Solutions to Assignment 1

Question 1

Everyone will have a different answer. Assuming the rightmost four characters of your BCIT ID were 4321, we could write this number as:

- a 16-bit binary number: 0001 0000 1110 0001 since $2^{12} + 2^7 + 2^6 + 2^5 + 2^0 = 4096 + 128 + 64 + 32 + 1 = 4321$.
- a hexadecimal number in network (big-edian) bit/byte order: 0x10E1
- with the two bytes in little-endian order the value would be 0xE110 and with bits within each byte also in little-endian order the bits would be 1000 0111 0000 1000 corresponding to hex 0x8708.

Question 2

The Unicode code point (the index in the Unicode code table) for the character: "ARABIC LETTER ALEF" is found in: `http://www.unicode.org/charts/PDF/U0600.pdf` as U+0627:

Question 3

The Chinese character for "Monkey" (the animal) is "猴" with Unicode value (code point) U+7334.

- the value of the code point U+7334 in binary is 0111 0011 0011 0100 so according to the UTF-8 encoding table it has to be encoded in 3 bytes (octets) using the third row of the table.
- the values of the three fields in the binary value are: zzzz = 0111, yyyyyy = 11 0001, and xxxxxx = 11 0011. The three bytes in binary are 1110 0111, 1011 0001 and 1011 0011 corresponding to the three hex values: 0xe7 0x8c 0xb4.

Question 4

The first byte in the document is 0x95 (1001 0101) which does not correspond to one of the possible initial bytes of the UTF-8 encoding of a character. The same applies for the second byte (0x94). However the third byte is 0xD3 (1101 0011) indicating a two-byte character of which the second byte is 0x82. The final byte is 0x81 which is not the first byte of a valid UTF-8 sequence and indicates another error. Thus we can only decode one character.

The UTF-8 encoding of this character is 0xD3 0x82 corresponding to 1101 0011 1000 0010 in binary. The fields in the UTF-8 encoding have values yyyyy = 10011 and xxxxxx = 000010. Concatenating these bits we get the value of the UTF-8 code point: 10011 000010 or 100 1100 0010 = U+4C2 which is "Cyrillic Small Letter Zhe with Breve" ("ӂ").

Question 5

The waveform created in Lab 1 contained one start bit, 7 data bits and four stop bits for a total of 12 bits for each of the 7 random (information) data bits. The duration of each bit (pulse) was 833 μs and it took 10 ms (0.833×12) to transmit 7 bits.

The source therefore generated 7 bits every 10 ms or 700 b/s. The average bit rate at the source is the same. Since the data is not compressible the information rate is also 700 b/s.

The baud rate is the inverse of the minimum time between signal level transitions. In this example it is the inverse of the pulse duration or 1200 baud.

Assuming no errors, additional overhead or retransmissions, the throughput would be the same as the data rate, 700 bps.

These are the values seen at the source. At other points these rates may be different. For example, if we were considering the output of the transmitter or the signal measured on the channel we might say that the bit rate was 1200 bps because the distinction between

information bits and overhead (start/stop) bits may not be relevant. Communication systems often quote the bit rate on the channel, inclusive of all overhead, and the data rate seen at the source or sink can be significantly lower.