state.

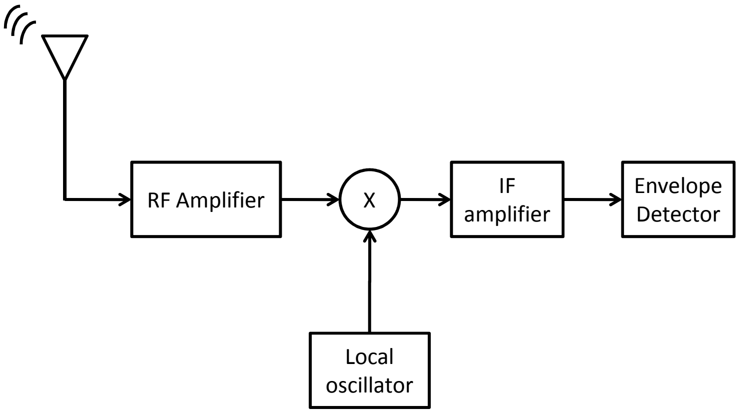


Figure 5.13 Block diagram of a receiver

(a) Why is the intermediate frequency amplifier desirable?

Intermediate frequencies are useful for three main reasons. At very high (Giga Hertz) frequencies, signal processing circuits perform poorly. Active devices such as transistors cannot deliver much amplification (gain) without becoming unstable. Ordinary circuits using capacitors and inductors must be replaced with cumbersome high frequency techniques such as waveguides. So a high frequency signal is converted to a lower IF for stable processing.

The second reasons is that in receivers that can be tuned to different carrier frequencies it is desirable to employ the same amplification and detector circuits for all the stations. This means, there should be a single intermediate frequency after the mixer component. It is difficult to build amplifiers, filters, and detectors that can be tuned to different frequencies, but easy to build tunable oscillators. Without using an IF, all the components of a receiver would have to be tuned in unison each time a different carrier frequency is chosen.

The third reason is to improve frequency selectivity. Often it is challenging to design narrow band filters that should operate at high frequencies.

(b) Suppose the receiver is used for receiving an amplitude modulated signal. How can the intermediate frequency be obtained?

The intermediate frequency (IF) can be obtained by mixing the received RF (carrier signal) with another RF signal that is produced by the local oscillator.

5.13 What type of roles do the transmitter's and the receiver's antennas play to enhance signal propagation and reception?

The gain of an antenna is proportional to the effective electrical area of the antenna, which in turn is a function of the wavelength and the physical cross sectional area. So by improving the effective cross sectional area of the antennas, it is possible to improve their efficiency.

5.14 Why is a considerable amount of power wasted at the power amplifier of a transmitter?

A power amplifier dissipates a considerable amount of heat which is the source of inefficiency.

5.15 Explain the trade-off between modulation efficiency and design complexity in quadrature amplitude modulation?

The modulation efficiency of a quadrature amplitude modulation increases if as much information as possible can be conveyed with a single discrete signal level. But as the number of distinct signal levels that can be decoded increases, the receiver's sensitivity (resolution) in discriminating between the discrete signal levels becomes severe as well. This requires a refinement in design which in turn has a cost implication.

6

Medium Access Control

6.1 What is the main purpose of the MAC layer and why is this challenging in networks with shared media?

The medium access control (MAC) layer is responsible for regulating access to the wireless medium. The wireless medium is a shared medium, i.e. multiple wireless devices use it for their communications, potentially leading to interferences among these devices. The responsibility of the MAC layer is to either prevent or appropriately respond to errors and interferences during communications. Further, the choice of MAC protocol affects the reliability and efficiency of such communications.

6.2 What are the advantages and disadvantages of contention-free and contention-based medium access strategies? Can you think of scenarios where one would be preferable over the other?

Contention-free medium access strategies avoid collisions by ensuring that when one device transmits, all other devices that could interfere with the reception of the transmitted data remain silent. Avoiding collisions can have positive impacts on communication latency and throughput. Most contention-free approaches are based on schedules, indicating exactly when a node can transmit or must listen. This also facilitate power management strategies, e.g. using such a schedule, a device knows exactly when it can power down its radio.

On the other hand, contention-based medium access strategies do not avoid collisions, but instead provide mechanisms to recover from collisions. An advantage of such a strategy is that a device does not have to wait (and delay), but instead can transmit immediately. In networks with low traffic loads, this approach may result in lower latencies. A contention-based strategy may also be easier to implement, since no schedules are required. Besides the potential collisions (which may lead to increased latency and reduced throughput), it may be more difficult to implement duty cycle techniques in a network using a contention-based MAC protocol.

Contention-based access strategies may be preferable in networks with light traffic loads and unpredictable (or sporadic) traffic patterns. Contention-free access schemes are useful when the traffic is predictable (e.g. sensor nodes periodically reporting their sensed data) and when it is important to maximize the sleep times of the wireless radios.

Fundamentals of Wireless Sensor Networks: Theory and Practice Waltenegus Dargie and Christian Poellabauer

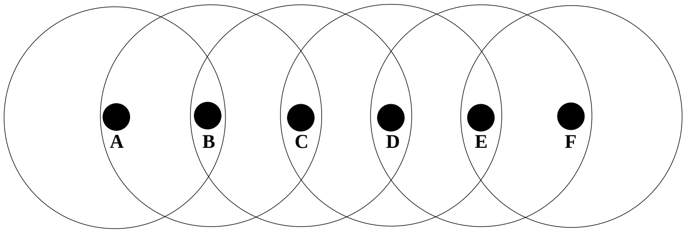


Figure 6.1 Hidden-terminal problem (Exercise 6.5)

6.3 The key idea behind CSMA/CD is that the sender detects collisions, allowing it to react correspondingly. Why is this approach not practical in wireless networks?

In the wireless medium, a collision occurs at the receiver of a transmission and the sender will be unaware of this collision.

6.4 What are "hidden terminals" and how do they affect the performance of wireless sensor networks?

When two sender devices A and B are both able to reach a third device C, but they are not able to reach each other, the situation may occur where both A and B attempt to transmit to C at the same time, unaware of the collision occurring at C. That is, A and B are hidden from each other. The consequences of such collisions are the need for recovery mechanisms, increased delays, and reduced throughput. A MAC protocol has to be able to either recover from such a collision or prevent such collisions by preventing devices A and B from transmitting to C at the same time.

6.5 Consider the network topology in Figure **??**, where circles indicate the communication and interference range of each node, i.e. each node can hear the immediate neighbors to the left and right. Assume that RTS/CTS is not being used.

(a) Node B currently sends to node A and node C wants to send to node D. Is node C allowed to do so (i.e. can it do so without causing a collision) and will it decide to do so?

Yes, node C will not cause a collision at node A. However, since node C can hear node B's transmission, it will not transmit.

(b) Node C sends to node B and node E wants to send to node D. Is E allowed to do so and will it do so?

Node E's transmission will not interfere with C's transmission to B, however its own transmission to D will not succeed since it will collide with node C's transmission. Node E is unaware of node C's transmission and therefore it will begin its transmission.

(c) Node A sends to node B and node D sends to node C. Which other nodes are allowed to send at the same time?

Node E is allowed to transmit to node F. Node F could also transmit to node E, but this transmission would collide at node E with node D's transmission.

(d) Node A sends to node B and node E sends to node F. Which other nodes are allowed to send at the same time?

Node C is not allowed to transmit since its transmission will collide with node A's transmission at node B. On the other hand, node D is allowed to transmit to node C.

6.6 Describe the problems in using CSMA as medium access control mechanism in a WSN.

CSMA does not address the hidden-terminal problem and collisions can be costly, particularly if the transmitted frames are large. Neighboring nodes also do not know how long transmissions will take, making it difficult to decide whether to power down a wireless radio (and for how long).

6.7 In a CSMA/CA network, nodes use a random delay before accessing the medium. Why is this being done?

In CSMA, nodes can access the wireless medium immediately after it has been sensed idle. CSMA/CA is a variation of the CSMA protocol, where the number of collisions are reduced by randomly delaying medium access. That is, a node with a short random delay may access the medium sooner than a node choosing a larger random delay. The node with the larger delay will therefore overhear the transmission of its neighbor and it will not initiate a transmission, thereby avoiding a collision.

6.8 Assume that the RTS and CTS frames were as long as the DATA and ACK frames. Would there be any advantage to using the RTS/CTS approach? Explain why or why not.

The reason RTS and CTS frames are used is to reserve the channel for a DATA frame. When a frame collides with another frame, it must be re-transmitted at a later time. If the DATA frame is large, it is preferable to exchange the much shorter RTS and CTS frames first, since their collision will result in less overhead. However, if the DATA frame is short (e.g. as short as the RTS frame), there is no benefit in first transmitting RTS and CTS frames. As a matter of fact, in such cases, using RTS and CTS frames introduces additional overheads because two additional frames must be transmitted.

6.9 How does MACAW extend MACA and what is the purpose of the additional control messages?

In MACAW, a receiver responds with an acknowledgment (ACK) frame to indicate a successful reception of the data frame. This allows neighboring nodes to learn that the transmission has finished and that the channel has become available. In addition, a node transmitting an RTS frame also transmits a Data Sending (DS) frame immediately after receiving a CTS frame for its RTS frame. A node overhearing the RTS frame, but not the corresponding CTS frame, has now confirmation that the reservation has succeeded and that it must remain silent to not interfere with the transmission.

6.10 What are the specific features of the IEEE 802.11 PSM (Power Saving Mode) and what are the main difficulties of using it in wireless sensor networks?

IEEE 802.11 offers the power saving mode for nodes operating in the PCF mode. A device can inform the base station that it wishes to enter a low-power sleep mode using special control messages. To ensure that incoming messages can be received, the device periodically awakens to receive beacon messages from the base station.

6.11 Does the NAV field in IEEE 802.11 networks prevent the hidden-terminal problem?

The NAV field does not prevent the hidden-terminal problem. It is used to reduce the need for continuously sensing the medium, thereby allowing a node to preserve more power.

6.12 Explain why the IEEE 802.11 standard uses three different "interframe spaces".

The "standard" interframe space DIFS is used to ensure that the medium has been sensed idle for a certain minimum amount of time before a transmission is attempted. IEEE 802.11 can use RTS and CTS frames to reserve the wireless channel, therefore, to ensure that no other node will begin a transmission when a reservation has occurred, a second (shorter) interframe space (SIFS) is used to separate RTS frames from CTS frames, CTS frames from DATA frames, and DATA frames from ACK frames. Finally, in the PCF mode, an even shorter interframe space (PIFS) is used to ensure that PCF traffic has priority over traffic generated by devices in the DCF mode.

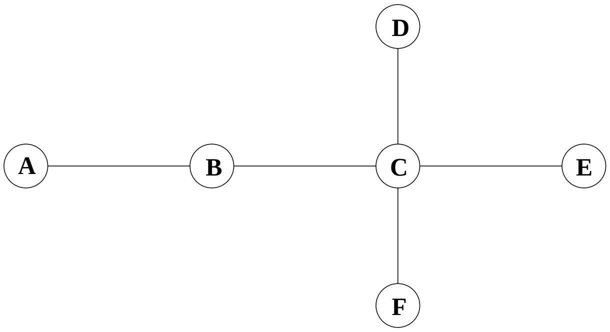


Figure 6.2 TDMA protocol (Exercise 6.13)

6.13 Consider the network topology in Figure **??**, where the lines indicate which nodes can communicate and interfere with each other. Assume a TDMA protocol with a frame size of 5 slots and that each node can only be sender or receiver during any time slot.

(a) Generate a schedule such that every node has an opportunity to communicate to all if its neighbors.

Table 6.1 TDMA schedule (Exercise 6.13)

Node	Slot 1	Slot 2	Slot 3	Slot 4	Slot 5
A	T	R	-	-	-
B	R	T	R	-	-
C	R	R	T	R	R
D	-	-	R	T	-
E	-	-	R	-	T
F	T	-	R	-	-

See Table **??** for an example of a possible schedule, where ‘R’ indicates a slot where a node is in the receive mode and a ‘T’ indicates a slot where a node is allowed to transmit.

(b) For your schedule, how many slots in a frame could each node sleep to preserve energy? What is your insight with respect to node density and energy preservation?

Nodes A, D, E, and F can each sleep for 3 slots in each frame. Node B can sleep for 2 slots. Node C must remain awake for the entire duration. From this example, it can be seen that node C will have much higher energy consumption than any other node in the network. That is, the more neighbors a node has, the more time it will spend in active mode to listen to its neighbors’ transmissions.

(c) Assume that node A sends a message to node E; how long (in number of time slots) does it take for E to receive the message using your schedule (explain your answer)?

Given the schedule in Table **??**, A can transmit the message to B in slot 1, B can repeat this message to node C in slot 2, and node C can repeat this message to node E in slot 3. Therefore, the transmission will only require 3 slots.

6.14 Why is the IEEE 802.15.4 standard preferable over the IEEE 802.11 standard for most wireless sensor networks?

The IEEE 802.15.4 standard was created specifically for low-power devices, with much lower data ranges compared to IEEE 802.11 and channels in the 868 MHz, 915 MHz, and 2.45 GHz ranges. Its focus on low power and low data rates much better meets the demands of wireless sensor networks than the IEEE 802.11 standard.

