Data Structures and Algorithms -- M. Richards

1 Describe Larsen's method of dynamics hashing that enables a record to be located on a disk given its key using just one disk transfer and only a modest amount of information held in main memory. [10]

In Larsen's method each key has associated pseudo random sequences of probe and signature values. Discuss what properties these sequences should have and outline an algorithm that could be used to compute the \$n^{th}\$ probe/signature pair for a given key, which you may assume is a character string. [6]

Briefly discuss why is Larsen's method is not used in most current filing systems. [4]

Answer: Bookwork The probe sequence should be a pseudo random permutation of the disk block numbers. The signature sequence shold consist of uniformly distributed random numbers in the range 0 to SigMax. Algorithm: first hash the key to give a positive integer, then add n*
big prime> and take the remainder after division by the number of blocks. The nth signature could take the remainder after division by the number (MaxSig+1). It would probably help to use a different large prime.

Larsen's method is good for reading from a disk, writing may be very expensive. It good for records that are all about the same size. It would not keep related files close together on the disk.

\newpage

2 Outline the mechanism used in the Burrows-Wheeler block compression algorithm, illustrating your description by applying it to the string ALFALFA [14]

Briefly discuss the advantages and disadvantages of the Burrows-Wheel algorithm compared to other commonly used compression methods. [6]

Answer:

ALFALFA. ALFALFA
LFALFA.AL ALALFA.ALF
ALFA.ALF ALFALFA.
LFA.ALFA FA.ALFAL
FA.ALFAL FALFA.AL
A.ALFALF LFA.ALFA
ALFALFAL LFALFA.AL
ALFALFAL LFALFA.A

Last column: AFF.LLAAAA

Use Move-to-front buffering then Huffman

Send new Huffman table every 100k bytes or so to make it adaptive.

Decoding can be done in linear time:

1[i] p[i] 0 0 0 F 1 F 1 L 2 1 .0 A0 A1 A2 F0 F1 L0 L1 A0 F0 .0 L0 L1 A1 A2 **A**0 L0 **A**1 Α F F1L1 L .0

PROSE

Needs larger blocks (100k to 4M) to be effective.