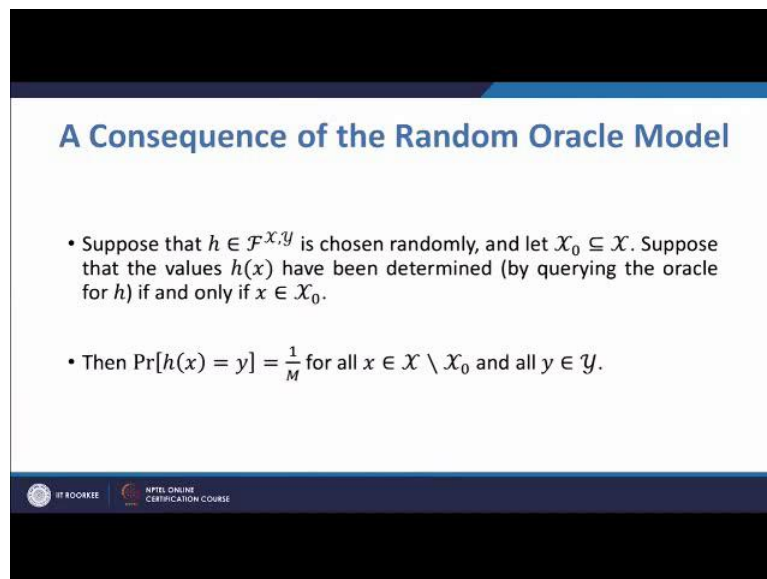


Introduction to Cryptology
Dr. Sugata Gangopadhyay
Department of Computer Science and Engineering
Indian Institute of Technology, Roorkee

Lecture – 18
Randomized Algorithm and its application on Preimage
resistance, second preimage resistance and collision resistance

Hello, welcome to week 4, Lecture - 3. We have been discussing cryptographic hash functions we have seen random oracle models and how the algorithms are constructed assuming random oracle model. Now we will look at these algorithms more closely. So, we go back once more to the consequence of random oracle model that we have discussed before we will just briefly make a recap.

(Refer Slide Time: 01:48)



A Consequence of the Random Oracle Model

- Suppose that $h \in \mathcal{F}^{\mathcal{X}, \mathcal{Y}}$ is chosen randomly, and let $\mathcal{X}_0 \subseteq \mathcal{X}$. Suppose that the values $h(x)$ have been determined (by querying the oracle for h) if and only if $x \in \mathcal{X}_0$.
- Then $\Pr[h(x) = y] = \frac{1}{M}$ for all $x \in \mathcal{X} \setminus \mathcal{X}_0$ and all $y \in \mathcal{Y}$.

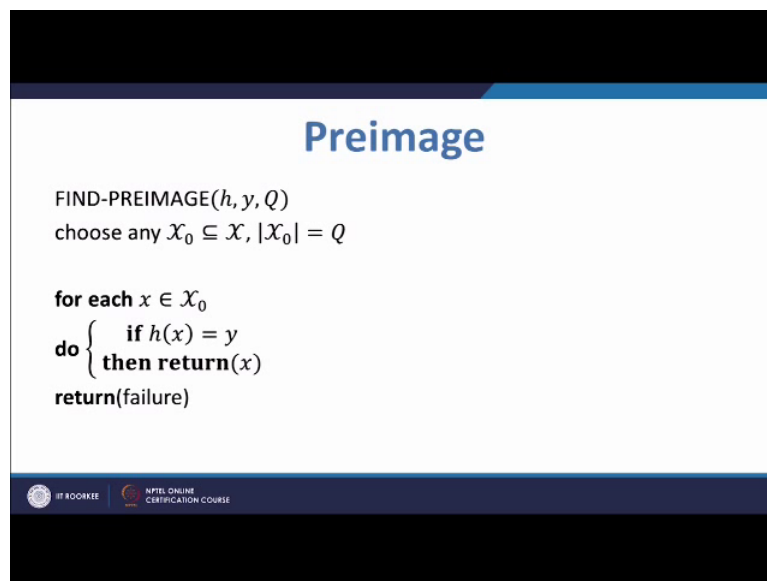
IIIT ROORKEE NPTEL ONLINE CERTIFICATION COURSE

So, when we are looking at a random oracle model, then we can assume that the hash function that we have got is chosen at random from the space of functions from x to y . And then we are only allowed to query h and get a value, and the value should look as if it is coming probably from the space y .