6.15 Describe how the design of the MAC protocol affects the energy efficiency of a sensor node.

Since the MAC protocol determines when a node may access a wireless channel for communication, its design determines when and for how long a node can power down its wireless radio to preserve energy. Some MAC protocols require nodes to be awake continuously to ensure that no messages are missed. Other protocols use schedules that determine exactly when a node must stay awake, allowing a node to sleep without the risk of losing messages. Other MAC protocol characteristics affecting energy efficiency include packet header overheads, reliability features (such as retransmission and error control mechanisms), and the number (and sizes) of control messages exchanged between nodes.

6.16 This chapter described five requirements of MAC protocols for wireless sensor networks: energy efficiency, scalability, adaptability, low latency, and reliability. Can you describe a concrete WSN application for each of these five requirements, where the requirement would be more important than the others?

Energy efficiency is most important in WSN applications with battery-powered sensors that must operate without human intervention for an extended period of time. For example, sensor networks that detect wild fires or monitor civil infrastructure must be able to operate for months or years before their power sources can be replaced or recharged. Scalability is important in sensor networks that cover large geographic areas, e.g. a sensor network measuring the flow of lava on a volcano or in many military applications. Adaptability is crucial in networks where changes in network topology, density, and traffic are common, including vehicular sensing applications and other applications where sensors are mobile. Low latency is important when sensor data must be reacted upon quickly, e.g. in networks that detect explosions, the presence of toxic fumes, or the impending collapse of a bridge. Finally, reliability is essential in networks where the loss of important sensor data can be catastrophic. For example, in military sensor networks, sensor data reporting enemy movements must reliably reach the command and control center.

6.17 The TRAMA protocol is an example of a contention-free MAC scheme. Answer the following questions about TRAMA.

(a) What are the advantages and disadvantages of the TRAMA protocol (compared to contention-based protocols)?

TRAMA reduces the probability for collisions and dynamically determines when a node is allowed to transmit (based on traffic), thereby increasing the throughput. Since TRAMA uses a time-slotted channel, it allows nodes to determine when they must stay awake and when they can enter low-power sleep modes. Nodes must be awake during the random-access intervals. As with other contention-free protocols, communication in TRAMA may experience larger latencies than in contention-based protocols since a node must wait for its scheduled slot before it can begin transmission.

(b) What is the difference between transmission slots and signaling slots?

The time slots of the scheduled-access intervals are assigned to individual nodes (for contention-free transmission) while signaling slots are used for random medium access (contention-based communication). Nodes can join a network by transmitting during randomly selected slots in the random-access intervals.

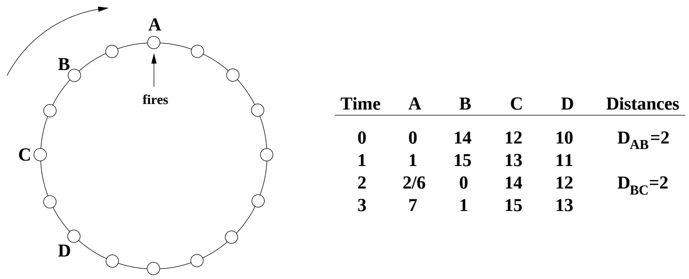
(c) What is the purpose of the NP component?

The Neighbor Protocol (NP) is responsible for propagating one-hop neighbor information among neighboring nodes, allowing these nodes to obtain consistent two-hop topology information.

6.18 What is the advantage of a receiver-initiated MAC scheme such as Y-MAC? What is the main disadvantage of Y-MAC that makes it unsuitable for most low-power and low-cost sensor nodes?

The main advantage of a receiver-initiated approach is that a node wakes up for a brief period of time to sample the medium for any transmissions and returns quickly to its low-power sleep mode if no traffic has been detected. Y-MAC further uses multiple channels (to increase throughput and to reduce latency). However, for most low-power sensor networks, it is currently infeasible to have multiple radios on each sensor node, because of the increased device cost and the increased energy consumption of a multi-radio design.

6.19 Demonstrate the concept of DESYNC using the example ring shown in Figure **??**. The ring has 16 positions [0..15], with node A currently in position 0 (the firing position), B in position 14, etc. Every unit of time, each node moves one position clockwise along the ring. The table indicates the positions of the four nodes, including the new distance information that is learned at each firing. Assume that node A has received D's last firing, indicating a distance of 10 between nodes A and D. In this table, at time 0, node A fires, allowing node B to learn its distance to A (i.e. 2). At time 1, no node is in the firing position. At time 2, node B fires, allowing node C to learn its distance to node B (i.e. 2). At the same time, node A now knows its distance to node B and its distance to node D. According to the description of the DESYNC algorithm in this chapter, node A can now find the midpoint between nodes B and D and jump to this new location (i.e. 6), which is indicated in the table at time 2. At time 3, again each node moves ahead one position. Continue this table using the DESYNC algorithm until time 19. Compare the average distance between neighboring nodes at time 19 with that of time 0.



Time	A	B	C	D	Distances
0	0	14	12	10	D_{AB}=2
1	1	15	13	11	
2	2/6	0	14	12	D_{BC}=2
3	7	1	15	13	

Figure 6.3 DESYNC ring (Exercise 6.19)

Each time a node learns about its distances to its immediate neighbors to the left and to the right on the ring (e.g.: D_{left} and D_{right}), it can compute its new position P_{new} as:

$$P_{new} = P_{old} = \frac{1}{2}(D_{right} - D_{left}) \tag{6.1}$$

The nodes' positions until desynchronization are shown in Table **??**. Every time a node's

Table 6.2 Node positions and learned distances (Exercise 6.19)

Time	Node A	Node B	Node C	Node D	Distances
0	0	14	12	10	$D_{AB} = 2$
1	1	15	13	11	
2	2/6	0	14	12	$D_{BC} = 2$
3	7	1	15	13	
4	8	2/2	0	14	$D_{CD} = 2$
5	9	3	1	15	
6	10	4	2/2	0	$D_{AD} = 6$
7	11	5	3	1	
...	...	...	...	...	...
12	0	10	8	6/4	$D_{AB} = 10$
...	...	...	...	...	...
18	6/4	0	14	10	$D_{BC} = 2$
19	5	1	15	11	

neighbor to the left (i.e. the following node) fires, the node can compute its new position. Such position changes are indicated as x/y in the table, where x is the old position and y is the new position. At time 0, the distances between neighboring nodes are 2, 2, 2, 10, respectively. That is, the average deviation from the target distance ($16/4 = 4$) is then $3 * (4 - 2) + (10 - 4) = 12$. At time 19, the distances are 4, 2, 4, and 6, respectively. That is, while the ring is still not fully desynchronized, the average deviation has been reduced to $(4 - 4) + (4 - 2) + (4 - 4) + (6 - 4) = 4$.

6.20 Discuss the cluster head election policy in the LEACH protocol and explain how LEACH can consider available energy on each node in this election process. What is the problem with this energy-aware election policy? Further, LEACH uses TDMA within a cluster; explain the advantages and disadvantages of this approach.

In the LEACH protocol, a node independently decides whether it becomes a cluster head or not. This decision is based on how long it has been that it has served as cluster head, i.e. a node is more likely to become a cluster head if it has not assumed this responsibility for a long time. The goal of this job rotation is to balance the workload among nodes and to prevent that a cluster head depletes its battery prematurely. Equation 6/6 in the book shows an approach to balance the cluster head responsibility among nodes based on the actual current energy levels of the nodes. The problem with this approach is that each node must know (or at least estimate) the sum of all nodes' energy levels. Using TDMA within a cluster provides contention-free communication between sensor nodes and cluster heads. To limit interference among clusters, LEACH uses direct sequence spread spectrum techniques. Using TDMA within a cluster has the "usual" advantages and disadvantages (e.g. with respect to latency, throughput, duty cycles) compared to contention-based techniques.

6.21 What does LEACH use the direct sequence spread spectrum technique for?

While collisions with a cluster are avoided using TDMA, communications within one cluster can still collide with communications in a neighboring cluster. Therefore, DSSS techniques are used, i.e. a cluster uses a spreading sequence that differs from the spreading sequences used in neighboring clusters.

6.22 How does the Mobile LMAC protocol handle changes in network topology?

In MLMAC, when a node leaves the radio range of another node, both nodes will realize that they do no receive any control messages from each other. As a consequence, they will remove each other from their neighbor lists. If a node X moves into the radio range of another node Z, node X's transmissions may collide with another node's transmissions to node Z. If this happens, node Z will mark the corresponding slot as unused. Node X will receive a control message from node Z, indicating that its slot is unused and node X will restart its slot selection mechanism.

6.23 Discuss why overhearing is a problem in a wireless sensor network and explain how PAMAS addresses this problem.

Overhearing can be a source of energy waste in a wireless sensor network when a node overhears and drops a packet that is not destined to this node. The node has to expense power to receive and decode the message unnecessarily. In PAMAS, when a node overhears a packet not sent to it, the node can power off its transceiver for the duration of the transmission (which can be embedded into the overheard RTS message).

6.24 Explain how the busy-tone scheme of PAMAS helps to avoid the hidden-terminal problem.

Typically, nodes overhear either RTS or CTS messages that indicate that another node is transmitting. However, it is possible that a node did not hear either message and therefore access the wireless medium, interfering with the ongoing transmission. In PAMAS, a busy tone is transmitted over the control channel during a transmission to indicate that the channel is busy and to prevent other nodes from accessing the medium.

6.25 How does the S-MAC protocol reduce the duty cycles of sensor nodes? How does the S-MAC protocol attempt to reduce collisions? How does it address the hidden-terminal problem? Name at least three disadvantages of the S-MAC protocol.

S-MAC uses virtual clusters, where nodes using the same sleep-wake schedule belong to the same cluster. Schedules are exchanged via SYNC messages to allow each node to learn about its neighbors' schedules. Transmissions are initiated using RTS messages and a node winning the medium can begin its transmission to the intended receiver, while all other nodes can return to the low-power sleep mode. Collisions are reduced by using a slotted approach, where a node randomly selects a time slots for transmission. The RTS/CTS approach also helps to address the hidden-terminal problem. S-MAC is a contention-based approach (with the same disadvantages as most other contention-based mechanisms), broadcast packets do not use RTS/CTS, which may lead to collisions among them, and duty cycle parameters (such as the lengths of the sleep and active periods) are fixed and may be inefficient.

6.26 Which shortcoming of S-MAC does T-MAC address? Explain briefly T-MAC's ability to adapt to traffic density.

S-MAC uses a fixed-size listening period, whereas T-MAC uses an active period that adapts to traffic density. In T-MAC, nodes awake briefly to listen for activity, but return quickly to the sleep mode if no activity is detected. Nodes transmitting, receiving, or overhearing a message, on the other hand, stay awake for a brief timeout period after the transmission has finished to see if more traffic can be observed. That is, when traffic is heavy, nodes will stay awake longer, but if traffic is light, they only awake for very brief periods of time.

6.27 What is the "early sleeping problem" and how does T-MAC address this problem?

The early sleeping problem is shown in Figure 6.11 in the book. The problem is that a node losing the medium, but wishing to transmit after the current transmission has completed will stay wake, while its intended receiver may be unaware of the node's intent and will go to sleep. Therefore, the future-request-to-send technique adds another control message, the FTRS packet, which is sent by the node planning a future transmission after it has observed the CTS from the receiving node of the current transmission. This FTRS packet will be seen by its intended receiver.

6.28 Describe the concept behind PMAC's approach to adapting a node's sleep durations to observed traffic.

In PMAC, nodes use "patterns" to describe their tentative sleep and awake times, where a pattern is a bitmask, each bit representing a time slot (a '0' bit indicates a sleep interval and a '1' bit indicates an awake interval). Further, "schedules" are used to describe actual sleep/awake sequences, where a schedule is a series of '0's, followed by a '1'. Comparable to TCP's slow start behavior, PMAC doubles the sleep interval whenever it has no data to send until a certain thresholds of '0' has been reached, after which the '0's are increased linearly. However, once a node has data to transmit, the pattern is immediately reset to 1, allowing the node to wake up quickly to handle its traffic load.

6.29 The use of duty cycles allows nodes to alternate periods of activity and low-power sleep intervals. However, this often introduces large communication latencies. In Figure **??**, node A wishes to send a packet to node F using the route A-B-C-D-E-F. Assume that the interval between two dashed vertical lines is one unit of time and each transmission requires exactly one unit of time of overlapping periods of activity of two neighboring nodes. The first transmission between A and its neighbor is already shown in the figure. Complete this graph to determine the end-to-end latency experienced by the packet. Further, explain how RMAC reduces these end-to-end latencies and what would this latency be in an RMAC network? Finally, explain how the RMAC protocol reduces collisions?

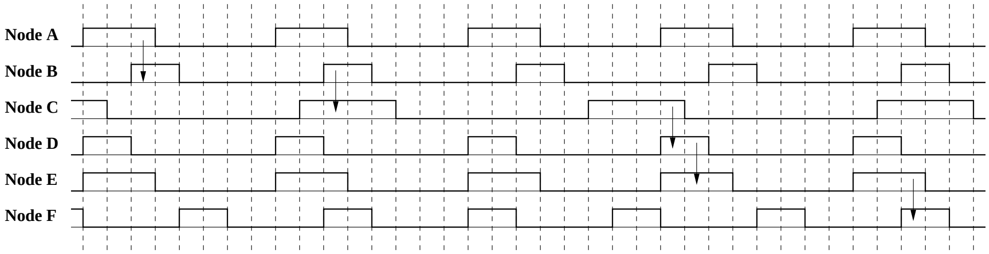


Figure 6.4 RMAC duty cycle pattern (Exercise 6.29)

Figure **??** shows the completed transmission sequence with an end-to-end latency of 33 time slots. RMAC aligns the sleep/wake periods of nodes along the path of sensor data such that data can be forwarded quickly. RMAC uses control frames that are sent along the route to inform nodes of an upcoming packet. Using this information, a node can then determine when to be awake to receive and forward the packet. In Figure **??**, the latency could therefore be shortened to 5 time slots. With respect to collisions, RMAC separates medium contention and actual data transfer into separate periods.

