Operating Systems Fundamentals
Prof. Santanu Chattopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture – 01
Introduction

So, welcome to this course on Operating System Fundamentals. So, in this course, we will see what are the components of an operating system and why is it needed. So compared to the basic computer hardware or basic computer architecture , so why do we need an operating system that is the first thing to understand and if you look into the platform like given a microprocessor , if you have done if you have knowledge about microprocessors, you know that it can take a set of instructions and execute them.

So anything that you want to do you can do it writing programs in the assembly language of that microprocessor and then doing some hand translation of that and finallysomehow ensuring that the microprocessor has got access to those instructions and then, the required job can be done. So if you user say wants to calculate say roots of a quadratic equation. So user can write a program in assembly language for getting the coefficients a, b and c of the quadratic equation and then, write down the program for doing the root calculation. This is fine.

But the amount of effort that the user needs to take so that actually makes it difficult. So a earlier computer systems, so somebody has to be an expert with the hardware and interfacing of it, that is how to talk to that basic hardware. So, that part one should be a master to get something meaningful done by the underlying hardware system.

So operating system actually evolved after that and it came as an interface between the user and the basic hardware. So, that the life of the user becomes simple. So, if the programs can be easily written; programs can be easily modified and all that and also the system resources that we have in a computer system. So, that gets utilized in a proper fashion. So, in this entire course what we are trying to see is how this evolution has taken place; what are the components of the operating system , why the individual components are important and how they interact with the underlying hardware.

(Refer Slide Time: 02:39)

So with that we will start this course and the concepts that we are going to cover in this course are like this. In this particular chapter is what is an operating system that is the most fundamental concept that we have to make it clear and fortunately or unfortunately,the space the precise definition of operating system is difficult. So different people will look at it from different angles. So you maybe if somebody looking into the user view perform the user perspective we will define it from the users perspective. Somebody looking it from the systems perspective will be giving information in the in terms of system perspective. So, that way, the definition of what is an operating system so, that is going to vary.

So, it is dependent on the computer system organization. How the computer system is , what are the components in the system? So many a times so the operating system design and its behavior is governed by the underlying computer system organization. There are the modules that or the hardware components that we have in the system.

Then the operating system is structured. So otherwise what will happen is that tomorrow you want to change the a part some of the features of the operating system, it becomes very difficult. So we have the over the years the structure has evolved for the operating system in terms of the component that it should have and the their interaction pattern.

So how much responsibility this individual components take so that is now more or less standardized and if you are designing a new operating system, then all then also we should try to think in that line. So we have to see like what are the components or what how the how are the individual components are generally designed and we should be guided by the design philosophy behind those components. Then we have got the operating system operations. How the system does operates? So as far as the comments are concerned or are it based on some graphical interface or command line interface so depending on that, so the operating principle of the operating system itself is another issue. Then individual responsibilities of an operating system we have got process management. So whatever the operating system does in a hardware system, so that is in terms of processes.

So process management is an issue then memory management. So, most of the computer systems that, we have today. So, they want to have large amount of memory. So now,having main memory in the range of gigabyte is not an issue and the having secondary memory of the disk in the in the range of terabytes is not an issue. So, those modern systems so they are having memories in that range. But having that much of memory is not sufficient. I should be able to utilize it properly. So that utilization pattern is this memory management and storage management when it comes to the secondary part so it to a secondary memory that is hard disk or this optical disk and all. So they are actually the storage part. So this memory part and the storage part, they are management plays a significant role.

And one thing we must understand that this memory and storage there is a basic difference in their management because when I talk about memory. So this is a semiconductor device. So their access time is pretty low and So they can be accessed very fast. So the speed gap between the processor or the CPU and the memory, this main memory that is RAM, ROM etcetera so that is not that much compared to the case when we have got the secondary storage in terms of the disk and say optical disk or say any other secondary storage devices. So there the speed is a factor.

So the how do we take care of the speed difference so that is a big issue. And the storage wise like we cannot have a steel to a as far as today's technology is concerned. So we cannot have main memory in the range of terabytes. So naturally if we are looking for

very large storage, we have to use the storage in terms of the secondary storage. So their management becomes an very important factor.

