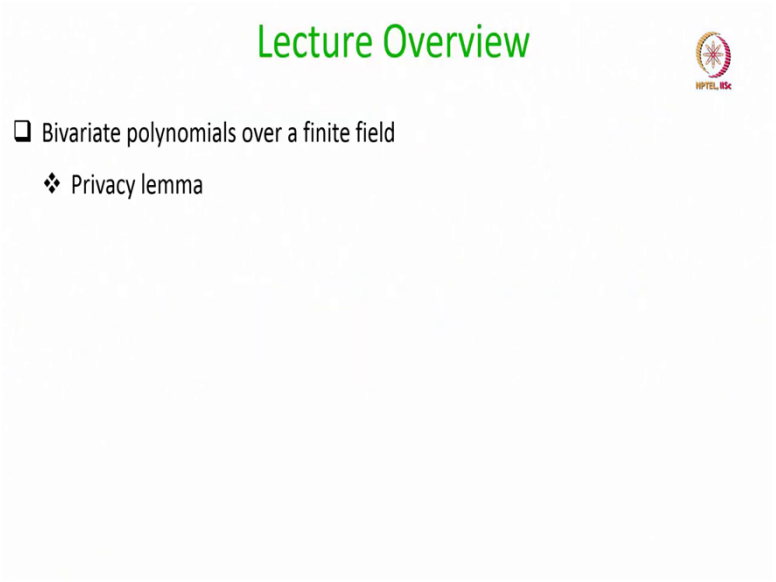
Secure Computation : Part II
Prof. Ashish Choudhury
Department of Computer Science and Engineering
International Institute of Information Technology, Bengaluru

Lecture - 36
Bivariate Polynomials Over Finite Fields: III

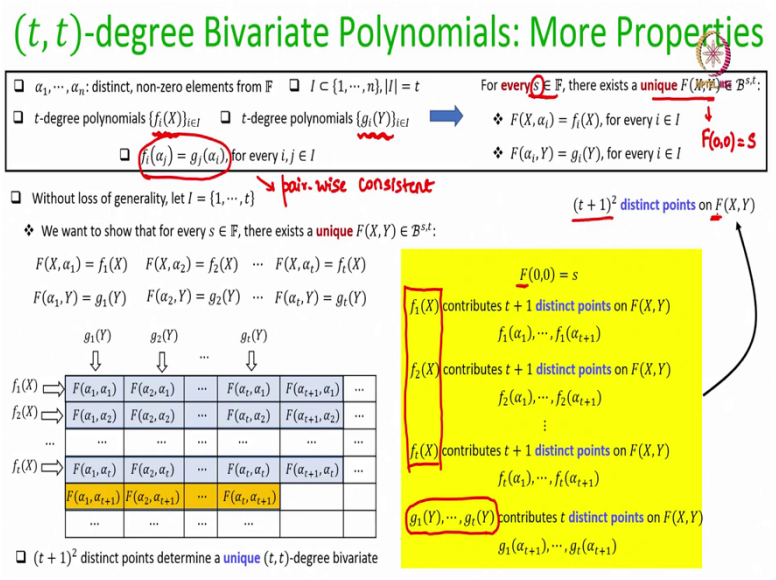
Hello everyone. Welcome to this lecture.

(Refer Slide Time: 00:24)



So, in this lecture, we will continue our discussion regarding Bivariate Polynomials over a Finite Field. And we will discuss an important lemma which we call as privacy lemma, and which will be later used when we design verifiable secret sharing schemes based on bivariate polynomials.

