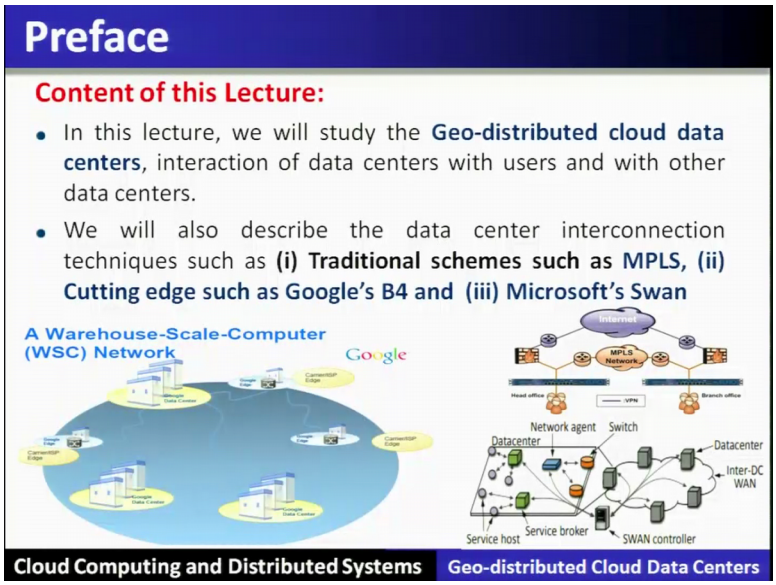
Cloud Computing and Distributed Systems
Prof. Rajiv Misra
Department of Computer Science and Engineering
Indian Institute of Technology, Patna

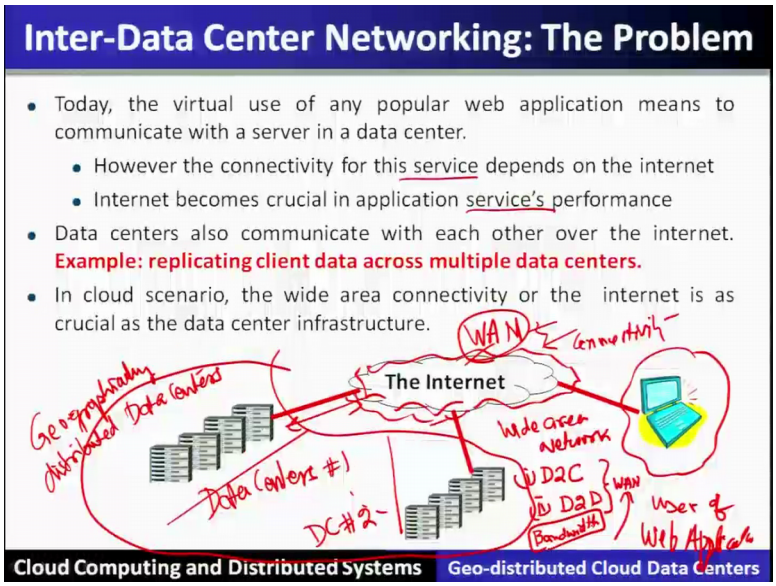
Lecture – 06
Geo – distributed Cloud Data Centers

(Refer Slide Time: 00:17)



Geo-distributed Cloud Data Centers; prefix, content of this lecture, we will discuss geo-distributed cloud data centers, the interaction with other data centers, and interaction of data centre with the users are the prime point for the discussion in this lecture. We will also cover data center interconnection techniques such as the traditional schemes of MPLS, versus the current edge schemes, which are applied in Google's B4, and Microsoft's Swans, and the current trends in the data center networking.

(Refer Slide Time: 00:59)



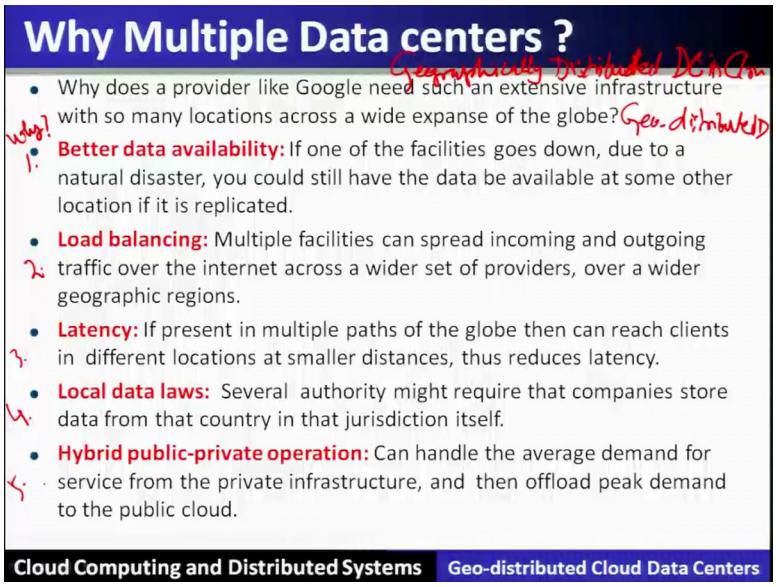
So, the inter-data center networking, what is it and what are the different problems, let us see and understand it. And then we will discuss the trends, and what are the traditional methods to overcome from this particular problems. So, in today's virtual use of any web based applications means that you are communicating with the data center in some form or the other, so that means if this is the user of web based application through the internet, this particular user will be communicating with the data centers. This is data center number 1 and so on; this is the data center number 2, which are at different locations. So, they are called geographically distributed data centers.

