

**Foundations of Cryptography**  
**Prof. Dr. Ashish Choudhury**  
**(Former) Infosys Foundation Career Development Chair Professor**  
**Indian Institute of Technology – Bangalore**

**Lecture – 28**  
**Cryptographic Hash Functions - Part II**

(Refer Slide Time: 00:32)

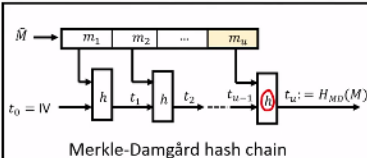
# Roadmap

## ❑ Constructing fixed-length compression functions

Hello everyone, welcome to this lecture. So just to recall, in the last lecture we had seen the Merkle-Damgård paradigm for constructing collision-resistant hash functions for any size inputs and in this lecture we will see the stage 1 of the Merkle-Damgård paradigm, namely how to construct a fixed-length compression function.

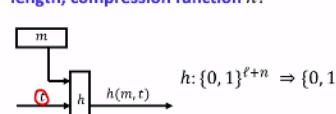
(Refer Slide Time: 00:49)

**Constructing Fixed-length Collision-Resistant Compression Functions**



Merkle-Damgård hash chain

❑ How to construct **collision-resistant, fixed-length, compression function**  $h$ ?



$h: \{0, 1\}^{l+n} \Rightarrow \{0, 1\}^n$

❑ **Two approaches** to construct the function  $h$ :

- ❖ Constructions based on **number-theoretic assumptions** --- not used practically
- ❖ Constructions based on **block ciphers** --- security proof in **unconventional model**

So just to recall, pictorially this is how a Merkle-Damgård paradigm will take any fixed-length compression function  $h$  which is collision-resistant and apply it iteratively to obtain a hash

function which can hash inputs of any size. So now the interesting question is how exactly you obtain this fixed-length collision-resistant compression function  $h$  at the first place? So pictorially, this function  $h$  takes an input of size  $\ell + n$  bits, which can be parsed as 2 inputs, an input  $m$  of  $\ell$  bits and another input  $t$  of  $n$  bits.

It will compress it and gives you an output of size  $n$  bits. So that is how you can interpret this fixed-length compression function and it turns out that there are 2 approaches to construct this function  $h$ . The first approach is to design a construction based on number-theoretic hard problem or number-theoretic assumptions. So when we will start our discussion on number theory and public key cryptography, we will come back and we will see how to use some number-theoretic hardness assumptions to construct this function  $h$ .

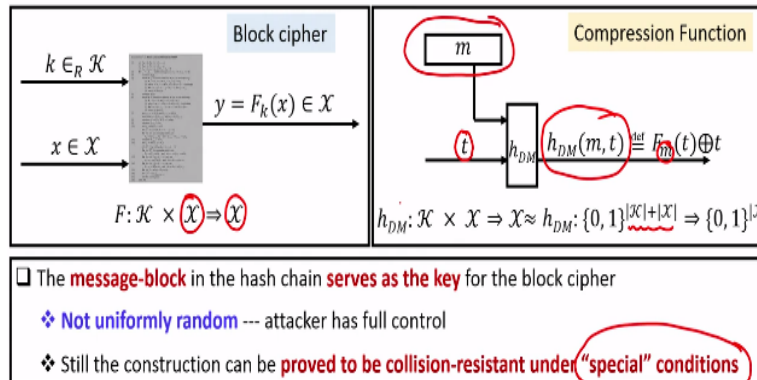
This approach is not used practically because even though the constructions, their running time, they are polynomial in the security parameter, the actual running time when deployed in practice is of order of several magnitudes and that is why we do not use the instantiations of function  $h$  based on number-theoretic hardness assumptions. Instead what we do is, we design constructions, we go for constructions based on block ciphers, we can take any of the existing block ciphers say AES, des or we can design dedicated block ciphers.

These are the constructions, these are the instantiations of the function  $h$ , which we use in practice. However interestingly the security proof of the constructions based on block ciphers, they are in very unconventional model. They are unconventional model in the sense, we make very strong assumptions from the underlying model and then give the security proof.

