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

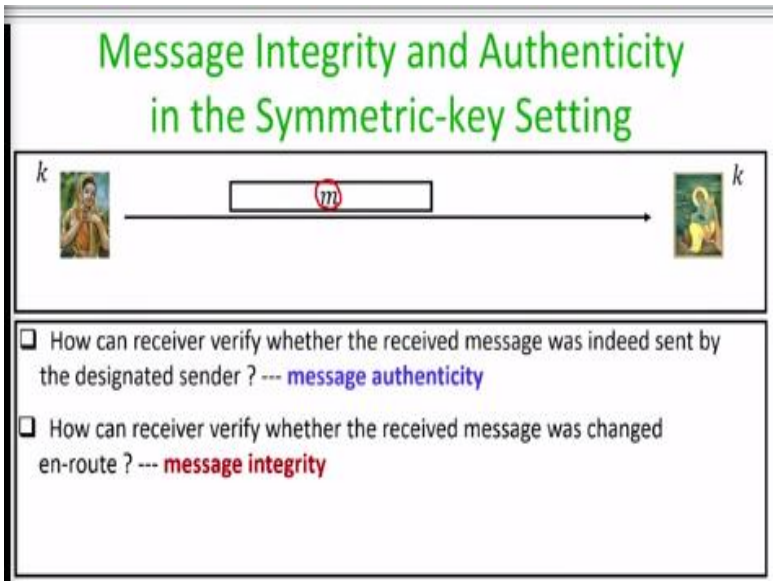
Lecture-22
Message Integrity and Authentication

Hello everyone, welcome to lecture 21, the plan for this lecture is as follows we will introduce the notion of message integrity and authenticity.

(Refer Slide Time: 00:36)

And we will discuss how to achieve these 2 notions using a cryptography primitive which we call us message authentication codes or MAC. And we will discuss the construction for max for fixed length messages using pseudo random functions.

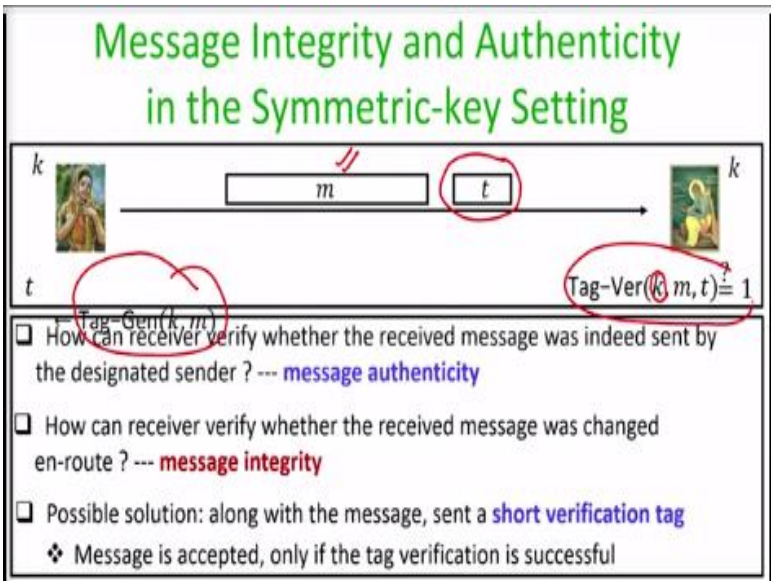
(Refer Slide Time: 00:53)



So, let us start with the definition of message integrity and authenticity in the symmetric key setting. So, the goal of the message integrity and authenticity in the symmetric key word is as follows. So, we assume that we have a sender and a receiver with pre shared random key known only to the sender and the receiver. And assume that there is a bit string, which sender has communicated with which over an insecure channel between the sender and a receiver. So, I stress here that m mean here denote just a bit string, it could be any bit string it need not be a cipher it is a just a bit string.

So, the problem of the message authenticity is as follows. When the receiver receives a bit string over this insecure channel, how can the receiver be sure that whether the received which string was indeed sent by the designated sender, that is a problem of message authenticity. Namely, the goal here is for the receiver to verify whether the contents of the bit strings that it has received over the channel has indeed originated from the so called sender. That is a problem of message authenticity.

