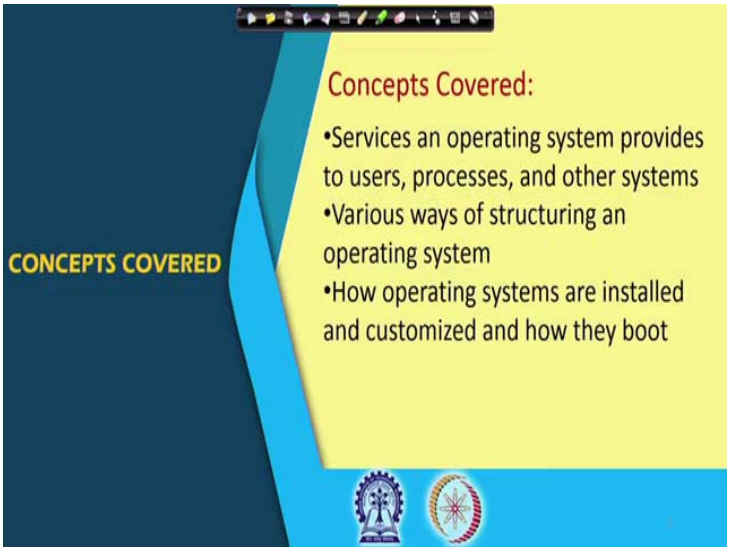
Operating System Fundamentals
Prof. Santanu Chattopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 07
Operating System Structures

Next we will be discussing on Operating System Structures. In which we will be discussing about how operating system normally organised so that different components of it can be utilised properly by the users the system.

(Refer Slide Time: 00:41)



Now, the concepts that we are going to cover is first of all the services and operating system provides to users, processes, and other systems. So, this is the first thing that we will discuss. Now, operating system must provide services to users, like as an user of the system. So, we can make some query to the operating system for some files exist, whether we want to open a file write something, read something from some device and all.

So, that is the services that are provided to the user, a user may want to login to the system. So, that is also another service, then processes that they also requires some services because processes are program under execution. So, they are actually the components that interacts with the operating g system and that will be doing the works done by the operating system.

So, that way the processes they will request for services, we will see what type of services they can request. And, the service request may come from other systems as well like on a network may be there is a client server situation, when the server it gets requests from the client operating systems.

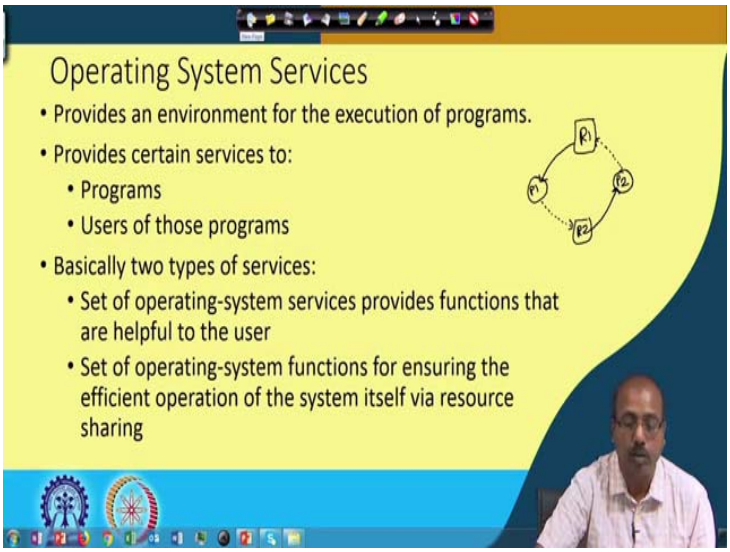
And, then they when they request for some content, or some files, or some database, or some such service, then that comes as a service request and the operating system of the host system of the host computer. So, it must respond to that. So, that way it provides service to other systems.