**(Refer Slide Time: 03:06)**

## Collision-Resistant Compression Functions from Block Ciphers

□ Davies-Meyer construction



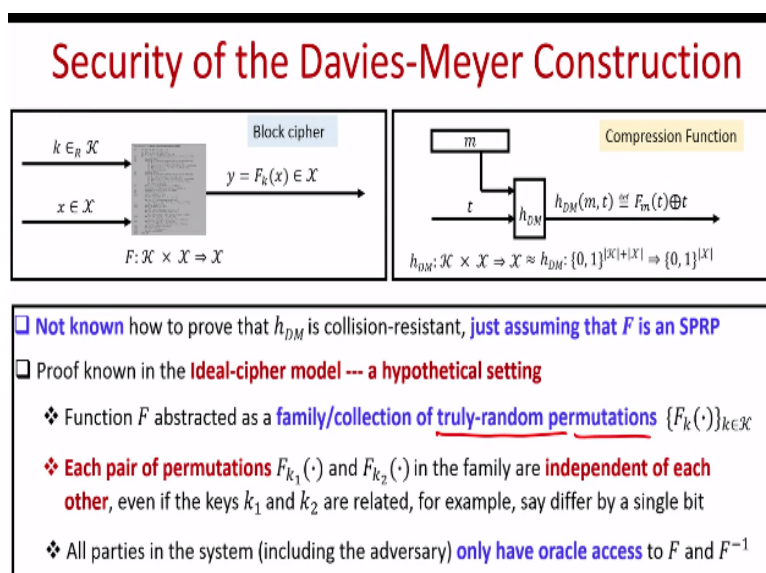
So let us see this approach of how to construct collision-resistant fixed-length compression function, from block ciphers. And there are several constructions available. We will see one of the constructions which is called as Davies-Meyer construction. So what you are given is a block cipher  $F$  which you can interpret as a keyed pseudorandom permutation, taking a key from the key space and input  $x$  from the block space and giving you an output, where the output and the block size are the same. Namely both of them belong to the set  $\mathcal{X}$  and using this we design a compression function  $h$ , which I denote as  $h_{DM}$ , namely we can call it as Davies-Meyer compression function, which takes an input of size  $|\mathcal{K}| + |\mathcal{X}|$ , namely it takes 2 inputs, 1 input is considered/interpreted as  $m$  and other input is interpreted as  $t$  and the output of this compression function is defined as  $F_m(t) \oplus t$ .

We evaluate the block cipher with respect to the key  $m$  and treating the  $t$  part as the block and the output is again XORed with the  $t$  part of the input of this Davies-Meyer hash compression function, so that is how this Davies-Meyer compression function is evaluated. So the interesting part here is that the message block in the hash chain (remember that this function  $h$  is going to be plugged in in the hash chain of the Merkle-Damgard paradigm right when we are evaluating the overall hash function  $H$ ). That is why pictorially I am representing this function  $h$  consisting of 2 inputs, one input will be the block part of the message which we want to hash in the bigger message and one will be the  $t$  part which will be coming from the outcome of the previous invocation of the  $h$  function). The way we are designing this Davies-Meyer compression function is basically the message block on which we are going to operate this Davies-Meyer instruction will be treated as the key for the pseudorandom permutation of the block cipher. It need not be uniformly random.

Because attacker could have full control over the bigger message which it want to hash as per the Merkle-Damgard paradigm and that means it gives the attacker full control over what exactly is serving as the key for the block cipher when a block cipher is internally used in the design of the Davies-Meyer compression function.

However even though the adversary has the full control over the key which is used in the underlying instantiation of block cipher in this Davies-Meyer compression function, it can be proved that the overall construction of this Davies-Meyer compression function is indeed collision-resistant under special conditions. That means there is nothing to worry about here, even though you might be wondering that we have discussed that block ciphers remain secure only if the key is not known to the attacker and the key is uniformly random, but the way we are operating or using the block cipher in this Davies-Meyer compression function is that, the block of the message which we are going to use or apply the hash function  $h$ , it is fully under the control of the adversary and that is actually serving as the key for your underlying instantiation of your block cipher and that need not be random. But still we can prove that under special conditions, the way the output of the Davies-Meyer construction is defined, overall the function in the Davies-Meyer construction that we have constructed here is indeed collision-resistant.

