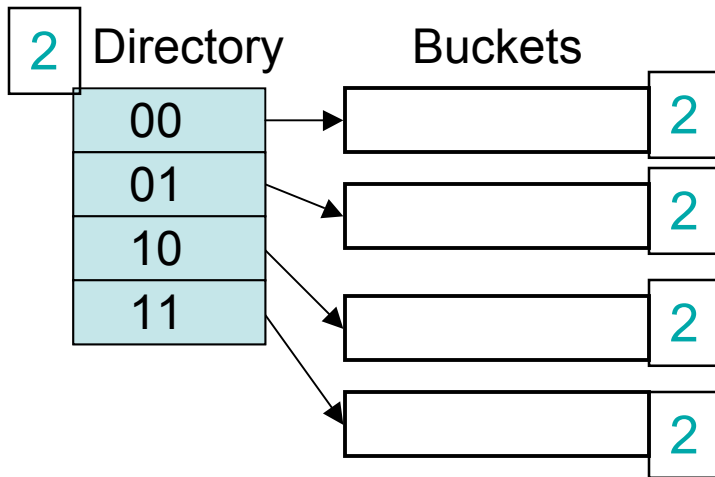


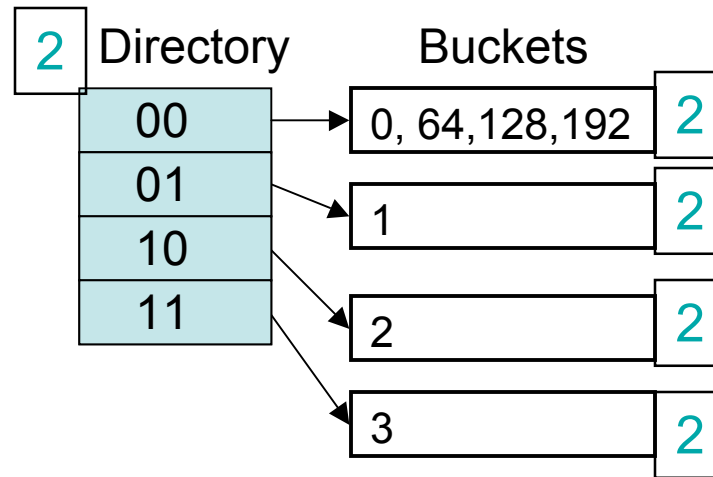
# COS 597A, Fall 2008: Problem Set 5, Problem 4 Solutions

## Part a:

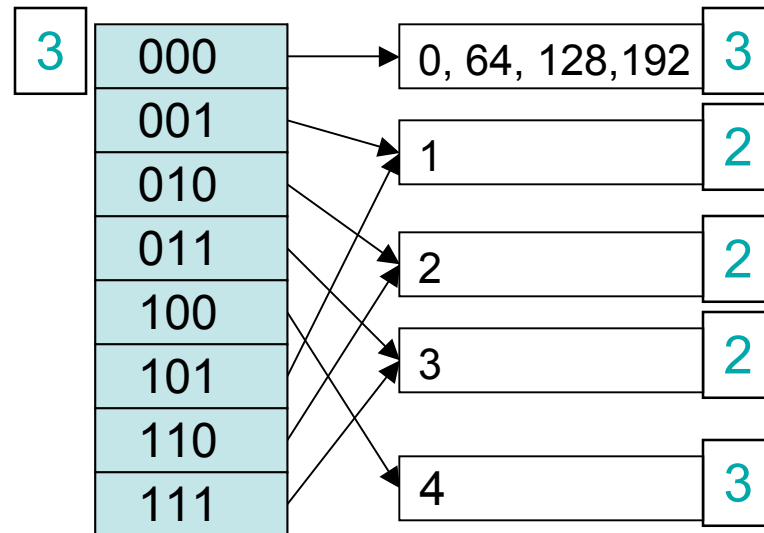
Initial:



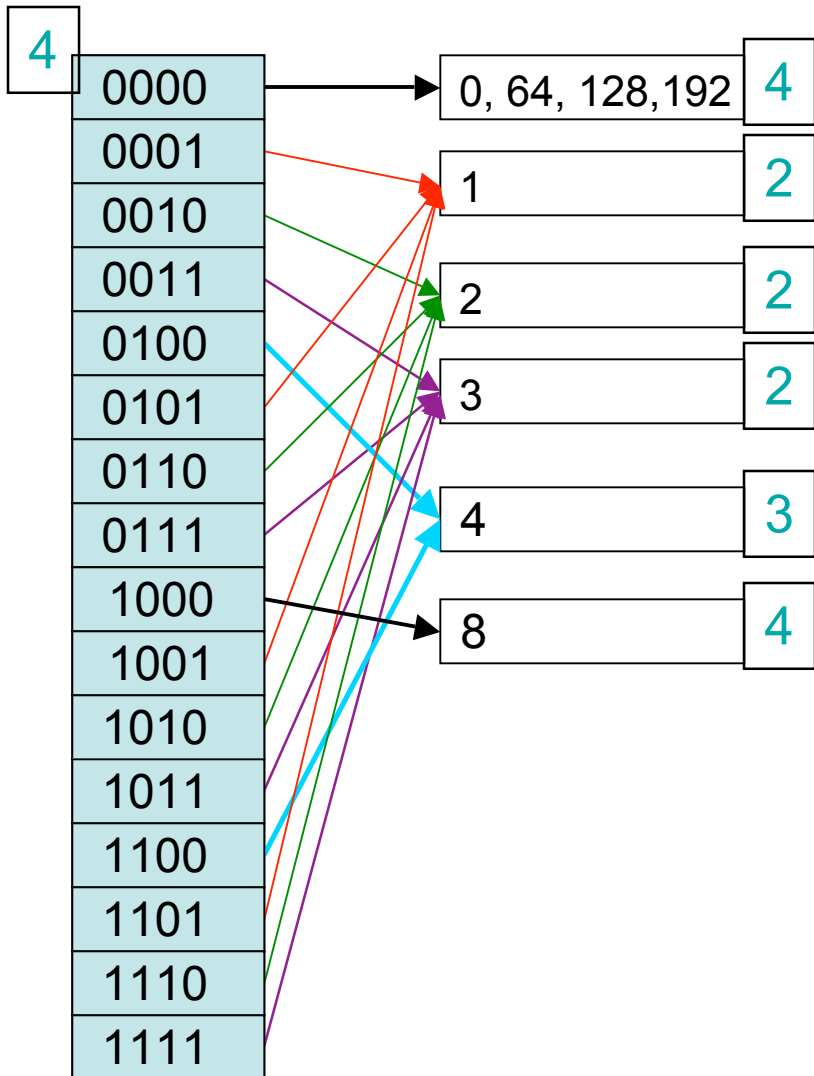
Just before first directory change:



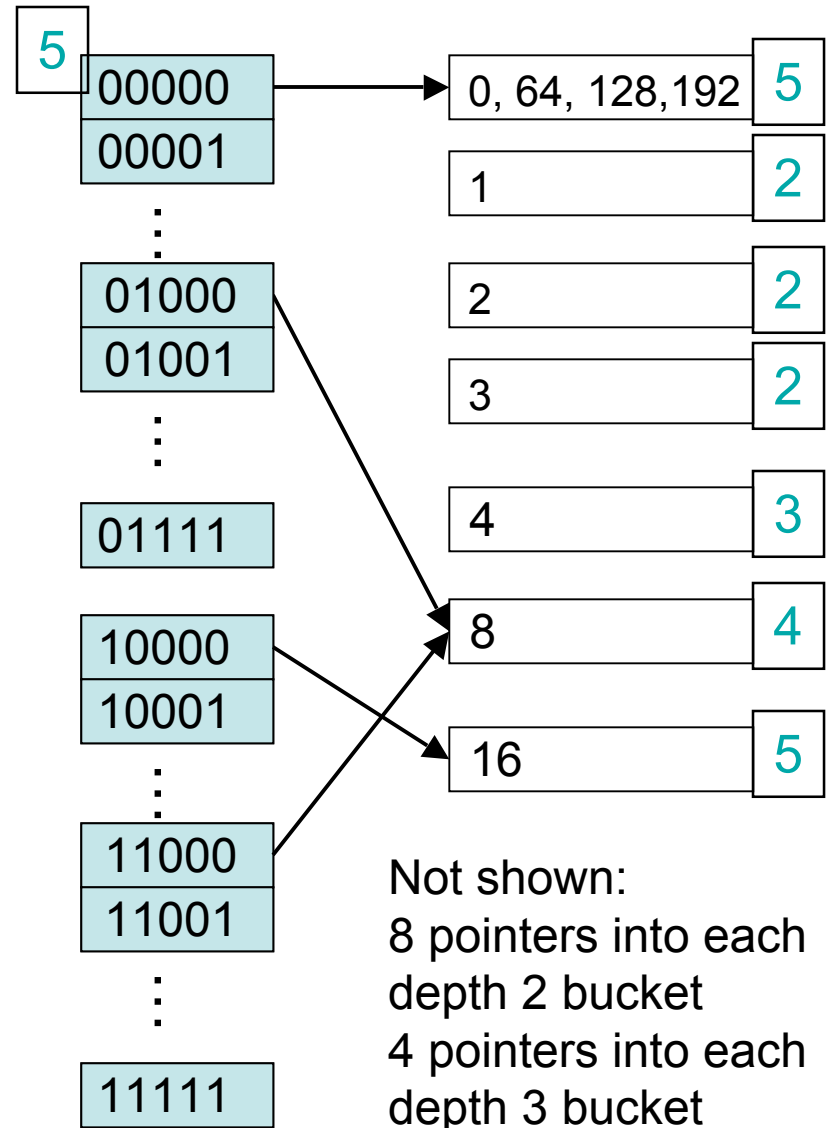
After first directory split



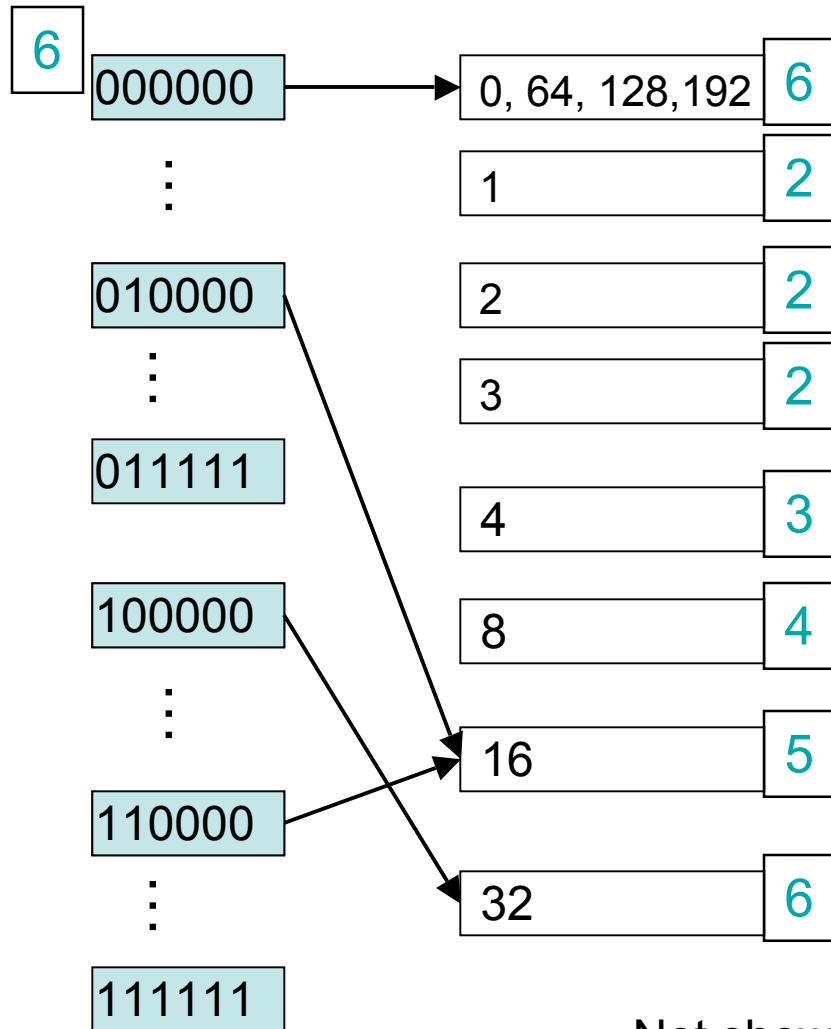
After second directory split



After third directory split\*



After fourth (final) directory split\*