6.30 For what type of WSN applications would you use the DMAC protocol?

The DMAC (Data Gathering MAC) protocol is optimized for sensor networks that rely on convergecast communications, i.e. data from the sensor nodes travels upward to the root of a data gathering tree. Such a tree could be used by static sensor networks, i.e. networks with little to no sensor node mobility to avoid frequent tree adaptations. Further, the DMAC protocols works best in scenarios where the traffic rates are known and stable.

6.31 Explain the problem of “idle listing” and describe how preamble sampling addresses this problem. How does WiseMAC improve upon “standard” preamble sampling?

Idle listening refers to a sensor node listening to a wireless channel to receive possible traffic, which leads to energy waste if the idle listening periods are large. The idea behind preamble sampling is that a device sends a preamble preceding the actual data to alert the receiving node of the incoming data. Sensor nodes periodically wake up to see if they can detect such preambles and return to the sleep mode if none are detected. Nodes detecting a preamble must stay awake. WiseMAC tries to further reduce the duration of the idle listening periods by letting a transmitter learn the sampling schedules of its receiver(s). Knowing the exact time a receiver awakes allows a transmitter to start the transmission of the preamble exactly when the receiver wakes up.

6.32 What is the advantage of having the receiver (instead of the sender) in control over the timing of transmissions (e.g. as in the RI-MAC protocol)? How does the RI-MAC protocol handle multiple contending senders?

When the receiver is in control when to receive data, it can minimize its awake time by indicating that it is ready for incoming data immediately after waking up. The receiver will have very little overhead due to overhearing. Collisions are detected at the receiver node and the receiver responds by issuing another beacon which indicates a window over which the contending senders must select their backoff values. The senders then retry after their chosen backoff values expire.

6.33 Explain how the Z-MAC protocol allows nodes to determine their own local time frames instead of using a single global time frame. What are the disadvantages of Z-MAC?

In Z-MAC, a node learns about its 1-hop and 2-hop neighbors during a setup phase and uses this information to select a time slot that only belongs to that node within its 2-hop neighborhood. Based on the neighborhood information, a node can then determine the periodicity of its assigned slot, i.e. the time frame to be used in its neighborhood. The consequence is that in denser neighborhoods (nodes with many neighbors), the time frame will be large, while in neighborhoods with low density the time frames may be very small. The main disadvantage of Z-MAC is that it requires an explicit setup phase, consuming both time and energy. Also, Z-MAC's ECN (explicit contention notification) messages add to the traffic overheads.

7

Network Layer

7.1 The previous chapter talked about MAC protocols, while this chapter introduced routing protocols. Can you think of examples how the choice of MAC protocol affects the design, performance, and efficiency of the routing protocol?

The choice of MAC protocol can have several impacts on the design or performance of a routing protocol. For example, communication latencies are determined by the type of MAC protocol (contention-free versus contention-based) and on the use of duty cycling at the link layer. The MAC protocol is responsible for error recovery due to collisions and other interferences, i.e. the routing layer may decide to rely on the MAC layer for reliable communications or it may implement its own mechanisms to ensure reliability. Some approaches discussed in the chapter on MAC protocol rely on a tight integration of MAC and routing protocols, e.g. the LEACH protocol (which uses clusters that affect both medium access and routing decisions) or the DMAC protocol (which uses fixed routes from sensor nodes to a sink).

7.2 What is the difference between a proactive routing protocol and a reactive routing protocol? Name at least two examples for each category.

A proactive routing protocol establishes and maintains routing information (typically in the form of tables) before it is actually needed. This allows a node to quickly make routing decisions. Examples include DSDV and OLSR. In contrast, a reactive routing protocol only establishes and maintains a route when it is actually needed, i.e. on-demand. This is typically done using an explicit route discovery process. Examples of such protocols include AODV and DSR.

Consider the following WSN scenarios and explain why you would choose either a proactive or a reactive routing solution:

(a) A WSN is used to monitor air pollution in a city where every sensor reports its sensor data once every minute to a single remote base station. Most sensors are mounted on lamp posts, but some are also mounted on city buses.
Because communication is frequent, a proactive approach may be appropriate. The routing protocol will ensure that table entries are updated whenever the topology changes, which are infrequent events since most radios are mounted on lamp posts.

(b) A WSN is used to measure humidity in a field, where low-power sensor report their measurements only when certain thresholds are exceeded.

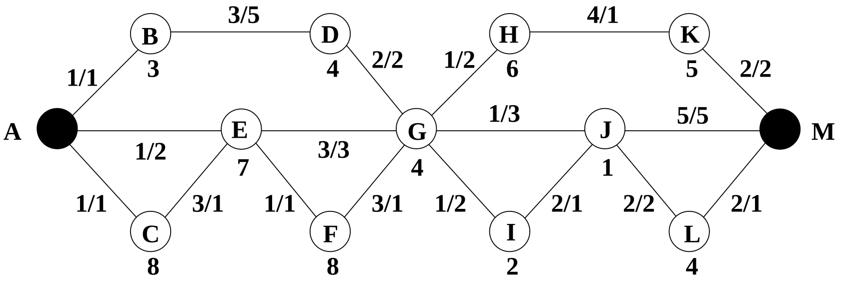


Figure 7.1 Topology for Exercise 7.5

In this example, sensor communications will be infrequent, therefore routing information will only be needed occasionally. The cost of maintaining routing tables may cost too much energy, therefore a reactive solution may be more appropriate. The type of sensor data (humidity) indicates that latency is not a concern, which also justifies the use of a route discovery phase.

(c) A WSN is used to detect the presence of vehicles, where each sensor locally records the times of vehicle detection. These records are delivered to the base station only when the sensor is explicitly queried.

Again, communication in such a network is infrequent and a sensor can establish a route whenever needed (i.e. when a query is issued) instead of continuously maintaining accurate table entries.

7.3 What is data-centric routing? Why is data-centric routing feasible (or even necessary) compared to routing based on identities (addresses)?

If the focus is on the data generated by sensors and not the identity of the sensors generating these data, then we call a routing protocol data-centric. For example, if sensor data describing the environmental temperature must be routed to all sinks interested in such data, irrelevant of the data originator (i.e. which sensors produced the data), the focus is on the gathered data itself.

7.4 Describe a WSN application for each of the following categories: time-driven, event-driven, and query-driven.

An example of a time-driven WSN application is environmental monitoring, where temperature, pressure, or humidity may be reported periodically to obtain statistics for weather and climate studies. An example of an event-driven WSN application is a network that detects and reports wild fires, i.e. sensor data is only generated and disseminated when events of interest occur. A query-driven WSN application could be a surveillance system, where law enforcement officers access sensors to detect or track a suspect in public places.

7.5 For the network topology shown in Figure **??**, identify the optimal routes for source A to sink M according to the following criteria (describe how you compute the cost for the

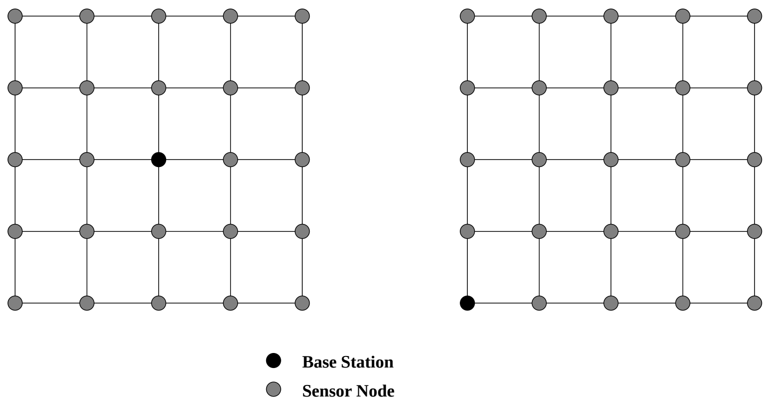


Figure 7.2 Topologies for Exercise 7.6

optimal route). The numbers X/Y along each link indicate the latency (X) and energy cost (Y) for transmitting a single packet over the link. The number Z under each node indicates the node's remaining energy capacity.

(a) minimum number of hops

The minimum hop route is A-E-G-J-M, which requires 4 hops (3 relay nodes).

(b) minimum energy consumed per packet

The route with the minimum energy consumer per packet is route A-E-F-G-H-K-M. The energy cost for this transmission from A to M is 2+1+1+2+1+2=9.

(c) maximum average energy capacity (eliminate hops that would result in a higher average but unnecessarily add to the route length!)

The route with the maximum average energy capacity would be A-E-G-H-K-M. The maximum average energy capacity for this route = (7+4+6+5)/4 = 5.5.

(d) maximum minimum energy capacity

The route with the maximum minimum energy capacity would be A-E-G-H-K-M (the same as the maximum average energy capacity). The minimum energy capacity on this route is 4 (node G).

(e) shortest latency

The shortest latency route would be A-E-G-J-L-M and the total latency for this route is 9.

7.6 A WSN is modeled as a 5x5 grid as shown in Figure **??**, with the base station placed at the center of the network (left topology) or at the bottom left corner (right topology). Assume that each node can communicate with only its immediate neighbors on the grid and that packet transmission or forwarding over a link costs exactly one unit of energy (packet reception and processing costs are neglected).

(a) For both topologies, find an energy optimal graph of routes, i.e. the energy cost for each packet traveling through the network is a minimum.

Figure **??** shows possible energy optimal routes for both network topologies.

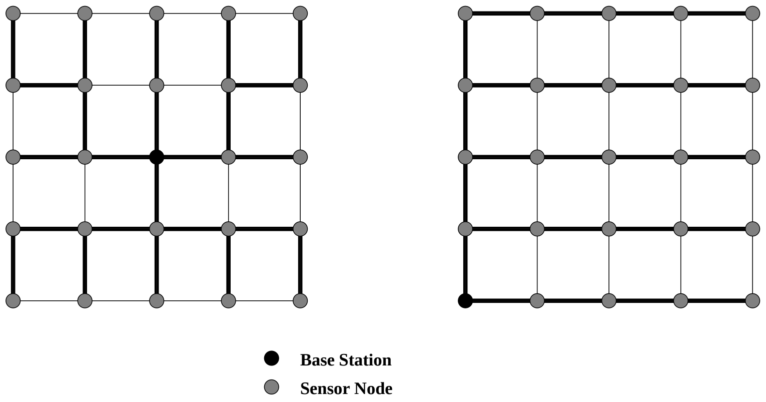


Figure 7.3 Result for Exercise 7.6

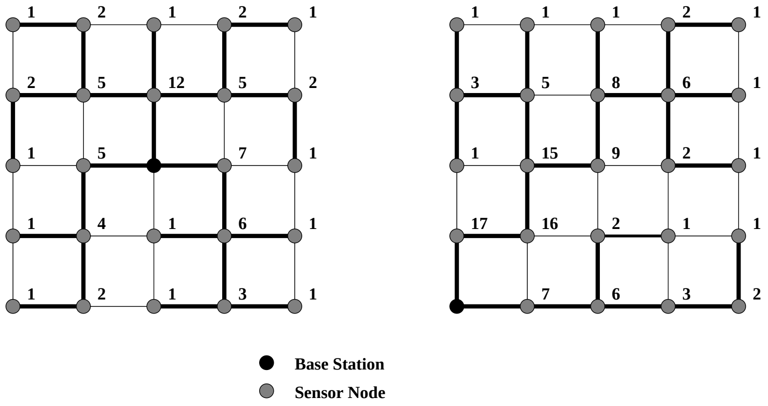


Figure 7.4 Topologies and routes for Exercise 7.6

(b) Consider the graphs shown in Figure **??**. What is the average and total load in the network, when the per-node load is defined as the number of routes a node has to service (including its own)? Do not include the base station in your calculations.

Figure **??** has been modified to also indicate the load for each node. The total load in the first topology is then the sum of node loads, i.e. 68. The average load is 68/24 = $2.8\dot{3}$. In the second topology, the total load is 112 and the average load is then 112/24 = $4.\dot{6}$.

(c) What is the lifetime of the network topologies in Figure **??** when during every second, each node generates and transmits its own packet and forwards all packets received during the previous second? Assume that each node has an initial energy budget of 100. Each transmission costs 1 unit of energy (there is no cost for reception, etc.). Consider the lifetime of a network to have expired once the first node depletes its energy budget. Compare the results and derive design principles for the network topology to optimize the lifetime of the network with respect to placement of the base station and the construction of routing trees.

In each topology, it is sufficient to focus on the node with the largest load, e.g. in the first topology, this node is the one right above the base station (load 12). During the first second, this node only transmits its own packet, i.e. its energy has reduced to 99. During the second interval, the node makes 4 transmissions (its own packet and the packets from its immediate neighbors), reducing the energy capacity to 95. During the third second, the node transmits its own packet and the packets from its 1-hop and 2-hop neighbors, reducing the energy to 87. Starting in the fourth second, the node reaches its maximum load, i.e. every second its energy capacity reduces by 12. That is, this node will expire during the eleventh second.