There can be various ways of structuring and operating systems, so that that we will see because after all operating system is designed by human being. So, there are different designs that are available and different computer designers they can they can follow 1 or other of this design paradigms and accordingly the overall operating system will be structured. Then, how operating systems are installed and customised and how they boot? So, this is another issue like if you have looked into a bit of history then, you will see that normally the old computer so they had to have a boot disk and that through that boot disk this operating systems has to start.

So, if that boot disk is corrupted or somehow that boot disc is lost then as a system cannot boot so, that was a major problem. After, that what has come is that we have got this the hard disk itself that is there in the system, then at the very beginning of the hard disk the sector 0. So, that contains the operating system parts. So, that when the basic input output services available in the ROM. So, it can ask the processor to read the content of the first sector of the disk.

So, that is another 1 type of mechanism. Another type of mechanism is there which is the now common like say thin client type of environment where, you client computers they do not have any operating systems. So, over the network they request the host to send the operating system and accordingly they boot. So, that way there can be different ways in which a system boots and what are the general structures. So, that we will have a look in this discussion.

(Refer Slide Time: 03:50)



So, to start with the services that are provided by operating system. So, operating system it provides an environment for execution of program. So, this is the primary goal of an operating system. So, it must provide and environment in which programs can be executed efficiently. And it can be executed, they can be executed easily. Now, provides certain services to programs and users of those programs like, where the program has to run.

So, that way it has to give enough resources to the program for run in terms of memory in terms CPU time in terms of bio devices etcetera. Second thing is that it should provide services to the users of those program. So, for example, the user wants to enter some data. So, that data entry should be facilitated by means of operating system and the devices. So, that way they must provide certain services to both the programs that are running in the system and the users for those programs.

There are two types of services that operating system can provide, first of all the set of operating system services provides functions that are helpful to the user. So, therefore example, user wants to login to the system, user wants to run a particular program, user wants to see what is the current status of a program that is running. So, that way the user has got a certain set of requirements from the system. So, that is the one type of services that operating system will provide to the user.

Another set of operating system function that ensures efficient operation of the system itself by a resource sharing. So there are several resources in the system. Now how do we allocate the resources? Ideally no process should be needed to wait when it requires a resource. So as and when a process is executing so if you find that at some point I need a resource that resource ideally it should be available at that point of time. Now that is a very serious question, because now how many copies of those resources do we have?

For example if 2 programs or 2 processes the time to print some data onto the printer so how many printers do you have connected to the system? So normally we have got 1 or at most 2 printers connected but there may be large number of programs that are trying to print sum data onto the output. So that way this so printer is a resource that has to be shared among a number of processes and when the sharing comes. So it is also the question comes like it has to be efficient.

For example, it should not be the case that 1 process has request for a resource and it is just waiting because others are using it. So that way the process is starving. Another situation that can occur in this type of shared resource environment is the problem of deadlock. So where all the processes they are waiting for some resource to be free. So it is such that process 1 is waiting for a resource, which is to be freed by a process 2. And process 2 is waiting for a resource which is to be finished by process 1. Now both these processes they are in a loop.

So they are waiting for a one resource which is currently held by another process. And looking for the another resource and looking for that resource and it is holding on one resource. So it is like this that if I have got say 2 resources R 1 and R 2. So this is 1 resource and this is another resource 1 and R 2. So, it may be the case that process 1 is currently having the resource with it and it is requesting for resource 2. So it is holding resource 1 and it is asking for resource 2. Similarly process 2 it is holding resource 2 and it is asking for resource 1 so in this situation so both the processes P 1 and P 2 so they will be in a deadlocked situation because none of them will be able to proceed because they are not getting the proper resource.

