Slide 4

The switch example. I can’t save a dictionary of all the macs and their results.

Slide 5

The challenge: on seeing I don’t know the total sum of numbers. On the other hand I can calculate the sum but I forget the numbers.

Slide 6

If the index was change to so clearly the index was chosen with the right probability.
Otherwise ( not changed), the index is selected with probability of:

 which is also correct

Calculating sum is ok –

Slide 8

Variance calculation:

Slide 10

Example: Each might represent a credit card number extracted from a sequence of
credit card transactions and we wish to determine how many distinct credit card accounts there are.

Lemma: If are of different sizes, then clearly this implies an error for one of the input sequences. On the other hand, if they are the same size, then if the next symbol is in but not, the algorithm will give the same answer in both cases and therefore must give an incorrect answer on at least one of them.

Slide 13

It follows that a set of hash functions H is 2-universal if and only if for all x and y in {1,2,…,m} ,x!=y, h(x) and h(y) are each equally likely to be any element of {0,1,2,…,M-1} independently

Proof of the example:


Slide 14





Slide 16



Slide 17

Slide 18

*Previously:* , =>
And then using Chebyshev’s inequality.
Chebyshev’s inequality: =>

Slide 23

If there is a need to prove the variance on board:


Minimization- In the important special case when , pick columns of A with probabilities proportional to the squared length of the

columns. Even in the general case when B is not , doing so simplifies the bounds, so we will use it