Applied Algorithms
CSE 589, Autumn 1997

Anna Karlin, 426C Sieg

Homework # 6
Due December 3

1. When you buy a ticket in the State Lottery, you choose six different numbers between 1 and 36. The lottery officials keep a dictionary keyed on the set of six numbers chosen on each ticket. After the officials pick the winning numbers, they access this dictionary to identify the winning ticket or tickets, if any. Since millions of tickets are sold, the officials have decided to keep the dictionary in external storage with a directory in an internal hash table. Their computer consultant has recommended that they use the hash function

$$h(x_1, x_2, \ldots, x_6) = (x_1 + x_2 + \ldots + x_6) modm$$

where m is the number of external buckets in which the records will be stored. Give a critique of this recommendation, and suggest a better alternative.

2. How long could it take in the worst case to insert N keys into an initially empty table of size M, using separate chaining with (i) unordered lists and (ii) ordered lists?

3. How long could it take in the worst case to insert N keys into an initially empty table of size M, using linear probing? How about in the average case?

4. Why might $h(K) = K \text{ mod } M$ be a better hash function when M is chosen to be prime?

5. What is the chief advantage of universal hashing over hashing with a fixed hash function?

6. Suppose that you want to compute fingerprints using Rabin's scheme of 2^{10} strings each of length at most 128 bits. How many bits should the fingerprint be in order to keep the probability of a collision below 2^{-12}?

7. Suppose you create a Bloom filter using one random hash function (as on slide 48). Suppose that the table has size M and that N keys have been inserted. (The table is initialized so that $B(i) = 0$ for all $0 \leq i < M$, and then, for each key, $B(h(K))$ is set to 1.) What is the probability that on an "Exists" query to a particular key K_1 which is in the set, the answer returned is false? What is the probability that on an "Exists" query to a particular key K_2 which is not in the set, the answer returned is false?