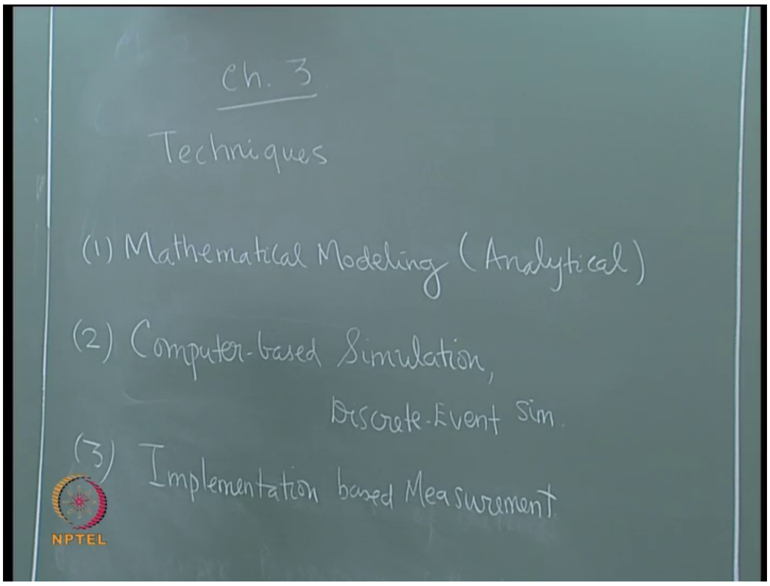
Performance Evaluation of Computer Systems

Prof. Krishna Moorthy Sivalingam
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture No. # 03

Selection of Techniques and metrics

(Refer Slide Time: 00:12)

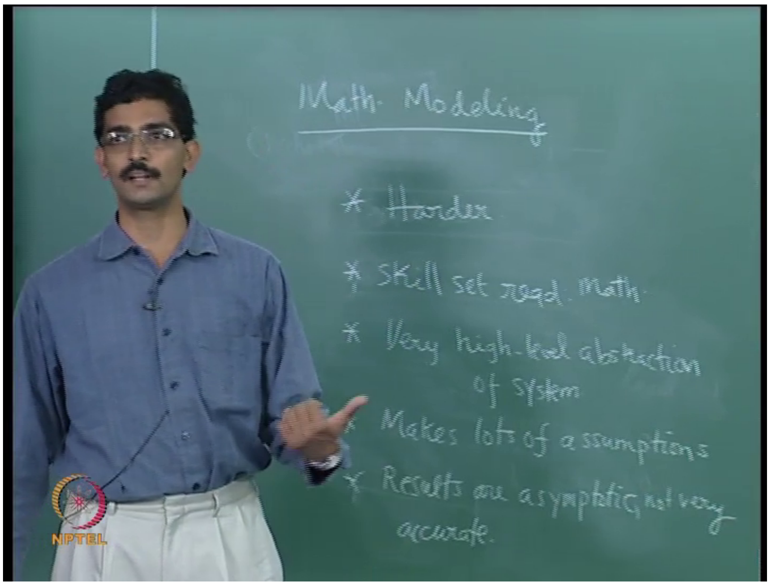


So, the first technique we briefly discussed this even yesterday, make of completion we will write it down. So, mathematical modeling is the first way to analyze understands the performance of the given system. And the second is simulation; again there is analog simulation, discrete simulation. Today, we talk mostly about computer based simulations. And in particular, we will talk about discrete event simulation.

And then of course, the third aspect is to implement, implementation based… So, the book actually calls this simply measurement, but we do measurement in any whether it is simulation also we measure. So therefore, I will simply add this term implementation, you basically implemented in a real system; this is the practical implementation. For example, some cash coherence protocol that you come up with, you should be able to, you can do analysis, simulation or mathematical modeling. And we sometimes also use the term analytical modeling here.

So, if we have taken ADSA or DSA, data structures algorithms probably have done some of these things in your labs. So, you have done in class, you have done your theoretical analysis of the algorithms, then we though probably than do this simply waits to when straight away to implementation. Because was mostly straight word algorithm is not a complex system which you are try to simulate and then implementation. But measurements you probably know the couple of values of sizes of list, but mostly worried about correctness. Functionality much worried about then scalability of system itself so, will talk a little bit about the nature that kind of task that is involved in a mathematical modeling system.

(Refer Slide Time: 02:57)



So, what comes to mind? What are the different things involved? And is it hard, easy time consuming, accurate, reflect and those entire thing let me know which is math modeling easy or hard?

(( ))

Hard it is interesting, because only now we start cannot go away from math modeling. Anytime somebody builds a system either you implemented or simulated and then you have done, nobody really wants to go the analytical model. But twenty-thirty years ago, this was not a real option; computer based simulation is where not really around. If you look at lot of the early papers in a given networks that use that in certain example. Most of the early paper was trying to do some set of math modeling; you have to have math model, if you have to have good paper. But then that is not end of it; somebody actually has to build the system. Why was TCP/IP successful, because somebody built it where it part of PSG and where it available to everybody.

So, it was mostly these two were the traditional options and those ways for the researchers. Now this is harder, if I certainly tell you go build a wide area network and test your routing protocols and that we simply cannot do that, even now it is hard. Even today we will see, I have seen cloud computing based experiments. Is it having different device, new algorithm for allocating resources on the cloud and then where is your cloud? I have four machines on a land; that may cloud; cloud varied every the cloud.

We should be actually have no account elsewhere or put your machines elsewhere and some an access there. So, implementation is always being hard, but this is always been easy. So, if you look at lot of the early work in computer system and so on, they really depended upon mathematical modeling.

But today we are (( )) wearing away from that, because we know that the there is an option. Otherwise you could simply proposed in algorithm and said I am done, you want be able to finish your work and go anywhere. So, either implemented or you analyze it. So therefore, this is say harder. So, the skill set involve is you need to have some knowledge of several math techniques. We have going to see only some of those in this class Markov chains, models and so on.

But there is a wide variety of analysis techniques, modeling techniques that are there and what usually what we do is, we have one tool I know Markov chains. Therefore, everything is a Markov chain; whatever system we give me Markov chain. We simply draw Markov chain for everything. So, we have to be willing to go beyond, even has talk and (( )) beyond that Markov say Markov. But though that is way you have a class of researchers try to keep on coming with new techniques.