(Refer Slide Time: 00:43)



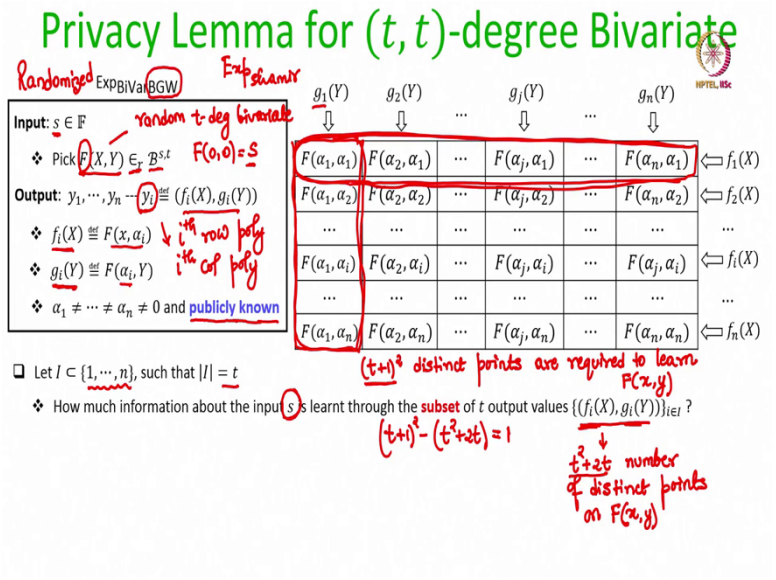
So, just to quickly recap, in the last lecture, we had seen an important property of t degree bivariate polynomials. And the property was that if I give you t pairs of univariate polynomials in X variable and Y variable which are pairwise consistent. Then, for every element s from the field you can find a unique t degree bivariate polynomial whose constant term will be s and over which these t degree X and Y variable univariate polynomials lie, ok, right.

And we have given a detailed proof for that. We have also illustrated that property. And on a very high level the reason this works is as follows. So, you want the constant term of the bivariate polynomial to be s. So, you are fixing one point on the bivariate polynomial on that unknown degree t bivariate polynomial.

And at the same time you would like the given t degree univariate polynomials also to lie on that same bivariate polynomial and you also want a given t degree Y univariate polynomials also to lie on the same bivariate polynomial. So, all together they fix $(t+1)^2$ points on that unknown bivariate polynomial and through $(t+1)^2$ distinct points we can pass only a unique t degree bivariate polynomial.

There cannot be more than 1, t degree bivariate polynomial passing through $(t+1)^2$ or more number of distinct points.

(Refer Slide Time: 03:06)



Now, based on this property we will derive the privacy lemma. And to understand this privacy lemma let us first formulate a randomized experiment. So, this experiment is a randomized experiment because the output will be determined randomly based on the input and the random coins which are tossed in the experiment.

So, even if you run this experiment multiple times with the same input s from the field, you are likely to get different outputs with different probabilities, ok. So, what exactly we do in this experiment? So, the input for the experiment will be an element from the field. And what we will do is the following? The experiment will randomly pick a t degree bivariate polynomial whose constant term is s.

So, random t degree bivariate polynomial whose constant term is s, right and this is the random step in the experiment. So, this bivariate polynomial is not fixed for this experiment; that means, it is not the case that for the input s, the polynomial F(X, Y) is a fixed polynomial. It is picked uniformly at random.

And what is the output? The output consists of $y_1, \cdots, y_n$ where the ith output is the pair of univariate polynomials $f_i(X)$ and $g_i(Y)$, where $f_i(X)$ is the value of the bivariate polynomial at Y equal to α_i and $g_i(Y)$ is the value of the bivariate polynomial at X equal to α_i, and where $\alpha_1, \cdots, \alpha_n$ are nonzero distinct entities from the field and they are publicly known. They are

fixed. They are not going to change during the different instances or during the different executions of the experiment. They are fixed, ok.

So, intuitively, the ith output is the pair of the ith row and ith column polynomial. So, output y i consist of the ith row polynomial and the ith column polynomial. And if you are wondering what are row and column polynomials, refer to one of our earlier lectures where we have seen the matrix view point of a bivariate polynomial, how we can interpret distinct points on a bivariate polynomial in the form of an n cross n matrix.

