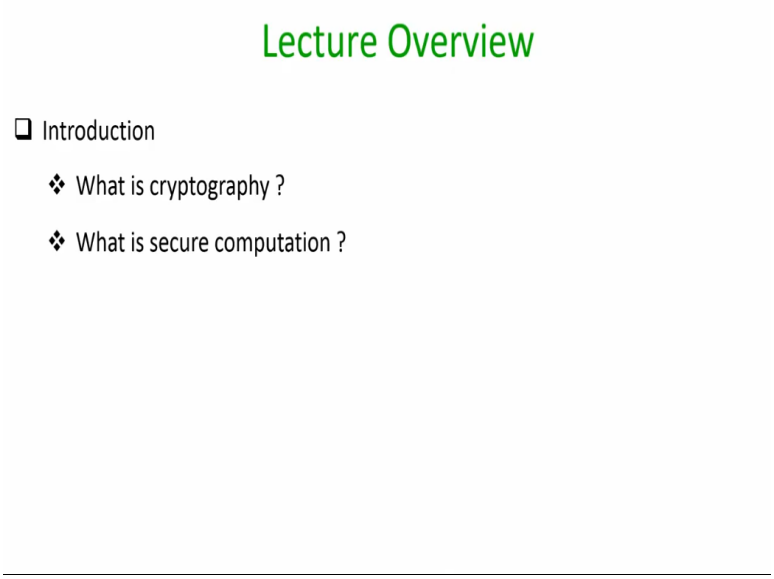
Secure Computation - Part I
Prof. Ashish Choudhury
Department of Computer Science
International Institute of Information Technology, Bangalore

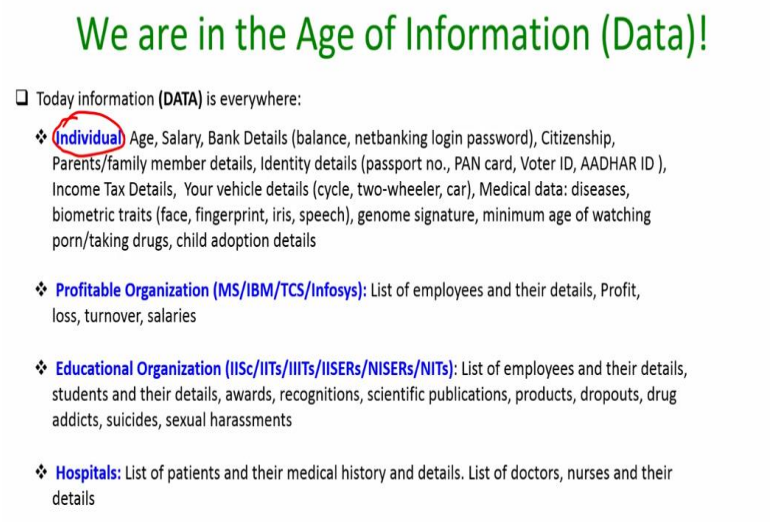
Module - 1
Lecture - 1
What is Secure MPC

(Refer Slide Time: 00:34)



Hello everyone. Welcome to this lecture. So, the plan for this lecture is as follows: In this lecture, we will discuss what cryptography is and we will have a very high level discussion on what is secure computation.

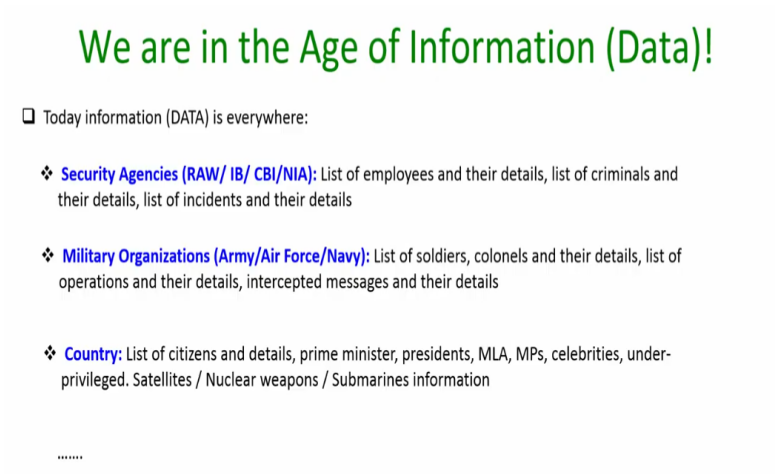
(Refer Slide Time: 00:47)



So, we all are in the age of information, and often, the term data is associated with information. Since we are in the digital age, data is present everywhere. Every individual has a variety of data associated with themselves. Say for example, the age, salary, bank details, both passport number, PAN Card, voter ID, Aadhaar ID, Income Tax details, vehicle details of the individual, the genome signature, etc.