Now this last line is very interesting it says essentially that if we have queried some several many times I mean let us say some capital Q many times covering a space or a subset of capital X , we denote by x sub 0. And then after that if we make the next query the probability

that the answer that we will be getting is a particular value y which is lying in capital Y is same that is 1 by capital M that is if 1 divided by the size of the set y ; that means, it is equi probable. So, the basic assumption of random oracle model leads us to this result, and we will use this result while analyzing the algorithms find pre image, find second pre image and find collision for average success probability.

(Refer Slide Time: 03:05)



Preimage

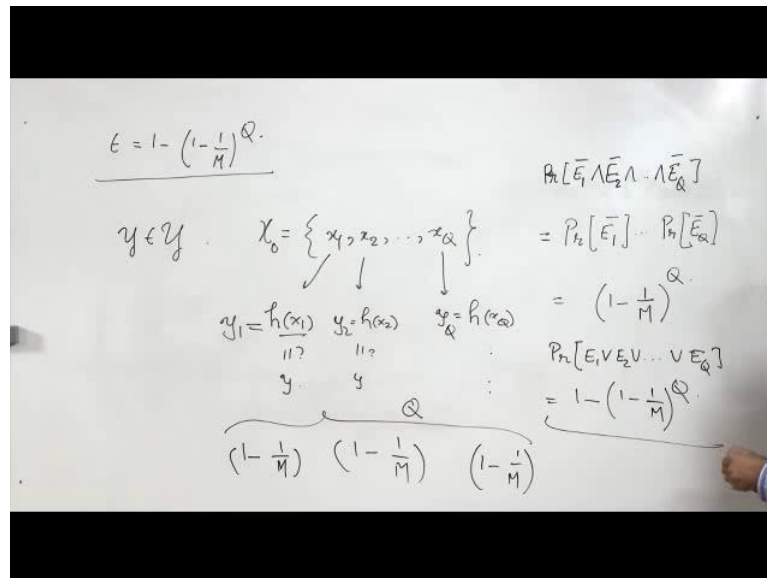
```
FIND-PREIMAGE( $h, y, Q$ )
choose any  $X_0 \subseteq X, |X_0| = Q$ 

for each  $x \in X_0$ 
do {   if  $h(x) = y$ 
      then return( $x$ )
}
return(failure)
```

IIIT KOOBEE NPTEL ONLINE CERTIFICATION COURSE

Now, we look at the algorithm preimage. In this algorithm preimage, we have input h , y , and Q ; Q is a number of queries that we are allowed to make. And then we choose a subset X_0 of X at random such that the number of elements is equal to Q , and we keep on querying. If we obtain a hit we say that the algorithm is successful, otherwise not; we would like to know what is the average case success probability of this algorithm; and let us look at this.

(Refer Slide Time: 04:05)



Now, this says that the average case success probability of this algorithm epsilon is equal to 1 minus 1 minus 1 by M raise to the power Q. Now we will get give the outline of the proof to have a reasonable understanding that why it is so. Now, at the first step, we have fixed a y and we have got the set X 0 on which we will query.

(Refer Slide Time: 04:30)

Success Probability of FIND-PREIMAGE

- For any $X_0 \subseteq \mathcal{X}$ with $|X_0| = Q$, the average-case success probability of FIND-PREIMAGE is $\epsilon = 1 - \left(1 - \frac{1}{M}\right)^Q$.
- Outline of the proof:
 - Let $y \in \mathcal{Y}$ be fixed and $X_0 = \{x_1, \dots, x_Q\}$
 - E_i is the event that $h(x_i) = y$.
 - $\Pr[E_i] = \frac{1}{M}$ and $\Pr[\bar{E}_i] = \left(1 - \frac{1}{M}\right)$. So $\Pr[\bar{E}_1] \dots \Pr[\bar{E}_Q] = \left(1 - \frac{1}{M}\right)^Q$.
 - $\Pr[E_1 \vee \dots \vee E_Q] = 1 - \left(1 - \frac{1}{M}\right)^Q \approx \frac{Q}{M}$.