(Refer Slide Time: 07:05)



So here is how we have constructed the Davies-Meyer compression function and as I am continuously saying that we can prove the collision resistance of the Davies-Meyer construction under special assumptions, this is because we do not know how to prove the

collision resistance of the Davies-Meyer construction that we have given, just assuming that the underlying block cipher is a strong pseudorandom permutation. It is not sufficient to just assume that the underlying function  $F$  is a strong pseudorandom permutation and give a proof that the Davies-Meyer construction is indeed collision-resistant.

However, interestingly the proof of the collision resistance of the Davies-Meyer construction is known in the ideal-cipher model which is kind of a hypothetical setting, because it makes very strong assumptions about your underlying block cipher. So let us see what exactly is this ideal-cipher model? So in ideal-cipher model, the underlying block cipher or the keyed permutation  $F$  is abstracted as a collection of several truly random permutations, where each member of this family or each member of this collection is indexed by the underlying key with which we are going to operate the function  $F$ .

So remember, the block cipher it takes a block input and a key input. So what basically we are assuming in this ideal-cipher model is that even though we are given the description of a single  $F$  function, as soon as we fix the value of  $k$ , that gives an instantiation of the  $F$  function, which can be treated as one member of a bigger family. So different values of the  $k$  which are used in your function  $F$ , will give you different members of the bigger family and that is why we are now considering that even though we are given the description of a single  $F$  function, when used with different  $k$ , it gives you different members from a bigger family and that is what we mean by function  $F$  abstracted as a collection of truly random permutations.

The second property in this ideal-cipher model is that once we have a family member  $F_{k_1}$  and another family member  $F_{k_2}$ , namely these are the instantiation of your block cipher with the key  $k_1$  and the key  $k_2$ , the assumption here is that the family member  $F_{k_1}$  and the family member  $F_{k_2}$ , namely the keyed permutations  $F_{k_1}$  and  $F_{k_2}$ , they are independent of each other even if the keys  $k_1$  and  $k_2$  are dependent or related to each other.

That means even if it so happens that  $k_1$  is almost the same as  $k_2$  except say the last bit, the assumption in this ideal-cipher model is that the function  $F_{k_1}$  will be completely independent in the sense, its output will be uniformly random and it will be completely independent from the outputs of the function  $F_{k_2}$ . So that is the second property in this ideal-cipher model. So till now there is nothing unconventional here that we are assuming about the function  $F$ .

Because when we are assuming that function  $F$  is a strong pseudorandom permutation, that means even though it is not a truly random permutation, it gives you pseudo randomness guarantee. That means we can assume that it behaves like a random permutation. But in the ideal-cipher model we are making slightly more stronger assumptions, we are assuming that each member of this bigger family is indeed a truly random permutation and independent of each other.

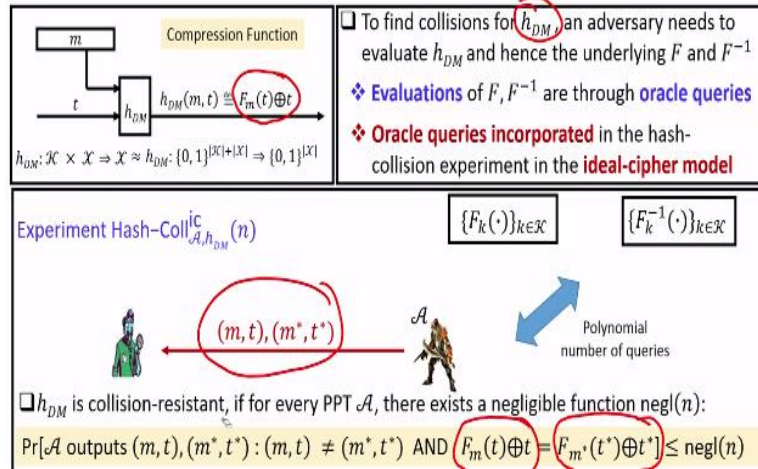
The strongest property or you can call it as the most unconventional property in this ideal-cipher model is that no party or no entity in the system who is going to use a cryptographic primitive, which makes use of this block cipher will have access to the code of the function  $F$  and the code of the function  $F^{-1}$ . That means if at all any entity wants to compute the value of the function  $F$  with respect to some key, it will make oracle access or it will make oracle calls to the keyed permutation  $F_k$  and it will just get back the value of the keyed permutation on the input for which it wants the value of the  $F_k$ .