(Refer Slide Time: 02:02)



And a related problem is that of message integrity, where the goal for the receiver is to verify whether the received bit string was changed en-route or not. That means assume for the moment that receiver ensure that the received bit string has originated was sent by the designated sender. Now, it wants to verify whether that bit string which it has received was changed during it is course, during it is travel from, during it is journey from the sender to the receiver, that is a problem of message integrity.

And, I would like to stress here that there is no notion or no issue of secrecy here which we are interested to solve. We are just interested to solve the issue of authenticity and integrity here. So, potential cryptographic solution to solve the issue of authenticity and integrity is as follows. Along with the bit string which sender would like to communicate to the receiver, sender can attach a short tag which is independent of the size of the message.

And which is computed as a function of the key and a message. And once the tag is associated with the message when the receiver receives the message along with the tag the receiver can run a tag verification algorithm with respect to the same k with which the sender has computed the tag. And then accordingly it can verify whether the tag verification algorithm output 0 or 1 and if it is outputs 1 that means the tag is successfully verified, then it can accept the message or else it can reject the message, so, that is the overall idea.

(Refer Slide Time: 03:38)

So, we will now discuss how exactly we go about designing such a tag generation algorithm and how do we design the tag verification, so that will be our next goal. But before going into that, let us discuss this 3 goals which we are trying to achieve using different cryptographic primitives. We have the problem of message integrity, message authenticity and message privacy.

So, the message in the problem of message privacy, the goal was to achieve the secrecy of the communication, right. Namely, whatever plain text sender is wanted to interested to communicate to the receiver. The goal of the privacy was to ensure that the message is not linked to any third party. Whereas for the integrity and authenticity, in the integrity and authenticity problem, the goal is the verification of the source and the contents of the communication.

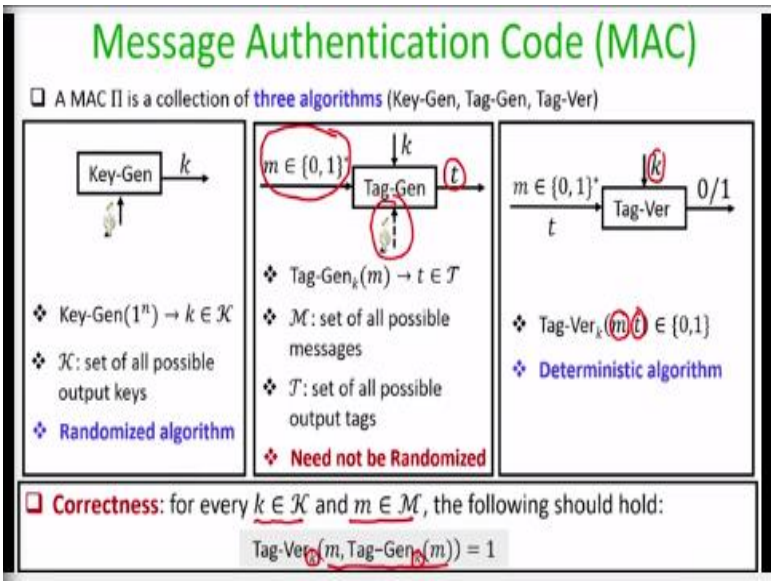
And it is fine if the privacy of the message is lost, that means we are not interested here to actually maintain the privacy of the countess. So, the issue of privacy and the issue of authenticity and integrity they are orthogonal to each other.

(Refer Slide Time: 04:49)

And it turns out that there are several application scenarios where the requirement is only to ensure the integrity and authenticity and not the privacy. For example, if we consider an RFID application, where a user shows a smart card before entering a building then there the goal is to ensure only the achieve the integrity and authenticity. Namely, the goal has to ensure that only the legitimate users are allowed to enter the building.

Whereas a user who is not supposed to enter the building should not give an access to the building, right. So, like that there are several real world applications where the goal is only to achieve integrity and authenticity and not the privacy. And that means we should now discuss how to design cryptographic primitives to solve the problems of integrity and authenticity.

(Refer Slide Time: 05:34)



And a common tool which solves both this problem is called as message authentication code. So a message authentication code or MAC in short, is a symmetric primitive which consists of 3 algorithms. The key generation algorithms output a uniformly random key from the key space, and it has to be a randomized algorithm. Because if it is a deterministic algorithm, then any third party in the world will know the key.