IIT KHARAGPUR
 NPTEL ONLINE CERTIFICATION COURSE

So, let us now define an event E i, which is the event that h x i is equal to y. So, I have fixed a y; in capital Y, and I have got a set of queries x 1, x 2 up to x Q. And I am evaluating in a

loop, these values y_1 equal to x_1 , y_2 equal to $h(x_2)$ up to y_Q equal to $h(x_Q)$. And each time I am asking let us go back here each time I will ask whether $h(x)$ is equal to y . Let us come back. So, I am asking each time whether this is equal to y whether this is equal to y and so on. Now, what is the probability and now the event is that if for example, the i 'th entry is equal to y that is image of the i 'th entry is equal to y then we say that E_i is true, E_i , we have written E_i over here.

Now we ask a question that what is the probability that E_i will hold. Now that means, now we come back to the probability result that we gave in the beginning of this lecture, so I can get any output with equal probability. Therefore, particularly with y that have fixed when I am querying the oracle with x_1 the probability that I will get y , this is $1/M$ so that the probability that I will get E_i , E_1 , E_1 will be true is $1/M$.

So, the probability that E_1 will not be true is $1 - 1/M$; and the probability that E_2 will be true that is thing will hold, so let me remove this portion probability that E_2 will be true is $1/M$ again and the probability that it will not be true is $1 - 1/M$. And like that I will come to the probability that E_Q is not true, this is $1 - 1/M$. How many times we are going here, so it is Q many times.

So, we are assuming that each of these events is occurring independent to the others. Therefore, if I am asking now that what is the probability that after Q steps, you are not going to get the answer that probability is simply the product of all the $1 - 1/M$'s. So, the probability that E_1 will not happen and E_2 will not happen and so on up to E_Q will not happen is a product of all the individual probabilities and that is $1 - 1/M$ raise to the power Q .

Now, the compliment of this event is that at least one of E_i 's happen. Now if at least one of E_i 's happen or it is true, so that means, that the algorithm is successful. So, probability that E_1 or E_2 or and so on E_Q is $1 - (1 - 1/M)^Q$. Now we see that this is independent of y . So, therefore, if we sum overall possible y 's and divide by the number then we are going to get the same number so it is in fact, the average case success probability. So, we know that it is like this.

(Refer Slide Time: 10:48)

The image shows a whiteboard with handwritten mathematical derivations. On the left side, the probability of success is calculated using the binomial theorem. On the right side, the probability of a pre-image being found is calculated as the product of individual probabilities for each element in the pre-image.

$$\begin{aligned} P_H[\text{Success}] &= 1 - \left(1 - \frac{1}{M}\right)^Q \\ &= 1 - \left(1 - \binom{Q}{1} \frac{1}{M} + \binom{Q}{2} \left(\frac{1}{M}\right)^2 + \dots\right) \\ &\approx 1 - 1 + \frac{Q}{M} = \frac{Q}{M} \end{aligned}$$
$$\begin{aligned} P_H[\bar{E}_1 \wedge \bar{E}_2 \wedge \dots \wedge \bar{E}_Q] &= P_H[\bar{E}_1] \cdot P_H[\bar{E}_2] \cdot \dots \cdot P_H[\bar{E}_Q] \\ &= \left(1 - \frac{1}{M}\right)^Q \\ P_H[E_1 \vee E_2 \vee \dots \vee E_Q] &= 1 - \left(1 - \frac{1}{M}\right)^Q \end{aligned}$$

Now, we can process this quantity a little more and will get something interesting. So, let us do that. So, I have got the probability of success is equal to $1 - \left(1 - \frac{1}{M}\right)^Q$. I can use binomial theorem and write $1 - \left(1 - \binom{Q}{1} \frac{1}{M} + \binom{Q}{2} \left(\frac{1}{M}\right)^2 + \dots\right)$. And if M is reasonably large then $\left(\frac{1}{M}\right)^2$ is very small, so this whole thing will be approximately equal to $1 - 1 + \frac{Q}{M} = \frac{Q}{M}$.

So, the average case success probability of pre image find pre image algorithm is approximately $\frac{Q}{M}$. So, when we are designing a hash function irrespective of whatever function we are constructing, we have to be careful about this. So, if Q and M are close then we will get a reasonably high probability of getting pre image. So, we have to assign we have to take M in such a way that we would not able to make queries so that Q and $\frac{Q}{M}$ is small.

