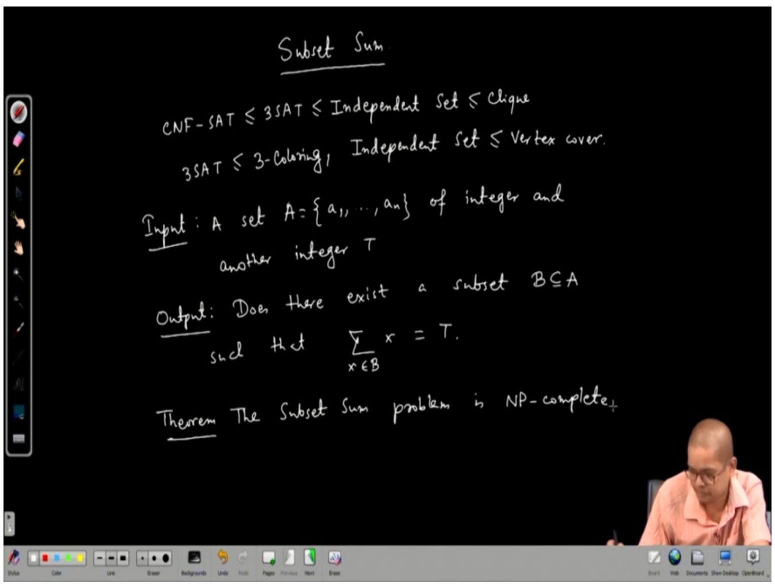
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 37
NP – Completeness of Subset Sum and Knapsack

Welcome, so from the last couple of lectures we have been seeing many NP complete reductions and we have seen Boolean satisfiability problem and various graph problems and today we will see a new problem a numerical problem and show that it is also NP complete.

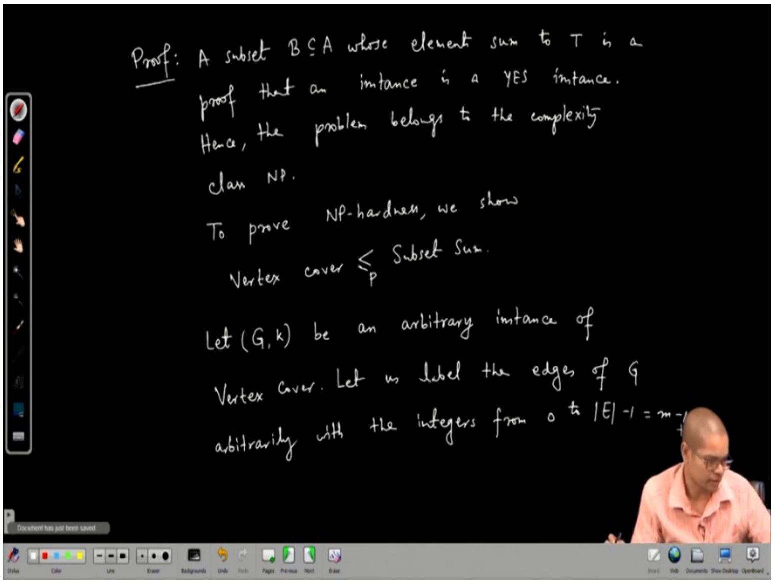
(Refer Slide Time: 00:47)



So, today's problem is subset sum but before that let us summarize what are the reductions we have seen so far, we have seen reduction from CNF-SAT to 3SAT to independent set to clique and we have also seen from 3SAT to 3 colour ability 3 colouring. So, today we see a reduction showing that the subset sum problem is NP complete or NP hard. So, what is the problem in problem formally? Input and we have also seen from independent set to vertex covered.

For subset sum the input is a set $A=\{a_1,\ldots,a_n\}$ of integers and another integer say T, output does there exist a subset B of A such that the sum of integers in B sums to exactly T. So, this is the subset sum problem and we will show today that this NP complete theorem. The subset some problem is NP complete.