The same way, every multinational company or profitable organization, for instance, Microsoft, IBM, TCS, Infosys, has a lot of data associated with it. For example, its list of employees, the profit, loss associated with the company, the turnover, and the salaries of various individuals. When we consider educational organizations of the country, like the Indian Institute of Science, IITs, IIITs, IISERs, NISERs, NITs or other universities, they have their own associated data. For example, they will have their own list of employees and their details, their students and their details, their awards, recognitions, scientific publications, products, etc. Similarly, every hospital has its own associated data. For example, the list of patients, their medical history and details, the doctors associated with the hospital, the nurses and several other details of that sort.

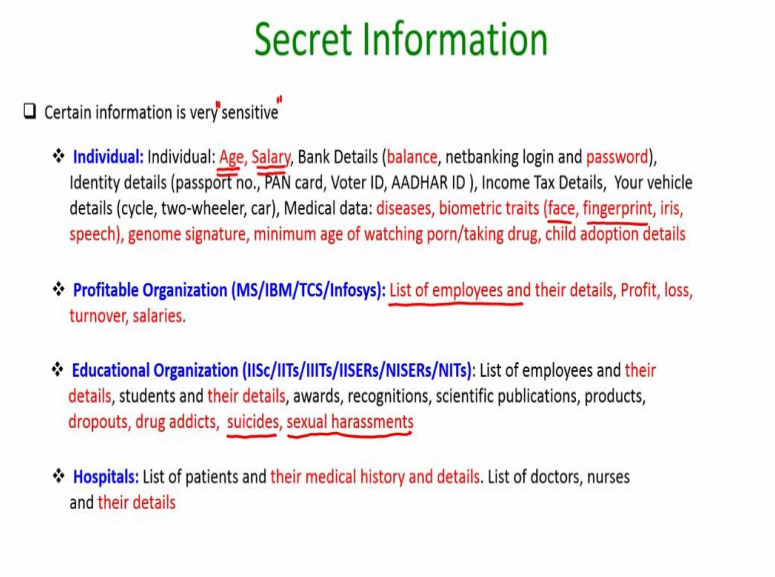
(Refer Slide Time: 02:45)



If you consider security agencies like RAW, IB, then, they have their own associated data like the list of employees, the list of criminals and their details, the list of incidents and their details and so on. If we consider any military organization, be it Army, Air Force, Navy, it has its own associated data like the list of soldiers, cardinals and their details, list of operations and their details, intercepted messages and so on.

Similarly, if we consider individual countries, then, they have their own associated data like the list of citizens, prime minister, president details, MLA, MPs details, celebrity details, what are the nuclear weapons lost by the country, what are the satellites launched by the country, submarine information, etc.

(Refer Slide Time: 03:37)



So, even though we are in the age of data, and data is associated with each and every entity that you can think of, certain kind of data is very sensitive. What do we mean by sensitive? By sensitive, I mean here that you cannot afford to make that information or data available in the public domain. For example, let us again take the case of individual and various kind of data associated with the individual.

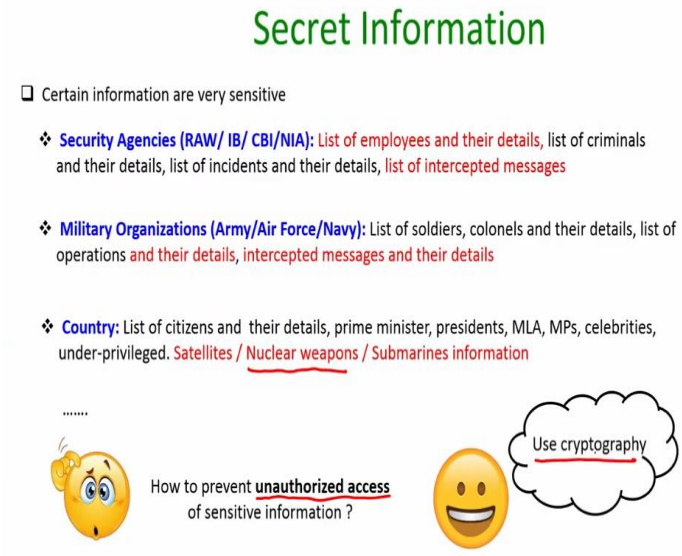
The terms which are highlighted in red color indicate the critical information associated with an individual. For example, people today wish to hide their age. As an individual, I may not be interested to give you the details of my salary information. In the same way, I might be having my net banking login; and the login might be known to everyone; but I may not be comfortable to share the password details with everyone.