In the second topology, the focus is on the node right above the base station. Applying the same method as used for the first topology, this node will also expire during the eleventh second.

Based on these results, it becomes clear that nodes serving as relays for other nodes have to carry a significant burden in terms of overheads and energy consumption. These nodes are typically closer to the base station. One design principle could be to “balance” the forwarding responsibilities among as many nodes as possible to avoid bottlenecks, i.e. nodes that have to carry much more traffic than other nodes in the network. Further, the less connected the base station is, the more load its neighbors must carry (as shown in the second topology). Therefore, it is advantageous to ensure that the base station is placed in a dense part of the network.

(d) Assume that the first topology in Figure **??** is used and each sensor transmits exactly one packet to the base station. Then the topology is switched to the second one and each sensor transmits one packet to the base station in the bottom left corner. Then the topology is switched back to the first one and the process is repeated. Explain why the network lifetime changes and what other design principle can be derived from this insight. (To facilitate the comparison, focus on the case where each node has already reached its maximum load.)

In the first topology, the “critical node” (right above the base station) carries a load of 12 per second, but only a load of 8 in the second topology. That is, its load will vary between change between 12 and 8, i.e. the average load is 10. That is, compared to the scenario in the first topology, its load has been reduced by 2 every second. Similarly, the "critical node" in the second topology now switches between loads 17 and 1, thereby resulting in an average load of 9. A possible design guideline could be that the presence of multiple base stations or mobility of a base station can be advantageous in terms of traffic load and network lifetime.

7.7 Flooding is a simple strategy for distributing data to one specific or all sensor nodes in a network. Answer the following questions:

(a) Explain the three challenges of flooding described in this chapter.

A node receiving a packet forwards this packet to all neighbors regardless of whether these neighbors have already received a copy of this packet. This is known as the implosion problem. The redundancy in sensor data from sensors that monitor similar physical environments is known as the overlap problem. Finally, flooding does not consider the resources available on all nodes, i.e. it is resource-blind.

(b) Which one(s) of these are addressed by gossiping and how are they addressed?

Gossiping addresses only the implosion problem, i.e. it uses a probabilistic approach to decide whether to forward data or not.

(c) For the topologies shown in Figure **??** and Figure **??**, what are good choices for the maximum hop count? Explain your answer.

In both networks, a good choice for a maximum hop count would be 8. The maximum hop count should be chosen large enough that every node can reach any other node in the network, but not too large to avoid packets from traveling too long (e.g. particularly in networks where a packet could travel in circles).

(d) How do sequence numbers contribute to reducing unnecessary transmissions? Are sequence numbers alone sufficient and if not, what other information is needed to use them correctly?

Sequence numbers allow nodes to distinguish packets from each other, i.e. to identify new packets or duplicates. Sequence numbers alone are typically not enough, they need to be combined with other information, such as the identity of the sender. Only then can a packet be identified uniquely.

7.8 Using the topology in Figure **??**, explain the problems of implosion, overlap, and resource blindness.

The implosion problem can be illustrated using nodes C and E, assuming that node A has initiated a flooding process. Both C and E will get the packet directly from A, but they will forward this packet to each other, unaware that they already have a copy. The overlap problem refers to nodes collecting overlapping sensor data, i.e. this typically means that these sensors are in close physical proximity. For example, the data collected by nodes C, E, and F could have some overlap. Finally, flooding is resource-blind, i.e. it will rely on nodes J and I to participate in the flooding process, even though their energy capacities are very low.

7.9 How does the SPIN family of protocols address the three challenges faced by flooding? What are the disadvantages of a negotiation-based protocol such as SPIN?

Table 7.1 G's routing table (Exercise 7.11)

Event	Distance	Direction
E1	3	F
E2	4	I

SPIN uses two techniques to address these challenges. First, to address implosion and overlap, SPIN nodes negotiate with their neighbors before they transmit. Second, to address resource blindness, SPIN uses a resource manager on each node, which keeps track of actual resource consumption and allowing the nodes to adapt routing and communication behavior based on resource availability. A problem with the negotiation-based approach is that it introduces a significant amount of control overhead, i.e. the cost of extra control messages could outweigh the benefits of SPIN in some scenarios. The negotiation-based approach also introduces additional delays into the communications.

7.10 Explain the concept of directed diffusion. Can you imagine at least three strategies or goals for reinforcement?

In directed diffusion, nodes request data by sending interests for named data to other nodes in the network. These interests are disseminated throughout the network and gradients are established that are used to direct sensor data back to the originator of the interest. Reinforcement is used to strengthen some gradients in the network based on some goals or metric. For example, a node could reinforce a neighbor from which the sink has received a previously unseen event, a neighbor that appears to have large energy capacity, or a neighbor that shares a very reliable link with this node.

7.11 Consider the network topology in Figure **??** and node G's routing table shown in Table **??**.

(a) Describe how node G would send queries towards events E1, E2, and E3 using rumor routing (note that node G does not have any routing table entries for event E3).

For events E1 and E2, G would forward the query directly to the next-hop neighbors indicated in its routing table, i.e. F and I, respectively. Since no routing entry for E3 exists, G will select a random neighbor to forward query E3, e.g. it will chose among neighbors D, E, F, H, I, and J.

(b) Assume that (i) I informs G that I can reach event E2 via 2 hops, (ii) F informs G that F can reach event E3 via 4 hops, (iii) E informs G that E can reach event E1 via 1 hop, (iv) D informs G that D can reach event E1 via 2 hops, (v) H informs G that H can reach event E2 via 2 hops, and (vi) D informs G that D can reach event E3 via 1 hop. What is the final table of node G? Can you identify the locations of all three events by the identity of the closest sensor?

The new routing table is shown in Table **??**. Based on the given information, E1 can be reached via F using 3 hops, via E using 2 hops, and via D using 3 hops, i.e. E1 appears to be near node A. E2 can be reached via I using both 3 and 4 hops and via H using 3 hops, i.e. E2 appears to be near M. Finally, E2 can be reached via F using 5 hops and via D using 2 hops, i.e. E3 appears to be near B.

Table 7.2 G’s new routing table (Exercise 7.11)

Event	Distance	Direction
E1	2	E
E2	3	I
E3	2	D

7.12 What are the concepts behind distance vector routing and link state routing and how do they compare to each other with respect to overheads for maintaining routing tables?

In algorithms based on distance vector, every node maintains a list of distances for each possible destination, including a neighbor which can be used to reach that destination. This information is stored in a routing table. Each node then broadcasts updates to the routing table periodically or whenever significant new information becomes available. DSDV uses two types of packets: a full dump contains the content of the entire routing table and an incremental packet contains only new information. In link state routing, each node periodically broadcasts its neighbor information using HELLO messages. These messages are flooded throughout the network so that every node can determine network-wide topological information. In distance vector, information is exchanged among neighbors only, but the amount of information exchanged can be significant in the case of a full dump. Updates in link state routing are flooded throughout the network, but these updates are typically small in size.

7.13 Compare a proactive routing protocol such as DSDV with a reactive protocol such as DSR with respect to overheads and route optimality.

A proactive routing protocol establishes routing tables that must be continuously maintained to prevent stale information. The costs of these table updates can be extensive. Further, these tables may contain many routes that will never be used (or at least not for a long time). In contrast, a reactive protocol only establishes a route when it is actually needed. This can reduce the overheads significantly. However, the process of establishing a route may be expensive since many protocols flood the entire network to find a route. The route discovery process also introduces a delay, i.e. data can only be transmitted towards the destination once the route discovery has completed. If a proactive protocol collects information from each node in the network, it is straightforward to identify the optimal route (e.g. in terms of number of hops or energy capacity). In reactive protocol, the optimal route can be found if the entire network is flooded with route discovery messages and it is possible that only a suboptimal route is discovered if some route discovery messages are lost.

7.14 Does DSR incur larger or smaller overheads for route discovery compared to the AODV protocol? Justify your answer.

The route discovery messages of DSR are typically larger than the messages in the AODV protocol since each forwarder node inserts its own information into the packet (to allow the protocol to collect entire paths within route discovery messages).

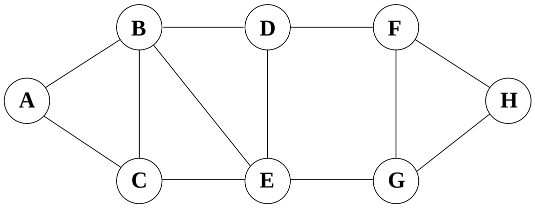


Figure 7.5 Topology for Exercise 7.18

7.15 In AODV, is it possible that route discovery packets travel in the network forever? Why or why not?

Each node is identified by its source and a sequence number (broadcast ID), allowing forwarding nodes to discard duplicate packets. Further, most routing protocols use time-to-live values (maximum number of hops a packet travels) to prevent a packet from traveling in the network for too long.

7.16 Asymmetric (or unidirectional) links occur when node A ca‘n hear node B, but B cannot hear node A. Explain whether this is a problem for the AODV protocol and if so, how this can be addressed.

This is a problem because the RREP message for the route discovery is sent back the exact same path taken by the RREQ that reached the destination and triggered the RREP message. As a consequence, the RREP may never reach the source if a link on the route is asymmetric. This could be addressed by enhancing the HELLO message approach, i.e. every node inserts its list of neighbors into its HELLO messages. When node A receives a HELLO message from node B, where node A is not in B’s list of neighbors, node A knows that the link from B to A is asymmetric. In such a case, node A could ignore RREQ messages coming from B to ensure that the asymmetric link B to A will never be used for a route that includes the non-existing link A to B.

7.17 What is the concept behind hierarchical routing and what advantages does it have over other techniques?

In hierarchical routing protocols, nodes are grouped together into clusters, where sensors communicate only directly with a leader node (cluster head). These cluster heads are often less resource-constrained than regular sensor nodes. This approach can reduce the energy and communication overheads for sensor nodes and it is also easier to regulate medium access within a cluster (e.g. a sensor could transmit only when queried by the cluster head).

7.18 Table **??** summarizes the routing information of all nodes in a WSN, i.e. each row indicates the routing knowledge of that particular node. For example, the first row shows that node A knows that it can reach nodes B and C via 1 hop and nodes D and E via

2 hops. Given this information, draw the network topology and determine the landmark radius for each node.

Table 7.3 Routing information for Exercise 7.18

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Landmark Radius**
A	0	1	1	2	2	-	-	-	1
B	1	0	1	1	1	2	-	-	2
C	1	1	0	2	1	-	2	-	2
D	-	1	2	0	1	1	2	2	2
E	2	1	1	1	0	-	1	-	1
F	-	2	-	1	2	0	1	1	1
G	-	2	2	2	1	1	0	1	2
H	-	-	3	2	-	1	1	0	1

Figure **??** shows the network topology that corresponds to Table **??**. The landmark radius is shown in the last column of Table **??**.

7.19 What is the advantage of using fisheye state routing in the LANMAR protocol compared to the basic landmark routing technique?

Fisheye state routing is a link state protocol, where the frequency of route updates depends on the distance, i.e. routes within a fisheye scope (a certain predefined distance) are more accurate than routes to more distant nodes. In LANMAR, routing updates are only exchanged with nodes in the immediate neighborhood and with landmark nodes. When a node needs to relay a packet, the packet will be forwarded directly to the destination if the destination is within the node's fisheye scope. The use of fisheye state routing reduces the overhead of maintaining routing tables.

7.20 Figure **??** shows a number of nodes as small dots. Each node has a radio range of 2 units. How would the gray node positioned at (0,0) route a packet to the gray node at position (9,9) using GPSR? Indicate the visited nodes.

The right graph in Figure **??** indicates the chosen route. The node located at position (4,4) uses the perimeter routing rule.

7.21 When does GPSR enter the perimeter routing mode and how does it use the right hand rule in this mode?

In "normal" GPSR mode, a forwarding node chooses the one neighbor that makes the largest progress towards the destination. Since this decision only considers the immediate neighbors of a node, it can happen that a packet reaches a void. GPSR uses a mechanism to route the packet around the void based on the right hand rule, i.e. when a packet arrives at node x from node y, the next edge traversed is the next one sequentially counterclockwise about x from the edge (x,y). That means that the rule traverses the interior of a polygon in clockwise edge order.

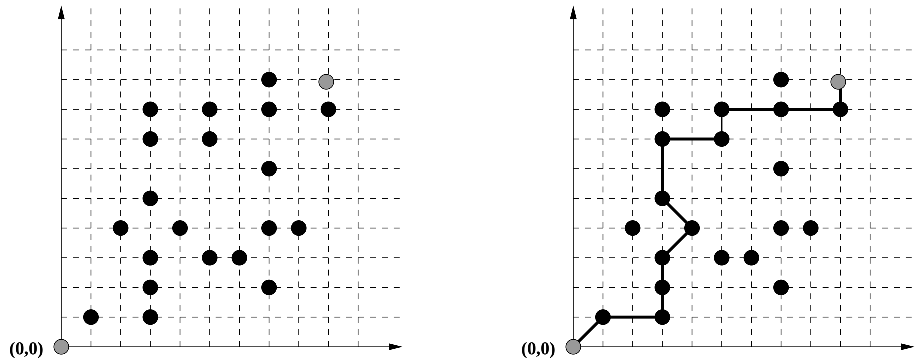


Figure 7.6 GPSR routing example (Exercise 7.20)

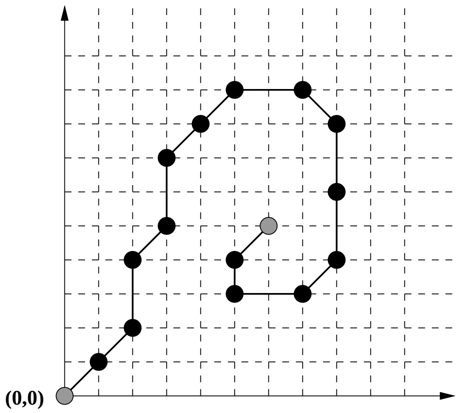


Figure 7.7 GPSR perimeter routing (Exercise 7.22)

7.22 Prove that it is false or show an example that the perimeter mode can cause a packet to traverse a network's entire outer boundary.