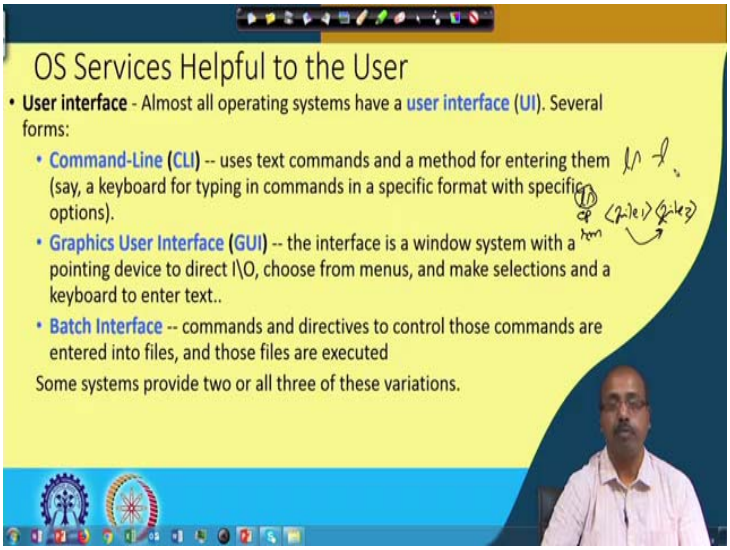
And, still there is no chance of getting the resource because until and unless P 1 finishes it will not release R 1 until and unless P 2 finishes it will not release R 2, but for P 1 to be finished R 2 is required for P 2 to be finished R 1 is required. So that way resource

sharing is a very important problems. So, we will dedicate a full discussion on this when we go to the chapter on deadlocks.

So, we will see that there is this is a very important problem that needs to be resolved. So operating system must provide functions. So you see that point is that when the operating system is trying to give resources to the users. So it has to ensure that the resource sharing is efficient. And also sometimes a process may be having lots of resources and it maybe in some sense it may malfunctioning. So that may affect the system performance significantly.

So we may like to has have a look like it is a maybe that process needs to be taken care of that process needs to be given some that process needs to be terminated or something like that. So that it becomes easy for other processes to continue.

(Refer Slide Time: 09:09)



Now the OS services that are helpful to the user. So one very important part of it is the user interface that is how does a user talk to the system? Like how is giving the commands and all so that is the first thing. So almost all operating systems have a user interface because otherwise there is no chance to talk to the system. Accepting the case where the operating system is an embedded 1 where from the outside we do not have any interaction with the user.

So that may be A 1 situation but in general any computer system. So that has got some user interface and it will be in different ways of the operating system is takes the information from the user commands from the user through this interface. Now, there are different types of interfaces that you can find in different operating system. The simplest one of them is the Command-Line interface or CLI interface. So it uses text commands and a method for entering them. So maybe like from keyboard with type some commands a specific format with specific options.

So the operating system manual will tell you what are the commands? Like if you are familiar with say Unix Linux operating system. So, apart so they have got a set of commands like, this you need to remember all those commands and the parameters for them . For example, in Unix system to copy a particular file.

So you have to give this command CP file 1 file 2. So file 1 will be copied to file 2. So, that way but there are so, this is fine. So, user has to remember all those commands. For example for deleting a file the command is RM for opening for let us see what are the files in my directory so, there is a command like LS.

So these are the things that you need to remember that is operating system specific ok. Now first operating systems or initial operating system so we have got this type of command line. This command line interface and they have got large number of arguments like, this l s command itself it has got many arguments, like you give simply LS . So, it will just list down the files.

So, if you give LS minus L so, that LS stands for long list. So it will tell you who is the owner of the file, when was the file created, what is the read write execute permission for different users of the system for this particular file. So, like that it gives a long list and fortunately or unfortunately there are so many such arguments for each of these commands or it is very difficult to remember all of them. And, only if you are practicing a lot or if you are a master in that interface then you will be able to do that.

Now today's system if you see then we will hardly find a system that is based on simply this command line argument because command interface. Because it is mostly what we have is a graphical interface or GUI based interface. Where you see some sort of a window on which we get lots of icons and we have got this mouse by which we can

select a particular icon, we can do something ways so menus will turn up and then from the menus we can select and all.

