COMP102 Fall 2006
Assignment 2 – Solutions

1. Lempel-Ziv Encoding [20]

[10] (a) The following message was compressed using the Lempel-Ziv encoding seen in class. Decode this message.

Below are the steps for decoding. Underlined is the pattern specified by the first bracketed expression to the right of the decoded message. In italics is this pattern in the decoded message. The number added after the italicized pattern is the 3rd entry of the bracketed expression.

0100101(4, 3, 0)(8, 7, 1)(17, 9, 1)(8, 6, 1)
0100101***010***0(8, 7, 1)(17, 9, 1)(8, 6, 1)
0100101010***0101010***1(17, 9, 1)(8, 6, 1)
0100101010001010101***00101010***0(8, 6, 1)
0100101010001010101001010101***101010***1

[10] (b) Here is a part of a message that was compressed using the Lempel-Ziv encoding. Based on the information given, how long was the original message?

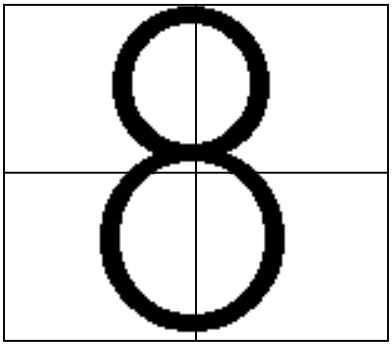
*#$(_, 3, #)(_, 6, $)
[Here "_" denotes an unknown number]

*The original message had length 3 (length of the original *#$) + 3 (length of the first pattern) + 1(the '#' to be added after the first pattern) + 6 (length of the second pattern) + 1(the '$' to be added after the second pattern) = 14.*

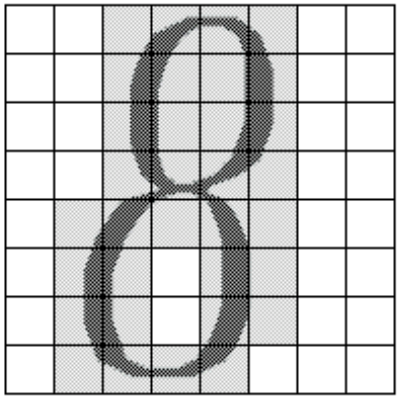
Note that while the first "_" could only be 3, the second "_" could be 6 or 7. This does not change the length of the decoded message.

2. Analog to Digital Conversion [20]

Suppose you were given the following hand-drawn image of the number 8 on the left:



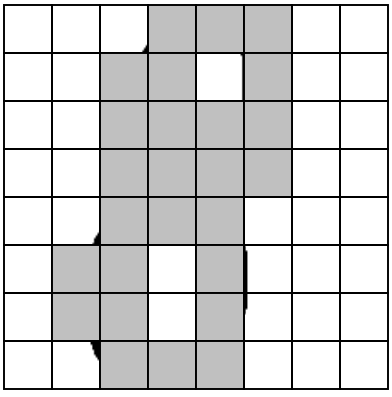
Your job is to convert this "analog" picture into a digital picture consisting of pixels. To do so, you will lay a square grid (like on the right) over the picture and then colour those pixels black that intersect any part of your drawing. Your digital image should be square (the image resolution should have as many height pixels as width pixels) and its width and height should be a power of 2 (2, 4, 8, 16, …, or 2^n for some n). What is the minimal resolution at which your digital image will capture exactly two of the holes in the "8", but not more? [10]



In the example above, we are using 8x8 pixels and only 1 hole is captured. 0 holes were present for resolutions 2x2, 4x4 and 2 holes will be present for the 16x16 resolution (as shown in class).

What different criterion could you use for digitizing your image that would capture both of the holes ~~using a smaller image resolution~~? [5] Show that your criterion produces a ~~smaller~~ *different* digital image of the drawing of the "8". [5] You can change the drawing of your "8" so that your answers work out. A digital hole is a region of white pixels surrounded by black pixels that are touching diagonally, vertically or horizontally.