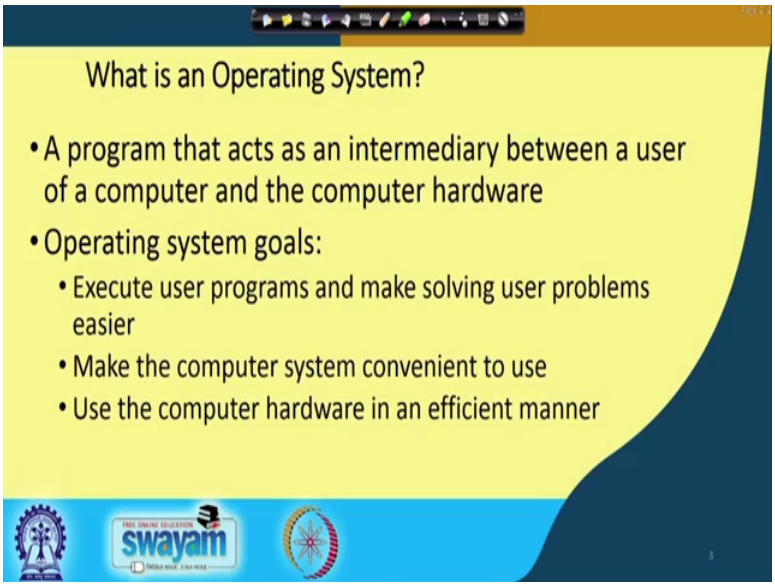
Then, the system protection and security like you most of the systems that we have today. So they are multi user system. So multi user system is good because several users can run their programs in they can give jobs to the computer system and the computer system can do that in due course of time. But the problem that comes is that everything is every user is not very , I should say there may there may be some users who are trying to say steal the data of some other program or there may be some misbehaving users or user processes that are created that tries to cause damage to the or damage to other user programs. So that protection and security is a very important issue in that case.

Then some of the data structures that we have in a computer system they are core to the basic operation of the system. So certain things that we need to remember. For example, the set of users who are currently running their programs ok. So, that is generally known as the process table. So, there we have got information about all the currently running programs that we have in the system. Similarly maybe at some point of time, there are 10 users and 10 users are doing manipulations on files and they have opened say 20 files for manipulation. So in my computer system there may be 100s of files , out of these 20 files are currently being operated on. So the system needs to remember which 20 files have been opened now.

So, in this way we can say that the operating system must remember what are the files that are open. So that way there are some important data or important data structures that the computer system will have and the operating system needs to protect them. So Kernel data structures are those data structures which are important for from the systems perspective and the operating system needs to protect their access and the computing environment that is provided.

So, finally, what type of environment that is provided to the user, so that becomes an important issue. So in this particular chapter we will try to look into we will like to have an overview of all these concepts and in subsequent chapters or subsequent discussions, we will be taking up the topics one by one and discuss in more detail about each of them.

(Refer Slide Time: 09:33)



So to answer the fundamental question like, what is an operating system. So these are some of the definitions that we can think about. So operating system is a program that acts as an intermediary between a user of a program and the computer hardware. So this is more of a Layman's description. So as a computer scientist we can have a much better definition than this. But to start with we will concentrate on this definition only that tells that this acts as an interface.

So as I already said even if this interface is not there. So if this intermediary is not there, then also an user can utilize the underlying hardware because the underlying hardware what it understands is nothing but a set of electrical pulses. So if you can make some arrangement by which these pulses are given to the basic hardware in proper time instants then we are done. So we do not have to do anything we do not have to we do not have to have any intermediary between the user program and the basic computer hardware.

So this operating system is an intermediary between a user of a computer and the computer hardware. So what are the goals? S, there are you we can think about a few goals of it. First goal is to execute user programs and make solving user problems easier. So again this definition like how do you what do you mean by making the problems problem solving easier? So this is basically the experience that we have. For example, if you are looking into today's interface today's computer systems, so all of them or almost

all of them have got some window based interface , graphical interface in which if you mouse clicks solves most of the problems.