The tag generation algorithm takes a message which sender would like to send to the receiver in an authenticated and a verifiable fashion. And it takes the key which is generated by the key generation algorithm and available with the sender as well as with the receiver and it could be a potential randomized algorithm. So it could have an internal randomness, but it is not necessary that it has to be a randomized algorithm and we will discuss this fact later, right.

So unlike an encryption algorithm which has to be randomized to achieve any meaningful notion of secrecy, when we come to the message authentication code since our goal is not to achieve the secrecy, but rather to solve the problem of integrity and authenticity, it is not necessary that your tag generation algorithm should be randomized. So the tag generation algorithm is a function of the message to be authenticated and the key with an optional randomness.

And it outputs the tag from tag space, and preferably the size of the tag should be independent of the size of the message ok. The tag verification algorithms takes 2 input namely it takes the

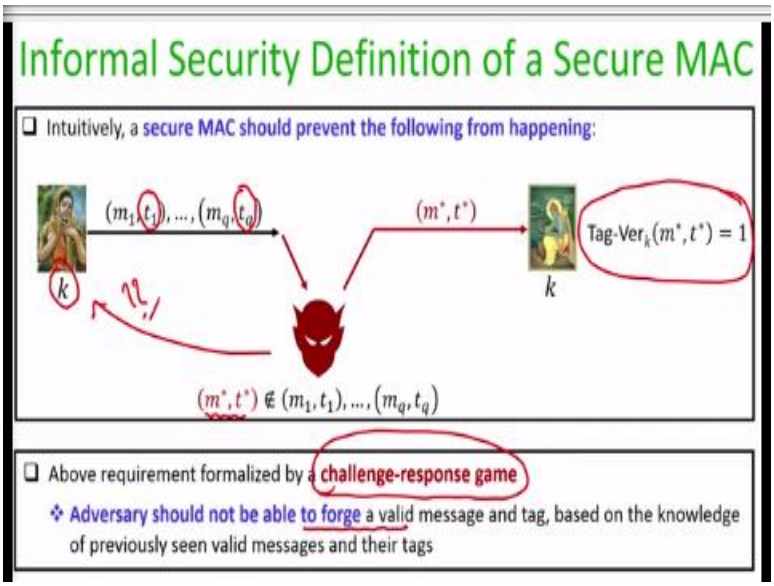
message to be verified along with the corresponding tag and a key is used which is the same key with which should be preferably the same key with which the tag has been generated.

And a tag verification algorithm either output 0 or 1, 0 means it rejects the message that means the tag is not satisfying with respect to the message. Whereas, the output 1 means the tag verification is successful and hence accept the message. And the tag verification algorithm is always a deterministic algorithm because we want the tag verification to be unambiguous. The correctness requirement from any message authentication code is as follows.

We require that for every k which is obtained by running the key generation algorithm and any message which has been authenticated using the tag generation algorithm. The output of the tag verification algorithm with the message and the corresponding tag generated by running the tag generation algorithm where the same k has been used for the tag generation and for the tag verification, the output of the tag verification should always be successful.

So, this is analogous to the correctness requirement of the decryption process in any encryption scheme, so the correctness requirement here is straightforward.

(Refer Slide Time: 08:22)



Now, let us proceed to the security definition of MAC, what exactly is the security property we require. So, before going into the formal definition, let us first try to intuitively understand what

exactly we want to achieve using a secure manner. So, intuitively a secure MAC should prevent the following from happening. Imagine we have a sender and a receiver who have shared a key which has been obtained by running the key generation algorithm.

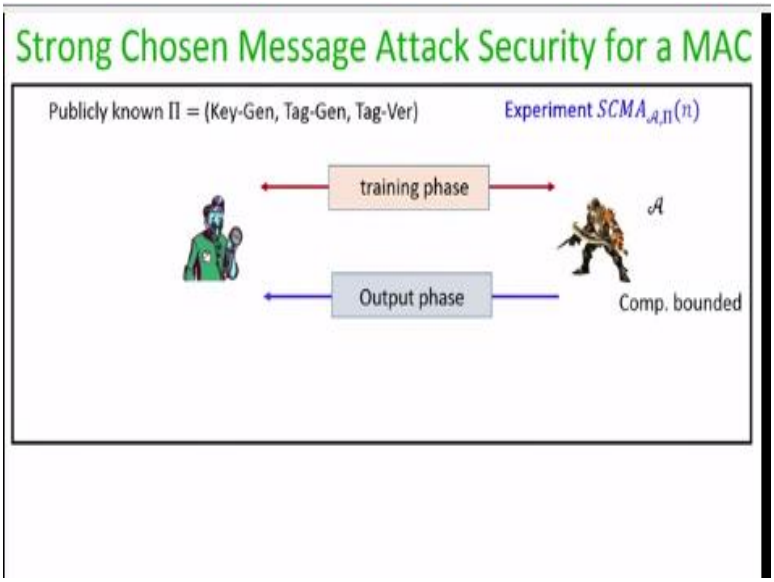
And say the sender has communicated several messages and it has attached the corresponding tags computed as per the tag generation algorithm with respect to the key which is known only to the sender and only to the receiver. And now we have a malicious adversary which intercepts all this message comma tag pairs, but it does not know the value of key. The key is unknown for this attacker.