Similarly, my face, my fingerprint details, whether I have taken drugs or not, whether I have adopted any child or not, whether I have watched porn or not - all these are examples of sensitive information associated with an individual which the individual may not like to share with any third party. If we consider a profitable organization, then all the data that I have talked about earlier, constitutes sensitive information.

Why would an organization be interested to make public its list of employees and their details? If we consider educational organizations, well, if you visit the website of the educational institute, you might come to know the list of faculty members, the office staff, what awards they have won, what papers they have published and so on. But there might be some information associated with these universities or educational organizations, which you may not find available in the public domain.

You may not be knowing the Aadhaar card details of the faculty members or the office staff of IISc or IITs. In the same way, you may not know the number of students who have committed suicide or what the various sexual harassment cases in that educational organization are. Those details, you may not find available in the public domain. That constitutes sensitive information. If you consider hospitals, then you might be find the list of doctors, nurses; those details, you can find out from their website. But, you may not be able to find the medical history of the patient, what kind of diseases they had, what treatment they have gone through etc. Those constitute of sensitive information associated with the hospitals.

(Refer Slide Time: 06:53)



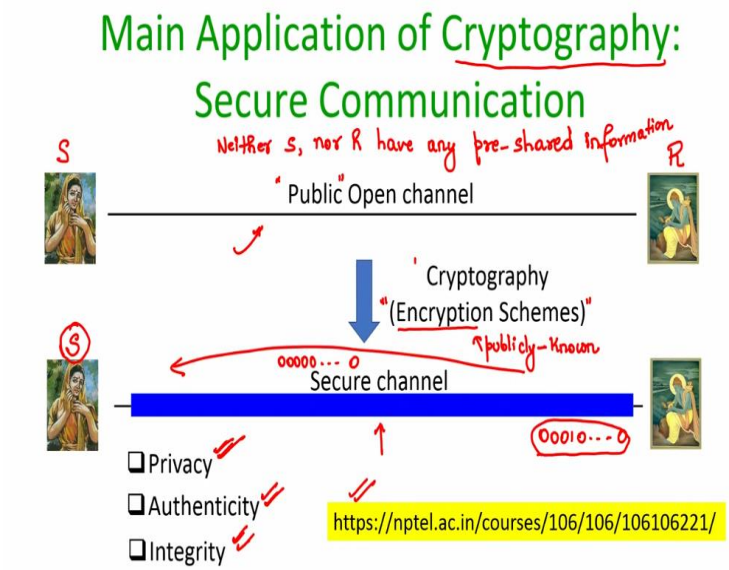
In the same way, if I consider security agencies, then, even the list of employees of RAW or IB is sensitive information, and you may not find those details available in the public domain. You might find a list of criminals and their details, but the employee details of these security agencies, what are the messages that have been intercepted by them - these constitute sensitive information associated with the security agency.

In the same way, if you consider military organization, Army, Air Force, Navy, then again, you may not find the details of the soldiers. You might somehow find out who is the head of the army, but what exactly are the personal details of the head of the Army or the individual soldiers, you may not find them available in the public domain. Similarly, what are the messages that they have intercepted and so on, that constitutes sensitive information.

The same way, if we consider countries, then the complete details of the nuclear weapons, the nuclear program, the satellites launched by the country, those details are very sensitive information. So, the question that comes to our mind now is the following: We are in the digital age where the data is associated with each and every entity, and this data is available somehow in digital form. How do you prevent unauthorized access of this sensitive information?

By unauthorized I mean, how do I prevent a third party who is not authorized to access this information to get access to this information? So, I would like to prevent this data from being accessed by any such third party. And one might say that, okay, if you want to prevent unauthorized access of the sensitive information, why not use cryptography? Because, typically, cryptography is a mathematical science which is used to prevent unauthorized access of sensitive information. So, what exactly is cryptography?