Figure **??** shows an example where several nodes route a packet around the outer boundary to handle voids in the network (assuming that each radio has a transmission range of 2 units).

7.23 Consider the topology in Figure **??**. Node A wishes to forward a packet towards destination L via one of its neighbors (its communication range is indicated with the circle). Which neighbor will A choose with each of the following forwarding strategies:

(a) greedy forwarding

G minimizes the distance to the destination L.

(b) nearest with forwarding progress

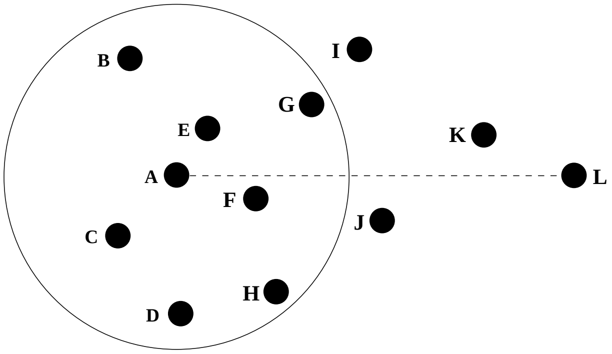


Figure 7.8 Forwarding strategies in GPSR (Exercise 7.24)

Node E makes positive progress towards L, but is also the nearest neighbor.

(c) most forwarding progress within radius

G also makes the most forwarding progress within A's transmission radius.

(d) compass routing

F is A's neighbor with the smallest angle between the line connecting the source and the destination and the line connecting the source with the neighbor.

7.24 The cell size of the GAF virtual grid can be pre-determined and each node knows to which cell it belongs. Discuss the consequences of choosing very large versus very small cell sizes.

An assumption of GAF is that all nodes within a cell can communicate directly will all other nodes in neighboring cells. Due to the transmission range limits of sensor nodes, this automatically puts a limit on the maximum size of a cell. The large the cell size, the larger the forwarding responsibility of the active node, leading to large energy consumptions for that node. Smaller cell sizes more evenly distribute this responsibility among more nodes. Very small cell sizes mean that more nodes in the network are active nodes and fewer nodes can be in the sleep mode.

7.25 How does the SPBM protocol ensure efficient multicast for large numbers of receivers?

SPBM uses a hierarchical group management scheme to maintain a list of all destinations for a particular packet. A network is represented as a quad-tree with a predefined number of levels and each node maintains two tables: a global member table, with entries for the three neighboring regions in each level, and a local member table, containing all members of the node's neighbors in level 0 (all nodes within level 0 are within each others' transmission ranges). A bit mask is used to encode multicast memberships such that a node can easily identify the regions of the network that contain multicast members for a particular packet. Packets are then forwarded towards the regions (instead of individual destinations), where each node independently decides whether to split a packet based on

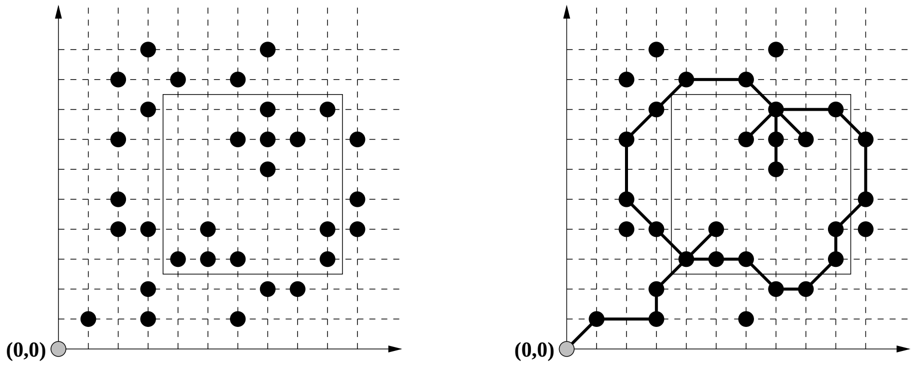


Figure 7.9 Geocast region with hole (Exercise 7.28)

characteristics such as the total number of nodes forwarding the packet and the optimality of the individual routes towards the destinations.

7.26 What is the concept of RBMulticast and how does it address the shortcomings of the SPBM protocol?

RBMulticast is a receiver-based multicast approach, i.e. a sender can transmit a multicast packet without specifying the next-hop node. RBMulticast also uses multicast regions (similar to SPBM), but it is a completely stateless protocol, i.e. it does not need to maintain membership tables. Instead, each region is represented by a virtual node and each forwarding node replicates a packet for each region that contains at least one multicast member. However, RBMulticast relies on an underlying MAC protocol, where receivers contend for channel access and nodes that make the most forward progress to the destination will contend earlier, thereby having a better chance to become the next-hop node.

7.27 The GEAR protocol uses two types of costs: learned and estimated. Explain how learned costs are used to route packets around holes (use a concrete example). What is the purpose of the estimated costs and what is the intuition behind calculating them as described in this chapter?

Figure 7.18 in the book shows a concrete example of learning routes around holes. When a node A receives a packet and if there are no neighbors that are closer to the destination, A knows that it is in a hole. It then uses the learned cost function to select one of its neighbors as the next hop (i.e. the node with the minimum learned cost). After forwarding the packet to node B, it sets its own learned cost to $h(B, R) + C(A, B)$, i.e. it takes the learned cost of B for region R and adds the cost of transmitting the packet from A to B, leading to an increased learned cost (allowing upstream nodes to avoid forwarding packets towards the hole).

7.28 Figure **??** shows a sensor network topology, where each node's transmission range is two units. The node at position (0,0) wants to disseminate a packet to all nodes within the rectangle. Show how GFPG routes the packet towards the region and how it distributes it to all receivers within the rectangle. Clearly indicate which nodes (inside and outside the geocast region) will receive the packet.

The right graph in Figure **??** show how the packet would travel inside and around the target region (redundant or unnecessary transmissions have been omitted).

7.29 Answer the following questions with respect to QoS-aware routing protocols:

(a) What advantages and disadvantages does multi-path routing have?

The availability of multiple routes provides fault-tolerance and quick recovery from broken paths. They can also help in achieving QoS requirements, e.g. different paths may have different latency or energy characteristics and packets can be sent along the path that best meets the QoS needs of a particular packet. Further, traffic can be balanced across multiple paths, reducing the traffic overheads and resource requirements for the individual paths.

(b) How does the SGNF component of SPEED work?

SGNF is the routing component of SPEED, which determines for each node i a forwarding candidate set of nodes for a particular destination. This set consists of all nodes from node i's neighbor list that are at least a distance K units closer to the destination than i. Packets are only forwarded to nodes in this set unless it is empty (than the packets are dropped). This set is further divided into a set with nodes that have a SendToDelay less than a certain single hop delay D and a set with all other nodes.

(c) How does the back-pressure rerouting component of SPEED work?

The back-pressure rerouting component is responsible for preventing voids that occur when a node fails to find a next hop node and for reducing congestion in the network using a feedback-based approach. The beacon exchange process is used to inform neighboring nodes of the delays experienced in packet forwarding. If these delays are large, the SGNF component of SPEED reduces the probability of nodes with large delays to being selected as forwarding nodes, thereby reducing the congestion around these nodes. When packets are dropped (because of a void), this also increases the delay, which is reported back to the upstream nodes. This back-pressure feedback can travel further upstreams if upstream nodes are also congested until the feedback reaches the source, which can then suppress further packets.

(d) Why does MMSPEED change the speed of packets as they travel along a route?

MMSPEED adjusts for low and high delays along the routes by adjusting the speed on downstream nodes. That is, if a node has experienced large delays in the beginning part of its route, its speed is increased in the remainder of the route to make up for the large latencies. Similarly, if a packet travels faster than expected, it can be slowed down in later parts of the route.

(e) How can latency and reliability considerations be combined in MMSPEED?

MMSPEED can first identify the required speed for a packet and then look for multiple forwarding nodes among those with sufficient progress speed such that the total reaching probability is at least as high as the required probability.

Part Three

Node and Network Management

8

Power Management

8.1 Give three reasons why dynamic power management is a crucial concern in wireless sensor networks.

(a) Given the complexity of their task, the nodes are small to accommodate large size batteries.

(b) In most cases, the network size is large which makes the task of manually changing, replacing or recharging batteries practically impossible.

(c) At present, taping other sources of energy is not a practical alternative.

8.2 What is the difference between local and global power management strategies? Give an example how a global power management can be realized at the link layer.

A local power management strategy focuses on a single node and it is usually implemented as a part of the operating system while a global power management strategy should take the entire network into account. This type of strategy is best implemented at the link and network layers.

8.3 Give two examples for accidental causes of power consumption in wireless sensor networks.

Accidental causes are unforeseen causes of energy consumption. For example a node attempts aimlessly to establish a link with a network even if its transmission range is insufficient. Another example is overhearing the communication of neighbor nodes.

8.4 How can a local power management strategy achieve an efficient power consumption in a wireless sensor node?

It can save energy by switching hardware components into a suitable power mode that is suitable to the current context of the node.

8.5 What is the main drawback of dynamic power management strategies that are based on a synchronous sleeping?

The cost of synchronization is very high.

8.6 Explain the idea behind power management strategies that are based on an asynchronous sleeping.

Fundamentals of Wireless Sensor Networks: Theory and Practice Waltenegus Dargie and Christian Poellabauer

In a medium access control protocols based on an asynchronous sleeping a node enters into a sleep state independently but for a specified duration. Any other node which is interested to communicate with the sleeping state transmits a preamble sequence for duration larger than the sleeping duration.

8.7 Explain the six different operational modes of the ATmega128L microcontroller.

(a) Idle – it stops the CPU but allows the SRAM, Timer (counter), SPI port and interrupt system to operate.

(b) ADC noise reduction – it stops the CPU and all I/O modules, except the asynchronous timer and the ADC.

(c) Power save – it allows the asynchronous timer to run but all the other components of the device enter into a sleep mode

(d) Power down – it saves the registers' content, freezes the oscillator and disables all the other chip functions until the next interrupt or Hardware Reset.

(e) Standby – it allows the crystal/resonator oscillator to run, but all the other hardware components enter into a sleep mode.

(f) Extended standby – it allows both the main oscillator and the asynchronous timer to continue to operate while the rest enter into a sleep mode.

8.8 What is a refresh rate of an active memory?

The refresh rate or refresh interval is a measure of the number of rows in the active memory that must be refreshed.

8.9 Explain the following terms in the context of RAM timing:

(a) RAS

RAS (*raw access strob*) is a signal that should be flagged before the processor subsystem accesses a particular cell in a memory. It enables the processor to determine the particular row or bank and then activate it.

(b) CAS

CAS (*Column access strob*) is a signal that activates a memory cell.

(c) t_{RCD}

The delays between the activation of a raw as well as a cell and the writing of data into or reading of data from the cell is given as t_{RCD}.

(d) t_{CL}

It is the time delay between the moment a processor orders the active memory to access a particular memory column and the moment the data is available to the processor.

8.10 The RAM timing of a certain processor is configured as $2 - 3 - 2 - 6$. Explain what it means.

The number $2 - 3 - 2 - 6$ refer to $t_{CL} - t_{RCD} - t_{RP} - t_{RAS}$. It describes the measure in clock cycles of the CAS latency, the time requires between RAS and CAS access, the time required to switch from one row to the next row, and the time delay between the precharge and activation of a row.

8.11 Explain briefly how the following DC-DC converters function.

(a) Flyback

It is a DC-Dc converter that changes the polarity of a DC voltage.

(b) Boost

It is a step-up DC-DC converter.

(c) Buck

It is a step-down DC-DC converter.

8.12 What is a rated current capacity?

Batteries are specified by a rated current capacity, C, expressed in Ampere-hour. A rated current capacity is a quantity that describes the rate at which a battery discharges without significantly affecting the prescribed supply voltage (or potential difference). In reality, the potential difference declines as the discharge rate increases.

8.13 Why do real batteries operate at a rate below the rated current capacity?

The available capacity of a battery depends upon the rate at which it is discharged. Because of the chemical reactions within the cells, the capacity of a battery depends on several factors including the magnitude of the current (which may vary with time), the allowable terminal voltage of the battery, and temperature. If a battery is discharged at a rate higher than the rated current capacity, the available capacity will be lower than expected.

8.14 What are the side-effects of drawing current at a rate greater than the discharge rate result?

Lower efficiency.

8.15 Describe the components of a typical DC-to-DC converter.

A typical DC-DC converter consists of a switching circuit and a low pass filter.