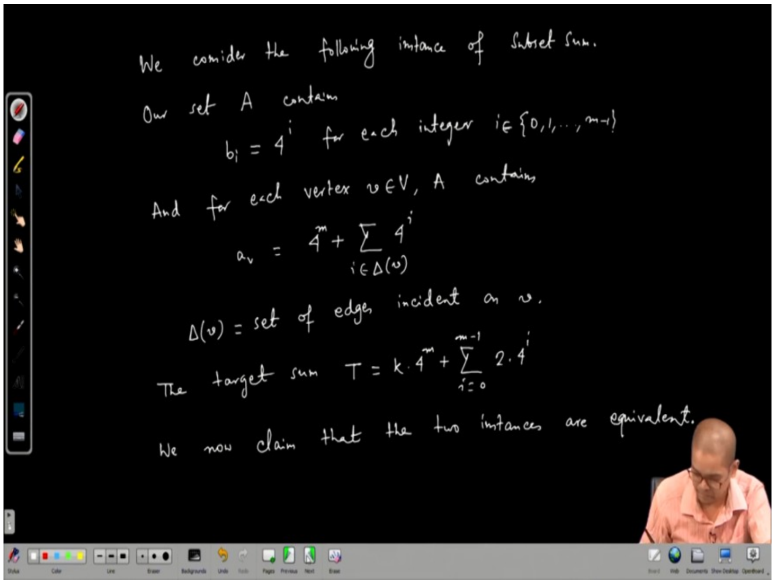
(Refer Slide Time: 04:34)



Proof, again if the instance is a YES instance a subset whose sum whose elements sum up to T can itself act as a proof for the easiness of the instance. Hence membership in NP is clear, subset B of A whose elements sum to T is a proof that an instance is a YES instance. Hence the problem belongs to the complexity class NP. To prove NP hardness, we exhibit many to one polynomial time card production from vertex covered.

To prove NP hardness, we show that vertex cover many to one polynomial time cover produces two, subset sum. So, let I need to start with an arbitrary instance of vertex cover and convert it to an equivalent instance of subset sum. Let (G, k) be an arbitrary instance of vertex cover. So, let us level the edges of G arbitrarily with the integers from 0 to say $|E|-1$. Let us suppose cardinality of E is m, so that is it is like m - 1.

(Refer Slide Time: 08:56)

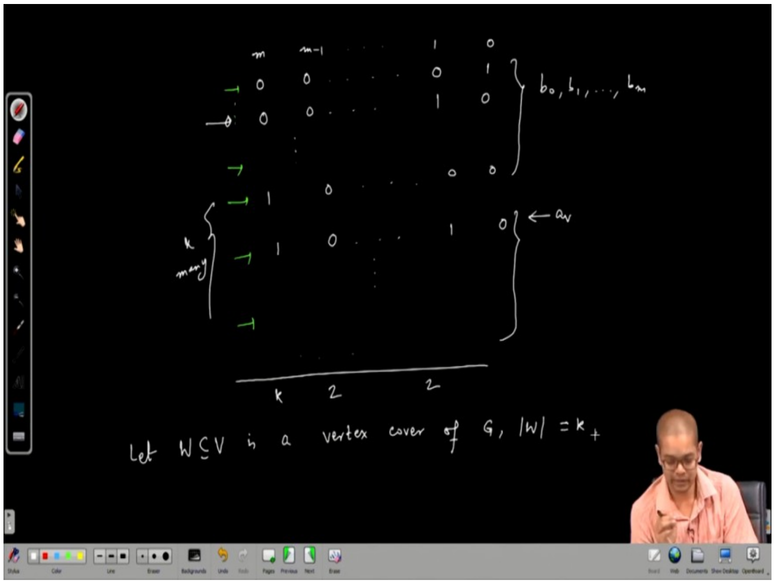


Now we have the following, we considered the following instance of subset sum. So, the instance contains a set of numbers and a target sum. So, what are the set of numbers? So, this set our set A contains $b_i = 4^i$ for each integer $i \in \{0, 1, 2, \ldots, m-1\}$. So, this is the set using which we have levelled the edges and for each such level of the edge I have an integer which is 4^i.