So, therefore the users of any web based applications are connecting or interacting with the geographically distributed data centers through the internet. Now, this particular service, which the user is trying to get through the web application highly depends upon the internet connectivity. Now, you know that from the background of the network, this internet is also having an underlying network, which is called a wide area network. Now, as far as these data centers are concerned, these data centers, which are at different locations they are also communicating with each other with the help of internet. Therefore, the wide area network that means to provide a good service, the performance of the service depends highly on the performance of the wide area network in this particular scenario.

So, in the clouds scenario, we will see here the importance of wide area connectivity or the internet, which becomes a crucial as the data center infrastructure. So, the problem is that there are two kinds of interaction, which we will focus that is from data center to the user. And the other is between to the data center with respect to the wide area network. Now, what is the issue, what are the problems this wide area network, which is nothing but the internet is highly dependent on the bandwidth.

And there is a huge cost if lot of communication is done on this particular WAN or the internet, then there will be a huge cost of this bandwidth. And if it is not utilized, then that cost will be overhead or a burden or a waste. So, how this particular bandwidth, which is primarily essential for the performance of running web applications is going to be optimized, so that the performance with the cost or with the low cost, it can be provided as the service in the cloud scenario.

(Refer Slide Time: 05:22)



Now, the question is why the multiple data centers are existing in the previous picture that is also called why the geographically distributed data centers are there in the cloud. So, this question we will see, why we need the geographically distributed data centers, and then will see the networking part, which is also an essential infrastructure for providing the data center to the cloud. Now, the question is why does the provider like Google needs such an extensive infrastructure, so that at many locations across the wide

globe, these data centers are residing, and their also called geographically or geo-distributed data centers.

Let us see this particular question, so the first region of having the geographically distributed data center is to provide the data availability, so that means, to provide the better data availability means, that if one of this data center facility is down, due to the calamities earthquake or some other region, then you will continue to get the data or the availability of the data. If it is replicated at more than one geographic location elsewhere, which is maybe active at that point of time.

The second region is called load balancing. So, the second region is due to the load balancing, so that means, if multiple facilities can spread the incoming and outgoing traffic over the internet across the wide set of providers, over the wide geographic regions, that means the entire traffic is uniformly distributed. Hence, the load balancing will give the more performance better performance, and also it is a cost effective way. So, load balancing is possible if you support the geographically distributed data center, that is why, the providers like Google has its data centers geographically, it spreaded over twelve different location across three different continents on the globe.

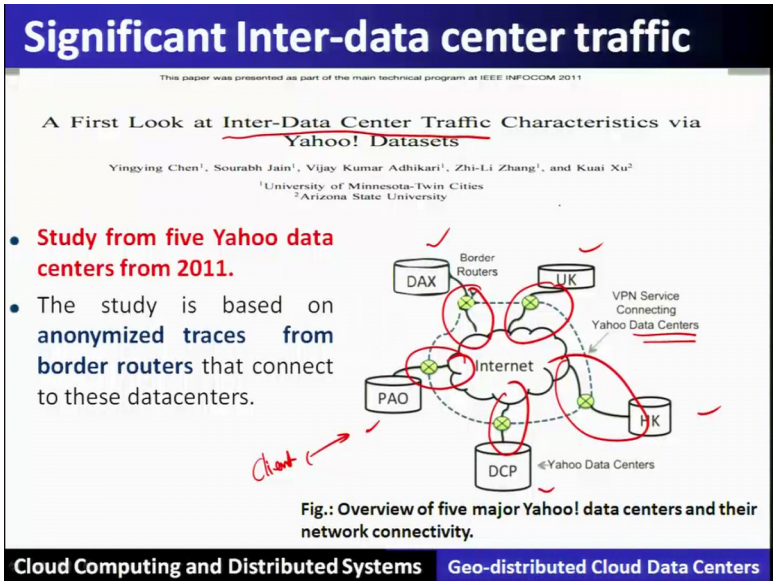
Third one is the latency. Now, if present in a multiple paths of the globe then can reach the clients in different location at a smaller distance. Hence, it reduces the latency meaning to say that if a client is located in the country India, and it is being served with the nearest data center located in India, then the traffic need not have to travel across a wide distance located in US or in some other country. Hence are a smaller distance is possibility, which will basically reduce the latency in accessing the application. So, hence the latency will be reduced, if the data centers are maintained in the form of geographically distributed sites.

Fourth one is called the local data laws. Now, there are different countries, which have different laws of for the companies, who stores that the data as per their jurisdiction of that country. So, therefore the data and its axis has to follow the laws of the land, wherever the data centers are maintaining those particular data sites. So, it depends upon the different applications, and how this laws can be useful in maintaining the geographically distributed data center to provide the services are also going to play an important role, that is why, this geographically distributed data centers are being made to

ensure the services, and also using the laws of maintaining the data as per the jurisdiction of different nation or the country.

Fifth one is called hybrid public-private operations. In the sense, that if there is a geographically distributed data centers, and also the private data centers are being maintained, then what people can do is that the average demand for the data or the service through the private infrastructure can be provided with the lower cost. And whenever there is a high demand in the peak axis or in peak period, then that kind of load can be shifted to public, hence it is called hybrid public private operations, can be supported with the help of geographically distributed data centers or multiple data centers, here in this case such as private and public clouds.

