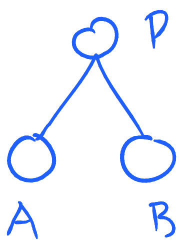
Problem 1 Hoffman Code

Thursday, October 20, 2016 1:00 PM

Given: P(A) > P(B), Where P(x) is the probability of x
Prove: L(A) <= L(B), where L(x) is the length of the Huffman encoding of x

Case 1 (Trivial)

When A and B are merged together

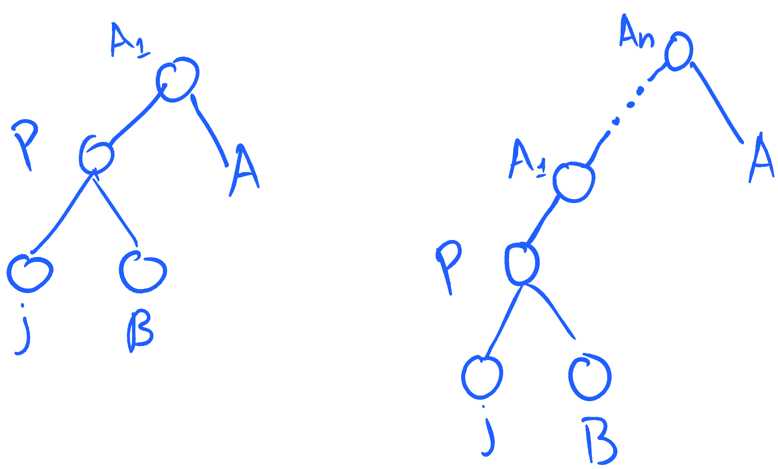


Where node P is the parent of A and B
L(A) = L(P) + 1
L(B) = L(P) + 1
L(A) = L(B)

Case 2 (Trivial)

When A is merged with the parent or any ancestor of B

Because P(B) is < P(A), we know that B will be merged before A (excluding Case 1)



From this case, we can clearly see that L(A) < L(B)

When A is merged with P:

L(A) = L(A1) + 1
L(B) = L(A1) + 2, which is > than L(A)

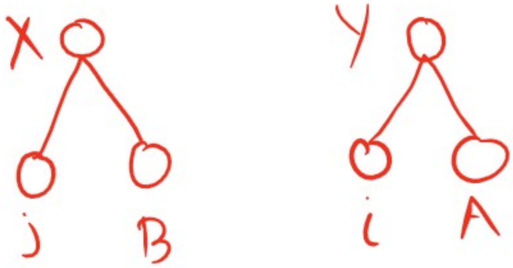
When A is merged with An:

L(A) = L(An) + 1
L(B) = L(An) + n + 2, which is also > than L(A)

Case 3

When A is not directly merged with the parent or any ancestor of B

Consider the parents of A and B



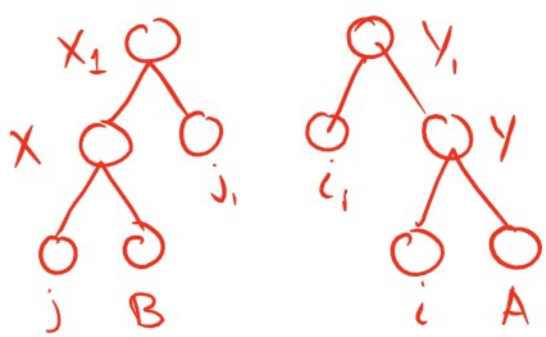
We will first prove that P(Y) > P(X)

- Facts: The Huffman algorithm merges the two symbols with the lowest probabilities at every recurrence.
- This implies that j and B were the 2 symbols with the lowest probabilities at one point of the code.
- Since we know that P(A) is greater than P(B), we can conclude that the merge including B occurred before the merge including A.
- This implies that when j and B were merged, i and A were still available symbols, and therefore have greater probabilities than j and B
- P(A) > P(B), P(A) > P(j)
- P(i) > P(B), P(i) > P(j)
- It is then easy to see that P(A) + P(i) > P(B) + P(j)
- Since P(Y) = P(A) + P(i), and P(X) = P(B) + P(j)
- Through substitution we have P(Y) > P(X)

Now we will prove by contradiction that L(A) <= L(B) by assuming L(A) > L(B)
Consider the lengths of the parents of A and B

L(Y) = L(A) - 1
L(X) = L(B) - 1

If we assume L(A) > L(B) then we can see that L(Y) > L(X)
We will now consider the parents of X and Y

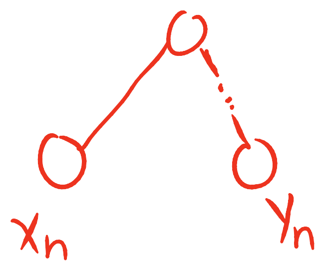


Since P(Y) > P(X) we can prove that P(Y1) > P(X1) with the same proof we showed earlier
We also know that L(Y1) = L(Y) - 1, L(X1) = L(X) - 1, and therefore L(Y1) > L(X1)

We will now propose an inductive hypothesis:

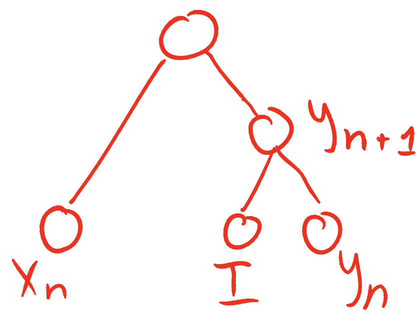
P(Yn) > P(Xn) and L(Yn) > L(Xn)
where Yn and Xn are the nth parent from A and B respectively

After a finite number of recursions, we will reach a point where Xn reaches the maximum value for n.
In other words, we will consider the point where the nth parent of B is a child of the ancestor of A and B:



At this point, we know through induction that P(Yn) > P(Xn) and L(Yn) > L(Xn) (through our assumption that L(A) > L(B)).

For the condition L(Yn) > L(Xn) to be satisfied, there must be at least n+1 parents from A.



We know that P(Yn) > P(Xn), however I is merged with Yn instead of Xn, forming a contradiction.
We can therefore conclude that L(A) !> L(B) which proves that L(A) <= L(B)