(Refer Slide Time: 09:02)



On a very high level, cryptography is the mathematical science, and the main application of cryptography is that of secure communication, namely, we want to solve the problem of secure communication using the help of cryptography. So, what exactly is a secure communication

problem? Imagine we have a sender S and a receiver R; and neither S nor R have any pre-shared information. They do not have any pre-shared information of any sort.

Say for example, they do not have any secure number known only to S and R; they do not have any kind of secure algorithm known only to S and R; no information of that sort. In fact, it could be the case that sender and receiver, S and R do not know each and other a priori, and they are interacting for the first time. They are connected by a public open channel, say for example, internet.

You can imagine S is a computer, R is a computer which is a part of a bigger network, and they are connected by a public channel which is known to everyone. Neither S nor R have any kind of pre-shared information. Now, on a very high level, cryptography helps the sender and a receiver to perform what we call as secure communication. That means, cryptography gives you a variety of algorithms which we typically call encryption schemes.

Using those encryption schemes, the sender and receiver, who do not have any pre-shared information, can talk over this public channel in a secure way. And when I say encryption schemes, they are also publicly known. So, it is not that I am talking about algorithms which are known only to sender and receiver. Cryptography will come up, and will provide you various kinds of encryption algorithms.

Using those encryption algorithms, the sender and receiver can exchange messages over this public channel itself. There need not be any private channel available between the sender and the receiver. If there is a public channel, by using the encryption scheme, the sender will convert whatever messages it would like to communicate to the receiver. Those converted texts or messages are called cipher texts.

And those cipher texts will be communicated over this public channel. The property of these encryption algorithms will be that it gives you the feeling as if sender and receiver are actually communicating over a secure channel. That means, you can imagine as if there is some kind of a tunnel, and only sender and receiver can pass messages through that tunnel. Any outsider who is viewing the communication that is happening inside the tunnel cannot make out anything about what is actually happening inside the tunnel. That is, intuitively, the way you can imagine the goal of cryptography. So, basically, it gives you the effect of a secure channel

emulated over a public open channel. And when I say secure channel, by that I mean that the communication that is happening between the sender and a receiver achieves 3 properties.

The first property is the privacy property, which guarantees that whatever communication or messages which are exchanged between the sender and a receiver by using these encryption schemes, any third party who is monitoring the communication, cannot make out what exactly are those messages. That is roughly what we mean by the privacy property. The second property of this secure communication is that of authenticity, namely, both sender as well as receiver will have the guarantee that the messages that they are receiving from the other party indeed belongs to the receiver or sender vice versa.

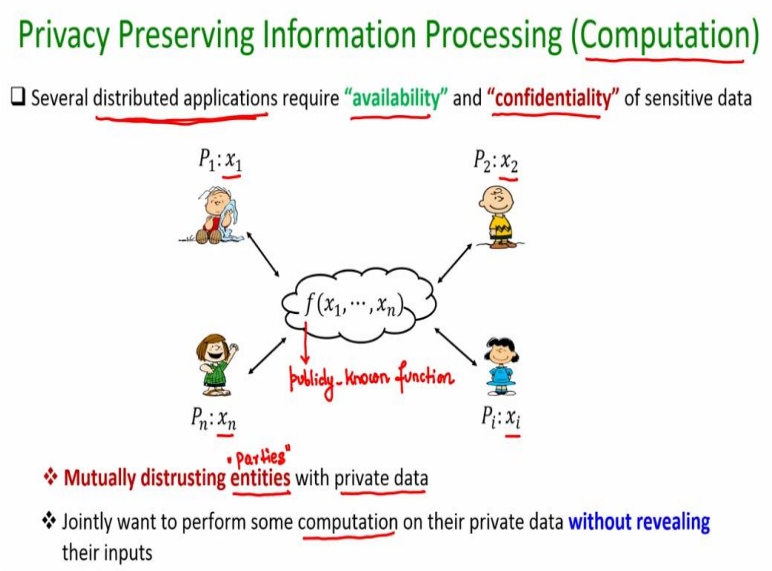
So, for example, if the sender is obtaining some messages from this receiver, then these encryption mechanisms will some verifying mechanism to this sender to verify whether those messages are indeed coming from the so called receiver. That means, it would not be possible for a third party to inject messages on the behalf of the receiver and forward it to the sender. That would not be possible. That is what these encryption schemes guarantee. And that property is called as the authenticity property.