And for each vertex $v \in V$, A contains a number a v which is 4^m, m is the number of edges in the graph plus summation 4^i. Where $i \in \Delta(v)$ where $\Delta(v)$ is the set of edges incident on v. So, these are the numbers and the target sum T is $k\, 4^m + 2 \sum_{i=1}^{m-1} 4^i$. So, this is the subset some instance.

We now claim that two instances are equivalent but before proving that there is there exist a pictorial view of these numbers.

(Refer Slide Time: 13:10)



That means if you write it in base 4 form then this is the m-th number m - 1 1 0, so for each i I have a number which is 4^i that means I have a number which is all these bits are 0 this and this is 1, this is 1 0 and so on here 1. So, these are you know this b_i's $b_0, b_1, b_2, \ldots, b_m$ and then a v is for each V you know some coordinates are 1. So, if the mth edge belongs to a vertex whose corresponding this is a v if m-th edge is incident on v.

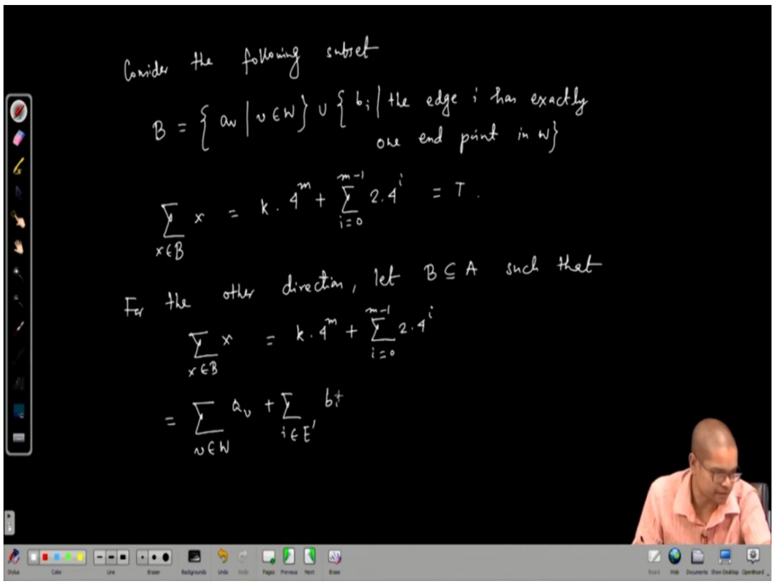
Then m-th coordinate is one this is also 1 m - 1 if it does not incident then it is 0 and so on. So, there is a bit vector a b can also be thought of as a bit vector of the of the edges indexed by edges which edges are incident on a v. So, there is a bit vector of $\Delta(v)$, so this for each vertex I have such thing. And I need to select you know k vertices exactly k vertices and that is when this sum here I will have k here k many ones here.

And you know each edge will be covered. So, all the edges will be picked you know and if each if this k vertices k many. If for an edge if both endpoints are in are on this k size set then I do not need to pick the corresponding edge from here on top $b_0, \ldots, b_{m-1}$. On the other hand, if at least one endpoint is incident on this, k vertices then I can pick the edge from top and that way you know in that column I will get 2, that is the target.

Now in each column for the edge I need to get 2, so that is the high level idea of the equivalence. So, now let us formally prove it. So, now to show the equivalence let so suppose that the vertex cover instance is the YES instance and let W subset of V is vertex covered of G and size of W is exactly k. If it is YES instance that means there exist a vertex cover of size at most k if it is less than k, W is a vertex cover less than k.

I can always add more vertices to it to ensure that the size of W is exactly k, we can assume without loss of generality that k is less than equal to n number of vertices.

(Refer Slide Time: 17:21)