8.16 Suppose the circuit shown below is used by a DC-DC converter. At what frequency is the voltage drop across the load resistor, R_L maximum?

The voltage drop at the load is calculated as:

$$V_{out} = \frac{R_L}{R_L + \frac{\left(\frac{1}{j\omega C} \times j\omega L\right)}{\left(\frac{1}{j\omega C} + j\omega L\right)}} V_{in}$$

Simplifying the above equation yields:

$$V_{out} = \frac{R_L \left(1 - \omega^2 LC\right)}{R_L \left(1 - \omega^2 LC\right) + j\omega L}$$

Substituting $j\omega L$ with s yields:

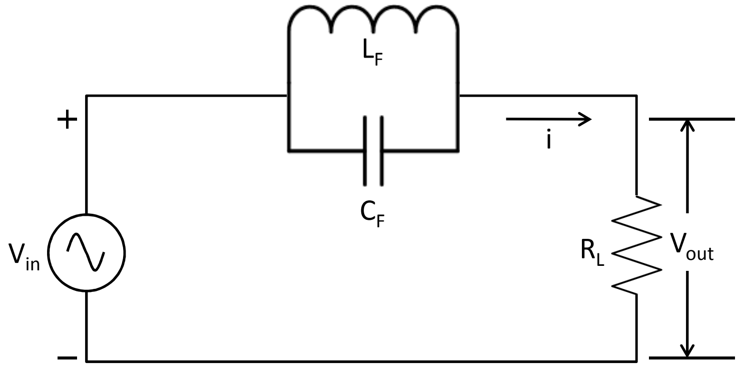


Figure 8.1 A conceptual architecture of a dynamic voltage scaling

$$V_{out} = \frac{s^2 R_L LC + R_L}{s^2 R_L LC + sL + R_L}$$

V_{out} will be maximum if the denominator is set to zero: $s^2 R_L LC + sL + R_L = 0$. Solving the quadratic problem yields:

$$s = \frac{1}{2R_L C} - \sqrt{\left(\frac{1}{2R_L C}\right)^2 - \frac{1}{LC}}$$

8.17 Why does a transition from low power mode to high power mode costs some power in the following subsystems:

(a) Processor subsystem

The processor subsystem has to load state information to resume operation from where it stopped before entering into a sleep state.

(b) Communication subsystem

The communication subsystem requires to calibrate hardware components before assuming full operation.

8.18 What conditions do justify the power transition costs?

Power transition from a higher power mode to a lower power mode is justified if the subcomponent stays in the lower power mode for a longer period of time.

8.19 Why does the performance of a switching transistor deteriorates at high operation frequencies?

At high frequency, the capacitance created at the junctions of a switching transistor increases. This in turn increases the transition time between the high and low voltages and vice versa.

8.20 How does the cumulative capacitance effect affect the switching time of a CMOS transistor?

The switching time is directly proportional to the cumulative capacitance and the bias voltage and inversely proportional to the drain saturation current.

9

Time Synchronization

9.1 Why is time synchronization needed in a WSN? Name at least three concrete examples.

Time synchronization is needed to achieve accurate temporal correlation of observed events, e.g. to detect the speed of a moving object. Time synchronization is also needed for many MAC-layer protocols, e.g. to accurately determine when a node can transmit in a TDMA-based system. Finally, many networks utilize duty cycling, i.e. devices are awake for brief periods of time and sleep otherwise. Accurate timing information is needed to determine when a node has to turn its radio back on.

9.2 Explain the difference between external and internal time synchronization and name at least one concrete example for each type of synchronization.

External synchronization means that the clocks in a network are synchronized with an external source of time (reference clock). For example, GPS can be used to achieve external synchronization. Internal synchronization refers to the process of synchronizing the clocks of sensor nodes in a network with each other, without the support of an external reference clock. To achieve internal synchronization, sensor nodes could elect a master node which periodically broadcasts its time to all other nodes in the network, allowing them to reset their clocks to correct for any drifts.

9.3 Consider two nodes, where the current time at node A is 1100 and the current time at node B is 1000. Node A's clock progresses by 1.01 time units once every 1 s and node B's clock progresses by 0.99 time units once every 1 s. Explain the terms clock offset, clock rate, and clock skew using this concrete example. Are these clocks fast or slow and why?

Comparing the two clocks, the clock offset is the difference in time between the two clocks. In this example, the current clock offset is 100. The clock rate indicates the frequency at which a clock progresses, i.e. node A's clock has a clock rate of 1.01 and node B's clock has a clock rate of 0.99. The clock skew indicates the difference in the frequencies of the two clocks, which is 0.02. Clock A is fast since its clock readings progress faster than real time. Similarly, clock B is slow since its clock readings progress slower than real time.

9.4 Assume that two nodes have a maximum drift rate from the real time of 100 ppm each. Your goal is to synchronize their clocks such that their relative offset does not exceed 1 s. What is the necessary resynchronization interval?

Each clock can deviate from real time by 100 μs per second in the worst case, i.e. it takes up to 10 000 s to reach an offset of 1 s. However, since both clocks have a drift rate of 100 ppm, the relative offset between them can be twice as large as the offset between a single clock and the external reference clock. Therefore, the necessary resynchronization interval is 5 000 s.

9.5 You need to design a wireless sensor node and you have three choices for clocks (1..3) with maximum drift rates of ρ_1 = 1 ppm, ρ_2 = 10 ppm, and ρ_3 = 100 ppm. Clock 1 costs significantly more than clock 2, which in turn costs significantly more than clock 3. Explain when one would choose clock 1 instead of clock 2 or clock 3 and vice versa.

The smaller the maximum drift rate, the less frequent re-synchronizations do have to occur. Re-synchronizations can incur large communication overheads, therefore more accurate clocks are preferred. However, cost is also an important factor for many wireless sensor networks. Therefore, in networks where cost is the primary factor, a cheaper and less accurate clock may be chosen. In networks where communication overhead due to synchronizations is a bigger concern than cost, more accurate clocks should be chosen.

9.6 A network of five nodes is synchronized to an external reference time with maximum errors of 1, 3, 4, 1, and 2 time units, respectively. What is the maximum precision that can be obtained in this network?

Since the error can go either way, i.e. a clock can be faster or slower than the external reference time by the amount of the error, the maximum precision is then the sum of the two largest errors, i.e. 3+4=7.

9.7 Node A sends a synchronization request to node B at 3150 (on node A's clock). At 3250, node A receives the reply from node B with a timestamp of 3120.

(a) What is node A's clock offset with respect to the time at node B (you can ignore any processing delays at either node)?

If $t_1 = 3150$, $t_2 = 3120$, and $t_3 = 3250$, then the offset can be calculated as

$$offset = \frac{(t_2 - t_1) - (t_3 - t_2)}{2} = -80 \tag{9.1}$$

That is, the two clocks differ by 80 time units.

(b) Is node A's clock going too slow or too fast?

Node A's clock is going too fast compared to node B's clock.

(c) How should node A adjust the clock?

One approach is to simply reset the clock by 80 time units. However, this can lead to problems since the clock repeats the last 80 time units, potentially triggering events

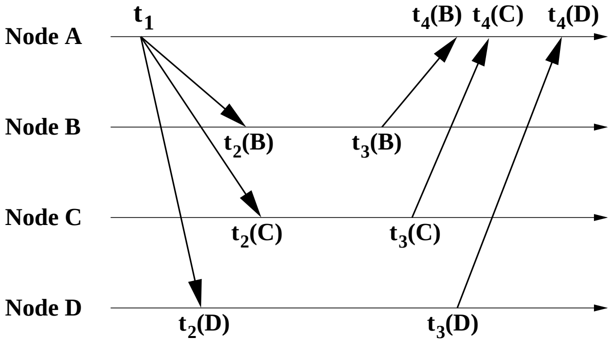


Figure 9.1 Pair-wise synchronization with multiple neighboring nodes (Example 9.8)

in the node that have already been triggered previously. Therefore, node A should slow down the progress of its clock until clock B had an opportunity to catch up the 80 time units it lags behind node A's clock.

9.8 Node A issues a synchronization request simultaneously to nodes B, C, and D (Figure **??**). Assume that nodes B, C, and D are all perfectly synchronized to each other. Explain why the offsets between node A and the three other nodes may still differ?

As indicate in Figure **??**, the times for the synchronization messages to travel between nodes can differ, e.g. based on the distances between senders and receivers. Besides propagation delays, synchronization messages also experience send, access, and receive delays that can differ from node to node, affecting the measured offsets.

9.9 Describe the reasons for non-determinism of communication latencies and why this non-determinism affects time synchronization.

There are four main contributors to non-determinism of communication latencies: send delays, access delays, propagation delays, and receive delays. Since these delays can depend on a variety of parameters, including the distance between nodes, system call interface latencies, context switches, and characteristics of the MAC protocol, the measurement of offsets and delays will be inaccurate, thereby affecting the quality of time synchronization.

9.10 Explain why the depth of the synchronization tree in centralized LTS should be small.

Errors resulting from the pair-wise synchronizations are additive and therefore increase along the branches of the tree as a function of the number of hops.

9.11 Discuss the differences and similarities in the design of the TPSN and the LTS synchronization protocols.

Both protocols rely on trees to organize a network and both are sender-receiver synchronization approaches using pair-wise synchronization messages. In both protocols, the synchronization error depends on the depth of the tree. TPSN uses time-stamping of packets at the MAC layer to reduce the error due to varying synchronization message latencies.

9.12 Explain the six different types of time stamps that characterize the communication in FTSP. How does FTSP remove the jitter of the interrupt handling and the encoding/decoding times?

The first time stamp t_1 occurs when the wireless radio informs the CPU (using an interrupt) that it is ready to receive the next piece of message to be transmitted. Once the interrupt handling has completed, a second time stamp t_2 is generated. At time t_3, the message is transmitted over the medium and received by the receiver node at time t_4. Decoding of the message at the receiver is concluded at time t_5 and the receiver radio issues an interrupt at time t_6 to inform its CPU that the synchronization has been received successfully. The jitter caused by the interrupt handling time can be removed by taking the minimum of the normalized time stamps.

9.13 Explain the concept behind the RBS protocol. How can RBS be extended to work in multi-hop scenarios?

RBS uses broadcast messages to a set of receivers, allowing them to synchronize to each other. The variability of broadcast messages is mostly due to propagation delays and the time needed by the receivers to receive and process the incoming messages. That is, it removes the non-deterministic synchronization errors caused by sender. In multi-hop scenarios, multiple reference beacons can be established, each with its own broadcast domain. These domains can overlap and nodes in both regions can support synchronization across domain boundaries.

9.14 Describe the term “post-facto synchronization”.

When nodes synchronize only when when events of interest happen, the post-facto synchronization scheme is being used. That is, nodes do not waste energy on synchronization messages when unsynchronized clocks are otherwise acceptable.

9.15 Compare the TPSN and RBS time synchronization protocols.

TPSN uses pair-wise synchronization along the edges of a tree connecting the nodes in a sensor network. The synchronization is affected by the non-deterministic communication latencies of synchronization messages. RBS is a synchronization approach based on reference broadcasts, allowing receivers of such broadcast messages to synchronize their clocks to each other. RBS removes the sender components of the non-deterministic communication latencies.

9.16 Compare the broadcast approach used by RBS with the pair-wise synchronization approach by TPSN and other protocols for the following scenarios:

(a) synchronization messages experience send and access delays with high variance and all other delays are negligible

RBS is a better choice since the only the time off arrival is important and everything before is irrelevant, i.e. send and access delays do not affect it (unlike TPSN).

(b) synchronization messages are sent using acoustic signals and the distances between nodes are unknown

When propagation delays are large (such as in acoustic communications), they have an important effect on the quality of synchronization. When two nodes have different distances to the sender, nodes using RBS will experience different propagation delays, which introduce an error into the synchronization. Protocols based on round-trip time measurements (such as TPSN) are more suitable in this case since they consider propagation delays in their synchronization process.

(c) synchronization messages experience send and access delays without variance and all other delays are negligible

RBS does not depend on send and access delays and TPSN works well when these delays do not vary, therefore both protocols will work well.

(d) synchronization messages experience significant receive delays that may differ from node to node

Both TPSN and RBS will be affected by the variation in receive delays and therefore both will be good choices.

9.17 Two nodes A and B use RBS to receive periodic acoustic synchronization signals from a reference node. Node A's clock shows 10 s when it receives the last synchronization beacon, while node B's clock shows 15 s. Node A detects an event at time 15 s, while node B detects the same event at time 19.5 s. Assume that node A is 100 m away from the synchronization source and node B is 400 m away from the synchronization source. Which node detected the event sooner and by how much? Assume a signal speed of 300 m/s.

When the acoustic signal arrived at node A at time 10 (on A's clock), the signal had to travel 300 m more to reach node B, which takes 1 s. That is, When node A's clock reading was 10 s, node B's clock reading was 15-1 = 14 s. That is, node B's clock is 4 s faster than node A's clock. Node A observed the event at 15 s, while node B observed the time at 19.5 s, which corresponds to 15.5 s on A's clock. That is, node A observed the event first by 0.5 s.

10

Localization

10.1 Why is localization needed in wireless sensor networks? Name at least two concrete scenarios or applications where localization is required.