Many user program execution to some system operation to everything , almost everything can be done by means of mouse clicks. If you go back by about say 10 years, then at that time this type of graphical user interfaces were not that common and people had to give comments in terms of in terms of in terms of keyboard. From the keyboard they have to type some comment and accordingly, the operations used to be done.

So basic operation remains same. In both the cases maybe you are creating a directory or creating a folder and or you may create by making a opening a new file or creating a new file like that. So basic operations were there as well as they are now, but, the way that the user feels today if somebody is given a system, where all the interfaces are by means of this by means of say these keyboard comments then perhaps nobody is going to use it. So that makes it difficult.

So it is the job of the operating system to execute user programs and make problem solving easier. The second goal is to make the computer system convenient to use. So this is again I should be able to do many things sitting at the terminal itself where today maybe you can see that even many displays they have got some USB interfaces in which you can plug in your pen drive and work from there. So you see this pen drive when you are attaching to this display device.

So, it is not that it is becoming a part of display device. So ultimately it is treated as a part of the basic computer hardware. But this extra thing that is done, so that helps in convenient usage of the system because your CPU box may be located somewhere else and you can just be operating with your keyboard and this display that you have.

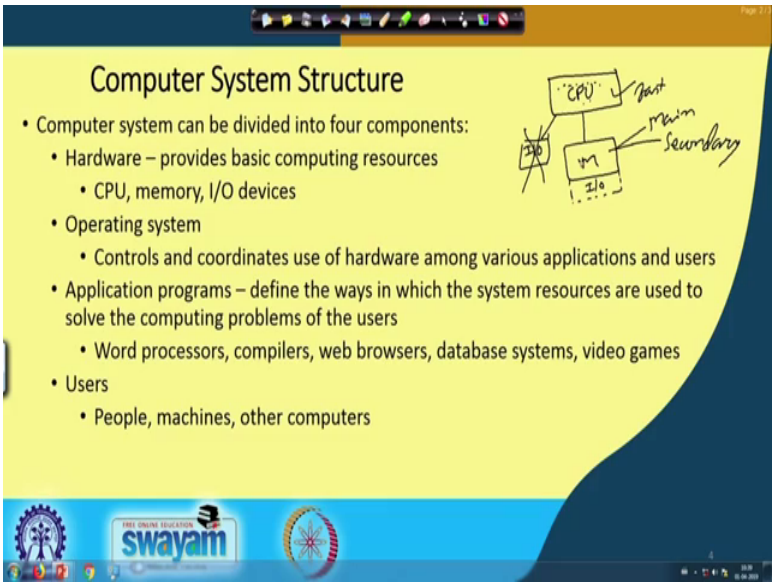
Then use the computer hardware in an efficient manner. So I should be able to use all the computer hardware in terms of the hard disk or the mouse keyboard all the devices that you have in the system. So I should be able to use all of them in a way in an efficient manner. So, this is again the tern efficient is a bit confusing it is not very clear at this point. So, when you go to the resource management, then it will be more clear. There can be several issues for example, the availability of the resources the resources should be available when I want it. Then, there may be that multiple users they want to use the resource simultaneously.

I will take a simple very simple example. Suppose there is a printer connected to a computer system. Now, this printer when if some program is or some process or program or some user is printing something on to the printer. So at that time, I should not allow any other program or process to access the printer and print something because otherwise what will happen is that on the printer you will get some pages from one user some other pages from other user and they will all be jumbled up. So there it is of no use to anybody.

So does it mean that you when I try in a program to print something on to a file on to the printer my program will be stalled? So this is a basic question. Ideally the answer is yes, but what the users will like to do is that, they would not like to wait ok. So,what most of the operating systems will do as we will see later is that the user is not made to wait rather whatever the user has to print on to the printer. So they are dumped onto a file onto the disk and later on when the printer becomes free then that content is put into the printer.