In the same way if some entity wants to know the value of  $F^{-1}$  with respect to some input key  $k$  and some input  $y$ , then it can make just oracle call to the  $F^{-1}$  oracle and it will get back the corresponding output, but it will not get any access to the code of the block cipher. So that is the most unconventional assumption that we are making in the ideal-cipher model because in reality when we are designing the Davies-Meyer construction and instantiating it, we need the actual block cipher which we are going to plug-in in this construction of the Davies-Meyer construction.

Every entity who is going to use that instantiation of the Davies-Meyer construction will know the code of the function  $k$  including the adversary, but when we are analyzing the security of the Davies-Meyer construction, the assumption that we are making in the ideal-cipher model is that no one will be knowing the code of the function  $F$ , it is just a collection of several random members, if you want to talk to one random member, just make an oracle call.

**(Refer Slide Time: 12:17)**

## Experiment Hash-Collision Against Davies-Meyer Construction in the Ideal-cipher Model



So assume now we are in the ideal-cipher model, what we are going to prove is that this Davies-Meyer construction of fixed-length compression function that we have given here indeed is a collision-resistant function. So remember that indeed it is a compression function because we are taking an input of size  $n + \ell$  bits and giving you an output of size  $\ell$  bits, so it is a compression function.

What we have to prove is that in polynomial amount of time, it is difficult for an adversary to come up with a pair  $(m, t)$  and  $(m^*, t^*)$ , such that the output of the Davies-Meyer function is the same where the Davies-Meyer construction is using a block cipher. So since we are going to analyze the security of this compression function in the ideal-cipher model, any adversary who would like to find out a collision for this Davies-Meyer construction, it needs to evaluate the Davies-Meyer function on several messages of its choice, then only it could come up with a collision.

To evaluate the Davies-Meyer function on messages of its choice, internally it needs to evaluate the function  $F$  and  $F^{-1}$  as well, because we do not know what exactly is the strategy of the adversary; it may evaluate the underlying  $F$  on several  $(m, t)$  pairs, it could evaluate the underlying  $F^{-1}$  also on several inputs of its choice, and then only it could come up with a corresponding collision.

But since we are now in the ideal-cipher model what we are going to assume is that even the adversary who would like to evaluate the underlying  $F$  and  $F^{-1}$  on inputs of his choice to come up with a collision for this Davies-Meyer function, will make oracle queries, because it

would not be knowing the code of the function  $F$  and code of the function  $F^{-1}$ . So that is why to incorporate these oracle queries from the adversary to find out the collision, these oracle queries are incorporated in the hash-collision experiment when we analyze the security of this Davies-Meyer construction in the ideal-cipher model.

So remember in the actual hash-collision experiment in the non-ideal-cipher model or in the standard model, adversary did not have to query anything, because it will be having access to the code of the function  $F$  or  $F^{-1}$  if we are using any  $F$  or  $F^{-1}$  in the design of the underlying hash function.

But now since we are in the ideal-cipher model, even adversary's interaction with the function  $F$  and  $F^{-1}$  will be through oracle queries and that is why they also need to be incorporated in the modified hash-collision experiment. So the modified hash-collision experiment is as follows. So we now assume that both the experiment as well as the adversary will have oracle access to the family of the functions  $F_k$  and the family of the functions  $F_k^{-1}$ .

So you can imagine these oracles  $F_k$  and  $F_k^{-1}$  are kind of lying in the sky, where no one can see what exactly is the code of the function  $F_k$  and  $F_k^{-1}$  inverse and if anyone wants to compute the value of the function  $F_k$  for any  $k$  of its choice, it just shouts to the sky that please give me the value of this function  $F_k$  on this input and it will see the value of  $F_k$  on that input and same way the interaction with  $F_k^{-1}$  is handled here. So that is why I am putting this family of functions  $F_k$  and  $F_k^{-1}$  inside a box here.