And the goal of the attacker is to produce a new message from the plain text space or from the message space. And the corresponding tag such that the message comma tag pair is different from all the message comma tags pair that it has intercepted or which has been communicated by the sender. Such that when it forwards the new message comma tag pair to the receiver, the tag verification outputs 1, that means the tag verification is successful.

So basically, the goal of the adversary here is to create a forgery right. Basically, based on the several legitimate message comma tag pairs which it might have seen in the previous sessions, which have been exchanged between the sender and a receiver under an unknown key. The goal of the attacker is to produce a new message comma tag pair such that that new message comma tag pair was never communicated by the sender.

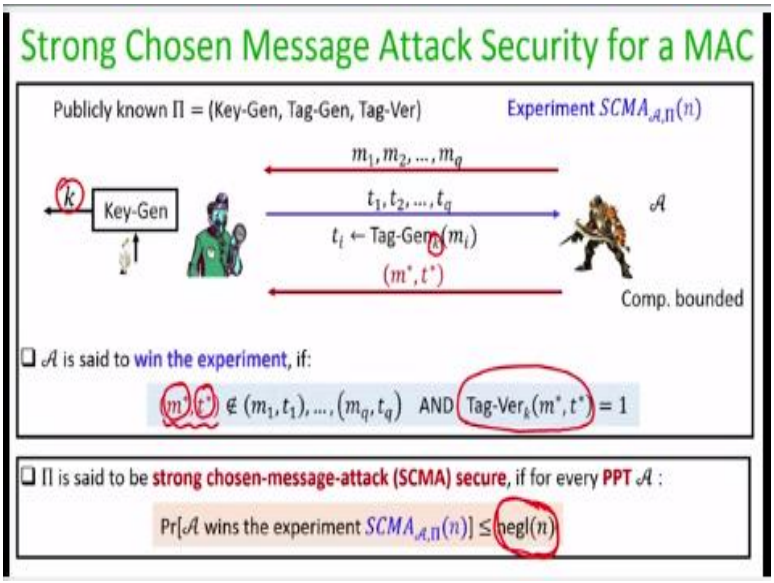
And forward that new message comma tag prior to the receiver such that when verified at the receiving end, the verification is successful. If that is possible, then our MAC is not considered to be secure. That is a intuitive goal which we want to capture through a formal definition. And as we have done for all the cryptographic primitives in this course, this intuitive requirement is going to be captured formally through a challenge response game.

(Refer Slide Time: 10:34)



Where, we have 2 kinds of security notions, so the first experiments corresponds to a stronger security guarantee which we call us strong chosen message attack security or SCMA in short. We have a computationally bounded adversary in this experiment, and hypothetical verifier and experiment basically consist of training phase and an output phase.

(Refer Slide Time: 10:58)



In the training phase the adversary adaptively submit several messages of it is choice and ask for the tags on those messages from the experiment, right. So, this corresponds to the fact the real world scenario where our adversary might have seen several legitimate message comma a tags payer communicated between the sender and a receiver in the previous sessions right. So here to

model that, we give the adversary the chance to train itself where we allow the adversary to submit any message of it is choice from the message space.

And see the tags on those messages under an unknown key k which is chosen by the experiment of the verifier. And the response from the verifier is the tags on those messages as per the tag generation algorithm and an unknown random key k which is not known to the attacker, right. So the number of queries for which the adversary can ask the tag is upper bounded by some polynomial function in the security parameter.