Those of the performance evaluation researches, there are people try to come up with new a math model and those it try to use them, and show as how to use them. And then, once somebody shows have to use them, it is simply use the same model over and over. This classic example is when the case of like optical network there is this, whenever the traffic request come in right path, come in get the established or that do not get established. So, there is a way to calculate the blocking probability. So, the blocking probability most of us there initially doing simulation based simulation, simulator and then show the results or do an ILP, integer linear programming formulation is the design problem.

What is the performance of the system? How do you design like plots and so on? And then I get the ninety nine-two thousand the paper came along that show, how to do blocking probability computation is fairly intricate know the probability and statistics; probability use to actually do this calculation. So, once they did that and that became the does the tradition. Everybody else we simply use that same modal and start computing in the same way. So, now and then people come and show you, how things can we differently. But otherwise, we strict to our known set of tools and then keep using that over and over again. So, this that the skill set read is purely mathematical and if we go to a math who there is say this is no big deal, we do this all the time.

Then, we come to computer science will say well we can do some it, but not all of it. So, skill set required is you know purely mathematical, but what else what other features? Is it good, is it bad? So, we said it is hard, forget it. Let us say, you have to do math model, I say nothing doing, without that you cannot get your PhD. Usually for PhD student, we initially start that. You better do a math model of a system, otherwise there is no rigor. Why does math model help you? Sometimes the math Modeling helps you, abstract the system away.

You are looking at system away very high level not at a detail level. Detail level worry about the entire variable, all the files that you have to include at the level of implementation; sometimes that is irrelevant. You really want abstract the system away and a very high level so, it gives you a (( )) give you of the system. So, that is the good point about it. So, you have a very high level abstraction of the system.

(( )) if it should be lower level abstraction, how to be (( ))

What is mean by high level? You are will say you are looking at queuing scheduling algorithm in an operating system and you have this multilevel feedback queue, you (( )). Where you start off with some have very low quantum, then you keep jumping to different queues and so on. And then you go to I/O queues also; you have the CPU queue, you go to some I/O queue and so on.

Now, if you look at it from real system point of view, you will start worrying about how do you build this queues study, use linked list, doubly linked list and all that. So, here if I look at this the math model, you want to know, what is the throughput of the system? How many jobs per unit time can the system complete?

You simply view the system as the collection of networks of queues. It simply a set of queues and that is about it, you do not worry about all the lower level details of implementation.

(( )) this is also various (( )).

Yes so, that is true. Yeah, that why it is the very abstract model which is good to start with, you do not worry about the details in usually, Because you first you want to get an approximate understanding of the behavior of the system. To get the approximate behavior, you start with the very I would say gross abstraction or not a detail abstraction.

(( )) your abstractions may be to go wrong.

True, so that is why this is the starting point. So, mostly the use this is the starting point. Even we talked about sorting to, you have to different algorithms is both of them saying n log n, which is better? What have to people use to fix all the time? You tell me, why is quick sort is more often? Quick sort is the more often than the others. Even though there are others merge sort is also n log n, why quick sort is use to more often?

(( ))

Yeah. So, that how did you know that? That is either you do analysis or you can also verify the temporary. So, they have both verifications to do, but you knew that I do not want to look at n squared algorithms. This eliminates for n squared algorithm space, all those in e where algorithms (( )). So therefore, you this helps to narrow down to a set of probable, most probable systems or techniques to use. So, that is that lose is there. So, that is why you are, you loose of this very high level abstraction is both the positive and the negative. Positive is that, let us you focus on the main characteristics not in the details, but sometimes details are also important.

So, you lose the details which you have to capture with the next level which is simulation based system. So, this very high level abstract makes lots of assumptions. So, you assume, you will always assume that and then you say, you let say want you to find out, what is the average delay of customers coming to the state bank or some bank, teller.

And either is it their daily and measure, measure or if somebody tells you that okay, I said there for an hour; I think about ten customers per hour, what is the average service time for customer? Okay about five to, five minutes for a customer is the average is often this one hour. Then you would say tell me, what is the average delay? So, you can simply go to a system with the known closed form solution which is say m m one or m g one and so on. If you make those assumptions, you first of all say that I assume that the arrival process coming to this queue is Poisson.

Nobody ever verify that is the Poisson, if it need not be Poisson. But that is the convenient assumption you make, because there are close form solutions. So, wherever you find convenient assumptions about the system characterization, you simply take those simply to get your close form solutions.

So therefore, that need not be holding true with what your action system is. Nobody when knows whether it is Poisson or not. We just say that well conveniently will assume that it is Poisson; many things is simply assume Poisson. Networks there were a long fight. Peoples that know calls are no longer Poisson; it is something like this Markov modulated Poisson, our self similar traffic, all those things came about. But fundamentally if I want this tell me approximately, you want the delay will be should have put ten servers or twenty servers. And for that sometimes m m one is good enough; your basic queuing model is good enough. So, you start from there, then you start going to more detail.

So, you do make a lot of assumptions which does not fully represent the system behavior; so that is the negative. And so, that is also because of that your results are also not accurate. Your results are only asymptotic results many times. It gives you sometimes there is accurate that you if you really system is m one, you will verify this. If I really simulate and exponential arrival and an exponential or Poisson arrival and exponential distribution time, you will get those close form solutions. It actually, you simulate those you verify there, but the question is whether the system are really behaves with those kind of assumptions.

So, the results are mostly, the results are asymptotic not very accurate. So, loss of accuracy is there, because we know assumptions are there, but I know. What else comes to mind with math modeling about math modeling? You may not be able to get very detailed performance in this.