And ith row means the points in the ith row and ith column means the points along the ith column. And all the values along the ith row of that matrix lies on a t degree univariate polynomial which we denote as $f_i(X)$. And all the points along the ith column, they also lie on a t degree Y polynomial $g_i(Y)$, ok. So, this is the matrix view.

So, what you can imagine is that the experiment is outputting the following: first output is the pair of first column and the first row. Well, in the experiment I have written that the first output will be $f_1(X)$ and $g_1(Y)$, but if I evaluate $f_1(X)$ at X equal to alpha 1, alpha 2, alpha n, then that is equivalent to saying that the experiment is giving me the first row of this matrix as the output.

And in the same way if I evaluate g sub 1 Y at Y equal to alpha 1, alpha 2, alpha n, then that will give me this first column. So, either I can interpret that the first pair of output is the polynomials f sub 1 X and g sub 1 Y or I can interpret that the first output consists of the first row and the first column, both forms are equivalent.

Now, what exactly is the privacy lemma here? The privacy lemma would like to assess the following. Suppose, I focus on a subset of t output values in this experiment, it could be any subset of t output values. Need not be the first t output values or the last t output values or it need not be t consecutive output values. Any subset of t output values, ok.

So, let us fix an index set I, where I is a proper subset of one to n such that the cardinality of I is t. We would like to assess that if someone is given only t pairs of X univariate polynomials and Y univariate polynomials, then how much information about the input s is learnt through these t pairs of X polynomials and t pairs of Y polynomials. That is what we want to assess now, ok.

Before going further, if you want to compare this experiment with the experiment Shamir which we had seen in the context of Shamir's secret sharing. Then, in the context of Shamir's secret sharing protocol the experiment was that input will be an element from the field. And to generate the output a uniformly random t degree univariate polynomial is picked whose constant term is the input and output consist of distinct points on that randomly chosen polynomial.

And in the context of that experiment, the name of that experiment was exps sub Shamir. In the context of that experiment we had proved that, any subset of t output values does not reveal any information about the input s, namely the probability distribution of any subset of t output values in that experiment was independent of the input s.

Now, what we are doing in this new experiment is, we have tried to generalize that Shamir experiment to bivariate polynomials. So, our input is still a single element from the field. And now instead of picking a univariate polynomial, we are picking a random bivariate polynomial whose constant term is this input. And now instead of outputting distinct points, we are outputting distinct univariate polynomials lying on this bivariate polynomial.

So, all these $f_i(X)$ and $g_i(Y)$ poslynomials, you can interpret them as univariate polynomials lying on this randomly chosen bivariate polynomials. And we would like to now assess that if I give you only t number of X and Y univariate polynomials how much information about the input s you will learn. Intuitively, the answer is that we learn nothing about the input s.

This is because the degree of the bivariate polynomial F(X, Y) is t. So, $(t + 1)^2$ distinct points are required, to learn the bivariate polynomial selected in the experiment. But we are given t number of X univariate polynomials and t number of Y univariate polynomials.

It can be shown that through these t number of X and Y univariate polynomials, one can learn $t^2 + 2t$ number of distinct points on that unknown bivariate polynomial F(X, Y) which has been selected in the experiment. So, that means, we require total $t^2 + 2t + 1$ distinct points on the bivariate polynomial F(X, Y) to uniquely identify what was the bivariate polynomial selected in the experiment.

But through the t number of output values which are given to us we will end up learning only $t^2 + 2t$ number of distinct points. Well, how we get this $t^2 + 2t$ number of distinct points? I

leave it as an exercise for you. You can use this matrix interpretation of the bivariate polynomial and the property which we had proved in the earlier lecture, ok.

So, we are falling short of how many distinct points. So, we require $t^2 + 2t + 1$, but we will learn only $t^2 + 2t$. So, we still lack one more distinct point to uniquely identify what was the bivariate polynomial F(X, Y) picked by the experiment. And since, that polynomial was randomly chosen it could be any polynomial and hence, its constant term could be any value s from the field.