(Refer Slide Time: 11:38)

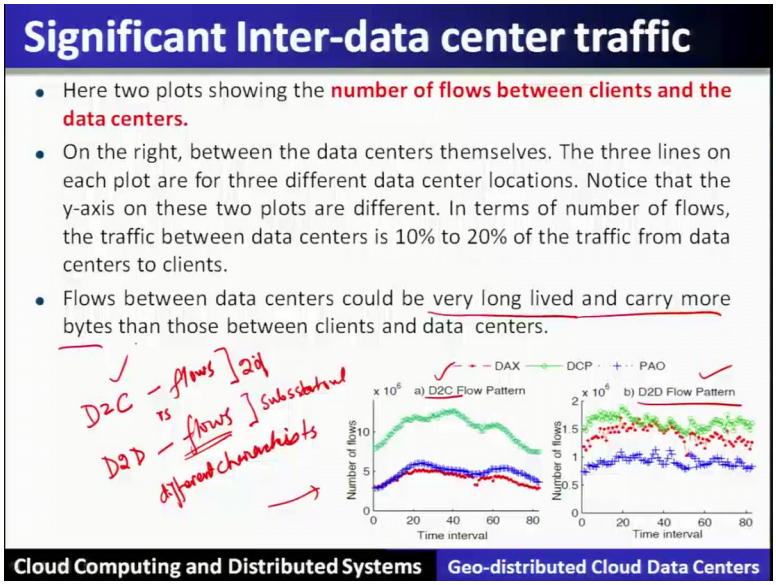


Now, we will see here that in inter-data center, what kind of traffic characteristics is being observed by different service providers, and what are the techniques to ensure the performance and efficiency. So, it was study of 2011, Yahoo data sets or data centers, which are maintained at five different locations. So, the study reveals that in the picture, five different locations are shown over here, which are nothing but, they are the data centers, these data centers are connected to the internet through the border routers, these are the border routers.

Now, as far as the inter-data center traffic is concerned that means, the traffic across the data centers. It is shown that these particular traffic, which will basically are going to be

a major data traffic between or among these data centers in contrast to the traffic, which is originated from the client to these data centers. If we compare, then this particular traffic is quite substantial, hence we will focus on the inter-data center traffic management, so that a better performance, and a lower cost can be achieved in this part of discussion.

(Refer Slide Time: 13:35)

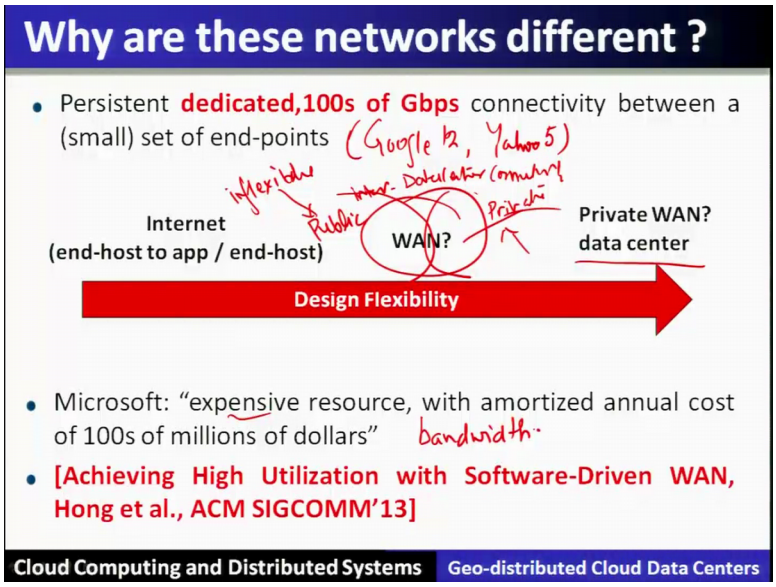


So, therefore a significant inter-data center traffic is observed in that particular study of the data sets of five different Yahoo data centers, that shows that the number of flows between the client and the data center is plotted in the two figures. So, this figure number 1 shows the traffic or a flow from between the data center and the client. Second figure shows the flow pattern of inter-data center traffic. Now, here you can see on the y axis the number of flows, which are being pointed. Now, one thing we can observe here is that the total number of flows that is the traffic between the data center is 10 to 20 percent of the data of the traffic from data center to the client.

So, here one more thing has to be observed that the flows between the data center could be very long lived and lived and carry more bytes than the flow between the client and data center, so that means the client and data center flows versus the flows, which basically has between data center is having a different characteristics. And this is quite substantial, this is only 20 percent of the total traffic, so 80 percent that is quite substantial is the flow or inter-data center flows. Now, as far as the characteristics of

these flows, if we can see here the traffic or the flows, which are there in the data center are very long lived and carry more bytes, that means, the traffic is a consistent between this data center or inter-data center traffic compared to the traffic or a data flow between the client and the data center.

(Refer Slide Time: 16:13)



So, now let us see that these particular flows, which are more substantial in case of inter-data center traffic has to access the wide area network. And this particular wide area network is nothing but 100s of Gbps connectivity between by small set of points, meaning to say that in the Yahoo only 5 different data center sites are there, Google there are 12 site, so hence or a set of end points basically will observe a huge basically traffic.