Once the adversary is sent, the goal of the adversary is to output of forgery, namely the goal of the adversary is to output a message comma tag pair. And we say that adversary has won the experiment, if the forgery which the adversary has produced namely the message comma tag pair, which it has produced is different from all the previous message comma tag pairs.

And at the same time, the tag verification of the new message comma tag pair which the adversary has submitted or produced a gives the output 1, that means that tag verification is successful. If this happens, then we say that adversary is able to win the experiment or adversary is able to produce a legitimate forgery even without knowing the value of k. And our security definition is we say that our message authentication code is strong chosen message attack secured or SCMA secure.

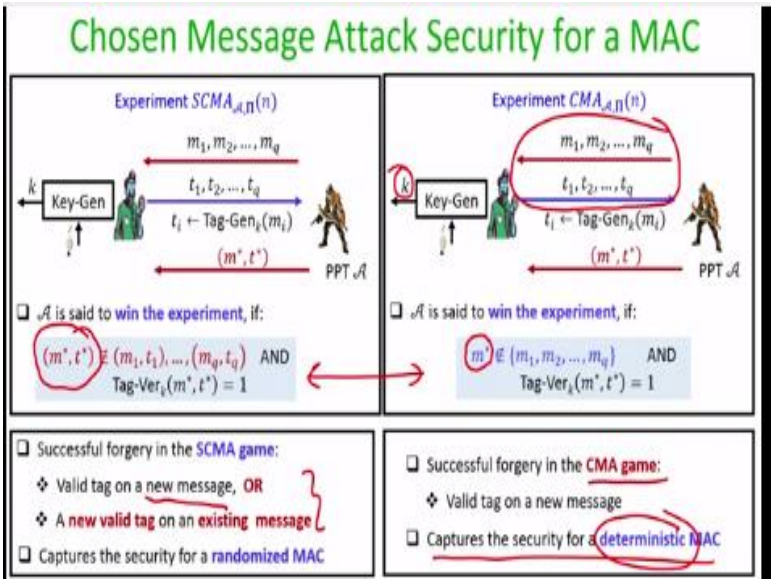
If for every polynomial time adversary participating in this experiment, the probability of adversary winning the experiment is upper bounded by some negligible function. Where the probabilities over the random choice of the key generation algorithm and the queries for which the adversary has ask for the tag service, right. So notice that unlike the previous security notions where we were actually requiring that success probability of that adversary should be upper bounded by half plus negligible.

There is no quantity like 1 by 2 here, because the goal of message authentication code is to prevent forgery not to achieve the privacy, right. That namely in this experiment there is nothing for which the adversary should be able to distinguish. The goal of the adversary is basically to

create a forgery. And there is always a guessing attack where the adversary can just guess a random message and a random tag.

Because it knows the description of the message space and the tag space and there is always a nonzero probability that the guest message and the guests tag indeed constitutes a legitimate message comma tag or a legitimate forgery. So that is why we can never demand in this security experiment that a success probability of the adversary should be 0. The maximum of the best that we can hope for is that the forgery probability of the adversary is negligible function in the security parameter.

(Refer Slide Time: 14:13)



So that is a definition of strong CMA security. Now, there is a related notion of security for the MAC, which we call just a CMA security. And here also the experiment basically consist of a training phase and the challenge phase wherein the training phase adversary submits several messages of it is choice. And sees the tags on those messages under an unknown key k and the goal of the attacker is to submit a forgery.

But the rule of the experiment here or the way we define the output of the experiment here is different. We say in this experiment CMA, that adversary has won the experiment. If the forgery which the adversary has produced is with respect to a new message m star, which is completely

different from the messages for which the adversary has seen the tag, right. So if you see these 2 notions of security, they differ in a very, very certain way.

The successful forgery in the SCMA game or the strong CMA game means that either the adversary has produced a valid tag on a new message or it has produced a new valid tag on an existing message. Because there the requirement is that the overall forgery namely the combined message and the tag when taken together, it should be different from all the message, previous message comma tag pairs that the adversary has seen.

And this condition that m star comma t star is different from all the previous m i, t i can occur in any of these 2 ways. Either the m star could be different from all the m i's that means that sorcery has creating a tag on a completely new message or it might be possible that m star is one of the existing messages. But the tag t star is different from any of the previous tag on the same message.