A potential different criterion is only including a pixel that is overlapped by some "significant" amount. The amount could be the area of the pixel that is overlapped by your drawing. When this area is (roughly) 1/8 of the pixel, in the example below, we see that 2 holes are present in the digital image for the 8x8 resolution. ***The important thing to learn from this question is that the choice of resolution and the technique for analog to digital conversion used affect the quality of the produced digital data.***



3. Peer-to-Peer [30]

In class we talked about pure P2P networks and Client-Server type networks. There is also a third type called hybrid P2P networks. A hybrid network combines certain aspects of both pure P2P and Client-Server. For example, in a hybrid file sharing network all the sharing could be done in the P2P style; however, the searches to locate the files could only be executed on specific dedicated servers.

(a) In such a network is there a need for a JXTA-type rendezvous super peer? Please explain. [10]

Best Answer: Rendezvous super peer will not be needed for searching for files but will be needed for other tasks, such as searching for peers or some other resources on the P2P network.
Less correct answer, but still good: No, since the searches are done centrally on dedicated server. Rendezvous peers are not needed.

(b) Suppose the creator of the network found out that his or her network was used for exchanging a few documents (say 5), which are copyrighted. Is it possible to prohibit this behaviour without completely shutting down the network? Explain how this can be done. [10] Would the same be possible in a pure P2P network? [5]

Yes. Since all the searches are done centrally, one can force the central server to return nothing in case there is a search for one of these 5 files. No, the same would not be possible in a pure P2P network. Since there would be no central server, all information, including search indexes, would be distributed among multiple peers over which you have no control.

(c) Suppose a very large number of edge peers join the network. Assuming that all peers would want to actively exchange files, how would the network performance be affected? Hint: Use your combined knowledge of what would happen in a pure P2P network and a Client-Server network. [5]

Searches will be slower since there is a fixed number of specific dedicated servers. The download of files would be just as fast or faster since there are more peers from whom one can download files (assuming most peers behave socially and don't just try to download without sharing with others).

4. Cache [30]

In class we talked about CPU cache, this cache is used to speed up computations. There are other kinds of cache. For example, hard disk cache is used to speed up reading from the disk. Similarly, Microsoft Internet Explorer stores Web pages to speed up Web browsing in a special folder called “Temporary Internet Files” stored locally on your computer. This is another kind of cache.

You are to come up with strategies for filling and maintaining this cache. Specifically, you are to suggest 3 strategies for what information needs to be stored in the cache and/or how it should be managed (added/deleted/updated). Describe each strategy. What are its advantages and disadvantages? [24, 8 points per strategy] Which strategy would you choose to use if you were the designer of this cache? Why? [6]

Here are some possible strategies for deleting data from the cache when it becomes full:

- *Discard the least recently used item first from the cache. A disadvantage of this strategy is that it requires keeping track of when each item in the cache was used. An advantage of this strategy is that it stores information that reflects your most recent interests.*
- *Discard the least frequently used item from the cache. A disadvantage of this strategy is that it requires keeping track of the number of times a website was accessed. It also doesn't reflect a change in habits. Assuming that the computer user's habits don't change, this strategy would work well.*

Selecting which items to add/update:

- *Keep those items that take a long time to download. This saves time if this item is downloaded occasionally from a slow network. A disadvantage of this strategy is that large items take a long time to download, but also take a lot of storage space.*
- *You may want to only store small items. This saves cache space and allows you to save many items. However, these items may not take long to download and may not reflect your browsing habits.*
- *In case of dynamic websites that change their content with time (a news website), you may want to update these pages periodically. This is good because the content doesn't become outdated. This is bad because the process of updating the cache is expensive (in terms of bandwidth and CPU cycles) to perform and schedule.*

The best strategy would probably combine many of the above ideas. For example, it would keep those items whose size is not too big and that took a long time to download that you recently used frequently.