So that way this graphical user interface so, this is the next thing that we have over and after this command line argument. So this interface is a windows system with a pointing device to direct I/O, choose from menus, and make selections and a keyboard to enter text. Sometimes we want to have to enter text like name of the file or something like that. So that may be entered through the keyboard or some data that is needed from the user. So that is through keyboard.

So here also keyboard interface is there CLI also keyboard interface was there, but here the user in GUI interface user did not remember what is a exact command format. So the user just by clicking a few mouse clicking mouse a few times will be able to execute some operation and there is batch interface.

So here the commands and directives to control those commands are entered into 2 files and those files are executed. So batch means in a batch will give a set of commands to be executed. So maybe we want 10 different operations to be done on a certain set of files or programs etcetera.

So in a file we write down those 10 commands and then tell the system that we execute commands from this file 1 by 1. So that is called a batch interface. So which was very common in earlier systems like what happens is that the program execution and all they take lot of time. So where suppose for example, say 50 users may have might have given their jobs for execution and because of the student priority is very low.

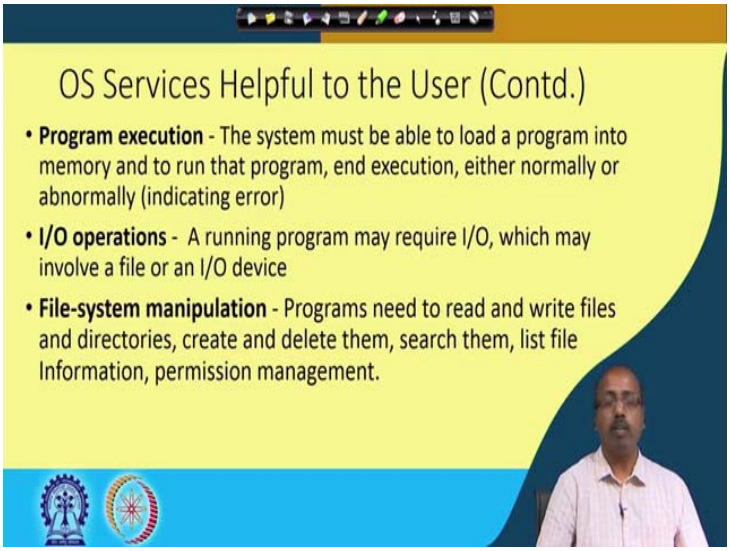
So they are run only at the night. Maybe daytime some more research problems they are getting solved like that. So at the night those batch files are executed and you get the output in the next day morning.

So that way we can have this type of batch interface then for example is payroll processing. So payroll processing job so that is maybe where it is done offline, in the sense that it is done after the office hours the payroll software maybe running and updating all the payroll data basis or taking backup of the system. So that is a backup background job. So that may be running at some time at the night at some periodic intervals. So that is also using some batch commands.

So, how to take that backup? So we can keep those commands in a file and then that file may be run and the backup is taken in a batch. So this way we can have the batch interface where we have got a batch file containing all the comments to be executed. Some systems provide two or all three of this variation ok.

So, there are many different situations where you can find only two of them or at least the GUI and CLI. So they are mostly there batch interface is there in some cases where it is required.

(Refer Slide Time: 15:37)



Then program execution so, this is another service that is provided to the user, because this is the most fundamental thing that the user be able to do. So, the system must be able to load a program into memory, and to run that program end execution either normally or abnormally indicating errors. So, these are the things that the user should be provided with the facility, like user should be able to run a program.

Now how to run a program? So when we compile a file. So this resides in the secondary storage, but as we know that processor can talk to only the main memory. So, program has to be loaded from the secondary storage into the main memory and then the program execution should start.

Now after executing the program for some time so maybe the program is over. So it will terminate normally or it may be that there is some error that has occurred in that case the

program will come out abnormally. So indicate the so the error code will be given back to the user. So that user can understand what was the problem with the program execution.

So, this is another thing. Then, the I/O operations any running program we may require some input output like for example if you are solving the roots of a quadratic equations then you have to give the values of this coefficients A, B and C. So that is coming from the user.