And this is possible only if your underlying tag generation algorithm is a randomize tag generation algorithm. And that is why when we discuss the syntax of the tag generation algorithm, I said it is not necessary that your tag generation algorithm should be randomized it. It could be potentially randomized or it may not be a randomized algorithm. So if your tag generation algorithm is a randomized MAC, then the strong CMA security gives you the guarantee.

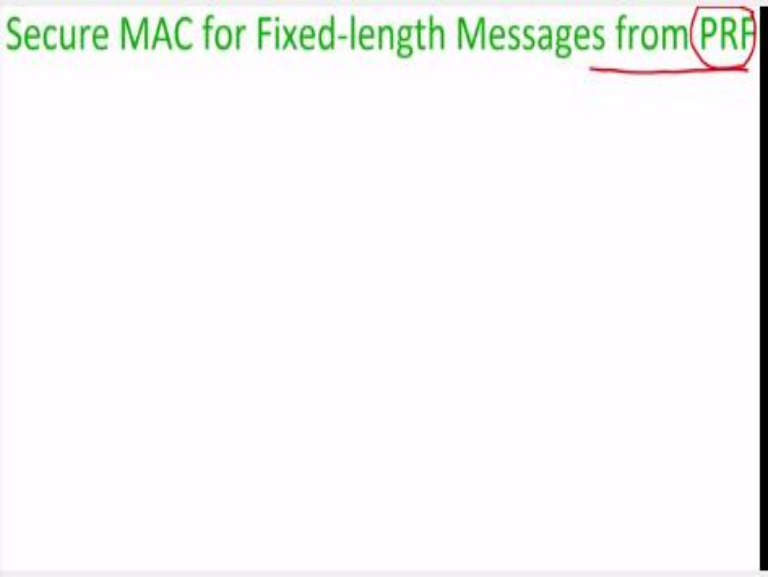
That even if the adversary wants to come up with a forgery on an existing message, the forgery should be with respect to producing a new tag for the existing message. Whereas if we see the forgery in the CMA game, we say a forgery is successful only if the forgery is on a new message which is completely different from all the messages for which the adversary has seen the tag in the previous session.

And this captures the security for deterministic MAC because if your tag generation algorithm is deterministic, then for every message there is a unique tag. And in that case, the forgery is possible only if the forced message is different from all the messages for which the adversary has

already seen the tag. So, that implies that if our tag generation algorithm is a deterministic process, then the strong CMA security as well as CMA security are same.

Because if our tag generation algorithm is deterministic, then this condition m star comma t star different from all the m i, t i can happen only if the message m star is different from all the m i's. In that case, the strong CMA security game becomes same as the CMA game. But if we are using a deterministic message authentication code, then this notion of strong CMA and CMA are completely different.

(Refer Slide Time: 17:49)

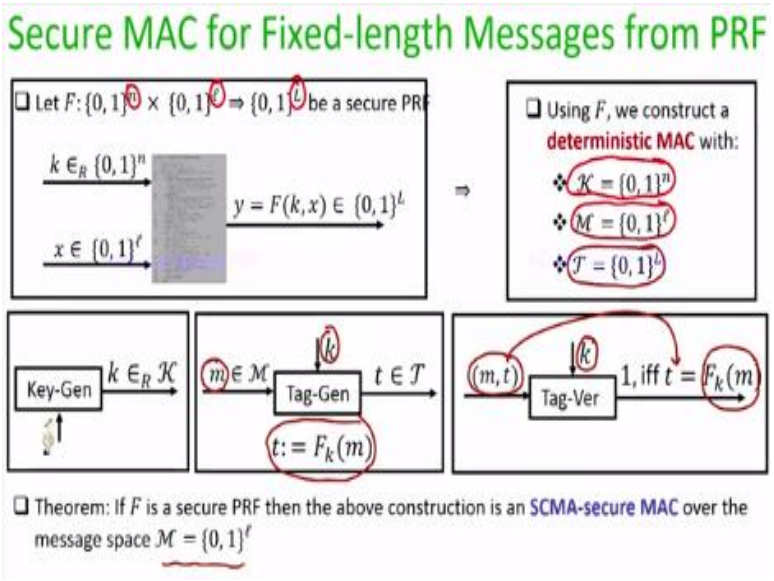


So now we have the definition of strong CMA security and CMA security. So let us see how we can construct a message authentication code for fixed length messages. In the next lecture, we will see that how we can construct message authentication course for arbitrary long messages. And interestingly to compute a message authentication code for fixed length message, we will use our old friend old symmetric key primitive old symmetric key friend namely pseudo random function.