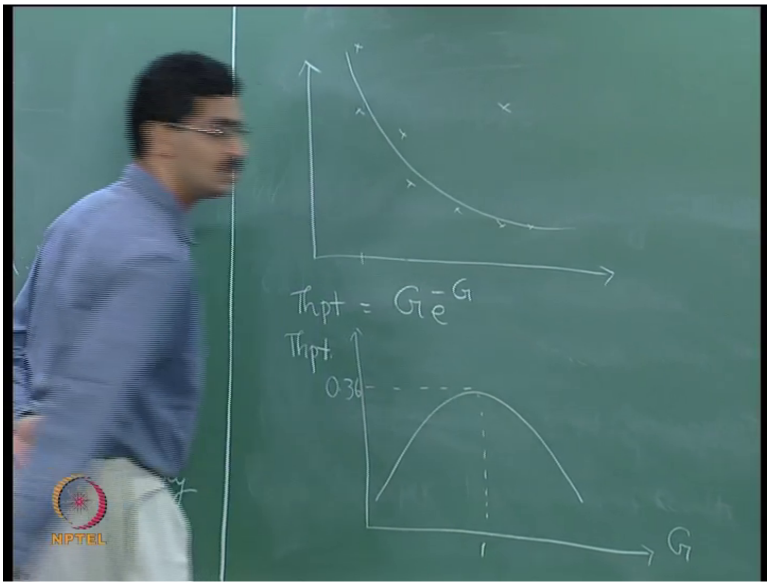
(( )).

Yes.

(( ))

Out layer may not be capture, yes. She is may not be able to yeah, some of those things that we saw many times when you look at these Math for the simple equation is simply plot them and say here is the nice smooth curve.

(Refer Slide Time: 13:36)



And for example, you are actually train to plot the inter-arrival times of packets coming to a system. If I have an exponential model, it will be simply something like this. Like this is the probability, let us your capturing the probability of actually packets come.

So, this would be something of these. So, packets of length this has. We actually, we do a c d f, but let say that these is the some sort of curve that represents like something that your probability of something that your measuring. So, this would be your Math model let say, this is the expected probability for these various types. But real as sticky your points might fall;

all over they could be one here, one here and so on. I could have simulation with different values on the x axis and then find that this is actually what I am getting by simulation by experimentation.

So yeah, those out layers are lost. So some detailed analysis is not there. That is in other disadvantage. And also, if I want to look at say per user throughput, let us look at aloha that is the classic example. If I use the, I can let say throughput of aloha, I can simply represent by this formula; G is the offer load and e power minus G. So, e is your e that we know from Math. And so, this is the offered load in terms of packets per second and this is the actual realized throughput by the system. This is the very simple is good enough for me to know that my… So, what will be if I plot this curve, what it will look like? So, this is G the load that we talked about; this load on this access.

And this is the throughput measured on this axis. What will it look like? You can verify by increase and decrease and the maximum again, you can find out. Differentiate this in a find this maximum G equals one which is therefore, the maximum throughput is one over e and that is 0.368t and so on. This is your 0.36 that is seeing when G equals one.

So, beyond offer load of one this means that we have offer more than one packet per second across the entire user. This is cumulative load offered by all the users of aloha system. If I do that, then the system starts. So, this is what I would get from a mathematical model. If I certainly ask you this is total throughput, how about the each user's throughput? So, what can happen in aloha and other case to you have so, called capture.

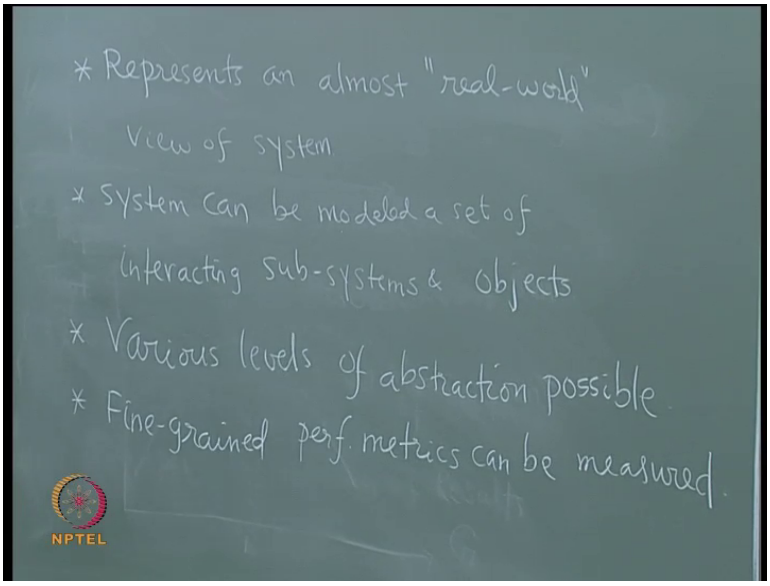
I am interested in what is the throughput in by each of the fifty users in the system across this entire simulation not this sometime spends? You may find that are problems with their. It is possible that one user as pretty much all the channel and got the entire throughput, others have not got anything. If I simply miss that throughput seen with the different users, one will be very high others could be very very low. We just plot the throughput per user on the graph. Now that kind of detail I will not get from a Math model. If I really want to see per user throughput, that is last. Sometimes you may, you would not you would not able to see this capture effect in Mathematical modeling like this.

That accuracy is got or if I want to say, what if I vary the load per user? Some users are 0.25 packets per second, some are 0.5, and some are 0.8 and so on, what will be the system behavior? Mathematical model say will handle anything about that I only have this one formula. I will simply give, you give me value of G, I will plot this curve for you that are about it. So, that again is lost is not really possible unless, you do even more intricate modeling. So, you sometimes just have this very high level picture, simply tells you at like a certain words they view, but this scheme will perform better or words them some of the other systems.

So, that detail asymptotic not very accurate, some of these things are also captures. Out layers are lost, very grew or very find there were details within the simulation which you can measure, very find measurements cannot be done. And let us say what the variance is? If I say what is the variance of delay for a particular packet? For each users delay, what is the variance? That you may not be able to get from this particular analysis so, there is the general characterization of what math modeling thus, but we always usually like to start with math models.

Because you get you think about the system, little bit more rigorous. You try to model as system, you find that you are as some rigor is added into your, the way you look at the system, the way you represent the system. They will define the system and so on. So, that is the nice part about you math modeling; it adds rigor to your think process.

(Refer Slide Time: 18:19)



So, now we come to favorite discrete event simulation. So, you all know what simulation is about? It is representing a system in the computer using some sort of modeling language.

Several modeling language is out there and simulation is not just our computer system could be even like a nuclear plan simulation, you can do that. People do financial simulation, the people do this climate simulation also lots of things; they try to they set that there will be know all those incorrect; some these predictions happen when use incorrect simulation model. So, will come to that later on, but in general you can capture you define so, this represents so, and this is almost a ... So, we will, but you have the system.

