

Count sort / bucket sort can be done  
 in  $O(n+m)$  time  $n$ : # elements  
 $m$ : # buckets

Best scenario  $n \sim m$

For sorting  $n$  strings with lengths  
 $l_1, l_2, \dots, l_n$ ,  $\max_i l_i = L$

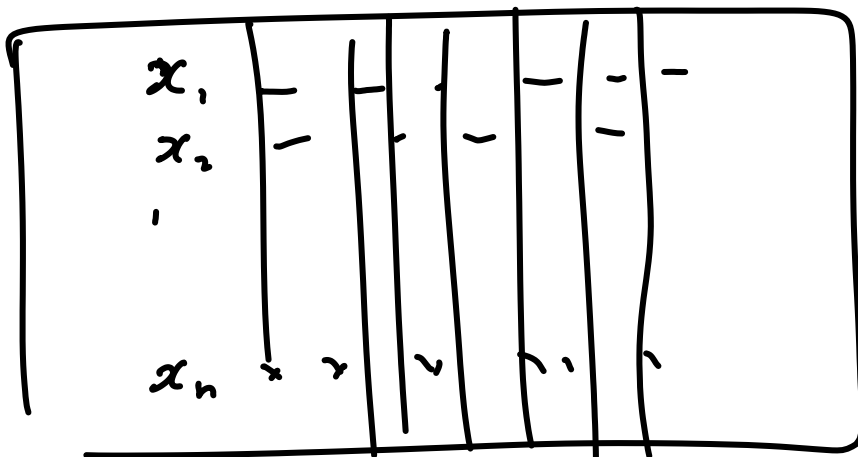
straight forward radix sort would cost

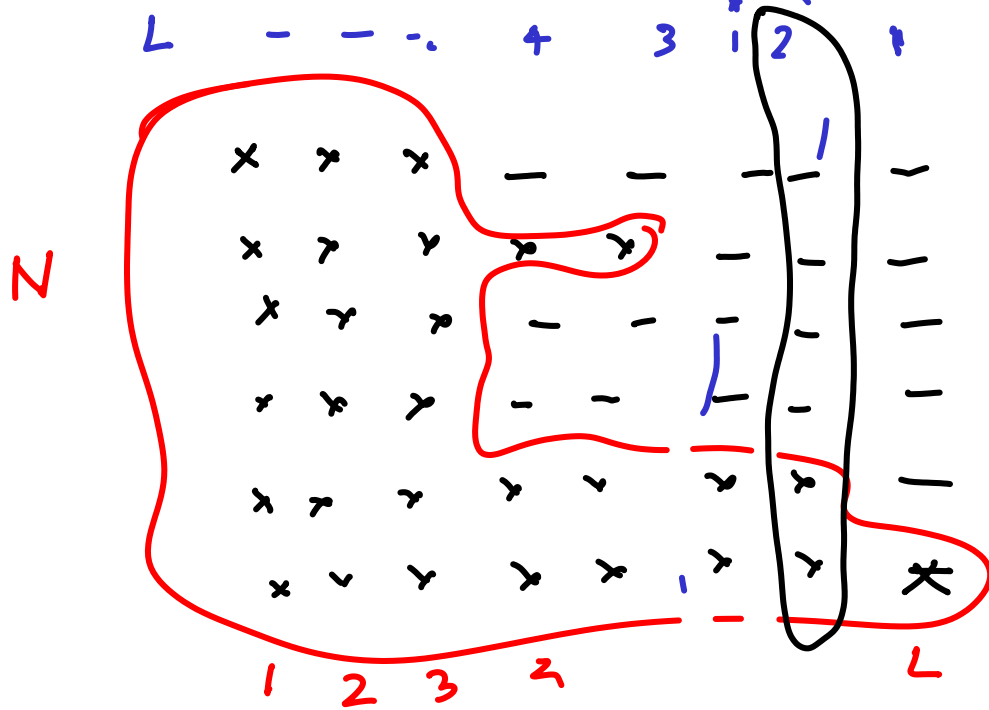
$$L \times (n + |\Sigma|)$$

Input size:  $\sum l_i = N$

Suppose  $l_1 = l_2 = \dots = l_{n-1} = 1, l_n = n$   
 $\Rightarrow L = n, \sum l_i = 2n - 1$

$$\sim n^2$$





Let the # of non-blank characters in position  $j$  be  $m_j$ . If we can sort in time proportional to  $O(m_j)$  for the  $j^{\text{th}}$  iteration of radix sort  $\Rightarrow \sum_{j=1}^L O(m_j) = O(\sum l_i)$