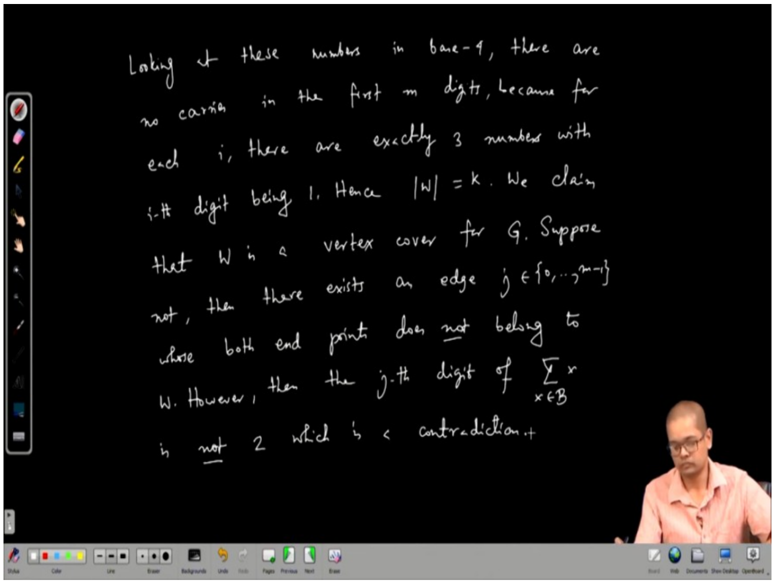


So, now consider the following subset. The following subset B equal to you know pick all those vertices pick all those numbers correspond to the vertices which are in the vertex cover and those integers whose exactly 1 end point is belongs to W. The edge i has exactly 1 end point belonging to W. So, now if I sum the integers in B, so you can see from this picture you know this mth column has k many ones.

So, this will be k times 4 to the power in plus, now each other column you know if it has exactly two ones. So, this is $k\,4^m + 2\sum_{i=1}^{m-1} 4^i$ which is exactly T. Hence, we have proved that if the vertex cover instance is at YES instance that means that implies that the subset sum instance is a YES instance. Now for the other direction, let B be a subset of A such that sum of x in B is $k\,4^m + 2\sum_{i=1}^{m-1} 4^i$.

Now again think of this as this number in base 4, so what is this B? This B is a subset of A, it contains some sub some numbers from the set a v and some numbers from b_i. So, this must be equal to v in W a v + i in some edges some set of edges b_i, this is how it should look like.

(Refer Slide Time: 21:41)

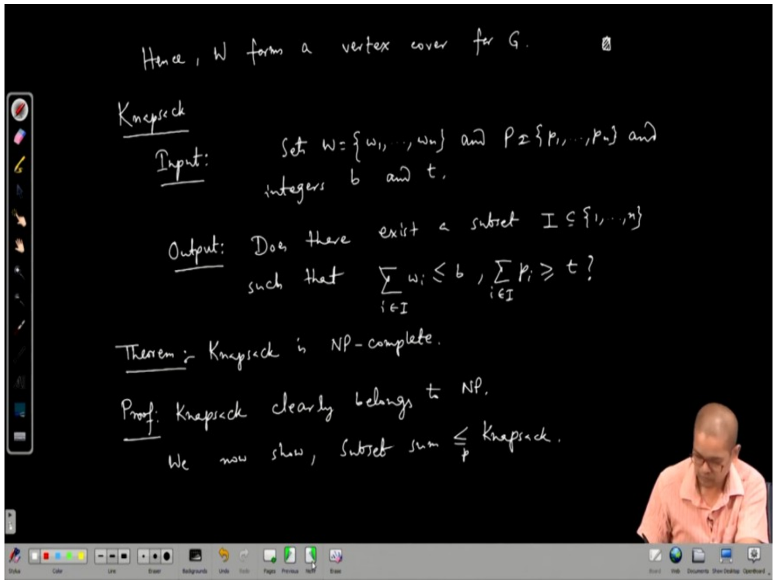


Now the crucial thing to observe that looking at this numbers in base 4, there are no carries in the first m digits, why? Because for each i, come back here these numbers for in each column how many 1s are there, there are 3 1s in the first m - 1 column 0 to m - 1 in each column there are exactly three 1s for every i. In b_i that number has a 1 there and this edge i is incident on two vertices. So, those two vertices will have one there.

So, each column has exactly three ones for each i the column i or let me write this way for each i, there are exactly three numbers with ith digit being 1. So, there cannot be any carry and hence for then for 4 to the power m, this coefficient for it to be k the number cardinality of W should be k. Hence cardinal will be W because there is no carry should be equal to k. Now we claim that W is a vertex cover for G, so it is a proof by contradiction.