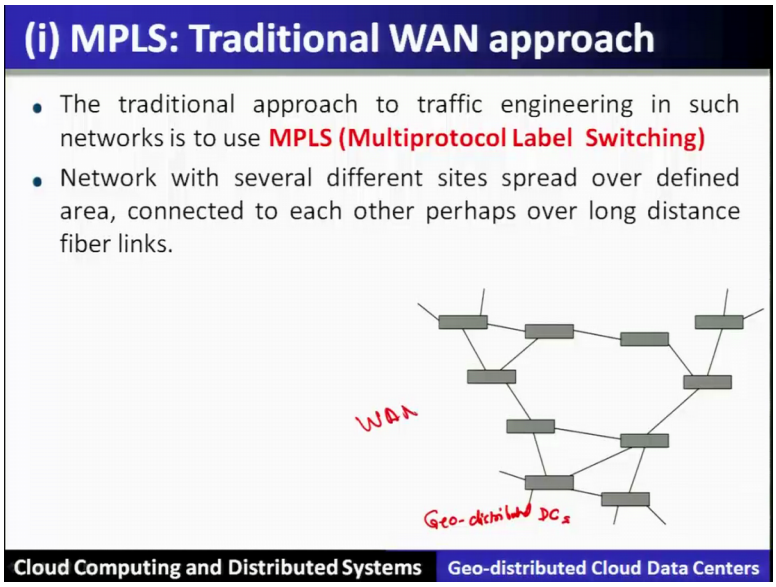
So, as far as we will see here, this particular WAN, which will provide the connectivity to the inter-data center connectivity through the wide area network. Now, the question is how much this of this part of WAN belongs to a private WAN, that means, the data center how much the data centers can lay their own fiber or can own that part of this WAN and rest, which is not own by them they can be a public, public versus the private, so that becomes a question.

So, public is as good as, private is as good as, it is owned by that particular company, but as far as the public is concerned the cost of usage is going to be paid and also, it is inflexible. Inflexible in the sense, the protocols, and algorithm, and the techniques, which are used in the public wide area network, that is in the internet is inflexible that means, it

cannot be much cannot be done. Whereas, if it is private then all the protocols, can be can be redesigned policies can be implemented.

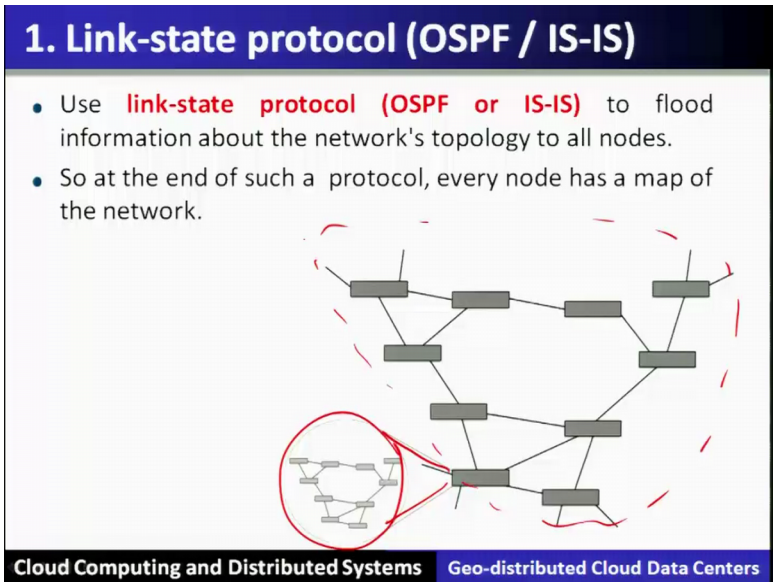
Therefore, Microsoft has said that the expensive resource, with amortized annual cost of 100s of millions of dollars is being is spend in the bandwidth. So, therefore we will see, how to achieve the high utilization with software driven wide area networks, so that is being published in ACM SIGCOMM conference in 2013.

(Refer Slide Time: 19:06)



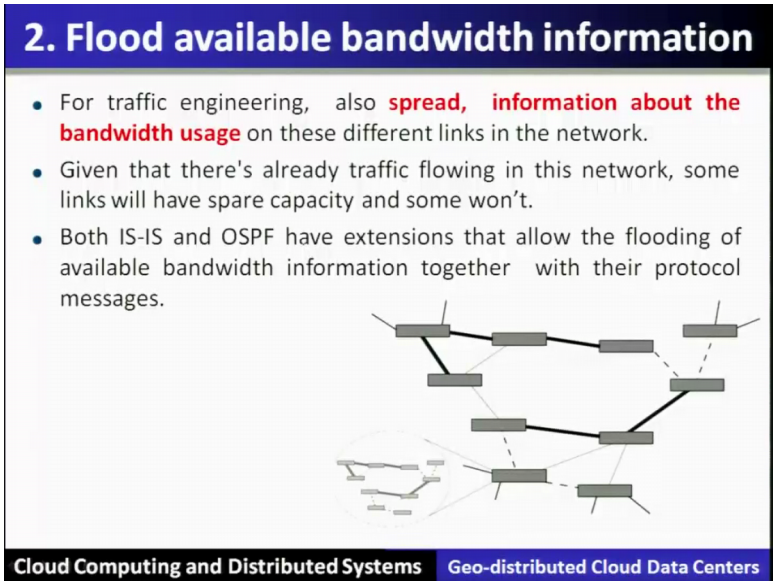
So, before we go into the details, let us see what are the approaches in this regard, that is what are the approaches to utilize the wide area network in a traditional way; so, traditional way of utilizing the wide area network efficiently was nothing but MPLS-Multiprotocol Label Switching, which ensures or which provides a traditional approach to do a traffic engineering in such networks using MPLS. Here, in this is key, this is the network with several different sites is spread over the defined area, so this is a geo-distributed data centers. So, they are connected over the long distance of fiber links either through the wide area network.