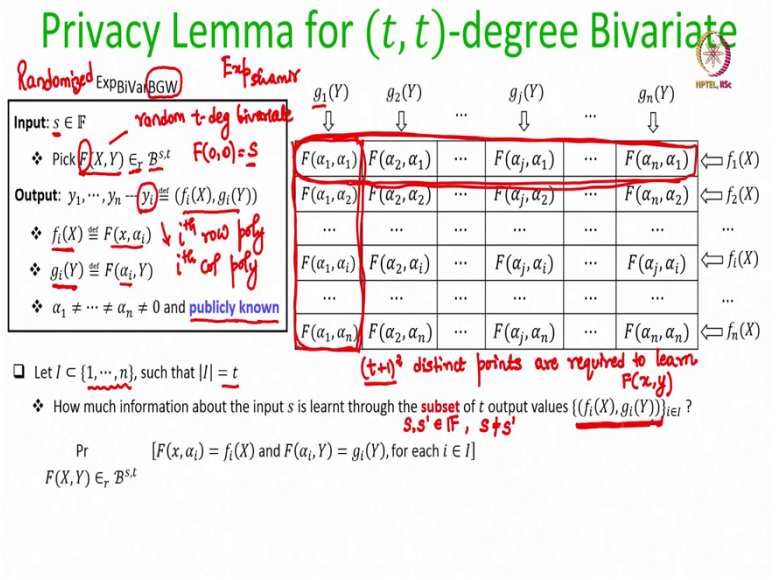
Again, this is a generalization of the argument which we used to prove that for the Shamir experiment, through the t output values, adversary or whoever is observing those t output values, does not learn anything about s, because in the context of the Shamir experiment degree of the univariate polynomial was t.

And t output values does not uniquely determine a randomly chosen t degree polynomial. So, there was one more point which was missing for anyone who is observing t output values. And hence, it cannot uniquely identify what was the polynomial picked in the experiment. And hence, its constant term could be any element from the field.

So, we are just generalizing that argument in the context of this bivariate polynomial based experiment. And why the name of this experiment is bi-var BGW? Why BGW because later on when we will design a perfectly secure verifiable secret sharing scheme, namely the VSS scheme due to the BGW, this experiment will be coming into picture.

So, intuitively through t pairs of X and Y univariate polynomials, we learn nothing about the value s, the input s of the experiment, but that is an informal argument. We now want to formalize it. So, let us see how we can formalize it.

(Refer Slide Time: 15:26)



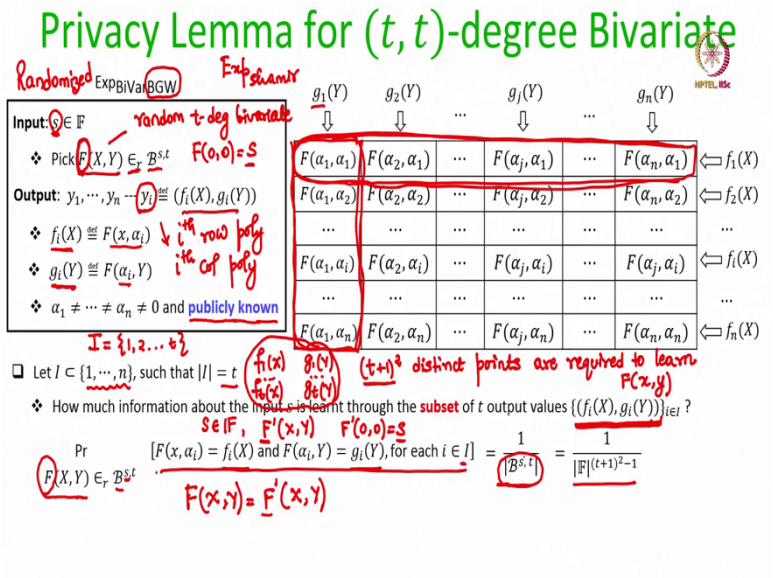
So, we will show that for every pair of candidate s and s prime from the field, where s is not equal to s prime, if we run this experiment with input s then that will result in these t number of X and Y univariate polynomials, with the same probability, with which the same experiment with input s prime would have resulted in the t pairs of output values consisting of the same X and Y univariate polynomials.