You could actually look at your simulation as an exercise that precedes implementation, because many times whatever code, all the logic that you implement in simulation can be directly imported to if it is a seen or if it is hardware then you have to recode so on. But in general, it gives you right so, you would system is model as a set of interacting. So, you give the system as a collection of sub-systems, you can implement those sub-systems; within the sub-system will have your objects and so on.

Just like your regular software engineering project, there you decompose the system and then identify what the objects are and so on. And these objects in turn will emulate your system, real system behavior, some component behavior, component behavior all that will be… And depending on, where you want to draw the line and you have various levels of abstraction possible.

So, you can make your system very detail, you can make your system very little bit below or little bit more detail than analysis, but not very detail. So as an example, your let say developing a new scheduling algorithm for some wireless network so, you have you can simply focus only on the scheduling part of your system. Simply say I am going to arbitrary generate request, there are the different types and different priorities and here is my scheduling algorithm, I will write just see how may scheduling algorithm allocate slots to the different users based on whatever factors, the amount of traffic they be allocated. So, for current priority level, total request and case and so on, quality of service and thinks like that.

.So, that is very just looking at the scheduling algorithm allow that that level of abstraction is possible. But then you might say that wait a minute, this is not enough; I really want to see, how this works with say TCP and IP running in the system. So, then you have to go an implement above a scheduling algorithm, the other network layer protocols stack implement IP/TCP or you go to NS to and just use NS to or whatever features that it offers. So, you can just use that, then have that is one more level of abstraction. You looking at the layers above when the protocol stack within you say there is so, much of channel modeling that is needed.

You say the wireless channel, I am simply assuming this to be a pipe that will either drop a packet or send a packet with some probability that I choose. But I really want to model a channel also. Wireless channel could be a simply a (( )) channel with some the carrier sense you can all that or it could be even more intricate. What kind of fading is there? Is this all (( )) fading model or some you have log number trading model. So, you have to start thinking about even those have also start coming to the picture, then you say have this full n to n system. And then if you lose simulation and then presented you will be results will be much more appreciated. So, that is where you are varying levels of abstraction coming.

But there also comes into the price, this level of too much of details sometimes, because now your systems have to many sub-systems interacting with each other. You may suddenly find that no packets are getting through, when you start caching a wait a minute until, only scheduling was there things was so perfect. Now you put this wireless channel model that I work well that, but no packets are going through. Any have to start looking at the parameters of the wireless channel may be your set your transmit power levels very low.

Therefore, no packets are going through, if we set it at point one milli watt, you find that nothing is going through and a distance between the base station and new set as like that two kilometers are something like that. So, adding detail gives you more results, but you better know how to set those parameters. So, configuration of parameters becomes even more important when you start bringing in other system together.

Because sometimes those systems are thinking that you did not right yourself. To the fashionable to always go to NS to get y max come v num together and run it, but nobody really goes in actually a part or nobody people, you have to go when actually start checking what those default values are? And why results are sometimes not Suring? So, the different parameter will change the way that your output is going to be. And this way we have experienced one right practically with using OPNET. OPNET is one of this paid commercial software that you pay there the large amount for. And open has this entire stack from generating video traffic, we can generate video sources, no voice, voice our I P everything can be generated at great detail.

And then, you can send the packet so, what the wireless channels in measure it and so on. But the number of bugs that are there, now this becomes a very large system. So, where simulation where suppose to be useful is no certain to making the simple or regular software engineering assignment, where there are too many sub-system which we do not know what is going on. In a simply trying to guess based on the AP as what they with do and without the source code sometimes with impossible to debug when failures take place. So, that is something that you have to watch for, but the nice thing is you can go to various levels of abstraction.

You can we look at the CPU you can simply say CPU is simply a box. I can simply emulate CPU instructions, let says add just which is to add means I emulate add as it in emulator; we have emulates as today. Then is known I want to one level further, I want to show even internally, what the registers are like that you want to go for that you know that is even look at that as this cashing. So, for memory simply I get data, I store data, you can even expand memory to its various level; one secondary memory, main memory cache in different levels of cache and so on.

So, that is the good thing about simulation levels of abstraction, but it also comes with the cost; that you better know. What happens with the different levels of abstraction? Anything else about simulation so, I can also do very detailed performance monitoring. I can go into every node; I can look at since everything is program. I can go and if you look at OPNET for example, the number of metrics it gives you is mind (( )). The long list of metrics and sometimes you are also incorrect and we have to figure out, why there is an incorrect? The remedy for example in voice calls, something call, mean opinions code.

That OPNET was giving as the numbers made no sense, the calls are going through when the sense that packets are getting delivered, but the MOS goes where very low. Out of scale of one to five there will giving it as the three. Then we found there were some mistakes in the way computes that the mean opinions code. So, that is, but at least we have the option of digging into very find levels of performance. So, we can get fine- grained performance metrics can be measured. So for, we are saying all the good things about. So, what are the bad things or negative things about using simulation?

(( ))

It is thrown to coding error.

Lot of assumptions, we make lot of assumptions.

You will make some more, you will be the assumptions in the system will be more relaxed compare to Mathematical modeling, but where the assumptions would fail is might not really capture what goes on a real system.

Real,

So, because in real system that we lot of other things happening that you do you not really cannot capture. For example, trying to measure some routing algorithm in or some scheduling algorithm; in the backbone you never know, what is going to the instantaneous load on the back bone router. How many connections are coming in? And depending on that your performance results will be there. So, you still only have set of representative set of results.

Simulation again gives you only trends, but not real values. You cannot say in simulation I got two seconds as the delay therefore, in real at which we two seconds. That is in right; that is to much of an extension. So, the real life is not fully captured; real life environment is not fully captured in the simulation. Simulation is the controlled environment. Lots of things are our control pretty much everything is in our control whereas, if you say that for example, simple thing why was channel modeling? You say I am modeling it with this fading channel.

You have this nice set of equations, you put them in MAT Lab and then we go to the system in you find that measurements, there are completely unrelated to what you thought where this, where this model that you had. So, hardly put together, put together so, on great difficulty. So, there is the assumptions are better than what you do in the case of analysis, but they have still done capture reality so, that is one serious problem.