(Refer Slide Time: 20:12)



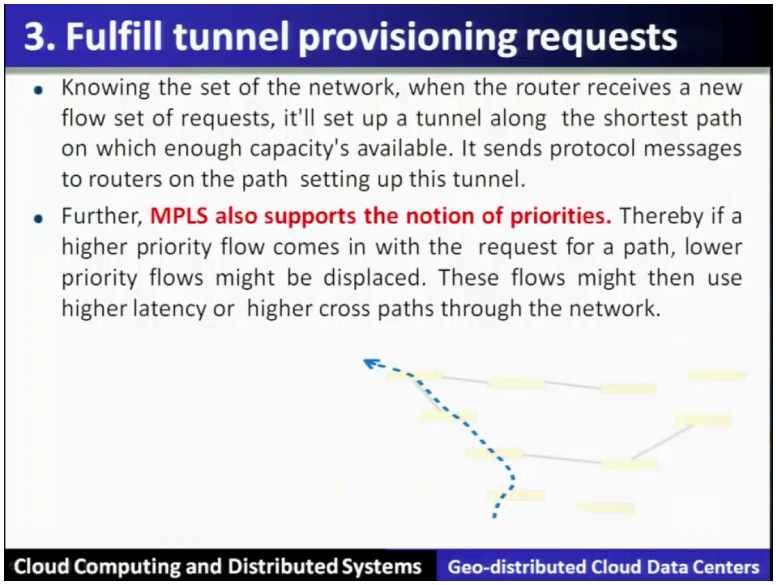
So, the information about this particular topology can be collected with the help of link-state protocol that is OSPF-Open Shortest Path Flooding, and IS-IS to flood the information about the network topology to all the nodes and when the protocol terminates, so every node will have the map of the network, which is shown over here. So, after the flooding of link-state, so at the end every node will have the complete picture of the network within it.

(Refer Slide Time: 21:03)



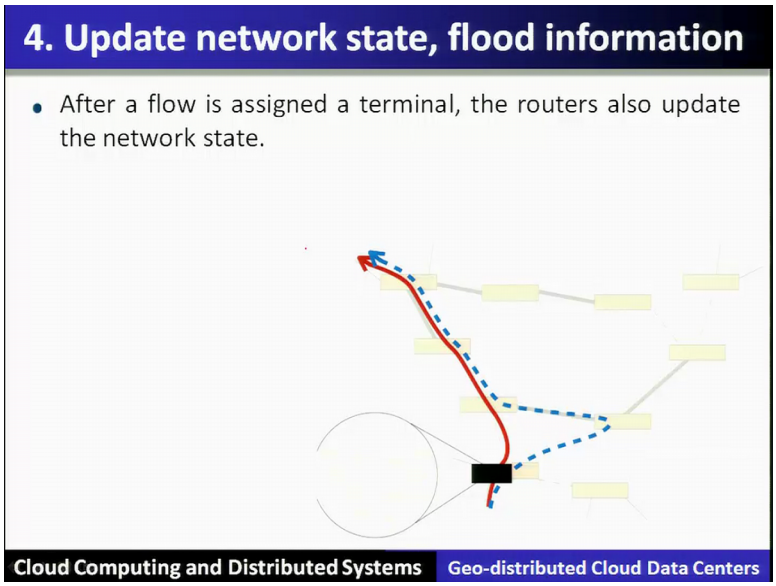
Now, let us see what are the use of this information. Now, for the traffic engineering requires to spread the information about the bandwidth usage on this links in the network, that means, how much is the bandwidth available required to be known at all the points. So, given that there is already the traffic flowing in the network, some links will have the spare capacity and some would not. So, both IS-IS and OSPF has extensions to allow the flooding of available bandwidth information together with their protocol messages. So, besides topology the information about the available bandwidth is also spreaded and now known at each point.

(Refer Slide Time: 21:57)



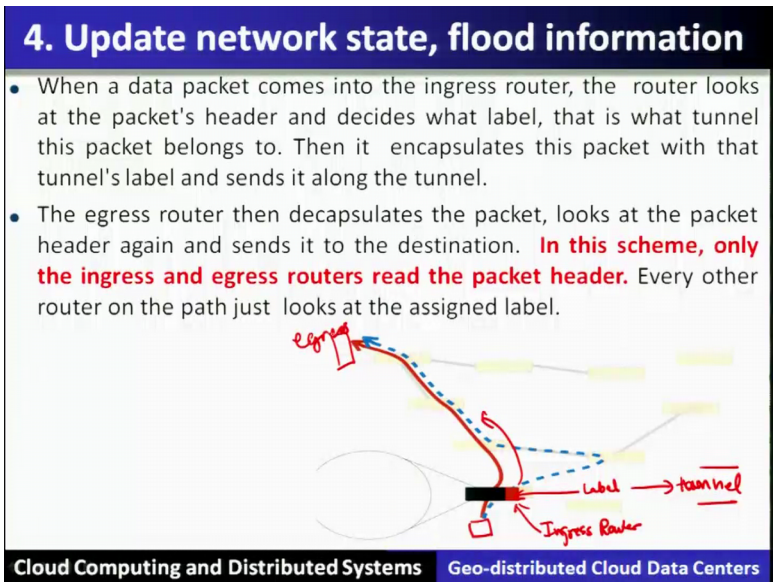
Now, third is to fulfill the tunnel provisioning request. Now, here knowing the set of networks, where the router receives the new flow requests, it will set up a tunnel along the shortest path on which enough capacity that is the bandwidth is available for supporting the data traffic. So, it sends the protocol messages to the router on the path setting up this particular tunnel. Further, MPLS also supports the notion of priorities. Therefore, if a even if the tunnel is established, and if a higher priority flow comes in with the request for a path in the lower priority flows will be displaced. And this lower priority flows and then use some other higher latency path to support the high priority flows, so that is all the possible using MPLS that is to fulfill the tunnel provisioning requests.