No one can see what exactly is happening inside the box, all the interactions are going to be through an interface, you supply some input and you get back the output. So the modified hash-collision experiment is as follows. So remember the goal of the adversary is, it knows the description of your Davies-Meyer hash function, namely it knows that the Davies-Meyer's hash function is basically  $F_m(t) \oplus t$ . The goal of the adversary is basically to come up with a pair  $(m, t)$  and  $(m^*, t^*)$ , which constitutes a collision with respect to this Davies-Meyer function. But since we are in the ideal-cipher model, any value of  $F$  which adversary would like to compute or any value of  $F^{-1}$  which the adversary would like to compute, will be made available through oracle queries. So imagine if adversary would like to know the value of  $F$  with respect to the key  $k_i$  and on the input  $x_i$ .



So, it can make an oracle query and what the oracle is going to return back is, it is going to return back the value of the function  $F$  with respect to the key  $k_i$  on the input  $x_i$ , and if the adversary sees this response, then basically what adversary sees is the value of the Davies-Meyer function on the input pair  $(k_i, x_i)$ , because the value of the Davies-Meyer function on this combined input  $(k_i, x_i)$ , will be the value of the function or the block cipher  $F$  on the input  $x_i$  with the key  $k_i$ , XOR with the input  $x_i$ .

But since adversary is in the ideal-cipher model, he cannot compute the value of  $F_{k_i}(x_i)$  and that is why it has made an oracle query. So that is how the adversary is going to compute the value of Davies-Meyer functions on messages of its choice. It turns out that this is not the only way by which the adversary could compute the value of the Davies-Meyer function on inputs of his choice.

It could take the help of the inverse oracle access as well, say for instance it can ask, hey inverse oracle give me the value of the inverse of the function under the key  $k_i$  on the input  $y_i$  and in response what the oracle is going to do is, it is going to return back the inverse of this function on the input  $y_i$  under the key  $k_i$ , sorry for the typo here, it should be under the key  $k_i$ , and once the adversary learns the inverse of this  $y_i$  under the key  $k_i$ , that gives him the value of the Davies-Meyer function on the input  $(k_i, x_i)$ . Because the way Davies-Meyer function is constructed, the output of the Davies-Meyer function on this input  $(k_i, x_i)$  would be like this.

So by making oracle queries to the  $F_k$  family and by making oracle queries to the  $F_k^{-1}$  family, we can assume that the adversary is now going to compute the output of the Davies-Meyer function on several messages of its choice, namely polynomial number of messages because our adversary is polynomially bounded and now finally he submits a collision, namely a pair of inputs  $(m, t)$  and  $(m^*, t^*)$ .

The security definition here is that, we will say that this collision-resistant experiment is successful for the adversary in the ideal-cipher model if in this model where adversary is now making or given oracle access to the  $F_k$  family and  $F_k^{-1}$  family, the probability that the adversary could come up with a collision is upper bounded by a negligible function and the

collision is now specific because now we are analyzing the collision with respect to the Davies-Meyer construction.

So, the output of the Davies-Meyer construction for the  $(m, t)$  message which the adversary is submitted will be  $F_m(t) \oplus t$ . And the output of the Davies-Meyer function for the message  $(m^*, t^*)$  that the adversary has submitted will be  $F_{m^*}(t^*) \oplus t^*$ . Our goal is to analyze what is the probability that the condition  $F_m(t) \oplus t = F_{m^*}(t^*) \oplus t^*$  hold. The security definition is we will say that the Davies-Meyer construction is collision-resistant in the ideal-cipher model if the probability that any poly-time adversary could come up with a special  $(m, t)$  and  $(m^*, t^*)$  satisfying this condition is upper bounded by some negligible function.

**(Refer Slide Time: 20:15)**

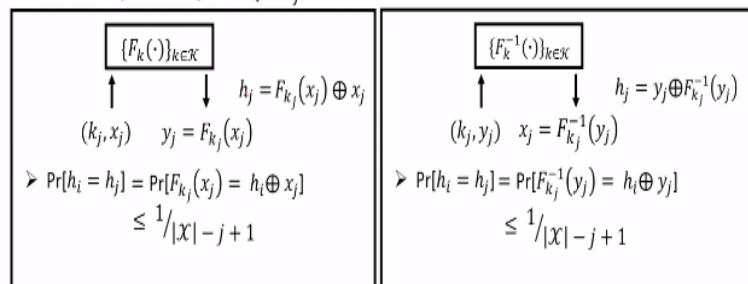
## Security Analysis of Davies-Meyer Construction

□ Theorem:  $h_{DM}: \{0, 1\}^{|\mathcal{K}|+|\mathcal{X}|} \Rightarrow \{0, 1\}^{|\mathcal{X}|}$  is collision-resistant in the ideal-cipher model, if  $|\mathcal{X}|$  is sufficiently large

Any  $\mathcal{A}$  issuing  $q$  ideal-cipher queries for  $(F, F^{-1})$  can find a collision with probability  $\leq \frac{q(q+1)}{|\mathcal{X}|}$

□ Consider query numbers  $i, j$ , where  $i < j$ , defining hash values  $h_i$  and  $h_j$

□ What is the probability that  $h_i = h_j$ ?