(Refer Slide Time: 12:39)

Second Preimage

```
FIND-SECOND-PREIMAGE( $h, x, Q$ )  
 $y \leftarrow h(x)$   
choose any  $X_0 \subseteq X \setminus \{x\}, |X_0| = Q - 1$   
  
for each  $x_0 \in X_0$   
do { if  $h(x_0) = y$   
    then return( $x_0$ )  
return(failure)
```

IT 800XEE NPTEL ONLINE CERTIFICATION COURSE

Now, we consider the algorithm find second preimage. Here we have a small change because we have a small change from what we did before in preimage we have got $h(x)$ and Q so we are allowed to make Q queries. Since, we are allowed to make Q queries and we have to make one query for the value of x , so we will lose one query. Therefore, in the first step, we are querying the oracle for x and we are getting y , we are storing that value y and we are choosing by M from something outside the set singleton x , so we are choosing x minus set x so X_0 is Q minus 1.

And then we are doing essentially the same thing as we have done for preimage, and if you look at these slides that the arguments are exactly same as before we are choosing here the only difference is that the number of queries is decreased by 1 because one query is already used. So, therefore, we have the probability of the average case success probability of find second preimage is epsilon equal to $1 - \frac{1}{M} - \frac{1}{M^{Q-1}}$ and the arguments are exactly similar only thing is that we will go $Q - 1$ steps rather than Q steps as before. So, the slides are available, you can go through step-by-step and check the proof.

(Refer Slide Time: 14:44)

Collision

```
FIND-COLLISION( $h, Q$ )
choose any  $X_0 \subseteq X, |X_0| = Q$ 

for each  $x \in X_0$ 
do  $y_x \leftarrow h(x)$ 
if  $y_x = y_{x'}$  for some  $x' \neq x$ 
then return( $x, x'$ )
else return(failure)
```

IT 800XEE NPTEL ONLINE CERTIFICATION COURSE

We now come to collision. In find collision, we are given h , we have to find two points which are distinct whose image is same so that is what we do here that we keep on evaluating h over the chosen points, and checking that whether we have got it, got a collision or not. And we would like to find out the success probability of this algorithm.

(Refer Slide Time: 15:28)

Success Probability of Collision

- For any $X_0 \subseteq X$ with $|X_0| = Q$, the success probability of FIND-COLLISION algorithm is

$$\epsilon = 1 - \left(\frac{M-1}{M}\right) \left(\frac{M-2}{M}\right) \dots \left(\frac{M-Q+1}{M}\right).$$

IT 800XEE NPTEL ONLINE CERTIFICATION COURSE

Now, let us look at the final result for X_0 subset of x and with cardinality Q , the success probability of find collision algorithm is epsilon equal to 1 minus M minus 1 divided by M M

minus 2 divided by M and so on up to M minus Q plus 1 divided by M. Let us look at the outline of the proof.

(Refer Slide Time: 16:03)

Outline of the proof

- $X_0 = \{x_1, \dots, x_Q\}$.
- E_i is the event that $h(x_i) \notin \{h(x_1), \dots, h(x_{i-1})\}$ for all $1 \leq i \leq Q$.
- We have the following probabilities:
 - $\Pr[E_i] = 1$
 - $\Pr[E_i | E_1 \wedge \dots \wedge E_{i-1}] = \frac{M-i+1}{M}$, for $2 \leq i \leq Q$.
- $\Pr[E_1 \wedge \dots \wedge E_Q] = \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \dots \left(1 - \frac{Q-1}{M}\right)$

$$\approx \prod_{i=1}^{Q-1} e^{-\frac{i}{M}} = e^{-\frac{Q(Q-1)}{2M}}$$

IIT ROORKEE NPTEL ONLINE CERTIFICATION COURSE

(Refer Slide Time: 16:03)

$X_0 = \{x_1, x_2, \dots, x_Q\}$
 $E_1: h(x_1) \notin \phi, \Pr[E_1] = 1$
 $E_2: h(x_2) \notin \{h(x_1)\}, \Pr[E_2 | E_1] = \frac{M-1}{M}$
 $E_3: h(x_3) \notin \{h(x_1), h(x_2)\}, \Pr[E_3 | E_1 \wedge E_2] = \frac{M-2}{M}$
 \vdots
 $E_Q: h(x_Q) \notin \{h(x_1), \dots, h(x_{Q-1})\}, \Pr[E_Q | E_1 \wedge \dots \wedge E_{Q-1}] = \frac{M-(Q-1)}{M}$