(Refer Slide Time: 23:02)



Now, the fourth point here in the traditional way that is in the MPLS way is to update the network is steady, and also the flood information. Now, after the flow is assigned a terminal, the routers also update the network a state.

(Refer Slide Time: 23:23)

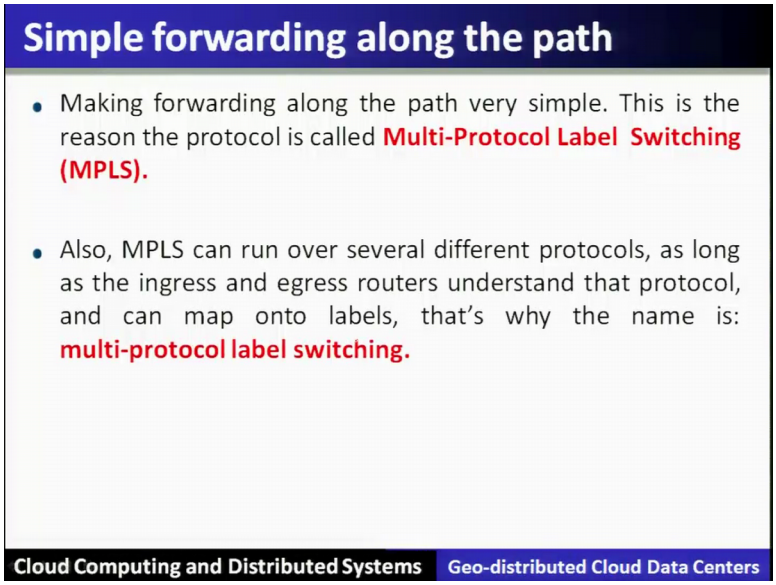


So, when a packet comes into the ingress router for sample pack end is originated from here, so this becomes an ingress router. The router looks at the packet headers and decides what label, that is which tunnel this packets belongs to here, it will assign the

label. Label means, which tunnel in then encapsulates this packet with that tunnels label, and send it along this particular tunnel.

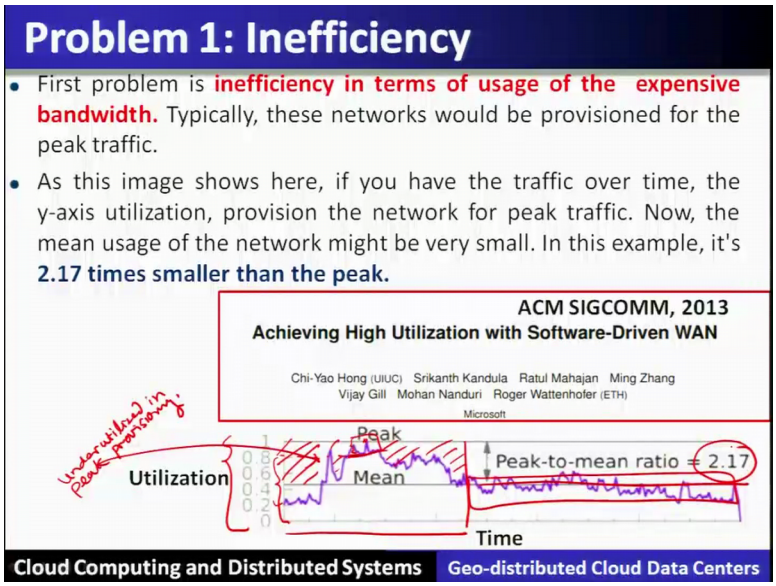
Now, egress router then decapsulates the packet, and looks at the packet header again and sends it to the destination. So that means only ingress router and egress routers are only looking inside the packet headers and rest are all using only the labels that means, they need not have to look into the routing tables and do not need much processing into the packet. Hence, once terminals are established the routing becomes quite simple, and therefore it is also safe and efficient also and also supports the virtual private network in this particular scenario. And using this particular tunnel lot of traffic engineering is possible in the traditional way.

(Refer Slide Time: 25:13)



Now, this simple forwarding along the path that is making forwarding along the path is now becomes very simple in Multi-protocol Label Switching. I have explained that only ingress and egress routers only. They have to basically consult their routing tables and rest of them are only in that tunnel, they have to only work on the labels. So, MPLS can run over several different protocols, as long as ingress and egress routers understand that protocol, and map on to the labels that is why the name is multi-protocol label switching.

(Refer Slide Time: 25:56)

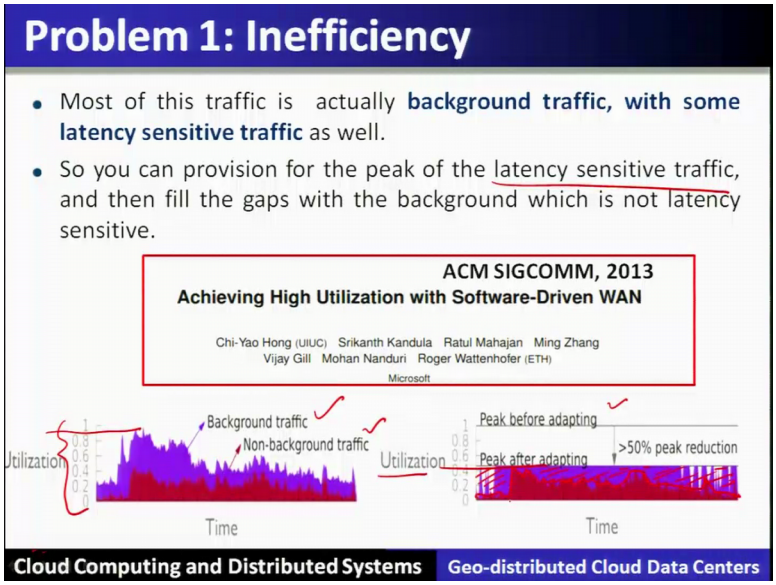


Now, in the traditional approach, let us understand the inefficiencies with respect to the cloud inter-data center networking. And then, we will see what are the latest approaches are going to be solved. So, the first problem is about the inefficiency so, the inefficiency in terms of usage of the expensive bandwidth. So, in the sense that whether the bandwidth is utilized fully or it is not that much utilized, so that is called inefficiency. So, typically these networks would be provisioned for the peak traffic.