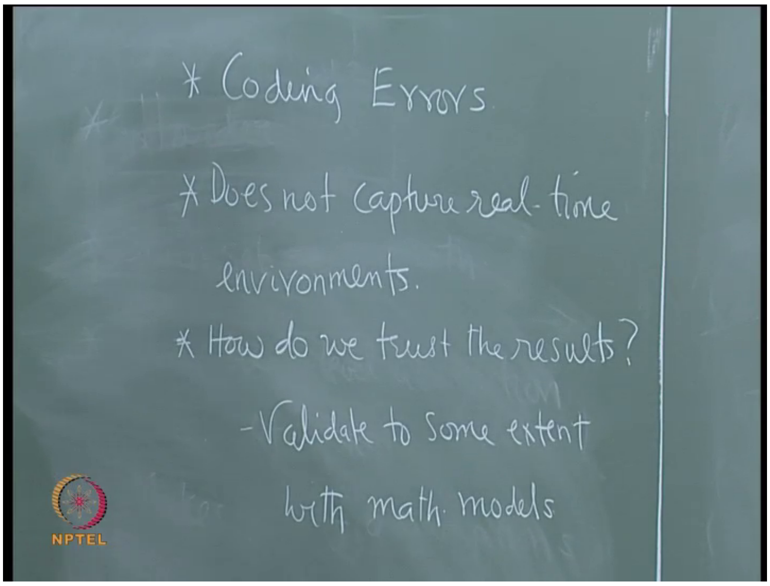
What else? I said coding errors so, if your system is prone to coding errors. And many times I have seen my own student for the past six months is come to me with four or five sets of results. MS student, in each time I found a coding error; I found that this is why, because there is for not matching what we expected. Some of the delay are not some ideal time measured is some another micro seconds. That is not true should be nanoseconds at therefore, something is wrong. So, we are to go back and to visit over and over again.

So, you cannot really be satisfied that your code is actually working correctly or wrong so, much of randomness. If it is sorting algorithm you know, at the end of it and I can simply verify. Then has to check, this is okay, yes this is indeed working correctly, no issues at all. Whereas it is a more complex system there is being a model, then you will never know with the results are correct or not.

If the system says delay is three point five two milliseconds, we are in an implementation of the queuing system, when it is something how do you trust that three point five two milliseconds. So, that is where you have to verify the code, verify the code and also look for other was to validate you are code. One is to make sure that the numbers from simulation your implementation is more or less accurate. At the same time, you need some other way to make sure that your values are correct and that is where analysis concept.

So, you use Math analysis as the way to substantiate your simulation results. So, many times you will see plots that compare, simulation and analysis, then they should be closed each other. They will not be exactly the same, but at least the trend should be more or less close to each other.

(Refer Slide Time: 29:39)



So, therefore, this still you are issues are coding errors and it does not capture or real world whatever environments. And it is how to we trust the results?

So, for that you should validate with and this is again to some extent. So, they save a mutually beneficial purpose that Math models can be verify by simulation and by source.

Which early a simulations are you may (()) as a model (()) and then, how do you go back again and bring your Mac model and then took care your results

Let us look at this queuing as an example. Queuing, I will I simply do my Markov chain analysis and I use my birth rate process and say delay equals one hour mu minus lambda. That is what the math model tells. I Implement this in a real system vary actually generate packets, where I am actually modeling the behavior of the system. Here only capturing in the Math model, I am not really capturing that.

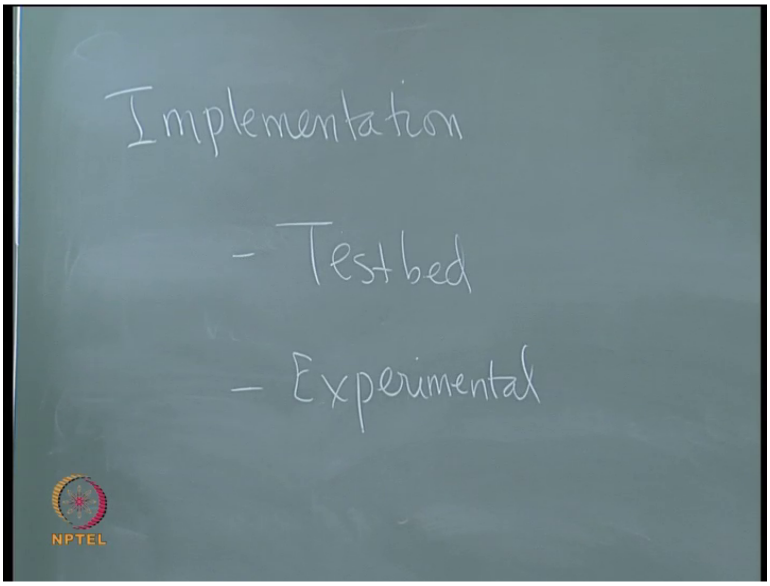
May be in like, even if you look at wireless channel model, your channel model is based on some formally to use and says that are this point in a time. This is the probability that this you know, bit will be filliped and so on. This will be the amount of noise in the system and therefore, that is listening probably that is simply adding noise and then do it; do it probability fillip there whether a bit is set is going to be change to one or zero or vice versa.

But whereas, you go to a real system in simulation also actually go and we do that. So, from Math to simulation, yes in that case, in some cases you are simply using the Math model as is, but in some cases when you really go to measuring on the field then that is totally different. There your noise is everything is not control by you at all. So, some parts of it yes, you are simply using the Math model also in the simulation, where there is no cross validation possible.

Only when we go to the real system, you would know. So, when physics again, you have theory that say this is the way the system should be behave. Then you have experiments that you show that whether your model is correct or not and then the key various changing those models. So, depending on which aspect of the system is model. If you are some parts of your simulation is also based on the Math model half say, channel or some other system. They knew as the same where as we have propagated, but where you can model elsewhere. But actually implement here in simulation, and then you will be able to differentiate.

The third part is experimentation which I guess not much to say about. You all know, how difficult it is? This is actually the hardest.

(Refer Slide Time: 32:45)



So, we say implementation, we use the term whatever test bits sometimes of if it is the large scale system, it is the small system implementation that several terms. Sometimes people also say experimental results, where the experimental they mean that, they actually build the system; the different terms you would find. We do experiments with everything is an experiment, whether it is simulation are real line. But they would use the term experimental results are present in the paper, which means actually have built the system and show the some results from that. So, what is then, what are the pros and concepts experimentation or implementation.

Takes time.