So, that means, whoever is observing these t, X and Y univariate polynomials, it cannot pin point whether the experiment was run with the input s or with the input s prime. With equal probability it could be the case that the experiment has been run with input s and that person who is observing those t output values is getting these polynomials.

And with the same probability it could be the case that the experiment has been run with input s prime, and even in that case it results in the same t number of X and Y univariate polynomials. We would formally prove that. If we formally proved that, then that is equivalent to showing that the probability distribution of these t pairs of X and Y univariate polynomials is independent of the actual input of the experiment.

So, let us first derive what is the probability that if we randomly choose a t degree bivariate polynomial whose constant term is s, then when that bivariate polynomial is evaluated at the evaluation points corresponding to the indices in this I set results in the row and column polynomials corresponding to this index set I, ok.

(Refer Slide Time: 17:51)



So, for simplicity, let us say we fix I to be the first t locations, ok. So, $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$. So, we now want to find out what is the probability that if I randomly choose a t degree bivariate polynomial whose constant term is s, then when evaluated at $\alpha_{1,\cdots,\alpha_t}$ will result in $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$.

And the probability of that is 1 over the number of t degree bivariate polynomials with s being the constant term. And the number of t degree bivariate polynomials with s being the constant term is this. We have already derived this in one of our earlier lectures.

Now, why this probability is 1 over the number of all t degree bivariate polynomials with s being the constant term? For this we can invoke the result in the previous slide. So, what we are given? We are given t pairs of X and Y univariate polynomials which are pairwise consistent namely $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$. They are pair wise consistent.

And we know that for s belonging to F, there is a unique bivariate polynomial, call it $F'(X, Y)$ such that the constant term of that bivariate polynomial is s, and these univariate polynomials $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$ lie on that bivariate polynomial.

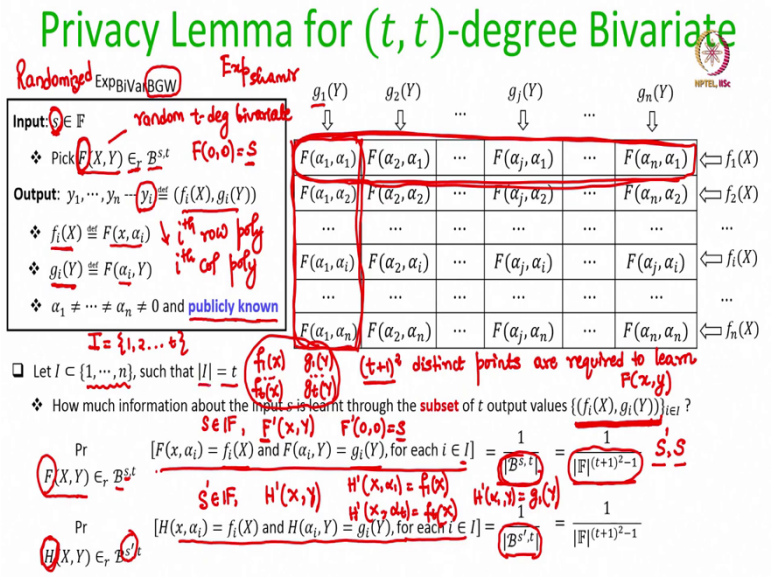
So, even though there are these many bivariate polynomials of degree t with s being the constant term. Among all those bivariate polynomials of degree t with s being the constant

term, there is only one bivariate polynomial, I am calling it $F^{'}$, whose constant term is s. And over which these given t pairs of X and Y variant polynomials lie, ok.

Now, we are trying to calculate the probability over all possible bivariate polynomials of degree t with s being the constant term, right. So, this condition will be satisfied only when the randomly chosen bivariate polynomial F(X, Y) is same as that unique bivariate polynomial $F^{'}(X, Y)$, ok. But in the experiment the bivariate polynomial was chosen randomly.