As shown here, in the in the figure below, you can see here the network is provisioned along this particular peak traffic. Now, you know that over a particular time, if you see how much is the percentage, which will be generating this particular peak traffic, so most of the time the traffic, which is generated is not peak. Therefore, this particular bandwidth of most of the time is underutilized. All the bandwidth is provided using the peak provisioning, but the peak traffic is not always consistent. It is coming in the form of a bus, so most of the time the bandwidth is basically underutilized.

So, now that can be seen here using the mean usage of the network. So, if you see the mean usage of the network, then we can see around the mean, the entire traffic is basically around the mean that is the mean the peak to the mean utilization is only 2.17. Hence, it is highly inefficient way of using the expensive resource that is the bandwidth.

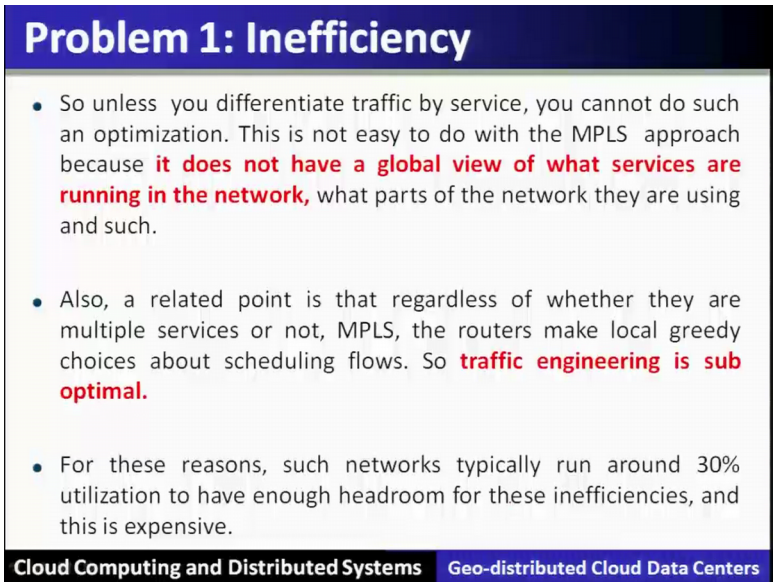
(Refer Slide Time: 28:16)



So, most of the traffic, now we will see into the more details that reveals that most of the traffic is actually the background traffic, with some latency sensitive traffic as well in the peak traffic. Even in the peak, if you see in the details, we will see that it is having two types of traffic. One is the background traffic, the other is non-background traffic. So, non-background traffic is basically the latency sensitive traffic, so, this is the latency sensitive traffic is to be provisioned. And this particular bandwidth to support this latency sensitive traffic cannot be compromised, and only that much of provisioning should be ensured.

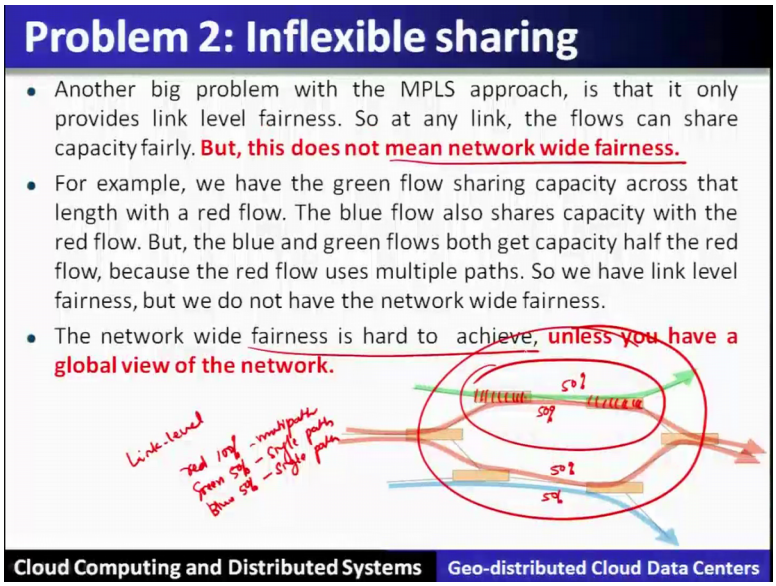
As far as, the background traffic is concerned, which is not latency sensitive and that can be basically fill the gap with the background, where the latency is not sensitive. If we see the traffic using these two classifications, then further there is a possibility of improving. So, here on the right side, if you see that the peak before the adopting, and we can further so that means, you can see here the peak is for the latency sensitive traffic. So, what we can see here, this is the provisioning, which is being allowed, and rest are all can be is filled by the background traffic, which is not latency sensitive. So, it can be spread, so this way there is a possibility of achieving high utilization with the software driven wide area network.