So, and that shows the significance of pseudo random function we had already seen that . If you want to design CPA secure schemes then you can use pseudo random functions in any mode of operation. And we have also seen that if you want to instantiate this PRF, then practice then you can always replace them with triple s E s or any practically known secure block cipher.

So, this pseudo random functions they are very, very significant primitive because they are not only used for designing cryptographic tools to solve the privacy problem. We are now going to see that how they can be used to solve a design message authentication codes to solve the problem of message authenticity and integrity right.

(Refer Slide Time: 18:59)



So assume you have a k pseudo random function with n bit key, little L bit block size and then output of size big L bits. And it is a secure PRF as per the definition of indistinguishability based game. And using this kth pseudo random function, we are going to now construct a deterministic message authentication code where the messages which are authenticated will be upsized little l bit strings.

The key for the message authentication code will be of little n bit string and the tag space or the tag size of the strings will be of size big L bits. And the idea here is very simple, the key generation algorithm of the MAC that we are constructing is going to do the following. It will output a uniformly random key for the underlying PRF that stuck key generation algorithm.

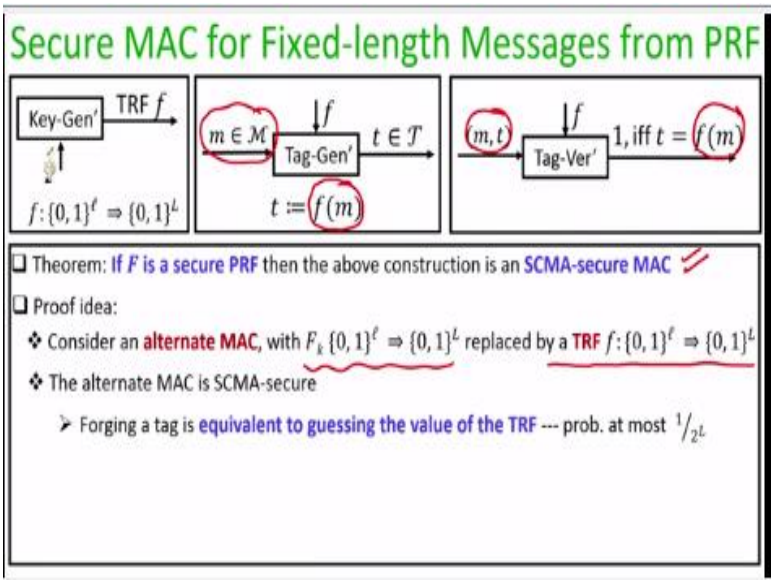
And the tag generation algorithm is a deterministic algorithm where to authenticate a message m of size little L bits under the key k. The tag generation algorithm computes a tag where the tag is just the value of the pseudo random function with the key k and the message m bring the block

input. And the tag verification algorithm is simple, it takes the message comma tag pair and the key k and it just recomputes the tag on the message m.

And compares it with the tag component t which it takes as the input, namely it recompute the tag of the message part under the key k. And compares it with the tag part and at the tag part matches with the recomputed tag, then the output is 1 otherwise the output is 0, right. So in a sense what basically this tag generation and tag verification algorithm is doing is the tag is basically the output of the pseudo random function, getting messages as the input.

And we can formally proved that if the underlying pseudo random function is a secure pseudo random function. Then this MAC construction that we have seen is a strong CMA secure MAC, for authenticating messages of size little l bit strings. And why it is strong CMA secure, because we are constructing a deterministic tag generation algorithm. And as we have argued that if our tag generation algorithm is deterministic. Then the notion of CMA security and strong CMA security are equivalent.

(Refer Slide Time: 21:30)



So this is the theorem statement that we want to prove and I would not go into the full formal details of the security proof here. I will just give you an intuitive argument why exactly this theorem holds. And I leave it as an exercise for you to convert this intuitive argument into a

formal reduction based argument. So for the moment, you consider an alternate message authentication code where all the instances of the key PRF.

Both in the tag generation process as well as in the tag verification process are replaced by an instance of a keyed truly random function mapping little l bit strings to big L bit strings. That means, in the alternate message authentication code, the key generation algorithm outputs a truly random function which will be available both with the sender and a receiver of course this is an inefficient key generation algorithm.