What we are going to prove is that indeed the Davies-Meyer construction that we have given constitutes a collision-resistant function in the ideal-cipher model provided the size of the  $\mathcal{X}$  is sufficiently large. More specifically, what we are going to prove here is that any poly-time adversary or any adversary  $\mathcal{A}$  issuing  $q$  number of ideal queries altogether to the  $F$  family and  $F^{-1}$  family can find a collision with probability upper bounded by  $\frac{q(q+1)}{|\mathcal{X}|}$ .

So here  $q$  is the total number of queries which the adversary is allowed to make to the  $F$  family and  $F^{-1}$  family. So it is not  $q$  for the  $F$  and  $q$  for  $F^{-1}$ . So if he is making  $s$  number of queries to  $F$  and  $t$  number of queries to  $F^{-1}$ , then  $s + t$  is basically  $q$  and what we are going to prove is that in the ideal-cipher model, any poly-time adversary making  $q$  number of queries, his

success probability of finding a collision in the Davies-Meyer function is upper bounded by  $\frac{q(q+1)}{|\mathcal{X}|}$ .

So what we are going to do here is we consider the query numbers  $i, j$ , where  $i < j$ . That means, assume that adversary has already queried up to  $j - 1$  queries altogether to the  $F$  and  $F^{-1}$  family and now he is making the  $j^{th}$  query and any time when it is making a query, the response for the query gives him the value of the Davies-Meyer function on some input pair, because that is we had already seen earlier.

So if he is making queries to the  $F$  family, then based on the response it gets the value of the Davies-Meyer function on some input or if he is making a query to the  $F^{-1}$  family, then again based on the response, it gives him the value of the Davies-Meyer function on some input. So let  $h_i$  and  $h_j$  denote the hash values, which adversary learn, based on the response of the  $i^{th}$  and  $j^{th}$  query respectively. We are interested to compute the probability that  $h_i = h_j$  holds. So now there could be 2 cases here. Assume that, the  $j^{th}$  query which the adversary is making is for the  $F$  family. So when he is making a  $j^{th}$  query to the  $F$  family, he is asking for the value of the family member  $F_{k_j}$  on the input  $x_j$ . So he will see the response which I denote as  $y_j$ . And based on the response he basically learns the value of the Davies-Meyer function on some input, basically he learns the value of the Davies-Meyer function on the input  $(k_j, x_j)$ . And that is denoted by  $h_j$  and this  $h_j$  will be  $F_{k_j}(x_j) \oplus x_j$ . So we are interested to analyze what is the probability that the  $j^{th}$  hash value which the adversary has learnt by making this query to the  $F_k$  family, it is the same as the  $i^{th}$  hash value which the adversary has already learned while making the  $i^{th}$  query. And remember the  $i^{th}$  query could be either to the  $F_k$  family or to the  $F_k^{-1}$  family.

So the probability that the  $i^{th}$  hash value which the adversary has learnt is the same as the  $j^{th}$  hash value which the adversary has just now learnt, is the same as the probability that the family member  $F_{k_j}$  when evaluated at the input  $x_j$ , gives output  $h_i \oplus x_j$ . If that is the case then  $h_j$  and  $h_i$  will be the same. But since we are in the ideal-cipher model right, what is the value of the family member  $F_{k_j}$  on the input  $x_j$ ? Well it is a uniformly random value independent of each other. It is a uniformly random value over the  $\mathcal{X}$  set and it is different from all the previous  $F_{k_j}$