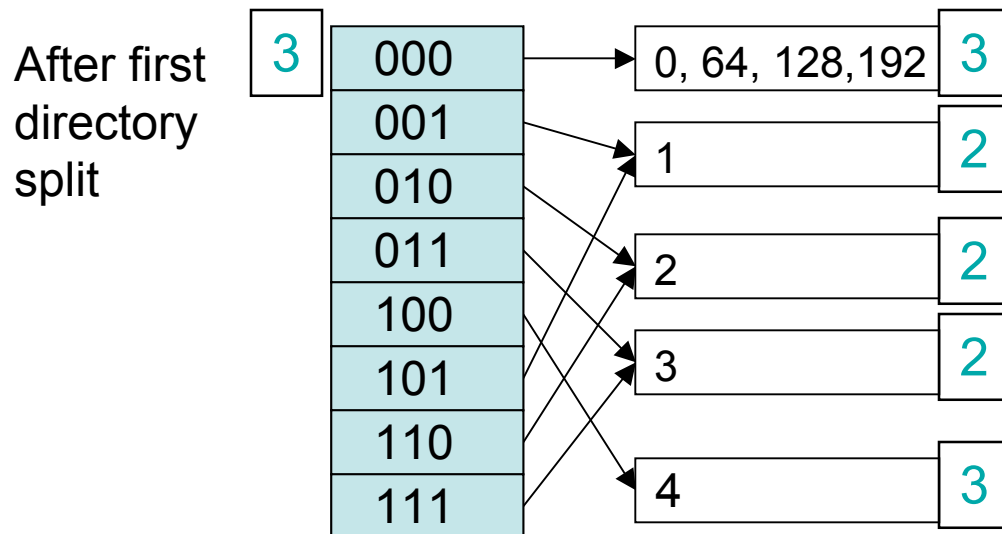
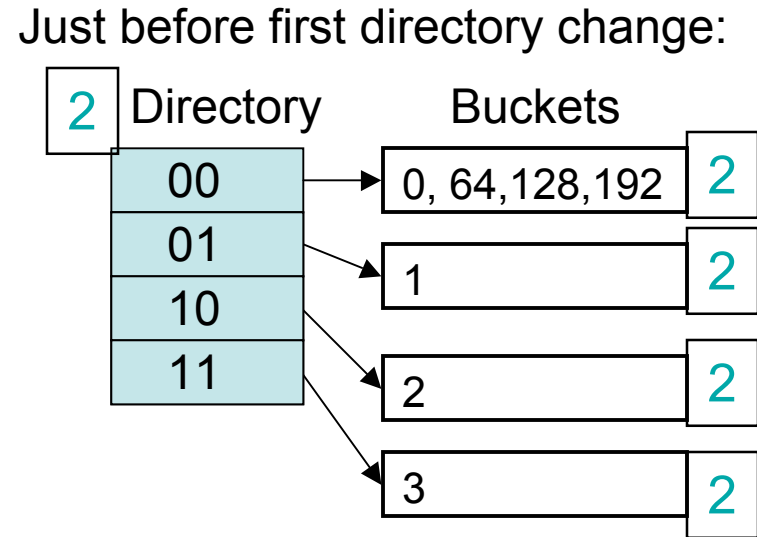
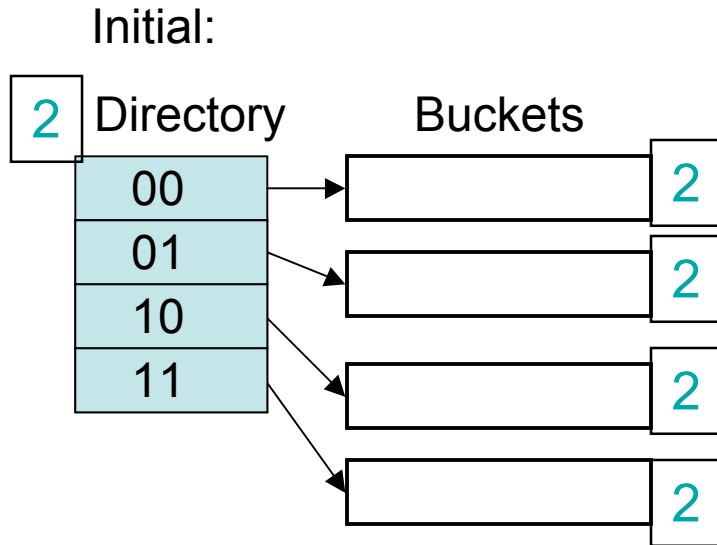
Number of blocks used:  
8 for buckets  
+ 1 for the directory  
= 9 blocks used