Takes time. Now, you have worry about lot of details you getting your systems, integrate getting those entire header files, and put together all the packet is working correctly. So, it is takes much longer to do implementation. What else?

It is capture on parameters this is the real world system

Yes. So, it will be affect operating in a real environment. Even that to in sometimes will never know. In the in the lab, you might be I will guess to do these. We had a system for doing zip code recognition or pin code recognition where in the lab everything works beautifully. Because your spread images and then you take it out to the field trial where the sponsor USPS postal service was giving some images as the system crashed. It does it may should have not crash, simply does not given in answer was system crash and I got the system, but anyway. So, there are some times, you are even in an experiment, yours still doing a lab experiment. It is not exactly your real world experiment.

So, whatever works well in the lab might not work for in the real world. When you to say for that the web server for example, web server in the lab you will be testing that all your local load, you works perfectly. Then once the open it out in the web, system simply starts crashing for various reasons. In compatibility is are just functionally the implementation error and so on.

But those are, but there you need that. If really want to take your product your system to become a product, and then you have to go into that. So, one it takes harder is definitely more time consuming. So, most time consuming of these three is that. And the second is you do usually find that and when you go to experiments, you always try to implement really simpler simple algorithms. You leave all the complex algorithm, complex mechanism how would because, especially in prototypes. When you go to a real system, then you have to in add all those wells and vessels. So, so that is one from the time and cost involve it is fairly high. In terms of results, it is more accurate or less accurate

More accurate,

More accurate will actually it is probably as bad as simulation too. You really cannot say that you are because whatever errors are there in simulation. We think about coding errors in simulation, coding errors can there is also in your system. So, your results here will be little bit more believable.

But we should not think that this is in the accurate. Sometimes we will find when as an experiment that you know, we are way that we lot of variations in have measured results. Because of variations load, variations in other background traffic or whatever it is in the system. So, it is little bit more believable, but I would not say that it is the most accurate, because all those errors still also propagate. So, but you generally it is a more easy to sell a system within implementation. If you had when Google implemented the page rank algorithm, probably able to show that see this works better than whatever searching algorithms of there.

And therefore, you should try to market the systems and so on. So, but it definitely gives you much more inside into the system and behavior and so on. And in terms of results, when if I want to vary amount in parameters, I can do that very easily in my simulation environment; that was I will say vary all this way n values, try this for another ten values see what happened? Okay, fine. In simulation like, because in simulation sometimes your (( )) times is constraint. I mean you are simulating it for say five minutes of system time, which is taking a few hours to run and so on.

So, you can vary parameters at will in the simulator environment. In a real environment you can, but it is going to be even more time consuming. So, mostly you will find that in a simulated paper that we like know the simulation based paper; you know twenty, thirty graphs, does not matter; vary, vary and get all those graphs put together. You go to a test bit waste paper you will find someone table that all. The measured it works here is the measurement, I am done. Because it is finally is shown to be working. So, usually takes much more time to get more lot more results out of the system, but just mostly matter of time.

So, that is why you do all the analysis with all the variables trying to figure out what are factors of being the system. You find out which is the more important one are? Then you focus on that when you trying do the analysis with the experimentation, you try to do that. So, that is the…

(( )) trying to implements something on the forms desires do not forming at some time we built in modeling

True.

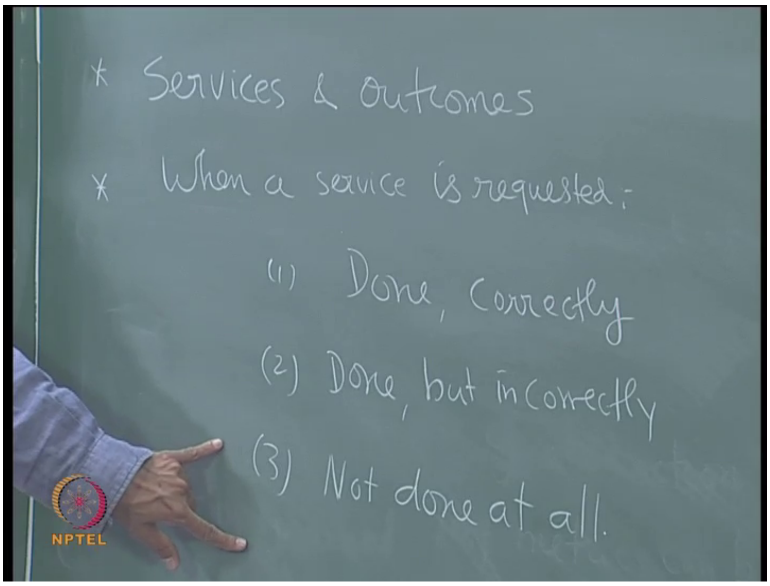
(( ))

That is almost like a feedback cycle that you are have your model, you have good implement, then you find that you have to go back and makes the changes your assumptions too. Yes, you find that some of the traffic is no longer Poisson; it is some other distribution. Therefore, I have to go and change that to some other model. I am not saying experimentation is inaccurate, but you have to be watch will about even experimental results. That you should not trust that completely. You should use again you should validate at with what your system tells you, with your other two systems tell you.

If there is a big (( )) and see let say and throughput that and so on. Then either it is purely environment dependent. Something is the environment this is not capture in those two models is what is causing in or it is may be the something wrong with implementation itself. So, it is there is like a magical bullet that says one is better than the other, but all these three systems usually tent to be tent together or right at least two of those should be there to that you has a good amount of faith. If you are, if you only have implementation, then sometimes you may find that there are too many systems base thing. For example, the amount of memory you give wired system should give you at a virtual box in a set the virtual box have only five hundred mega bytes are so.

And then, in you run you are testing on that five hundred MB system when you say well and this is very slow, this algorithm is very bad; you put into a 4 GB system suddenly, that the system is better. If you did think about that the fact when you miss that so, you have to think about those are also. What is the environment, if there anything in the environment that this causing the system to behave differently from how would be hell in the other two methods. So, these are three high level descriptions of the techniques. We will next look at some of the common performance metrics and that will pretty much doing as to the end of chapter three. And then that is the end of all the introductive material, then will start getting into random variables and so on.