values for which the adversary might have already asked the output value. So, remember we are in the case where  $i < j$ , we are analyzing the probability with respect to the collision for the output of  $i^{th}$  query and  $j^{th}$  query. So in the worst case it may so happen that for all the previous,  $j - 1$  queries, the adversary might have asked for the value of the function family with respect to the key  $k_j$ . In response, the adversary might have seen the output of the family member  $F_{k_j}$  for various inputs and since  $F_{k_j}$  is a permutation, when the oracle is giving him back the response for the  $F_{k_j}$  family on the  $j^{th}$  input, it has to be different from all the previous  $j - 1$  outputs, that oracle has already returned to the adversary. So the probability with which this condition, namely the output of the  $j^{th}$  query under the key  $k_j$  is equal to this, is always upper bounded by  $\frac{1}{|\mathcal{X}| - j + 1}$ . So that is the probability with which, by making the  $j^{th}$  query to the  $F_k$  function, the adversary can hope that it gets a collision with respect to the output of the  $i^{th}$  query, which he had made in the previous interaction with the oracle.

On the other hand, it may so happen that the adversary's  $j^{th}$  query is for the  $F^{-1}$  function. So it could ask for the value of the inverse, it could ask for the oracle service for this input, from the  $F^{-1}$  oracle and in response it gets back the output, basically it sees the inverse of the input  $y_j$  under the key  $F_{k_j}$  right.

So he is basically asking for the access to the  $F_{k_j}^{-1}$  member on the input  $y_j$  and it gets back the output  $x_j$  and based on that it learns the  $j^{th}$  hash value, namely the value of the Davies-Meyer function and that the output  $h_j$  will be  $y_j \oplus F_{k_j}^{-1}(y_j)$ . Now in this case, what is the probability that the  $j^{th}$  value which the adversary has learnt here is the same as the output of the  $i^{th}$  query which the adversary has already learnt in the previous interaction?

Well, the probability of that is the same as the probability that your  $F^{-1}$  function under the key  $k_j$  on the input  $y_j$ , gives you an output which is same as XOR of  $h_i$  and  $y_j$  and again we can run the same argument that we have done for the earlier case. Remember that in the ideal-cipher model, each of the family members  $F_k$  and  $F_k^{-1}$ , they are truly random and independent of each other.

So that is why we can now safely conclude that the probability that  $h_i = h_j$  happens where  $h_j$  is now defined like this, by adversary making an oracle call for the inverse function is upper bounded by  $\frac{1}{|X|-j+1}$ . So in both the cases, irrespective of whether the  $j^{th}$  query is for the  $F_k$  function or the  $j^{th}$  query is for the  $F_k^{-1}$  function, the probability that the hash value which adversary learns due to this  $j^{th}$  query matches the output of the  $i^{th}$  query is upper bounded by  $\frac{1}{|X|-j+1}$ . So that is the fact we have established now.

(Refer Slide Time: 28:58)

### Security Analysis of Davies-Meyer Construction

□ Theorem:  $h_{DM}: \{0, 1\}^{|\mathcal{X}|+|\mathcal{X}|} \Rightarrow \{0, 1\}^{|\mathcal{X}|}$  is collision-resistant in the ideal-cipher model, if  $|\mathcal{X}|$  is sufficiently large

Any  $\mathcal{A}$  issuing  $q$  ideal-cipher queries for  $(F, F^{-1})$  can find a collision with probability  $\leq \frac{q(q+1)}{|\mathcal{X}|}$

□ Consider query numbers  $i, j$ , where  $i < j$  defining hash values  $h_i$  and  $h_j$

➤  $\Pr[h_i = h_j] \leq \frac{1}{|\mathcal{X}| - j + 1}$

□ By the union bound, the probability that  $h_i = h_j$ , for any  $i, j \in \{1, \dots, q\}$ :

$$\Pr[\text{collision}] \leq \sum_{j=1}^q \sum_{i=1}^{j-1} \Pr[h_i = h_j] \leq \sum_{j=1}^q \frac{j-1}{|\mathcal{X}| - j + 1} \leq \sum_{j=1}^q \frac{j-1}{|\mathcal{X}| - q} \leq \frac{q(q+1)}{2(|\mathcal{X}| - q)}$$