So that may be given by a keyboard that is an I/O device or it may be by means of some file. In a file the values may be written and from there are the values are taken for execution. So this way a running program may require input output and this must be provided by means of file or IO device, then file system manipulation.

So this is another important facility that has to be provided by operating system. The programs need to read and write files and directories create and delete them, search them, list file information and permission management. So these are the operations that we may need for file management ok.

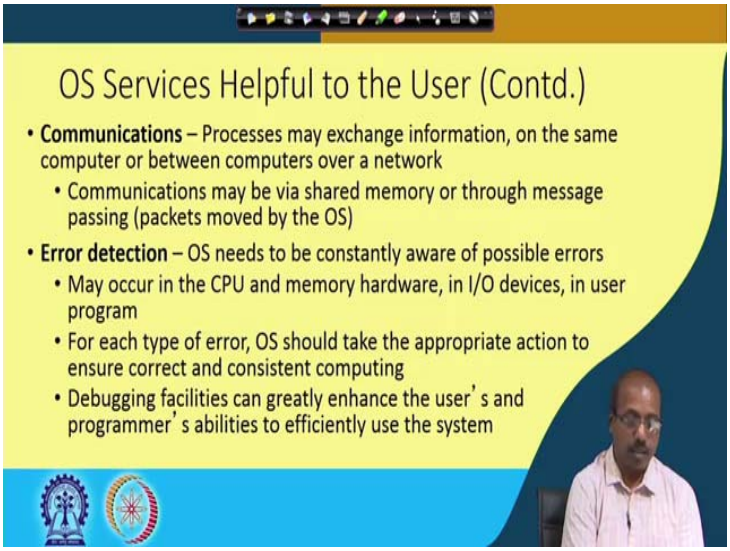
So for example all the files they may be organised into some directory structure. So, that the related files they are kept in the same directory. So, that for an user so it becomes easy to remember where the file may actually be there. So, that way user may like to create some new directories and subdirectories and within that directory. So, it they can like to create some file etcetera.

So, they should be able to delete file, then delete directory is they should be able to search the directory ok, then the list of file that is there. So, that information may be may be required to be printed, then what are the permission like a file in a single user system of course, there is no problem, because all the files are under the control of a single user, but for a multi user system. So the there are files from different users. So, they should not be mixed up and they should one user should not have unrestricted access to the programs of other users or files of other users.

Now, sometimes we need to give access to all of them and that access pattern will also vary. For example, the a student record may be viewed by a student, but it may not be modifiable by the student. Similarly, a student record may be viewed by a teacher and is

also modifiable by a teacher. So, that can happen so, depending upon the user community that we have. So, this file permission management has to be done.

(Refer Slide Time: 19:19)



So, this type of facilities are to be provided. Then communications so processes may exchange information on the same computer or between computers over a network. So, another fundamental facility, that must be provided to the processes or the users of the system. So, as a users or maybe we developing a program, but that program is not a standalone one.

So, it may be that it needs to talk to some other programs in the system, may be some part of the some of the variable values are computed by another program and those values and variables are required in my program and vice versa, or something that is computed in my program is needed on the other program.

So, these two programs they should be able to talk to each other, there must be some exchange of information. And, the operating system must provide facility for doing that. As I have just said that one users data should not be accessible by other user, in a multi user system it should not be available in by the other user, but in this case we are telling just the opposite, we are telling that something that is created by one user is going to be visible by another user. So, you see we need to have that facility at the same time we need to have restriction.

So, this permission for exchange of information that is very important and how this whole thing is managed so, that is important. So, this communication is important facility that has to be provided with proper care; so communications maybe via shared memory or through message passing. So, normally what we do is a part of the main memory. So, that is your mark to be shared memory. So, that it can different processes are able to read or write on to that memory.