Sensor nodes are often deployed in an ad-hoc fashion, without knowledge of their exact deployment position. Localization is then necessary to provide a physical context to sensor readings, i.e. most sensor readings are meaningless without knowledge of where the readings were obtained. For example, localization is required in sensor network that detects wild fires to ensure that firefighters can quickly locate the affected areas and to predict the spread of the wild fire. Similarly, in surveillance applications, location information is required to be able to guide law enforcement officers quickly towards an intruder.

10.2 A node's position in two-dimensional space is $(x, y) = (10, 20)$ with a maximum error of 2 in the x direction for 95% of all measurements and a maximum error of 3 in the y direction for 90% of all measurements. What is the accuracy and the precision of this location information?

The accuracy of the location information is 2 in the x direction and 3 in the y direction. The precision is 95% in the x direction and 90% in the y direction.

10.3 Explain the difference between physical and symbolic positions and name at least two examples for each type.

Physical positions are positions measured within a reference frame, which can be both a global reference frame (e.g. GPS) or an arbitrary coordinate system or reference frame (e.g. positions expressed as distances to other sensors). Symbolic positions are not associated with a reference frame, but instead describe locations such as "office number 112", "basement", or "intersection of Adams and Fir Road".

10.4 Define the terms anchor-based localization and range-based localization.

Anchor-based localization refers to the use of reference nodes (anchors) with well-known locations. Communication between sensor nodes and these anchors can be used to estimate the locations of the sensors. Range-based localization estimate locations based on range measurements, i.e. measurements of the distances between nodes.

10.5 Time of Arrival (ToA) is one example of a ranging technique. Answer the following questions (assume a propagation time of 300 m/s):

(a) What is the advantage of two-way ToA over one-way ToA?

The one-way ToA requires that sender and receiver are accurately synchronized. In the two-way ToA, the round-trip time is measured using the clock on only on device, thereby removing the need for clock synchronization.

(b) In a synchronized network with unknown synchronization error, an anchor node periodically broadcasts an acoustic signal to sensor nodes in its range. At time 1000 ms on the anchor node's clock, the anchor node issues a beacon, which is received by node A at time 2000 ms (on node A's clock). What is the distance that A can now compute?

The propagation delay of the acoustic signal is 1000 ms and the signal travels at 300 m/s, i.e. the distance is 300 m.

(c) Instead of computing the distance itself, node A also responds with an acoustic signal issued at time 2500 ms, which is received by the anchor node at time 3300 ms. What is the distance computed by the anchor node? What can you say about the synchronization of anchor node and node A?

The round-trip time is $(2000 - 1000) + (3300 - 2500) = 1000 + 800 = 1800$. The distance is then the round-trip time divided by 2 and multiplied by the velocity, i.e. 270 m. Since the propagation time of the signal from anchor to sensor is 1000 ms and the propagation time from sensor to anchor is 800 ms, the offset between the sensor node's clock and the anchor's clock appears to be 100 ms.

10.6 What is the main disadvantage for both TDoA and AoA ranging techniques?

Both techniques requires extra hardware (e.g. two radios for TDoA and an array of antennas for AoA), which may be infeasible for many low-cost and low-power sensor networks.

10.7 RSS-based localization techniques are often combined with a process called RF profiling, i.e. the mapping of the effects of objects in the environment on signal propagation. Why is this necessary and can think of examples of such objects?

The signal strength of radio-frequency signals is affected by objects (size, shape, thickness, etc.) between sender and receiver, causing effects such as reflection, refraction, scattering, etc. RF profiling allows localization techniques to take such effects into consideration. Objects include trees, buildings, traffic signs, mountains, etc.

10.8 Two nodes A and B are known to be positioned at locations $(0, 0)$ (node A) and $(0, 1)$ (node B) in the two-dimensional space. A third node C wishes to determine its position using trilateration. Based on ranging techniques, node C knows its distances to node A

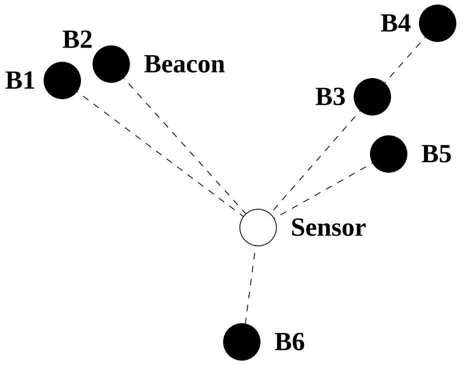


Figure 10.1 Example 10.11

($d(A, C) = \sqrt{0.75}$) and node B ($d(B, C) = \sqrt{0.75}$). What are the two possible positions of C?

Since both A and B are positioned on the x-axis and C's distances to A and B are identical, x_C must be $d(A, B)/2$ for both possible positions, i.e. $x_C = 0.5$. The y coordinate of the first possible position can be computed as $y_C = \sqrt{d(A, C)^2 - x_C^2} = \sqrt{0.5}$. The second possible position is the first position mirrored on the x-axis, i.e. the two possible positions are $(0.5, \sqrt{0.5})$ and $(0.5, -\sqrt{0.5})$.

10.9 Three nodes A, B, and C are known to be positioned at locations $(0, 0)$, $(10, 0)$, and $(4, 15)$, respectively. Node D is estimated to be a distance of 7 from A, a distance of 7 from B, and a distance of 10.15 from C. Determine the location of D using trilateration.

Since A and B are positioned on the x-axis and D's distances to A and B are identical, x_D is half the distance between A and B, i.e. $x_D = 5$. The y coordinate of the first possible position can be computed as $y_D = \sqrt{d(A, D)^2 - x_D^2} = \sqrt{24}$. That is, the two possible positions are $(5, \sqrt{24})$ and $(5, -\sqrt{24})$. Since the distance between node C and D is less than the distance of C to the x-axis, the location of D must be the first one, i.e. $(5, \sqrt{24})$.

10.10 Consider the two-dimensional topology in Figure **??**. The sensor node in the center can select three of the six anchor nodes as basis for trilateration. Which nodes should the sensor node select? Justify your answer, i.e. what guideline for anchor selection should be considered? What would this guideline be in three-dimensional space?

A possible selection would be B1, B3, and B6. The guideline for this selection is that the chosen anchors should be non-collinear in the two-dimensional space (or non-coplanar in the three-dimensional space) such that unique locations can be obtained. Nodes B1 and B2 and similarly B3, B4, and B5 are close together such that it may be impossible (as is the case if B3 and B4 were chosen) or at least difficult to identify one exact and unique location.

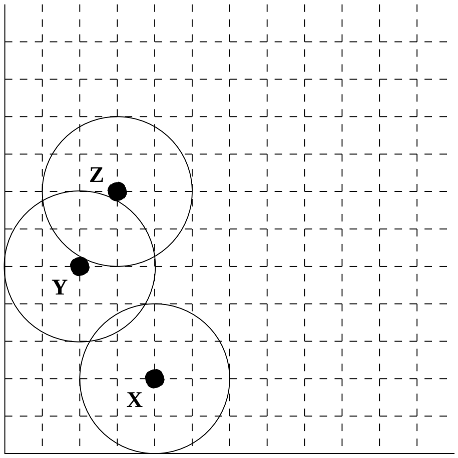


Figure 10.2 Example 10.11

10.11 Two nodes A and B do not know their own positions, but they can hear beacons in their proximities. Node A can hear beacons located at $(4, 2)$ and $(2, 5)$. Node B can hear beacons located at $(2, 5)$ and $(3, 7)$. All nodes have a radio range of 2 units.

(a) Are either $(3, 3.5)$ or $(3, 4.5)$ possible locations for node A?

Figure **??** illustrates the locations and transmission ranges of the three anchor nodes. Node A can hear the anchors marked as X and Y. Therefore $(3, 3.5)$ is a possible location for node A. However, $(3, 4.5)$ is not a possible location, since it would be within Y's transmission range, but not within X's range.

(b) Are either $(2, 6)$ or $(4, 5)$ possible locations for node B?

Node B can hear the anchors marked as Y and Z. Therefore, $(2, 6)$ is a possible location since it resides within the intersection of both anchors' radio ranges. However, $(4, 5)$ would place the node outside of Z's radio range and it is therefore not a possible location.

10.12 What are the differences between iterative and collaborative multilateration?

The iterative multilateration process extends the lateration technique to locate nodes that are not within transmission range of at least three anchor nodes. Instead, a node that uses lateration to find its own position can serve as an anchor for other nodes. The downside of this approach is that the localization error accumulates with each iteration. The collaborative multilateration technique also allows nodes without at least three anchors to determine their locations, but it uses a different approach than the iterative multilateration technique. Here, a graph of participating nodes (i.e. nodes that are anchors or have at least three participating neighbors) is constructed and a node can then try to estimate its position by solving the system of over-constrained quadratic equations relating the distances among the nodes and its neighbors.

10.13 Explain the concept of GPS localization and answer the following questions:

(a) Why are three satellites enough to obtain a position on the globe?

Even though the use of three satellites leads to two intersection points of the three spheres, one of these two intersection points can typically be eliminated easily, e.g., because it would place the receiver device far out in space or deep below the ground.

(b) Why is it preferred to have at least four satellites available for localization?

A fourth satellite allows to obtain a more accurate location. This is due to the stringent time synchronization requirements, i.e. receiver devices and satellites are supposed to be tightly synchronized. However, the receiver clocks are much simpler and cheaper than the clocks in the satellites, thereby leading to synchronization and consequently also to localization errors. The fourth sphere should ideally intersect the other three spheres at the exact location of the receiver. Because of the timing errors, this may not be the case. If the spheres are too large to obtain an intersection point, their sizes can be reduced by adjusting the clock until the spheres are small enough to intersect in one point (and similarly we can address the case where spheres are too small). This is possible because the timing errors are the same for all measurements and a receiver can calculate the required clock adjustment to obtain a single intersection point. Finally, a fourth satellite also allows the receiver to obtain a measurement for the elevation.

(c) What is the purpose of the monitor stations and the master control station?

The monitor stations constantly receive the data sent by the satellites and forward this information to a master control station (MCS). The MCS uses this information to compute corrections to the satellites' orbital and clock information, which are then sent back to the appropriate satellites via ground antennas.

(d) Why is typically not feasible to have all wireless sensor nodes equipped with a GPS receiver?

GPS receiver are both costly and energy-hungry devices, which may make them unsuitable for use in low-power and low-cost sensor networks.

10.14 Explain the difference between range-based and range-free localization?

Range-based localization techniques are based on distance measurements using ranging techniques (RSS, ToA, etc.) between nodes, e.g. between sensor nodes with unknown locations and anchor nodes with known locations. Range-free localization techniques do not use distance measurements based on ranging techniques. The advantage of range-free techniques is that they typically do not require extra hardware and are therefore more cost effective.

10.15 Figure **??** shows a network topology with three anchor nodes. The distances between anchors A_1 and A_2, anchors A_1 and A_3, and anchors A_2 and A_3 are 40 m, 110 m, and 35 m, respectively. Use the Ad hoc Positioning System to estimate the location of the gray sensor node (show each step of your process).

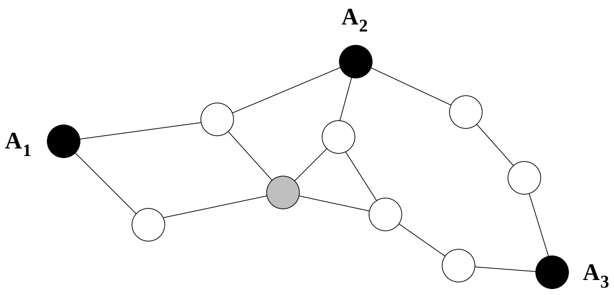


Figure 10.3 Topology for Exercise 10.15

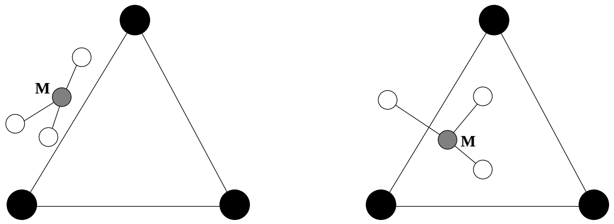


Figure 10.4 Examples for Exercise 10.16

First, each anchor determines a correction factor for each other hop based on the formula:

$$c_i = \frac{\sum \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}}{\sum h_i} \tag{10.1}$$

where h_i is the distance to the anchor in hops. A_1 computes a correction factor of $\frac{40+110}{2+5} = 21.43$. Similarly, A_2 computes a correction factor of $\frac{40+35}{2+3} = 15$ and A_3 computes a correction factor of $\frac{35+110}{3+5} = 18.13$. These correction factors are distributed in the network and a sensor node uses the correction factor from the closest anchor, e.g. in this example, the gray node could use the correction factor from either A_1 or A_2. If it uses the one from A_2 (15), it computes its distances to the three anchor nodes by multiplying the correction factor with the hop distances, i.e. its distance to A_1 is $2 * 15 = 30\,m$, its distance to A_2 is $2 * 15 = 30\,m$, and its distance to A_3 is $3 * 15 - 45\,m$. If the locations of the anchors are known, triangulation or trilateration can be used to find the position of the gray sensor.

10.16 For the APIT test, can you show a concrete scenario where a node M would come to the wrong conclusion that it must be inside a triangle? Use a scenario where node M has at least 3 neighbors. Can you also show an example where node M would come to the wrong conclusion that it must be outside a triangle?