So, what is the probability that randomly chosen bivariate polynomial was $F^{'}(X, Y)$, that special polynomial? Well, it is 1 over the sample space size. So, that is the probability with which an instance of the experiment with s being the input can produce the outputs $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$ as the first t output pairs.

(Refer Slide Time: 22:00)



Now, using the same argument we can conclude that if we try to compute the probability over all possible t degree bivariate polynomials whose constant term is now $s^{'}$. Then, what is the probability that any randomly chosen polynomial from such set when evaluated at $\alpha_1, \cdots, \alpha_t$ will result in $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$.

Well, again corresponding to this $s^{'}$ belonging to F, there is a unique bivariate polynomial of degree t whose constant term is $s^{'}$, let us call it $H^{'}(X, Y)$, such that $H^{'}(X, Y)$ evaluated at $\alpha_1, \cdots, \alpha_t$ will result in $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$.

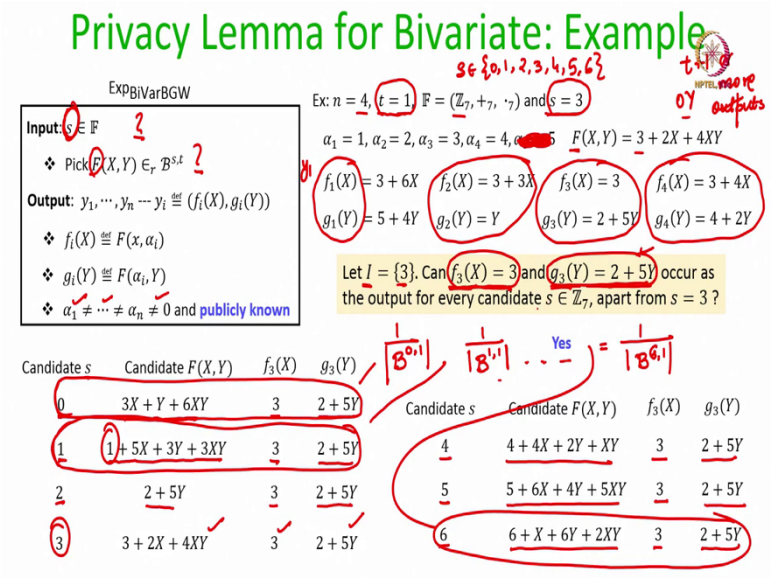
So, now we are choosing the bivariate polynomial randomly from the bigger set, namely the set of all t degree bivariate polynomials whose constant term is s prime. And it is only for one bivariate polynomial from this set, that this given condition will be true. It will not be true for any other bivariate polynomial from this set.

Now, what is the probability that randomly chosen bivariate polynomial in the experiment takes the value $H^{'}(X, Y)$,? Well, the polynomial in the experiment is randomly chosen. So, the probability that indeed during the run of the experiment, the chosen polynomial was $H^{'}(X, Y)$ is 1 over the sample space size.

And the sample space size, namely the set of all t degree bivariate polynomials with s prime being the constant term is same as the set of all possible t degree bivariate polynomials with s being the constant term. And that is why both these two probabilities are equal. And we have thus proved that if you take any pair of candidates, s and s prime from the field, and run this experiment with input s prime and with input s, then with equal probability they will result in the first t output being $f_1(X), \cdots, f_t(X)$ and $g_1(Y), \cdots, g_t(Y)$.

Well, again I stress that for simplicity the t output values for which we want to prove this lemma, I have fixed those t indices to be 1 to t, but this lemma holds for any subset I which is a proper subset of the index set 1 to n, such that the cardinality of the index set I is t, ok.

(Refer Slide Time: 25:43)



So, now let me demonstrate this privacy lemma with an example here. So, for the purpose of demonstration I am taking n to be 4, t to be 1 and I will be performing all the operations over the field Z 7, where all the addition operations will be addition modulo 7. And all the multiplication operations will be multiplication modulo 7. And suppose we run this experiment with the input s being 3.