So that I am simultaneously not allowing more than one user to access the printer, but at the same time I am not making the users to wait for other users printing to be over. So this way I can try to think in terms of operating system operation like how can I utilize this resources in an efficient manner. So the term efficiency is a bit difficult to define like how much efficient is the system. So it is a bit difficult to define but we have to do something so that this efficiency is ensured.

(Refer Slide Time: 16:00)



Coming to the computer system structure so there is they can be divided into 4 main components. Any computer system overall structure can be defined a divided into 4 main components. One is the hardware part that provides the basic computing resources like CPU memory and IO devices. So we know that we have got interfaces like this. So in any computer system that you have is that there is a CPU, the central processing unit and we have got the memory. So this is from our school days we know that a computer system will have the structure like this and there are several IO devices that are connected. So we have got several IO devices and these IO devices may be say the keyboard display etcetera. So there can be different IO devices.

So in some sense, you can also think that, this IO device is part of memory because what you can do to a memory is to do some read and write operations to individual locations of the memory the same thing is true for IO device. So if it is an input device you cannot write on to that device but you can read from that device similarly if it is an output device you can write on to that device but cannot read from the device and if some device is both input and output in nature for example, some disk etcetera. So in that case it can do both read and write operation.

So in some system we can conceptually think that that there is no separate IO but this IO is a part of the memory itself. So many a system so they also visualize it like this that we have got this IO as a part of memory. So this memory and IO devices sometimes they are take a they are discussed together sometimes they are discussed separately. Of course, with the difference that when I talk about semiconductor memory. So this memory devices they are much much faster compared to the IO devices which are often electromechanical in nature . So that takes more time for communication.

So this entire thing is the hardware. So out of this CPU if this is very fast ok. So in a computer system CPU is the fastest unit that we have then comes the memory part. So, memory, you as we may be as we are familiar this memory can be divided into the main memory and secondary memory. This main memory is fast, but it is not as fast as CPU because this processor technology and memory technology, they are not advancing at the same rate. So this memory though density is improving but access time and all. So that is still an issue and when it comes to the secondary storage then access becomes even further an issue.

Another very important reason that why this access time is becoming important is that when this processor or the CPU is accessing some variable, some content of the locations which are inside the CPU, then it is doing on chip communication. But if the CPU chip and memory chip, they are separate they are connected by a bus the address bus, data bus control bus and so etcetera. So in that case for getting a content of a particular location, it has to access through this external bus and whenever we go for off chip bus the off chip bus is at least 10 times slower than the on chip signal lines. As a result, this off chip communication of the memory access becomes slow compared to the accesses within the CPU.

So that classifies this hardware devices that we have in a computer system. So though everything is hardware but everything does not operate at the same speed. So we have to look into them more carefully and we have to factor out the timing needed or the timing characteristics of this individual devices or they are to have a good utilization of the system.That is the first part. So we have got this in a computer system architecture, we have got this hardware devices. Then we have got another component which is called the operating system.

So it controls and coordinates use of hardware among various applications and users. So this is more of a administrative job and it is a coordination job. So operating system is a piece of software which will be utilizing which will be trying to exploit the feature of the underlying hardware and get the user jobs done. So, these controls and coordinates use of hardware among various applications and users. So third part is the application programs. So application programs they will define the ways in which the system resources are used to solve the computing problems of the users.

Now once the hardware and operating system is provided now it can be left to the user. So user can now start writing own program for doing something. But how do I write a program? For writing a program, I need to for example, suppose I have got a c language program. I have written it on a piece of paper. Now I have to enter it into the computer system. Now here comes the difficulty. So how do I enter it? So there may be some specific software by which I enter it, but then the question comes who designed that software? somebody has put that software into the system. So it is attached on top of the operating system that helps us in developing further programs.

So, it is again a set of programs that allows us to develop further set of programs that may be doing more specific computation. So like this word processors, compilers, web browsers, database systems, video games. So they are actually the next level of programs that are there with you on top of the operating system and they will be making the user program development easy.