The third property is integrity which guarantees that any communication which is tampered with by a third party, when it is going from one side to the other, is detected at the receiving end. Suppose that the sender has forwarded a bit string, say 00000, and there is a third party who has changed say one of the bits from 0 to 1. Then, there will be a mechanism at the receiving end provided by these encryption schemes to verify whether indeed the bit string that it has received is a changed bit string or whether the same bit string was communicated by the sender. So, that is, roughly, what the integrity property is. Secure communication means privacy, authenticity and integrity of the communication.

The main goal of cryptography is to provide you with encryption schemes using which 2 unknown parties who do not have any pre-shared information can perform secure communication by interacting over a public channel. That is a main goal of cryptography. Of course, now the domain of cryptography has expanded tremendously and you have varieties of other problems which can be solved by cryptography apart from the classical problem of secure communication.

If you want to know more about encryption schemes, the types of encryption schemes, their security properties, etc, you are referred to this course on Foundations of Cryptography, available on NPTEL.

(Refer Slide Time: 16:40)



In this course, our goal is to solve the problem of secure computation. Even though the main application of cryptography is secure communication, our main goal in this course is to show how to actually use cryptography to perform secure computation. Namely, we will encounter several distributed applications where there will be multiple agents, multiple parties with their own sensitive data, and together they would like to perform some computation while keeping their data private.

Namely, we will be considering several distributed applications where each individual will have some sensitive data, and we would like that data to be available for doing the computation. But at a same time, we would like to not reveal the data. And this might look like a conflicting task, a conflicting requirement. On one hand, I am saying that you want the data associated with individual entities to be made available, so that you can perform computation on the data. But on the other hand, I am demanding that the data should remain private. That means that the data associated with the individuals should not be learnt by a third person. This looks like a conflicting requirement. And that is precisely what we would like to achieve. We will see a variety of mechanisms which will help you to solve this problem. So, now, let us try to understand what exactly I mean by saying that you have several distributed applications where you require the availability and confidentiality of sensitive data.

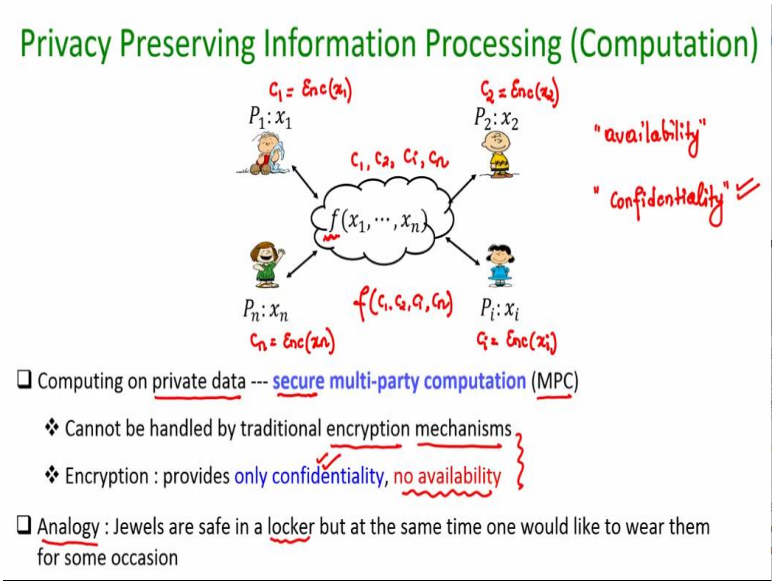
I am talking about the following scenarios. Imagine you have mutually distrusting entities. They do not trust each other. Entities are also called as parties. So, imagine you have n parties, $P_1, P_2, \ldots, P_i, \ldots, P_n$. Each entity has its own private data. Again, I am abstracting out the data and using the notation x_i to denote the private data associated with the ith party.

So, the data associated with P_1 is x_1, the data associated with P_2 is x_2, the data associated with P_i is x_i, the data associated with P_n is x_n. Again, this is an abstract data. It could be any kind of data. So, for instance, if the parties are the hospitals, then this $x_1, x_2, \ldots, x_i, \ldots, x_n$ could be the medical database. If these parties are individuals, then this $x_1, x_2, \ldots, x_i, \ldots, x_n$ could be their Aadhaar number, PAN card details and so on.