Now, there are many polynomials of degree 1 with 3 being the constant term, ok. By the way let me fix the evaluation points alpha 1, alpha 2, alpha 3, alpha 4. So, there are many bivariate polynomials of degree 1 in X and Y whose constant term will be 3. So, suppose in the experiment this polynomial is chosen, ok.

So, you can see this is a bivariate polynomial whose constant term is 3 and of course, here is no term of the form Y raise to power 1, it is there with coefficient being 0, right and so on. So, this is a 1 degree bivariate polynomial. And as a result output generated in the experiment will be this. So, this pair of output will be y 1, ok. So, this is y 1. This is the second pair of output. This is the third pair of output and this is the fourth pair of output, ok.

Now, suppose I gave you only one of the 4 outputs, ok of your choice, suppose you asked for the third output, namely the index set is 3 for you. And index set I denotes the indices corresponding to which you see the output. So, since t is equal to 1 here, you are allowed to see only one output.

So, suppose you asked for the third output, third output means, the third row polynomial and the third column polynomial. So, I give you the third row polynomial and third column polynomial, but I do not tell you what was my input. Well, you know that my input could be either 0 or 1 or 2 or 3 or 4 or 5 or 6 you know that information because you know the steps of the experiment.

What exactly is not known to you is the value of the input which I have used in the experiment. So, I am not unfair to you, I am giving you one of the outputs which you asked for, ok. Now, I challenge you that can you tell me what exactly was my input based on this output. Now, I am going to show you that for you it could be the case that I have run the experiment with s being 0 with the same probability as I have run the experiment with s equal to 1.

And with the same probability it could be the case that I have run the experiment with s equal to 2 and with the same probability it will be the case that I have run the experiment with s equal to 3 for you. That means, you cannot pinpoint whether I have run the experiment with 0 or 1 or 2 or 3 or 4 or 5 or 6.

With equal probability it could be the case that I have run the experiment with any of the 7 candidate s values, and in each case it would have resulted in the third output being $f_3(X)$ being 3 and $g_3(Y)$ being 2 + 5 Y. That is what I am going to show. And if I can show this then that is equivalent to showing that you cannot pinpoint whether my input was 0 or 1 or 2 or 3 or 4 or 5 or 6.

By the way I am fixing here the number of output, which you are allowed to be seen, to t, namely t equal to 1. If I give you t + 1 or more number of outputs, then of course, you can find out what exactly was my input.

It is only when I restrict to t number of outputs, only when I give you t number of outputs, this claim holds. You cannot pin point what exactly was my input, right. So, what I am going to show here is the following. We will take different candidate s value. So, this is some kind of mental experiment, mental calculation to be more precise. This is the mental calculation which you are now going to do. You do not know what was my input, so that is question mark.

You do not know what was the bivariate polynomial I have chosen in the experiment, that is also a question mark. What is known to you? alpha 1, alpha 2, alpha 3, alpha 4, they are known to you. And the third row polynomial, and the third column polynomial they are known to you. Now, you might try to calculate that is it the case that I have run the experiment with input being 0, such that the third row and the third column polynomial in that run will be the polynomials 3 and 2 + 5y respectively.

And the answer is yes. It is quite possible because if I turned up running the experiment with an input 0. And if I would have selected this bivariate polynomial, then it would have resulted in the third row polynomial to be 3 and the third column polynomial to be 2 +5 Y, respectively, right.

So, you cannot rule down, rule out this possibility. Or, it could be the case that I have run the experiment with input 1 and I have chosen the polynomial 1 + 5 X + 3 Y + 3 XY whose constant term is 1, in which case the third row and the third column polynomial which you would have seen would be the polynomials 3 and 2 + 5 Y, respectively. Again, this possibility also you cannot rule out.