So, that is the shared memory access. So, that may be provided or there may be there may be message passing. So, one process may send a message to another process and another that way the message exchange can take place. So, packets are this movement is taken care of by operating system so that the information goes from one process to another process.

Then error detection so OS needs to be constantly aware of possible error like, when a program is under execution. So, before a program execute so, we do not know whether the program will give rise to some error or not. For example, there may be a division operation in my programs z equal to x by y. So, until and unless the value of y is computed we do not know, whether y becomes equal to 0 leading to a divide by 0 error. So that way it is most of the errors are difficult to predict statically. So only when the program or process is executing. So we can have this type of errors occurring.

So the operating system should be constantly aware of these possible errors, the error occur in the CPU and memory hardware in IO devices or in user program the example that I have taken is in the user program or it may so happen in the memory there are certain locations will become faulty ok. So that way it becomes a problem. Or in the CPU also some fault may occur. So, that way during program execution there may be some problem.

So, IO devices also all on a sudden printers stops working. So, that way there are maybe some errors. So the error code that this devices send. So, this they must be recognize recognizable by the operating system and operating system should accordingly inform the user that such error has occurred.

So for each type of error OS should take appropriate action to ensure correct and consistent computing. So some of the errors they are they are transient maybe if you try after sometime then the error will not be there. So the operating system designer may

decide that is a certain type of errors occur then I will just retry the operation after some time. So that way so the for ensuring that computing goes on fine so, this message is this action is taken by the operating system, where it takes care of the error.

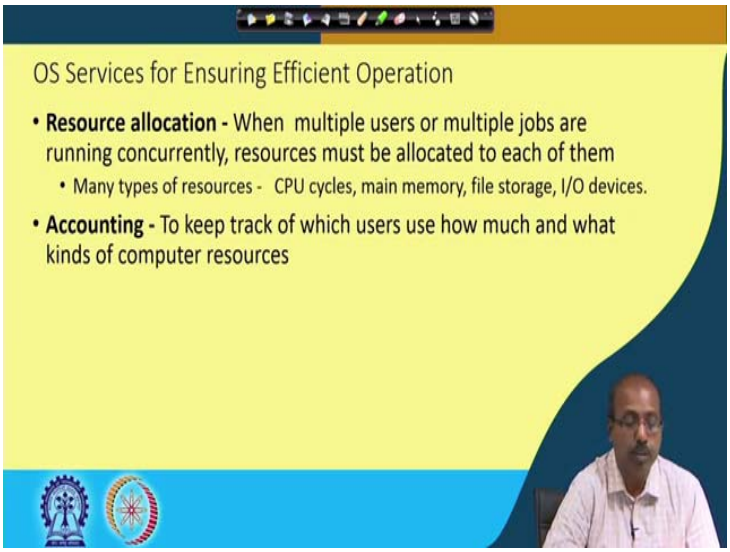
Then another important thing that we should have is providing debugging facility. So we should be able to debug the program like, if a program is working fine, then it is fine, but normally like a program may have the different ways in which it proceeds execution through it. So, it is not possible like while doing the testing. So, it may not be possible to check for all different types of execution sequence that the program may follow in it is lifetime.

Now, when the program is executing maybe some error occurs. For example, if a telephone exchange is operating then at some point of time it may develop a fault ok. So, that way so, that way it is difficult to catch those errors statically. So when the exchange is not operating. So, it is difficult to catch those errors.

So we should be able to debug that problem ok. So, that debugging is a very important thing. So, on the logical errors that a program or process can develop so, they are checked by this debugging facilities. So, that can help you the by giving enough information about the content of CPU registers memory is a memory locations and all. And from there the user can possibly find out what is the problem. What was the difficulty that had occurred due to this for this for which this failure has a come.

So, this way this debugging facilities are another important issue and since the program is executing under operating systems control so, if the operating system does not provide enough information for debugging then or enough facility for debugging. So, it is very difficult for the user to do the debugging operation.