The important thing here is that the parties do not trust each other. And that is why they are not willing to provide their private data to the other parties. However, they are interested to perform some computation on their private data without revealing their inputs. Now, again, that computation could be any kind of computation. And we abstract that computation by publicly known function f.

That means, the details of the function are publicly known. And the goal here is basically to design a mechanism which allows these n parties to learn the output of the function f on the inputs x_1 to x_n, without revealing the inputs x_1 to x_n to the individual parties. That is what we are interested to do here. That is what we mean by privacy preserving computation here. You want to maintain the privacy of the data, but at a same time, would like to perform some computation on the joint data.

(Refer Slide Time: 21:34)



This problem of privacy preserving computation or privacy preserving information processing is also called as computing on private data. Why it is called computing on private data? Because you would like to perform some computation on the data which is private to individuals. It is not the case that $x_1, x_2, \ldots, x_i, \ldots, x_n$ is available in the public domain. The data belongs to individuals and is private, but you would still like to perform some kind of computation.

This function f could be any kind of computation. In the next lecture, we will see a variety of real-world computations which can be abstracted by this function f. Since we would like to perform some computation on private data which is available with individuals, we call this problem as computing on private data. And this problem is also called as secure multi-party computation or MPC in short.

Why multi-party computation? Because you have multiple parties associated and the computation is performed on the data associated with the individual parties. And you would like to preserve the privacy of each individual's data. That is why this computation is called secure multi-party computation. Now, remember, our goals are the following: We want the availability of the data and we also want to achieve confidentiality.

Because, only when the data $x_1, x_2, \ldots, x_i, \ldots, x_n$ is available, the function f can be computed. Because, without knowing x_1 to x_n, how can you expect the output of the function to be made available? At the same time, I am demanding here that no one should learn the inputs of the other parties. So, if you are thinking that let us try to solve this problem using cryptography or encryption algorithms, then that may not help you to solve this problem.

Because, if each individual party encrypts its data and provides the encrypted data to other parties, say for instance, P_1, it computes an encryption of its data. So, let us say P_1 has computed a cipher text c_1 and that is obtained by encrypting x_1. Similarly, c_2 is the result of encrypting x_2; c_i is the result of encrypting x_i; and c_n is the result of encrypting x_n. Imagine that each party encrypts its data.

And now, the encrypted data,, $c_2, .., c_i, .., c_n$, is available in the public domain. So, encryption or encrypting the individual data will help you to achieve confidentiality. You have made your data available in the public domain, but since it is available in an encrypted fashion, no one can make any sense of it. That is fine. So, one of the goals is achieved. But how can you perform computation on the encrypted data?

Namely, you cannot compute$f(x_1, x_2, .., x_i, \dots, x_n)$, and expect that the result gives you the output of the function on the clear inputs $x_1, x_2, \dots, x_i, \dots, x_n$? That may or may not be possible. Because, as soon as you encrypt your data, that becomes jargon. And then you cannot perform a computation on the jargon and expect that the result is actually the result that you expect from the computation.

That means, confidentiality is achieved by encrypting the data, but the goal of availability is not achieved. And the analogy here is the following: If you are thinking that encrypting the data will help you to solve the problem, then, that is not the case. An analogy here is that, if you have a jewels, which is a very precious item, the analogue of data is the jewels. Imagine that each party has some associated jewels.

So, if you want to ensure the security of the jewels, then you can keep the jewels in a locker and keep the lock of the locker with yourself. That will ensure the confidentiality; that will ensure that the jewels are safe. But at the same time, on some rare occasions, say for instance, if there is a party or some social gathering, then you would like to wear those jewels and show it to everyone.

And that is what the analogy is here. If you are thinking that you can solve this problem by asking each individual entity to encrypt its data, just encrypting the data and making the

encrypted data available in the public domain is not going to solve your problem. So, with that, I end this lecture. Just to summarize, in this lecture, I introduced the problem of secure multi-party computation.

We have not given the precise formal problem definition, but we have roughly seen what exactly we want to achieve, what exactly we want to solve. Our problem is the following: We have a set of mutually distrusting parties. Each party has some private data associated with itself. And we want a mechanism which allows these parties to perform some joint computation or compute some publicly known function of the private data associated with the individual entities without revealing the data to the other parties. That is roughly is our goal. Thank you.