So, suppose not, then there exist an edge let us call it edge $j \in \{0, \ldots, m-1\}$, we have labeled the edges from 0 to m - 1. So, there is an edge whose both endpoints does not belong to does not belong to W but then how come the j-th, how come the j-th column has two ones because now look at this sum each column must have two ones and even if I pick b_j you know the other two numbers where the j-th is one those are not picked but then however then the j-th digit of the sum is not 2 which is a contradiction.

(Refer Slide Time: 27:36)

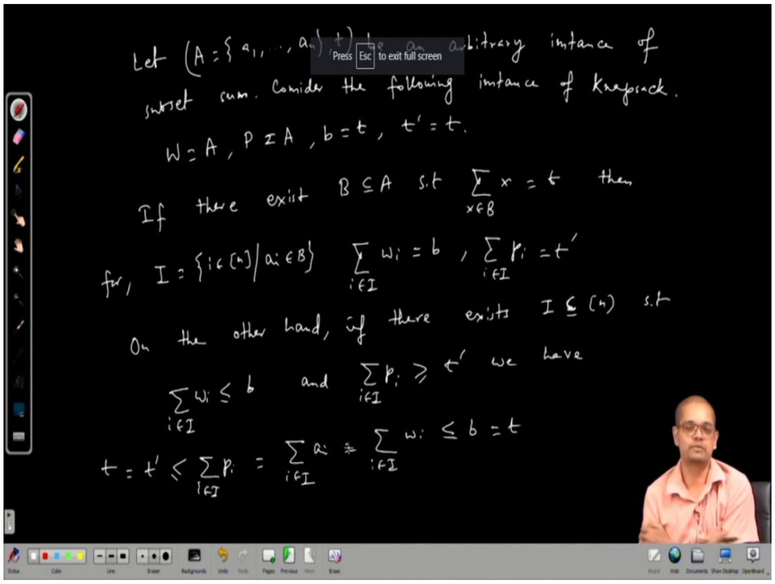


So, hence W forms a vertex covered for G, hence these two instances are equivalent this concludes the proof that the subset sum problem is NP complete. Now using subset sum we can we prove another problem, another popular problem which is called the knapsack problem to be NP complete. So, what is the problem instance? The input is set of objects, an objects with weights a set of the weights the integers are set of weights $W=\{W_1,\ldots,W_n\}$, so this is one set.

So, let us call it sets W these are the weights and profit $P=\{p_1,\ldots,p_n\}$ and integer integers b and this is the back size and T targeted profit. Output you know does there exist subset I of 1 to n such that $\sum_{i\in I} W_i \leq b$, their profit $\sum_{i\in I} p_i \geq T$. So, again this problem is NP complete and this can be easily shown by a reduction from subset sum.

So, theorem knapsack is NP complete, again membership is obvious any subset of items which with two sum of weights is at most p_n sum of profit is at least T is a proof for Yes. So, knapsack clearly belongs to NP, we now show that subset sum many to one Karp reduces two knapsack.

(Refer Slide Time: 32:17)



So, let $A=\{a_1,\ldots,a_n\}$, t be an arbitrary instance of subset sum, then consider the following instance of knapsack. What is that instance? That W weights is A and profit is also A and but back size is t and target t prime is also t. Now it is easy to see that if there exist of B subset of A such that $\sum_{x\in B} x$ is t, then you know the index I which is i in n such that a_i is in B, then for this we have $\sum_{i\in B} w_i=b$ and $\sum_{i\in I} p_i=t$.

Hence if the subset sum instance is the YES instance, then knapsack is a YES instance, on the other hand if there exists I subset of n such that $\sum_{i\in I} w_i$ is less than equal to b and $\sum_{i\in I} p_i\geq t$. We have $\sum_{i\in I} a_i$ is $\sum_{i\in I} w_i$ this is less than equal to b but b's value is t this t' and this is $\sum_{i\in I} a_i\geq t'$ but t' is also t.

So, hence all this i have a chain of inequalities where both endpoints are same that means $\sum_{i\in I} a_i$ must be equal to t all must be equal to t and hence the subset sum instance must be a YES instance which concludes the proof. So, we will stop here today.