**Algorithm Engineering**

**4 June 2013**

**Exercise [rank 6]**

Describe the reservoir-sampling algorithm, comment on its correctness by showing that items are sampled with probability m/n.

**Exercise [rank 6]**

Assume you are given 6 strings (aa, ab, bb, bc, ca, db) and you wish to construct a minimal ordered perfect hash function (MOPHF). Assume that rank(x)=1,2,3,4 for the characters x=a,b,c,d respectively. Given a string x’x’’ of two characters, we let the two random functions required by the design of MOPHF as h1(x’ x’’) = rank(x’) \* rank(x’’) mod 11 and h2(x’ x’’) = rank(x’) + rank(x’’) mod 11 (so m=11). Design a proper g(t) to construct the final h(t), if this is possible given h1 and h2 above.

**Exercise [rank 6+3]**

Describe the Snow Plow algorithm, comment in which sorting algorithm is used and how it is used, and quantify the expected advantage in terms of I/Os it induces over the sorting-algorithm that uses it.

Simulate the working of Snow Plow over the sequence 2,5,4,7,1,4,2, and show which sorted blocks it forms with a memory of size 2.

**Exercise [rank 6+3]**

Describe the LSD-Radix sort, prove its correctness, and indicate its time complexity as a function of the key length b (in bits) and the number of keys n.

Simulate its working over the sequence of 4 strings (1010,0011,0111,0000) by assuming to group digits 2-bits at a time.

**Exercise [rank \*]**

Prove that the expected reconstruction time of cuckoo hashing is O(n), if it is built over n items and we consider a sequence of a\*n insertions, where a is a constant.