So if I have to make a document and keep it in the computer system. So previously so a suppose the word processors were not there, then how do we do it. So it is a big question. So, how do I do it. So maybe I have to write a program in that case which is similar to word processor and that accepts input from the keyboard as the user is typing the document and storing it into the computer system. So basic operating system is providing as facilities like file and all. So it is actually the word processing software that I start writing and when I end so I have developed a word processor as well as I have written the document which I want to save.

Similarly, if I want to say I have written a C program, I want to convert it into machine code. Then how do I do it either I do it by hand. So I know the machine address machine instructions of that underlying computer CPU and then the underlying processor and accordingly I write down the machine language program for the which corresponds to the high level program.

Now this thing if it was not very there if the compilers were not there then, I have to do this hand assembly. Now presence of compilers they are actually helping us I do not have to do that. So there is a piece of software which translates my source language program into machine language program. So they are actually all intermediaries that who will be helping the user to develop the programs.

So these programs they will be using services from the operating system to get some jobs done and again acting as an interface between the users of the system and the operating system and the basic hardware. So as you can think of newer and newer applications, so every day some other such software some application program is getting added to the system and that way it is going on. So this procedure is continually going on.

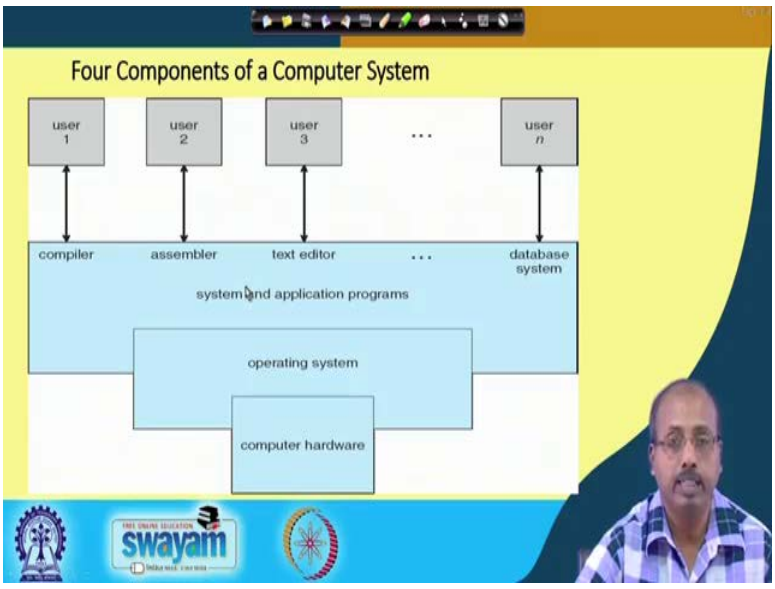
So at the last level we have got users who are the people machines or other computers. So when I say people. So I say people who are trying to use the computer system to get some maybe somebody is browsing the net. So getting weather information somebody is

writing a piece of letter, so that is using word processor somebody developing some program for some scientific computation. So accordingly using the compilers for doing the translation job so like that people may be using the computer system.

Then, the machines so, there may be some machine, which is in terms interfacing with the computer system and it is sending some comments to the computer and accordingly the computer is searching that were searching the database coming up with the answer and giving it back to the machine. So particularly for say if I have got a robotic movement. So, robots they are trying to do this mechanical movement and then, at any point of time so it is sending some command to the underlying hardware through the operating system and they are actually responding with the answer and then that is taken and sent to the machine.

And other computers may also send these requests. So we can have , say some web browser that is through with some other processor other computer so they are linking it. I may have something like say in a distributed system some other computer may be needing some information from this computer. So, as a result it sends a request over the network and that has to be responded too. So, like that we can think about different computer system architectures different computer system components that we have in the computer.

