ELEX 3525 : Data Communications
2020 Fall Term

Solutions to Midterm 1

Question 1

A communication system can transmit five different messages. Four of these are transmitted a probability of 1/8. This system transmits each message using a 3-bit word at a rate of 10,000 messages per second.

1. What is the entropy of this source in bits per message?

   The entropy is given as:

   $$H = \sum_i -P_i \cdot \log_2 P_i$$

   Since the probabilities have to add up to 1 and the probabilities of the first four messages add up to $4 \times 1/8 = 1/2$, the probability of the fifth message must be $1/2$. Thus the entropy is:

   $$4\times\frac{-1}{8}\log_2\frac{1}{8}+\frac{-1}{2}\log_2\frac{1}{2}=\frac{4\cdot 3}{8}+\frac{1\cdot 1}{2}=\frac{16}{8}=2$$

   bits/message.

2. What is the information rate in bits per second?

   The information rate is given by the message rate times the entropy or $10,000 \times 2 = 20$ kbps.

3. What is the data rate in bits per second?

   The data rate is the number of bits transmitted over the channel each second which is $10,000 \times 3 = 30$ kbps.

Question 2

A Unicode character is encoded using UTF-8 as the three bytes (in hexadecimal): `E3 81 AC`

1. What are the values of the `z`, `y` and `x` bit fields as defined in Table 3-6 of the Unicode standard?

   The three bytes in binary are:

   ```
   E3 = 1110 0011
   81 = 1000 0001
   AC = 1010 1100
   ```

   Table 3-6 defines the three-byte UTF-8 encoding as:

Scalar Value	First Byte	Second Byte	Third Byte
zzzzyyyy yyxxxxxx	1110zzzz	10yyyyyy	10xxxxxx

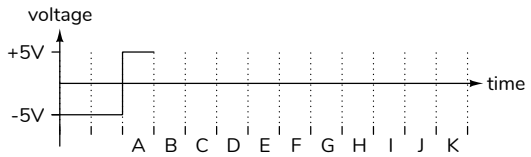
   From which we can extract the bits:

   ```
   zzzz = 0011
   yyyyyy = 00 0001
   xxxxxx = 10 1100
   ```

2. What is the Unicode code point of this character?

   The code point is the scalar value in the table above. Combining the bits: 0011 00 00 01 10 1100 or in hexadecimal, `U+306C` (ぬ, "Hiragana Letter Nu").

Question 3



The diagram above shows the outline of an RS-232 waveform. In the space below, write the voltages that would appear in each time interval if a byte with value 0x49 was transmitted using 8 bits per character and odd parity. The interval marked A is the start bit. For each letter B through K write the voltage for as H (for high, >5V) or L (for low, <-5V). Please ensure you answer is unambiguous (e.g. A:H, B:L, C:H, etc.):

- The start bit is always high and is given in the question.

- 0x49 in binary is 0100 1001. Writing the bits starting with the least-significant bit and converting 0 to H and 1 to L the sequence of eight transmitted values would be: `LHHL HHLH`.
- There are three '1' bits so the parity is already odd and so the parity bit should be 0 (`H`).
- A stop bit is always low.
- The sequence of transmitted levels should thus be:
  - `A:H` (start bit, given),
  - `B:L`, `C:H`, `D:H`, `E:L`, (first four data bits, l.s. bit first)
  - `F:H`, `G:H`, `H:L`, `I:H`, (second four data bits)
  - `J:H`, (parity bit)
  - `K:L`, (stop bit)

Question 4

The signal received by a receiver makes an error when the noise voltage is more negative than -1 V. The noise has a zero mean and an RMS value of 0.33 V. What is the error rate?

The noise has a mean of zero ($\mu = 0$)and a standard deviation of $\sigma = 0.33$ V. The normalised threshold is $t = \frac{-1-0}{0.33} \approx -3$. The probability that the noise voltage is less than (more negative than) this is P(-3) which can be found from the graph in lecture 3 as approximately 0.1% or using a calculator as $\approx 1.35 \times 10^{-3}$.