(Refer Slide Time: 25:04)



Then, resource allocation so, there can be multiple users or multiple jobs that are running concurrently and resources are to be allocated to them. So that way we have got this resource allocation problem and there are different types of resources like CPU cycle. So, that itself is a resource like for executing a program we need the CPU to be given to it. So, for how long time CPU is given to a process that determines whether the time is sufficient or not.

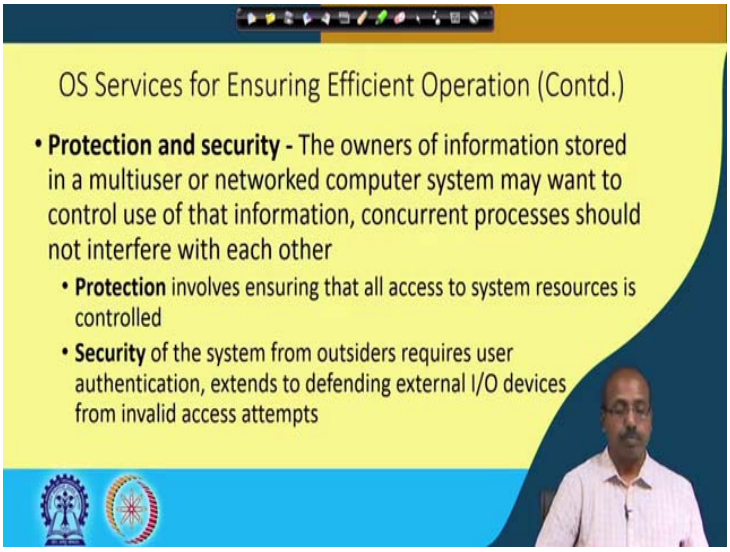
So, it may be that the system the operating system designer assume makes it like this that it will be in a fashion that for in a round robin way it will execute the programs one program for sometime, maybe some 2 second also then the CPU will be given to another process or a program that will also is equal 2 second.

Then to the third process so it goes on in a round robin fashion across all the processes that we have in the system. So that way is a CPU cycle allotment that is a resource that we have then main memory is a main memory is finite. So out of that different processes for execution so they need certain amount of memory so that main memory has to be provided.Then this file storage how much where to store the files. So, disk space is also divide a shared across a number of processes. So, this how this disk storage is given to different processes how are they are they allocated? Then the IO devices that are that are given.

Then, accounting to keep track of which users use how much and what kinds of computer resources? So, that is important. So normally when we are working in an educational organization maybe we are not paying for the computation that we do by the system, but in a commercial organization we have to pay for that. And that payment will definitely depend on the CPU time that is used by the program the memory that is used in the disk space that is used for each of the resources that is there in the system, how much of it is used by the program. So, that will determine the overall charge that we have to take from the user of the system.

So that way the accounting information is very important. So we must provide some mechanism by which this accounting is done properly and you understand that there are so many processes running and this time values are so, small. So it is not possible without the help of the operating system to tell, what are the exact values how much memory it was it has use, how much disk it has, how much CPU time it has used for process. So, it is very difficult to do it manually.

(Refer Slide Time: 27:45)



Then, another service that often provided is the protection and security. So this is particularly true when we have got multi user system or a network computer system. So, there we want to control the access of the information of corresponding to one user or one computer from other users or other computers.

So we have to have protection to ensure that all access to system resources is controlled. So it is not that two the two process is there allowed to modify one particular shared variable. For example, if it is a printer and if two processes are given access simultaneously to the printer, then the lines are all jumbled up in the printer. So, that is the protection. So, we must provide that at one point of time so, the resource is given to one process only.

So, the access of the system resources is controlled and security will ensure that from the outsiders requires user authentication like, I will only valid users of the system normally almost all the systems, now they have got login and password. So extend using those login and password values only it can go for one particular user.

And it extends to defending external IO devices from invalid access attempts. Somebody tries to access some IO device. So until and unless it is authenticated so it is not permitted. So, that way security becomes an important concern today and operating system must provide facilities by which the security and protection is provided to the users.