Or, it could be the case that I have run the experiment with input 2, by choosing the bivariate polynomial 3 plus 5 Y. And in that case the third row and the third column polynomials would have turned out to be again 3 and 2 plus 5 Y, respectively. Or, it could be the case that I have run my experiment with input 3 with this being with the bivariate polynomial, in which case the third row and third column polynomial would have been the polynomials 3 and 2 plus 5 Y, respectively.

Or, it could be the case that my input was 4, this was the bivariate polynomial I have chosen. And if that would have happened then it would result in the third row and the third column polynomial to be 3 and 2 plus 5 Y, respectively. Or, it could be the case that my input was 5. I have run the experiment by choosing this bivariate polynomial. And you would have seen the third row and the third column polynomial to be 3 and 2 plus 5 Y, respectively.

Or, it could be the case that my input was 6, I have chosen this bivariate polynomial and in that case the third row and the third column polynomial would have been this. So, now let us try to calculate each of the probabilities individually. What is the probability that this was the case? Well, this could be the case with probability 1 over the size of all bivariate polynomials whose constant term is 0 and whose degree is 1 over the field Z 7.

The probability that this is the case, the second possibility is again 1 over the size of all possible bivariate polynomials with 1 being the constant term and where the degree of the bivariate polynomial is 1 and where the bivariate polynomial is over the field Z 7. And like that if I take the last case, the probability of this occurring is 1 over the cardinality of all possible 1 degree bivariate polynomial over Z 7 with 6 being the constant term.

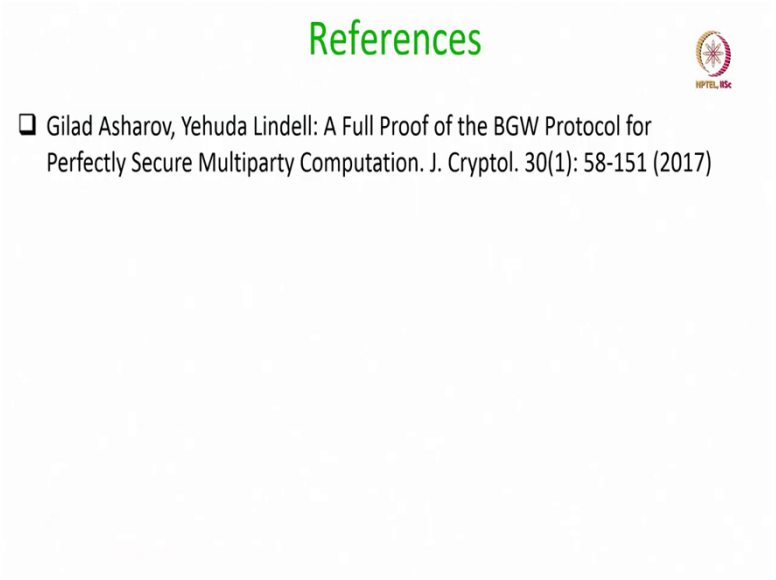
Now, in the denominator all these sets have the same cardinality. Namely, the number of bivariate polynomials of degree 1 over Z 7 with 0 being the constant term is same as the number of bivariate polynomials over the field Z 7 of degree 1 with 6 being the constant term.

That means, when you are doing this mental calculation, you cannot say that, ok s being 0 is more likely to be the case compared to s being 1 with equal probability my input s could have been 0 or 1 or 2 or 3 or 4 or 5 or 6 and would have resulted in the third row and the third column polynomials being 3 and 2 plus 5 Y, respectively.

And by the way this mental calculation you could have performed even using unbounded computing time. Even if you have unbounded computing time resources here you cannot pinpoint what exactly was my input, ok. So, that is a very important point here. So, even if I give you unbounded time resources just with t output values, namely t row and column polynomials you will completely fail to find out what exactly was my input.

And that shows that from your viewpoint the probability distribution of the t row and column polynomials which you see will be independent of the exact input of this experiment.

(Refer Slide Time: 37:22)



So, with that, I end this lecture. This is the reference used to prove today's Privacy Lemma.

Thank you.