Now, again we are taking X_0 equal to x_1 up to x_Q . So these are the queries that we are making all right. And then we are defining a sequence of events E_i 's. Now let us see how we are defining that E_i is the event of $h(x_i)$ not belonging to $h(x_1)$ and so on up to $h(x_{i-1})$ for all i between 1 and Q inclusive of the end points. So that means that E_1 means that $h(x$

1) and $i - 1$ does not exist here, therefore, $h(x - 1)$ does not belong to ϕ the empty set. Of course, the $h(x - 1)$ does not belong to the empty set, nothing belongs to the empty set, therefore you will always get, it is always true therefore, probability of E_1 is always 1.

Now, what about E_2 , E_2 is $h(x - 2)$ not belonging to the set $h(x - 1)$. Now definitely there is a point $h(x - 1)$ in the domain so that means, that $h(x - 2)$ can be anything other than that wreck point. So, once we know the value of $h(x - 1)$ then we have got some $M - 1$ many choices to find out $h(x - 2)$. And the probability of that is going to be probability that E_2 happens that is E_2 is like this is given that of course I can write that given that E_1 is true E_1 is always true, so that is going to be $M - 1$ divided by M .

Now, let us look at E_3 $h(x - 2) \neq h(x - 3)$ is not in $h(x - 1)$ comma $h(x - 2)$ so that means, the corresponding probabilities, yes, probability that E_2 happens given that E_1 and sorry it is this is E_3 probability that E_3 happens given that E_1 and E_2 holds. That is so that means, I have checked the first element of course, E_1 holds; I have checked the second element second element is second image is not equal to the first image, so E_2 holds, so that probability is this.

Then probability that E_3 , so we have a conditional here that it is not these two, so this conditional gives me $M - 2$ because two elements have already been checked and I am looking at the third element. So, I am building a sequence of probabilities that that can be calculated reasonable easily, so this is what I am doing here. And then ultimately I would like to know the probability that E_1 and E_2 and so on E_Q happens. Now if that is so; that means, that I have got a failure.

So, let us look at this again. So, this means that of course, this is ϕ when I say this, this is happening and what does it mean that I have a failure because before that also I have not a success. And here also I do not have a success and that probability is that E_1 must hold and probability of E_2 given that E_1 has hold E_1 hold E_1 holds always, so $M - 1$ by M that is ok. Now, if you come to this third row and you will find that here, I am assuming that E_2 holds E_2 holds means that I have a failure here and E_1 holds E_1 always holds, so I have got this one.

And I am then asking that what is the probability that E_3 holds, the probability of E_3 means that $h(x - 3)$ is not equal to $h(x - 1)$ and $h(x - 2)$ and so I have a failure up to third attempt. So,



this is so when I when I go up to the Q'th place then I have got E Q probability of that means, that this $h(x_Q)$ is not a member of $h(x_1) \dots h(x_{Q-1})$, this is what we have. And so probability that E Q holds given that E1 Q minus 1 holds, and so that probability is M minus 1 over M. It is better to write like this, $\frac{M-1}{M}$ that is Q I take out Q minus thing Q minus 1 element, so I can choose among M minus Q minus 1 element and divided by M. And should I take the product of all these things I get this probability that probability of E 1 and E 2 and so on up to E Q and that probability is this whole product.

(Refer Slide Time: 23:31)