□ Case I: If  $q \leq \frac{|\mathcal{X}|}{2}$

❖  $\Pr[\text{collision}] \leq \frac{q(q+1)}{|\mathcal{X}|}$  holds

□ Case II: If  $q > \frac{|\mathcal{X}|}{2}$

❖  $\Pr[\text{collision}] \leq \frac{q(q+1)}{|\mathcal{X}|}$  holds trivially

How many queries adversary is actually making here? We are making an assumption that adversary is making total  $q$  number of queries, and what we have analyzed is that out of those  $q$  queries, what is the probability that a pair of distinct queries gives him a collision for the Davies-Meyer construction in the ideal-cipher model, that probability we have calculated just now. So now what we have to do is we have to apply the union bound, namely we have to take this probability and sum it up over all pairs of distinct  $i, j$  where  $i$  and  $j$  ranges from 1 to  $q$ .

So let collision be the event that by making  $q$  number of queries to the family  $F$  and  $F^{-1}$ , adversary obtains at least one pair of collision. That means adversary obtains at least one  $i, j$  pair, where the  $h_i$  value and  $h_j$  values are the same. So it turns out that by applying the union bound, the probability of collision can be easily upper bounded by  $\sum_{j=1}^q \sum_{i=1}^{j-1} \Pr[h_i = h_j]$ .

What I can do is that, I can substitute the value of the inner summation because now I know that the value of each of the probabilities that is present in the internal summation is upper

bounded by  $\frac{1}{|\mathcal{X}|-j+1}$ , and since  $i$  ranges from 1 to  $j-1$  and there are  $j-1$  such terms, so in the numerator we get  $j-1$  and each of those probabilities is  $\frac{1}{|\mathcal{X}|-j+1}$ . So that is what we obtain and if we further simplify, then we can replace this inequality by  $\sum_{j=1}^q \frac{j-1}{|\mathcal{X}|-q}$ , namely what we have done here is I can always replace each of this  $-j+1$  by  $-q$ , because  $j$  is always upper bounded by  $q$ .

If I take out this summation finally, since  $j$  ranges from 1 to  $q$ , I can upper bound the probability of event collision by  $\frac{q(q+1)}{2(|\mathcal{X}|-q)}$ . So now let us analyze this final inequality for 2 cases. If the number of queries which the adversary has made in this ideal-cipher model is upper bounded by the size of  $\mathcal{X}$  over 2, then we can do some simplification here and end up showing that the probability of collision is indeed upper bounded by  $q$  times  $q+1$  over the size of  $\mathcal{X}$  set.

On the other hand if the number of queries what the adversary has made is greater than the size of  $\mathcal{X}$  over 2, then trivially it so holds that probability of collision is upper bounded by this. That means it does not matter whether we are in case 1 or in case 2, in both the cases the probability of collision by making  $q$  queries is upper bounded by  $q$  times  $q+1$  over the size of  $\mathcal{X}$  and that establishes the fact that we have stated in this theorem.

That means if  $q$  is say some polynomial function in the security parameter and if  $|\mathcal{X}|$  is some significantly large quantity, say some exponentially large quantity in the security parameter, then overall this probability  $q$  times  $q+1$  over the size of  $\mathcal{X}$  turns out to be a negligible function in the security parameter and that is why we can now safely conclude that the Davies-Meyer construction is collision-resistant in the ideal-cipher model and adversary by evaluating the Davies-Meyer function for  $q$  number of messages in the ideal-cipher model could not come up with a collision except with a negligible success probability.

So that brings me to the end of this lecture. Just to summarize in this lecture, we had discussed how to construct the fixed-length compression function and there are 2 approaches for that, one approach where we actually use number-theoretic hardness assumptions and we get provably secure guarantees in the standard model, but we do not opt those constructions in practice because the amount of computations which are involved in those constructions are of order of several magnitude.

Rather we use constructions based on dedicated block ciphers or the existing block ciphers and we had seen one of those constructions namely the Davies-Meyer construction which is very simple, but it turns out that even though practically no attacks have been reported on that construction, we cannot prove the security of that construction just assuming that your block cipher is a secure block cipher or a strong pseudorandom permutation.

We have to make very strong assumptions in our model, namely we have to assume that we are in the ideal-cipher model where block cipher is basically a family of several truly random permutations where the access to the block cipher is through oracle calls even for the adversary. Only if we make this ideal-cipher assumption model, we can prove the security of the Davies-Meyer construction. Thank you.