1. How do we identify the strings that have non-blank characters in the  $j^{\text{th}}$  index
2. We should show the non-empty buckets corresponding to  $j^{\text{th}}$  index  $|\Sigma|$  is constant.

$$O(m_j + |\Sigma|)$$

Observation: for string  $s_i$  with length  $l_i$ , it will start participating from round  $L - l_i$

Consider string <sup>(1)</sup> bat, <sup>(2)</sup> cave, <sup>(3)</sup> at

- (1)  $\rightarrow$  bat
- (2)  $\rightarrow$  cave
- (3)  $\rightarrow$  at

Classifying - them according to lengths implies bucketing them in 1, 2, ... L

Bucket sort :  $O(n + L) < N$

~~bat : <sup>1</sup>(3,b) <sup>1</sup>(2,a) <sup>1</sup>(1,t)~~  
~~cave : <sup>2</sup>(4,c) <sup>2</sup>(3,a) <sup>2</sup>(2,v) <sup>2</sup>(1,e)~~  
~~at : <sup>3</sup>(2,a) <sup>3</sup>(1,t)~~

Sort - the pairs using radix sort (2 digits)

#pairs : N  $\leftarrow$   $O(N + |\Sigma|) + O(N + L)$   
 is  $O(N)$  provided  $|\Sigma| < N$

~~<sup>2</sup>(1,e) <sup>1</sup>(1,t), <sup>3</sup>(1,t) <sup>1</sup>(2,a) <sup>3</sup>(2,a) <sup>2</sup>(2,v), <sup>2</sup>(3,a) <sup>1</sup>(3,b)~~  
~~<sup>2</sup>(4,c)~~

Should be left adjusted

4 3 2 1  
 bat : (4,b) (3,a) (2,t)  
 cave : (4,c) (3,a) (2,v) (1,e)  
 at : (4,a) (3,t)

$$^3(4,a)^1(4,b)^2(4,c)^1(3,a)^2(3a)^3(3,t)^1(2,t)^2(2,v)^2(1,e)$$

Each round can be done in time

$$O(m_j)$$

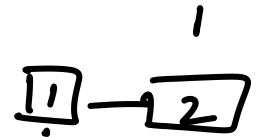
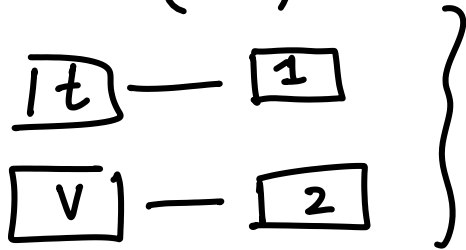
since

$$\# \text{ buckets non-empty} \leq m_j$$

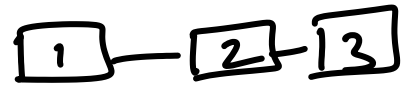
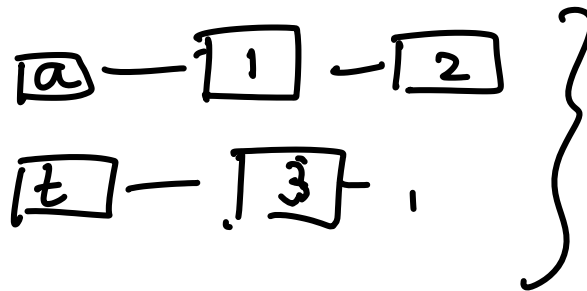
1<sup>st</sup> round

(1,e)

2<sup>nd</sup> round



3<sup>rd</sup> round



4<sup>th</sup> round