The example on the left in Figure **??** shows a scenario where none of the three neighbors of M is either closer to or further away from all three anchor nodes simultaneously. M therefore concludes that it must be inside the triangle. In the example on the right, M's leftmost neighbor is further away from all three anchor nodes than M, therefore M concludes it must be outside the triangle.

10.17 A sensor node in a WSN using the lighthouse approach for localization detects the first beam of light at time 0 s and the second beam of light at time 0.25 s. The next time the first beam of light is detected is 7 s. The distance of the two light sources (beam width) is 10 cm. What is the distance of the sensor to the light emitter?

The distance is computed as:

$$d = \frac{b}{2sin(\alpha/2)} \tag{10.2}$$

where α is computed as:

$$\alpha = 2\pi\frac{t_{beam}}{t_{turn}} \tag{10.3}$$

Since $t_{beam} = t_1 - t_2 = 0.25 - 0 = 0.25$ and $t_{turn} = 7$ s, $\alpha = 0.224$ and $d =$ 25.5 m.

11

Security

11.1 Describe the CIA security model.

Confidentiality refers to the need for security mechanisms that ensure that only the intended receiver can correctly interpret a message and unauthorized access is prevented. Integrity refers to the need for security mechanisms that ensure that a message cannot be modified as it propagates from the sender to the receiver. Finally, availability refers to the need for security mechanisms that ensure that a system is able to perform its tasks without interruption.

Which service(s) described in this model do you think are essential for the following scenarios. Justify your answers.

(a) A WSN that allows emergency response teams to avoid risky and dangerous areas and activities.

Confidentiality may only be a minor concern in such a scenario. However, integrity and availability are important since modified messages or interrupted service can have catastrophic consequences.

(b) A WSN that collects biometric information collected at an airport.

Biometric information must be protected, therefore confidentiality is important. Integrity is also important since the collected should not be compromised. Availability is desirable, but probably somewhat less of a concern compared to the other services.

(c) A WSN that measures air pollution in a city for a research study.

None of the services is absolutely essential for this type of application, although as much uptime (availability) as possible may be desired.

(d) A WSN that alerts a city of an impending earthquake.

The time-criticality of this type of application may make confidentiality a minor concern. However, integrity (preventing of false alarms that could have severe consequences) and availability (ensuring that the alerts arrive reliably) are important.

11.2 What is a man-in-the-middle attack? Can you imagine a concrete WSN scenario where such an attack could be catastrophic?

Fundamentals of Wireless Sensor Networks: Theory and Practice Waltenegus Dargie and Christian Poellabauer

The man-in-the-middle attack refers to attacks where an intruder positions itself between the sender and receiver such that the sender's messages can be intercepted, modified, and retransmitted to the receiver, making the receiver believe that these messages came directly from the original sender). In the earthquake scenario from the previous exercise, if an intruder changes the messages such that the receiver believes wrongly that there is no risk of an impending earthquake could have severe consequences.

11.3 Explain the concepts of symmetric and asymmetric keys. This chapter mentioned a *shift cipher* as a simple example of a cryptographic technique. Is this cipher a symmetric or an asymmetric key cryptography technique? What are the problems with such a simple cipher?

Symmetric key cryptography means that both parties in a communication use the same secret key for encryption and decryption. In asymmetric key cryptography, two keys are used, one for encryption and one for decryption. The shift cipher is an example for symmetric key cryptography. Besides the need for secure distribution of the secret key, the shift cipher is very simple with only a limited number of possible keys. Also, the shift cipher does not hide patterns, beginning and end of words, punctuation, etc., which makes it easy to break them. For example, 'e' is the most common letter in the English alphabet. It is easy to find the most common letter in the encrypted message, assume that it is the letter 'e', and derive the shift key from this knowledge.

11.4 Why do you think authentication can be a particularly significant problem in a WSN?

Many sensor nodes are placed in publicly accessible areas, i.e. it is often easy to gain physical access to sensor devices. An attacker may then be able to modify or replace the device or analyze the device for its content (e.g. to learn about security keys and algorithms). It is often necessary to share cryptographic techniques and keys with all nodes in the sensor network, making it easier for an intruder to obtain such information. Authentication is particularly important since it is difficult to distinguish between a new node that joins a network or a a node that was temporarily disconnected from the network from a sensor node that has been inserted into the network or compromised by an intruder.

11.5 Explain some of the characteristics of a WSN that make routing security difficult to implement.

The resource constraints of wireless sensor networks make it easier to use attacks with the goal to exhaust a device's or network's resources. It is also more difficult to use resource-intensive security measures (e.g. CPU-intensive algorithms). The lack of centralized control in a WSN means that many security measures must also be decentralized. The remote location of sensor nodes makes it difficult to protect them from physical access by an attacker. Communication in a WSN is error-prone, which may interfere with security-related communications.

11.6 While "typical" computers are in homes, offices, labs, etc., wireless sensor nodes are often placed in places that are publicly open and accessible. What kind of attacks could an adversary initiate by accessing a single sensor node in a large-scale WSN?

A variety of attacks are possible once an attacker gains access to a device. It can be used to learn about encryption algorithms and keys used in the network. It could be used as a starting point for a variety of denial-of-service attacks and man-in-the-middle attacks. Attacks on the routing layer (e.g. selective forwarding attacks, etc.) or on data aggregation are also possible.

11.7 What is "data freshness" and why is it important in sensor networks?

Data freshness ensures that sensor data is recent and no old recording of such data are being replayed. For example, this is important for key distribution schemes to prevent later replays of key distributions.

11.8 What is a denial-of-service attack?

The goal of a denial-of-service attack is to stop a network from functioning or to disrupt the services a network provides.

Explain the following attacks:

(a) Jamming attack
A jamming attack is a DoS attack at the physical layer, where an adversary interferes with the radio communications of sensor nodes by continuously transmitting a strong signal that lowers the signal-to-noise ratios on the sensor nodes.

(b) Exhaustion attack
The goal of an exhaustion attack is to prematurely deplete a sensor node of its energy resources, e.g. by increasing the overheads and workloads of a sensor node. For example, an adversary could exploit the error recovery mechanisms of a sensor node by continuously interfering with transmissions, thereby triggering retransmissions of entire sensor messages.

(c) Tampering attack
A tampering attack occurs when an adversary obtains physical access to a sensor node, allowing the attacker to destroy or modify the device.

11.9 Consider routing attacks such as selective forwarding, sinkhole, blackhole, Sybil, rushing, and wormhole attacks. Describe briefly each type of attack and discuss how these attacks could take place in the following types of networks:

An adversary positioned on the route between a sender and receiver could drop packets that meet certain characteristics (selective forwarding attack) or even all packets (blackhole attack). A sinkhole attack occurs when an adversary attempts to position itself on the route of as many sensors streams as possible, thereby drawing traffic towards the adversary, who can then disrupt the traffic or modify packet content. A Sybil attack occurs when an attacker claims to have several identities (or locations), thereby convincing more sensor nodes to choose the attacking node as forwarding node. In a rushing attack, an adversary quickly forwards route request messages towards the destination to increase the attacker's probability to be on the sensor stream's route. Finally, in a wormhole attack, two collaborating attackers attempt to convince other nodes that they have a good link between them such that they will be chosen as forwarder nodes for as many routes as possible.

(a) A network using a table-based routing protocol such as OLSR.

In table-based protocols, the goal of an attacker is to appear as next-hop neighbor in as many tables as possible. Therefore, the attacker can pretend to have low forwarding costs to all destinations. Once the attacker receives packets to forward, it can modify or drop them as desired.

(b) A network using an on-demand routing protocol such as DSR.

In an on-demand routing protocol, the goal of the attacker is to be on the path chosen by either the receiver or sender device. Therefore, the attacker could either rush RREQ packets towards the destination or it could modify the content in the RREQ packet headers such that the receiver believes that the route with the attacker is the one with the lowest cost.

(c) A network using a location-based routing protocol such as GEAR.

In location-based routing protocols, the goal of the attacker is to make its neighbors believe that it is closer to the destination than all other neighbors or that its cost for forwarding towards the destination is lower that those of all the other neighbors. For example, this can be achieved by pretending to have multiple identities and locations or by changing the location information sent to its neighbors. For example, if an attacker node sees that one of its neighbors frequently sends packets towards a destination, but not via the attacker node, the attacker can "change" its location such that it becomes the better forwarding choice towards the destination.

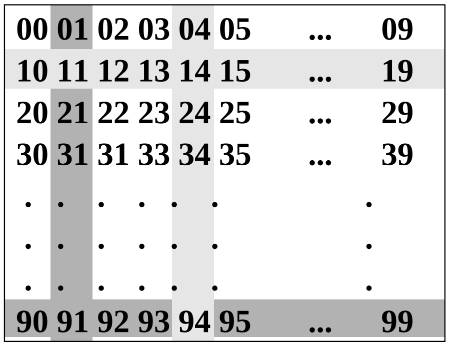
11.10 In this chapter, data aggregation functions such as *average*, *sum*, and *minimum* were called "insecure". What does this mean and which technique can be used to increase the resilience of aggregation functions?

Insecure means that the aggregation function can be easily influenced by an attacker such that a network's behavior can be altered. The delayed aggregation and delayed authentication scheme delays both aggregation and authentication to make it more difficult to modify the result of an aggregation.

11.11 Consider the virtual ID space for the PIKE scheme in Figure **??**. In this example, how many options does node 3 have to establish a key with node 15? Describe each option.

Node 3 has two options: it can establish a key with node 15 either via node 13 or node 5 (the intersections of the rows and columns of nodes 3 and 15).

11.12 What is a "nonce"? How does SPINS use them and what services are provided by the SNEP protocol?



00	01	02	03	04	05	...	09
10	11	12	13	14	15	...	19
20	21	22	23	24	25	...	29
30	31	31	33	34	35	...	39
.	.	.	.	.	.		.
.	.	.	.	.	.		.
.	.	.	.	.	.		.
90	91	92	93	94	95	...	99

Figure 11.1 Virtual ID space in PIKE (Exercise 11.11)

A nonce is a random number so long that an exhaustive search for all possible nonces is infeasible. SPINS uses nonces to ensure strong data freshness, i.e. a node randomly generates a nonce which is send along with a request message to another node. The second node then returns the nonce with the response message in an authenticated protocol, which ensures that the first node knows that the second node generated its response after the first node made the request.

11.13 What are the security models provided by IEEE 802.15.4. What is the purpose of the trust center in ZigBee?

IEEE 802.15.4 has four basic security models: access control, message integrity, confidentiality, and replay protection. The trust center in ZigBee (a responsibility typically assumed by the ZigBee coordinator) is responsible for the authentication of devices wishing to join a network, maintaining and distributing keys, and enabling end-to-end security between devices.

12

Sensor Network Programming

12.1 Describe the difference between node-centric and application-centric programming.

Node-centric programming focus on the development of sensor applications and software for each sensor device, whereas application-centric programming considers and develops software for the networks as a whole.

12.2 Explain the difference between *provides* and *uses* interfaces in nesC.

The provides interface is a set of method calls that are exposed to higher layers. The uses interface describes the use of some kind of service.

12.3 What options does nesC provide developers to prevent race conditions?

Race conditions can be prevented by using synchronous code, which is always atomic to other synchronous codes. If asynchronous code is used, one option is to convert code that modifies shared state into synchronous code. Another option is to use atomic sections, i.e. brief code sequences that nesC will always run atomically.

12.4 A common strategy to ensure atomicity is to disable interrupts in an operating system as long as atomic operations are being executed. What is the danger of disabling interrupts?

Important events that trigger interrupts, may not be reported (either at all or only after a delay, i.e. after interrupts have been reenabled), which can have severe consequences.

12.5 What are the main advantages and disadvantages of thread-based programming models?

The thread-based programming model allows multiple tasks to make progress in their execution without the concern that a task may block other tasks indefinitely. However, thread-based approaches increase the operating system complexity and may also require more complex synchronization support in the operating system.

12.6 This chapter introduced several macroprogramming models. Contrast how these different models are able to address multiple (or all) sensor nodes simultaneously.

Fundamentals of Wireless Sensor Networks: Theory and Practice Waltenegus Dargie and Christian Poellabauer

Abstract regions groups sensors together using certain neighborhood relationships, e.g. “the set of nodes within distance d”. Discovery of region members can be achieved using periodic advertisements. In EnviroTrack, groups are formed by sensors which detect certain user-defined entities in the physical world, with one group formed around each entity. Groups are then identified by context labels, which are logical addresses that follow the external tracked entity around in the physical environment. Database approaches treat a wireless sensor network like a distributed database that can be queried to obtain sensor data. That is, the network can be represented logically as a database table that has (as an example) one row per node per instant in time and each column corresponds to a type of sensor reading.

12.7 Why is it necessary to provide the opportunity to dynamically reprogram a sensor network? What is challenging in distributing a new program to all sensor nodes in the network?

Reasons for reprogramming a sensor network are that details of certain applications and application characteristics may not be known until after deployment, sensor applications may need upgrades or bugfixes, and usage scenarios of sensor networks may change during their lifetimes. Challenges in reprogramming include the need for reliable code distributions (all nodes must receive all pieces of the code), quick code dissemination (to limit downtimes), and energy-efficient dissemination. Reprogramming should interfere with the goals of the sensor network as little as possible. Another challenge is that during reprogramming, different sensors may run different applications or versions of applications. In such scenarios, it is important to prevent failures and miscommunications due to version mismatches.