(Refer Slide Time: 39:25)



So, commonly used performance metrics so, when you think of performance metrics for a system, you think about all the services that the system offers, whatever it is delivering a packet, computing some in some computation something and then the expected outcome for that. So, if you give, if you it is the routing simulation, then your user generates the packets gives into the TCP layer. You expect that the network will your TCP model simulation model says, I will take care of all the necessary behavior that will deliver this packet to the n node that includes queuing, intermediate routers, de-queuing and finally, getting deliver to the end over. So, you talk about services and outcomes.

So, that is for this the when we list the ten things to do, for doing performance evaluation this is one of them. What are the services offered by the system and what are the expected outcomes of the services? So, this is more or less your requirement so, you start here. And in general the outcomes for many cases are when you request a service either there are three possibilities, when you request a service from a system. And these is not only system as are even sub-system level, some sub-component also might be doing certain things.

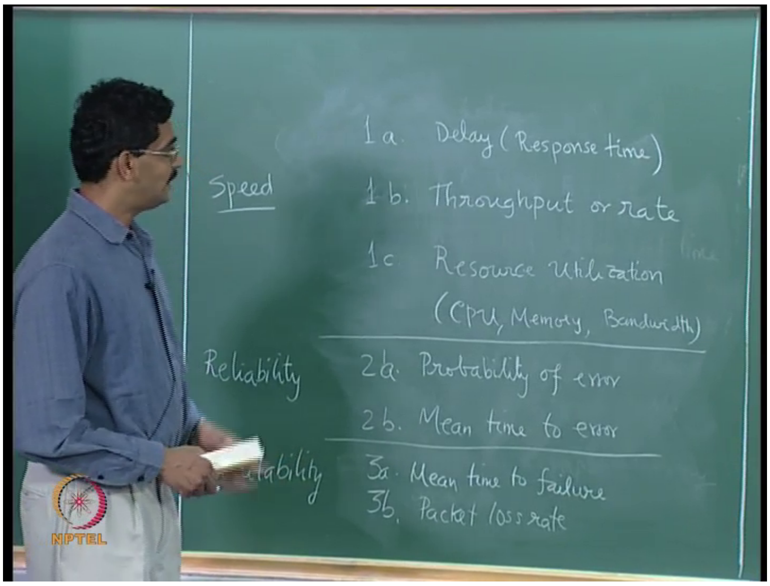
So, you will have to define for each sub-system, each sub-component some these things. So, when a service is requested, there are three things that can happen. One it is done correctly; second it is done, but incorrectly and third it is not done at all your request was rejected for whatever reason. Now these will make sense in some system, it would not make sense more. If you are talking about a sorting algorithm, it will either sort are not sort there is no question of in that the system not get.

Your system might crash, if you give it very large list of say two hundred to two hundred million entries the system crash, then that is an example of if we did not get that. That is because of some memory leak in your system. So, that if functional issues from a network point of view it related to capacity and errors and so on. So, if I give a packet to a router either it forwards it or does not forward it. If we do not forward it simply means that the packet got dropped. Because a buffer got full so, there for there are some capacity constraints that cost this to happen.

So, what is the probable? Always this is system itself if we look it at as when says web server. Web server is sometimes you request a server some http messages sent; you expect the response back from the server. The server could either give you a response or the server could simply not respond at all, because it is overwhelmed or it is down currently. So, but unlikely you will get a different web page if we go to a server. So, it depends on the service that you have to look at see which of these outcomes are possible.

But in the case of network accepting a packet, you can deliver the packet correctly destination which is here, what one or it to be delivered, but still because a where are as you might be delivering an incorrect packet of the destination; that is possible. Errors are introduce in your somewhere in the system some errors are happen and the third is the router has in the simply drop the packet. So, these there are three. So, depending on which of these happens, we have again some classification of metrics. So, if a packet is in the delivered correctly, then we look at set of three metrics.

(Refer Slide Time: 43:32)



So, I would say 1 a is delay or response time. So I made a request, what is the response time of the system? When did this request get satisfy? So, that is your first metric. We know that the packet is very easy; I sent it when these get delivered and it is courier by speed post when these get delivered. So, that is the… So this is a user specific metric. What did this user see? What did I see in terms of response time? Then from the system perspective, I would like to know how many packets per second I am able to deliver? That is representing the capacity of the system. So, that is usually throughput or rate.

So, that is in some service request per second; that is to capture. The third would be resource utilization. Resource could be many things like CPU, cycles consumed memory or you know bandwidth consumed etc. So, during for this example, I am sending several packets on a link utilization of the link is fifty percent, hundred percent for the same load. If I am able to use the resource is less than that is better or when you writing a searching algorithm, sorting algorithm if I use less memory that is better. So, you look at resource utilization as in other way of capturing.

So, if you might be able to get the same delay, same response time or same throughput. But if one system use is less memory than the other, then that is better so, utilization of resources is less then this particular system.

So, that is what, so these three would happen, whenever the service is being done correctly. The service is done therefore I am more interested in one of these three. Of course, other of these is also possible variance of this. So, these are the high level metrics in within that we would say whether it is mean or mean plus variance of this that there. So, that is the first category. Then the second category is when something gets done incorrectly. So, simple thing is packet gets to deliver, but there are errors in the packet, corrupted packet is being delivered.

So, you would like to know, what is the error to say 2 a probability of error? And this probability of error is if I if look at bit error that is simply that metric. Simply bit error is simply what is the probability the bit error is corrupted, when it sent on a link or packet error rate then the packet travels this multiple links what happens? What is affect to packet error rate?

So, therefore, that is those are metrics that will capture the probability of error and set of related to this on some system they also say, what is the mean time to or mean time between or mean time to error? So, this is will tell you, how long the system will function typically, correctly before errors take place. That say for hundred milliseconds there will be no error and then error take places; every hundred milliseconds there is an error. So, this might be possible to compute, might not be possible to compute, but that is in other way of looking at the system character of thus metric characterization.