Outline of the proof

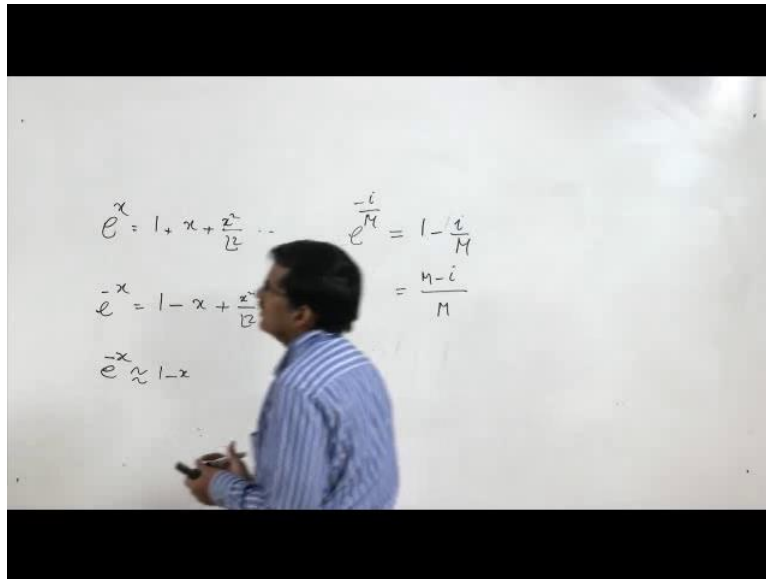
- $X_0 = \{x_1, \dots, x_Q\}$.
- E_i is the event that $h(x_i) \notin \{h(x_1), \dots, h(x_{i-1})\}$ for all $1 \leq i \leq Q$.
- We have the following probabilities:
 - $\Pr[E_i] = \frac{M-i+1}{M}$
 - $\Pr[E_i | E_1 \wedge \dots \wedge E_{i-1}] = \frac{M-i+1}{M}$, for $2 \leq i \leq Q$.
- $\Pr[E_1 \wedge \dots \wedge E_Q] = \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \dots \left(1 - \frac{Q-1}{M}\right)$

$$\approx \prod_{i=1}^{Q-1} e^{-\frac{i}{M}} = e^{-\frac{Q(Q-1)}{2M}}$$

And now we will process this product. If we multiply all these terms, then we will find that all of these things each individual factor approximates to e to the power minus i by M.

(Refer Slide Time: 23:46)





This is because we know that e to the power x is equal to 1 plus x plus x square by factorial 2 and so on, so e to the power minus x is 1 minus x plus x square by factorial 2 and so on. And now if x is small then we can even say that e to the power minus x is approximately equal to 1 minus x and this is what we are using here. So if you write e to the power minus i by M , so that is going to be 1 minus i by M that is equal to M minus i and that is precisely the i 'th term in this product. And so I approximate in this way, so i will run from 1 to Q minus 1 of this, and then I will sum the exponents since the exponents are in arithmetic progression, so I will sum and I will get is minus Q into Q minus 1 divided by $2 M$.

(Refer Slide Time: 25:09)

Outline of the proof

- $\epsilon = 1 - e^{-\frac{Q(Q-1)}{2M}}$
- $e^{-\frac{Q(Q-1)}{2M}} = 1 - \epsilon$
- $\frac{-Q(Q-1)}{2M} = \ln(1 - \epsilon), Q^2 \approx 2M \ln\left(\frac{1}{1-\epsilon}\right)$
- $Q \approx \sqrt{2M \ln \frac{1}{1-\epsilon}}. \quad \text{If } \epsilon = \frac{1}{2}, \text{ then } Q \approx 1.17\sqrt{M}.$

 IIT ROORKEE
  NPTEL ONLINE CERTIFICATION COURSE

Now, I have got an estimate of the success probability, which is epsilon equal to 1 minus e to the power minus Q into Q minus 1 by twice M. And if we go few more steps as given in this slides, we will ultimately find that Q is proportional to square root of twice M natural log of 1 by 1 minus epsilon. And here if we put epsilon equal to 0.5 that is half then we will get this quantity Q is approximately equal to 1.17 root over M. This means that if Q is 1.17 times of the square root of the cardinality of the set y, then it will be more or less half the probability will be more or less half that we will get a collision by querying this many times.

What is interesting over here is that this is not a very large number for example, if you are so let us see what we get 1.17 square root of M. Now what is M, M is cardinality of y. Now suppose we take y to be cardinality of y to be 2 to the power 40 this if this is M then Q is 1.17 times 2 to the power 20, and this is not a very large number, we can make this many queries easily. So, if you are making this many queries then the probability are half or more than half that we will arrive at a success. So, this is what the find collision algorithm tells us under the assumption of random oracle model. This means that no matter how well we build a hash function; if we can make this many queries then we will get a success with probability at least half.

That is the end of this lecture. We will discuss constructions of hash functions in the next lecture.

Thank you.