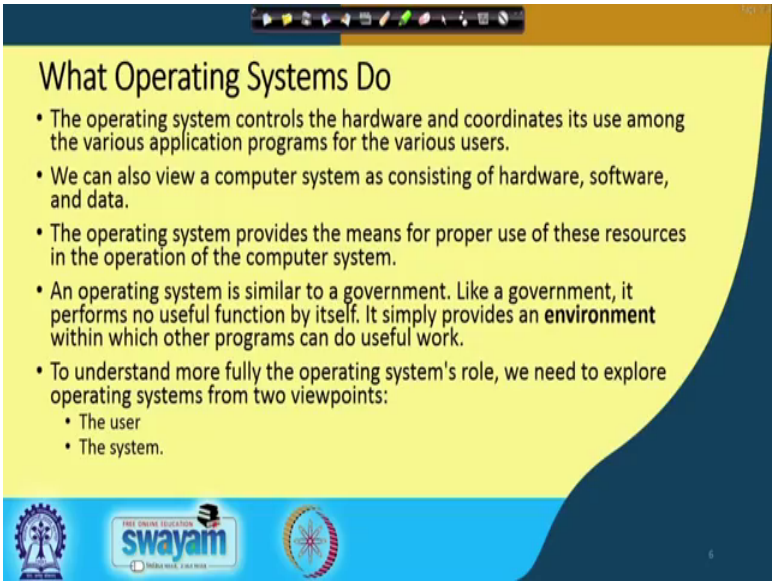
(Refer Slide Time: 26:43)



Now, if you look into in a more detailed fashion, so we can think of this computer system to be consisting of we can think of this particular diagram. So at the lowest level we have got the computer hardware. On top of that we have got the operating system. On top of operating system, we have got system and application programs and this application programs system.

And application programs or how many application programs that we can have. So, there is no limit. So, there can be I can have many things like compiler as assembler, text editor up to databases. These are some representative example. So you can have many more such application programs and users will be interacting with these system and application programs for getting their job done.

(Refer Slide Time: 27:26)



So what the operating systems can do? So the operating system they will control the hardware and coordinates a coordinate its use among the various application programs for various users. So this is the most important job that the operating system does. One thing is controlling the hardware. So that is one part and the other part is to coordinate between the users. So these are the two important part. So some it may so happen that the one user is running two programs and these two programs they need to coordinate between them. So some value some data value computed by some program has to be sent to some other program.

So naturally until and unless the first program has computed the data value the second program cannot use it. So there is a question of synchronization between these two programs. Then we can have this so that way the so coordination is very important and controlling the hardware is important because otherwise all the devices they will start sending information to the basic computer and the that may be getting flashed on to the some on to some IO device. So users will get confused. So, how are they going to be controlled ? So this operating system has to do this control and coordination job.

So you can think about a computer system as hardware software and data. So this is also possible because I have got a piece of hardware, which will be which can do some manipulation on the data. We have got some software which are controlling the hardware and doing the coordination and there are some data elements on which these control and coordination works. So we can think about computer system as consisting of hardware, software and data. It provides the means for proper use of the resources in the operation of the computer system.

So we will see that how this resource management is done. So this is very important as I said that we need to make this resource available to the appropriate processes at appropriate time, at the same time we should ensure that they are not given to more than one process in some situation if it is not possible. So that way the coordination has to be done.

So, it in some sense it is similar to a government like a government, it will be performs no useful function by itself ok. So what is the job of a government? So it is actually providing facility to the citizen of the country. So it is like that. So it is actually a facility provider or creating an environment in which the citizens can live peacefully and all. So similarly here also the operating system is providing an environment in which other programs can do some useful work. So this is providing the facility.

So, normally when something is not happening. So if I have written a program and the program is not working. So many a times, we first what we do is that we blame the underlying system that this system, it is not work. Many times what happens is that you try a program in system 1, there it is running properly. You take it to system 2 with a different operating system and there it is not working. So we start blaming the operating system in system 2. But there must be something very specific in my program and which

does not match with the requirements of the second operating system. So, that way it is creating the problem. So that we have to understand.

So we have to see like this environment that is provided is proper. We have to understand more fully the operating systems role, we need to explore the operating system from 2 different viewpoints.The user viewpoint and the system viewpoint. As I said the operating system can be viewed from different angle. As a user I see what are the facilities provided to me, as a system designer so I have to see what are the facilities that I have to provide to the users. So these are the two different angles from which an operating system can be looked into.

So, we will continue with this in the next lecture.