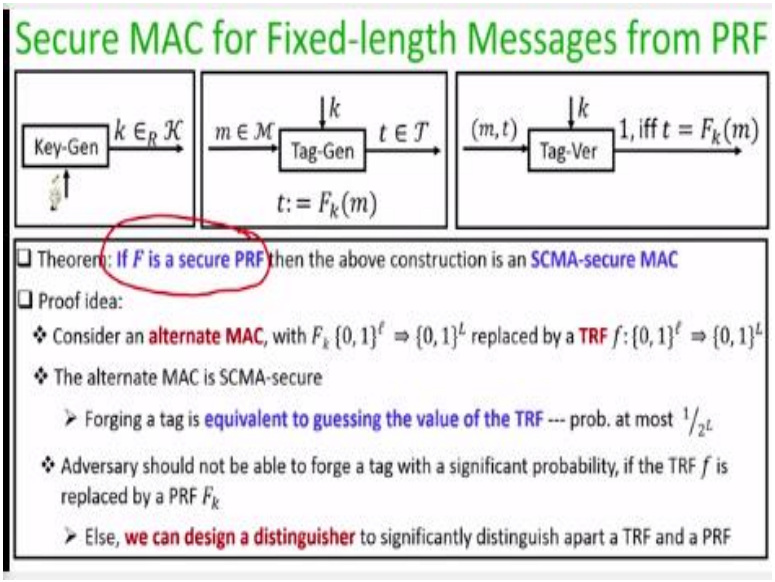
But we are not interested in the efficiency aspect here. The reason I am introducing this truly random function based message authentication code is that it will help us to understand the underlying intuition behind a proof of this theorem, which we want to prove here. So that is a key generation algorithm for the truly random function based message authentication code.

And the tag generation algorithm, for this alternate message authentication code is as follows to authenticate a message, little m. The tag basically is the value this truly random function on the message as input. And the tag verification algorithm of the alternate MAC namely is analogous namely to verify a message comma tag pair. The receiver recomputes the value of the tag by evaluating the truly random function on the message part and compares it with the tag part.

And if the verification is successful then the output is 1 otherwise output is 0. Now, it is easy to see that the truly random function based MAC is definitely strong CMA secure, right. The reason is that if an adversary has seen the value of the tag on several messages of it is choice in the CMA game in the past. And now if it wants to forge a tag on new message basically it has to guess the value of the truly random function on the new message on which it wants to generate the forgery or create the forgery.

And the probability with which it can guess the value of the truly random function on a message m for which it has not seen the output of the truly random function is at most 1 over 2 to the power L, because little f is a truly random function. So that is a simple argument based on which you can state that the truly random function based MAC is definitely strong CMA secure.

(Refer Slide Time: 24:25)



Intuitively, the same should hold namely the strong CMA security should hold even if we replace all the instances of the truly random function by a keyed pseudo random function where the key is not known to the attacker. Because that is what is the security property of a secure pseudo random function. Namely, if the adversary was not able to successfully forge a tag on a message on the truly random function based MAC.

Because it was interacting with a truly random function, the same should hold even if it is actually now even if the sender and receiver are using a message authentication code where instance of truly random function a keyed pseudo random function is used. Because, if at all it is possible now for an adversary to successfully do the forgery with a significant probability which is not negligible.

That means we now know an adversary or we can design an adversary who can differentiate apart an interaction with a truly random function and an keyed pseudo random function. But that will be a contradiction to the assumption that we are assuming the F to be a secure pseudo random function. So that is a overall intuition of this construction, I am leaving the formal details of the reduction proof as an exercise for you.

So that brings me to the end of this lecture. In this lecture, we have introduced the problems of message integrity and authenticity. And we have seen 2 equivalent notions of message authentication code, security notions of message authentication code. Namely, we have seen the definition of strong CMA security which holds against randomized message authentication codes.

And we have seen the security definition, CMA security definition for deterministic message authentication codes. And I stress that the goal of the message authentication code is not to solve the problem of privacy. But rather to solve the problem of integrity and authenticity where a receiver would like to verify whether it is receiving a message from a designated sender. And also it would like to verify whether the contents which it has received from the designated sender has changed en-route or not. I hope you enjoyed this lecture, thank you.