Not shown:

16 pointers into each depth 2 bucket  
8 pointers into each depth 3 bucket  
4 pointers into each depth 4 bucket

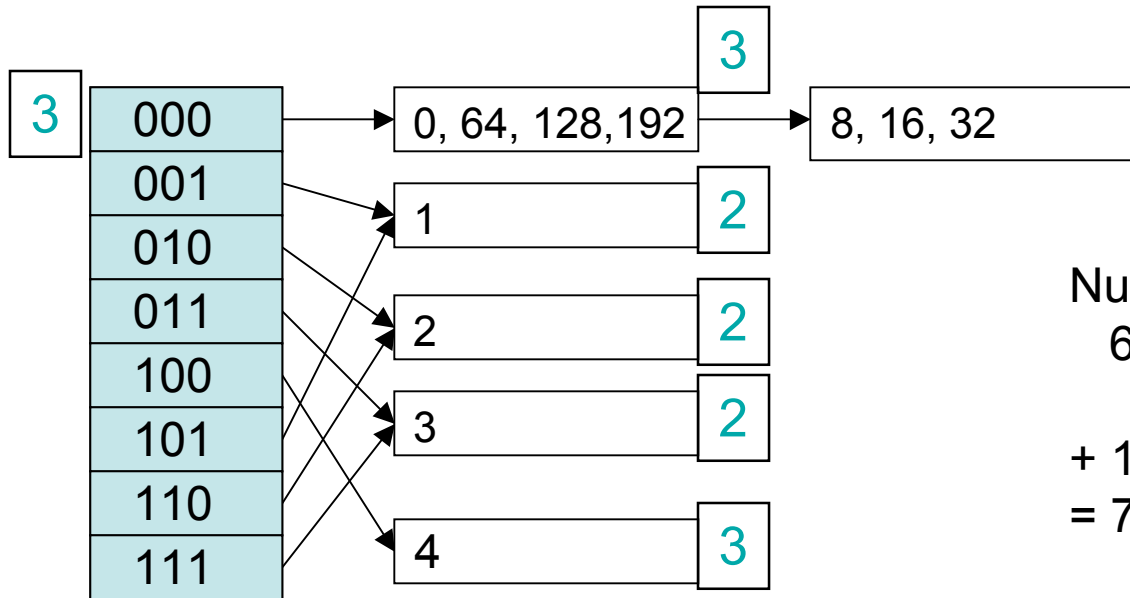
## Part b:

The first split occurs as for Part a:



At this point, 3/5 of the buckets are at depth 2, less than the depth (3) of the directory. The next three insertions all fall in bucket 000, but will not cause a split.

After inserting 8, 16, 32:



Number of blocks used:  
6 for buckets, including the  
overflow block,  
+ 1 for the directory  
= 7 blocks used.

In this example, the average number of blocks read per look-up (excluding the directory block) is  $(8+6)/11= 1.3$ . This is a fair trade if disk space is a concern. However, the criterion might be too biased against splitting -- allowing long chains of overflow blocks. I would consider modifying the criterion to split a bucket even if over half the buckets have depth less than the directory in the case that the bucket would have more than one (or two?) overflow blocks if not split.