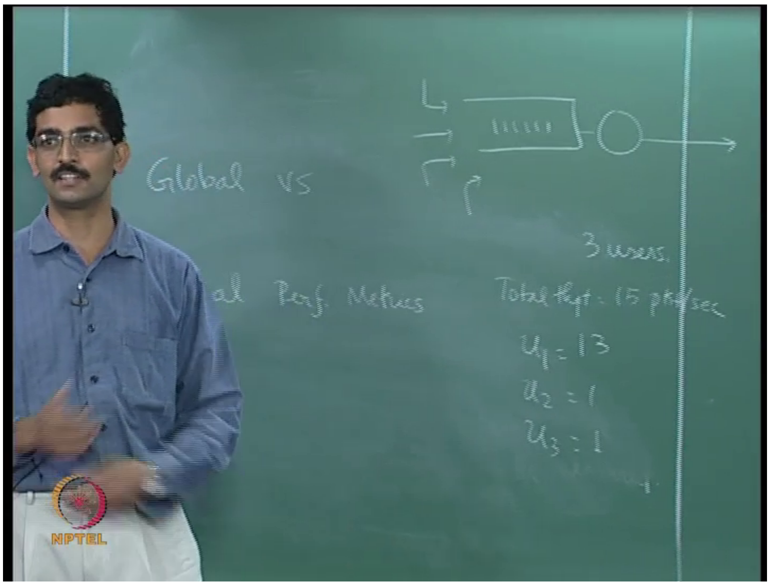
Sir, we cannot maintain to the care of.

Maintain that would be in the case of system, Yeah so the system face so, their usually come into reliability. Yeah this is Yeah mean temp repair is also there, but on a link that is kind of not there. Because these are transient errors; they cannot of come and go; they are not fixable errors. But if it is a system that goes down that would come and will more availability of the system. The third part is then something is your requested is not at all done. So, not done can be in the case of network it is simple, what is the packet drop rate? How often this is happen or it could be something so, this is usually you know, called as speed reliability and the third category is availability.

So, a variable is depending on what system it is, though common metric that we would see is like mean time to failure. This is to of machinery; there we say what the mean time to failure is than the mean time to repair also comes in. But that is not the behavior of the system itself; that is like a separate process. The repairing process is independent, simply take this machine of the shop and get it repaired also or in the case of networks and so on it is usually the packet loss rate, the something that will capture with the fraction of packets that your drop is not service at all. This is loss in the sense of dropped packets, because buffers are full is not because of link errors.

It is because of buffer was full, it was dropped by the system therefore, the system load is very high or the system did some random RED can of algorithm which top packets on the way. And then, which an each of these metrics I can have mean like a set variance and all the moments that I want but that is the high level. And then one more categorization of metrics is whether they are global versus local.

(Refer Slide Time: 50:00)



So, global metrics would be something for example, in the let us look at throughput. Thus earlier example we saw there is aloha systems are that is the router; that is sharing packets; that is a servicing packets from several queues. So, if I represent a queue like this, like a there are several users sending packets to this queue, this is the queue. So, packets are buffered here. So, that the buffer that is your server, packets are entry from different sources. So, as a global metric, what can you think of? Global in the since of system level, I would say global some system level metric. What is system level metric?

Through put,

So the through put throughput utilization. So from at this, I can say how many packets per second are being delivered, are being delivered by the server? That is throughput and the second is utilization. How often is the server being utilized? It is ten percent, twenty percent, fifty percent and so on. So, those are system level metrics. Now if you spend the money to install a server, you want those to be high. You want throughput to be high, because your able to service more customers, you want to put faster server, you also been the want to keep the server is busy the most of the time. Because you are them, paying in this installing the server and so on. So, you want to utilization to be very high.

So, from that perspective I want to keep those to these two numbers is very high; throughput as well as utilization. What from the user perspective, which is important?

Delay,

Delays, response time I do not care of others get service or not. My packet should go through first, I should be able to go, people do that right, you stand in line, always come in front of the line of this try to jump you. So, that is the, so therefore, they only care about response time.

So, the user's time is one thing in response time and the second is from throughput perspective, I want my fancier of throughput. So, fairness is also an important consideration. So, if say that this system let say had though 3 users, we saw this is in networks. If the total throughput is fifteen packets per second and user 1 had thirteen and user 2 had 1 and user 3 had 1 that is not fair. But system is very happy, because it is able to service fifteen packets per second.

System capacity is eighteen packets per second, system is been wonderful is serving everybody or at least in terms of number of packets per second; system has been able to resize to be very high number. But in terms of we look at fairness, and then it is not fair. So, there is disparity in the amount of service that is been given to the different customers. So therefore, when you try to design a system you have to see which of this metrics you have to the measure. So, simply reporting total throughput will be misleading. If there is this big descriptive and this also happens sometimes in simulation, sometimes you never pick up I get from a particular connection.

For different you had your some metric to compute and it say end of not selecting source at all in which case this is the unfair. So, from the user perspective this is important system perspective this is important. So, you have to look at a system if this is not at all important, you does not matter. These are all VIP connections; therefore they will end in get more connection that is different. But if you are interested in fairness, then you should also measure this other metrics.

You should not simply stop with total throughput as the only metric of a system itself.

(( ))

Yeah. So, it depends upon the on the impact of those metrics on these system. So, in the case of packet delay mean is alone not adequate. We all probably know that because variance also matters, variance in terms of range. What is the range of… If I tell you that your packets are going to be having delays from 1 millisecond to 5 seconds, very long range in delays and you are watching a video and so on, then your video catching and your video buffering algorithm has to buffer almost that much amount of five seconds worth of data has to be buffered. And therefore, it affects it was the affects you or you if you do not have a buffering on what happens packets will start. It will have a packet, the no packet for a long time other packet.

So, it will be very jumping. So, it is depends upon how it impacts the application. And that is again coming from the intuition of the expert, who is looking at the system. In some cases, it is really does not a matter; in some cases it could be important. So, throughput if you look at throughput of a user, measure over various point in time, if it varies a lot if that could be let us say over every 10 seconds, they vary the measure throughput.

And I always get one kbps throughout those 10 second intervals, then the user is do will not care. If we suddenly goes to 10 and comes down to 0.5 and so on, the user might not if the user does not know anything differently. That the delay absorbed is does not vary as much, then it does not really matter. So, usually for some metrics, you look at second third moments and so on.

But most of the times we simply look at mean of course, not only mean, we will look at this much later on and you can also look at your regular arithmetic mean or something like a geometric mean, harmonic mean, all those things also come into play depending on which system it is sometimes some of these will make more sense. Then for example, ratios comparison you might not found to do a simple arithmetic mean, but like a some other mean will more useful or even median, mode all those things also remember from your basics of probability, our statistics. So, sometimes it is also median mode, those are things are hardware to compute with a Mathematical model, but easier I compute the simulation for a really depends on the system.