(Refer Slide Time: 30:21)



So, unless you differentiate the traffic by the service, you cannot do much on the optimization of the bandwidth. So, this is not easy to do in the MPLS approach, because it does not have the global view of what services are running in the network, what part of the network they are using and such. So, also a related point is that regardless of whether they are multiple services or not, MPLS routers make greedy choices about the scheduling of flows. So, the traffic engineering is suboptimal with these reasons, such networks typically run around 30 percent utilization to have enough headroom for these inefficiencies and this is also going to be very expensive, due to the inefficiency.

(Refer Slide Time: 31:08)

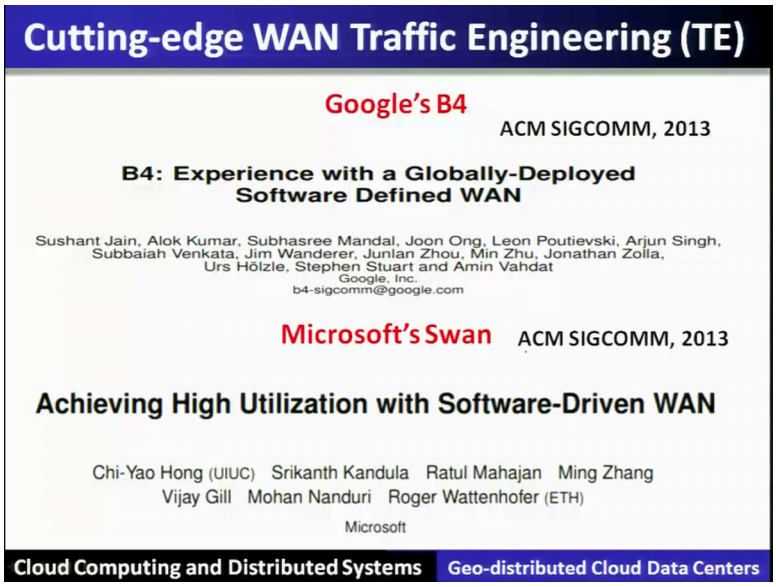


Another problem is called inflexible sharing. For example, here in MPLS, you can see the link level sharing for example the link level sharing means, as far as the flow, the green flow is concerned is using these two kind of bandwidth sharing the common bandwidth across these two devices, with the red one. So, 50 percent of the bandwidth will be given to this one green one, and 50 percent of this bandwidth will be on this flow to the red one.

As far as red is concerned, it is also using another path. And in this path 50 percent, it will get as far as the bandwidth. As far as the blue is concerned, so as far as the red one is concerned; red one will get 100 percent bandwidth using multipath; and the green flow will get only 50 percent, due to the single path; and the blue also will get 50 percent, due to the single path.

Now, with the link level; link level does not ensures the global view. Hence, it does not cover the network wide fairness to ensure the network wide fairness. Here, it requires the complete information or the complete picture of the global view. So, the network wide fairness is hard to achieve unless, you have the global view of the network. Therefore, the sharing is inflexible that means, at means that the link level sharing is not achieving the network wide fairness, this is another problem.

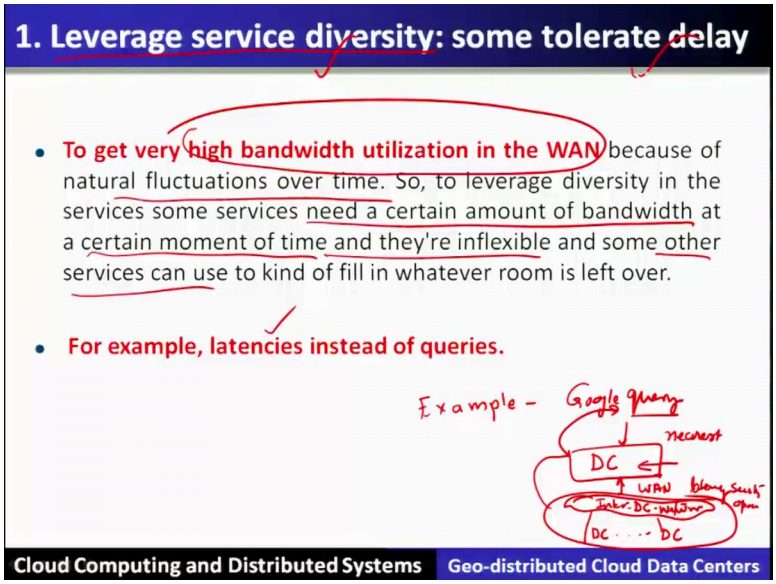
(Refer Slide Time: 33:06)



So, now with this traditional approach of using MPLS has two main problems. One is the inefficiency in the bandwidth utilization; the other is the global view of network wide

fairness is not ensured. Now, using this particular two problems, now let us see the cutting-edge solution of using wide area network traffic engineering in the modern times that it is the recent times. So, Google B4 has shown his way of solving these problems using software defined wide area network. We will see how these problems, Google's B4 is going to solve. We will also see the Microsoft swan, which will also showcase its particular software defined wide area network for achieving high utilization of the bandwidth.

(Refer Slide Time: 34:15)



So, let us see in the newer approaches, what are the main points for the design, which they consider, and then how they are going to address in the implementation in their cloud data center or how they are managing the wide area network for inter-data center networking.

So, the first one is to leverage the service diversity, let us see what does this means to leverage the service diversity. Now, here we can see that to get very high bandwidth utilization wide area network. We have to basically see that there are some fluctuations, which are happening over the time in different diversity of the services. For example, some services requires a certain amount of bandwidth at a certain moment of time and they are inflexible, whereas some other kind of services, who are not very rigid about the requirement of the bandwidth over a particular time can fill in wherever the rooms are

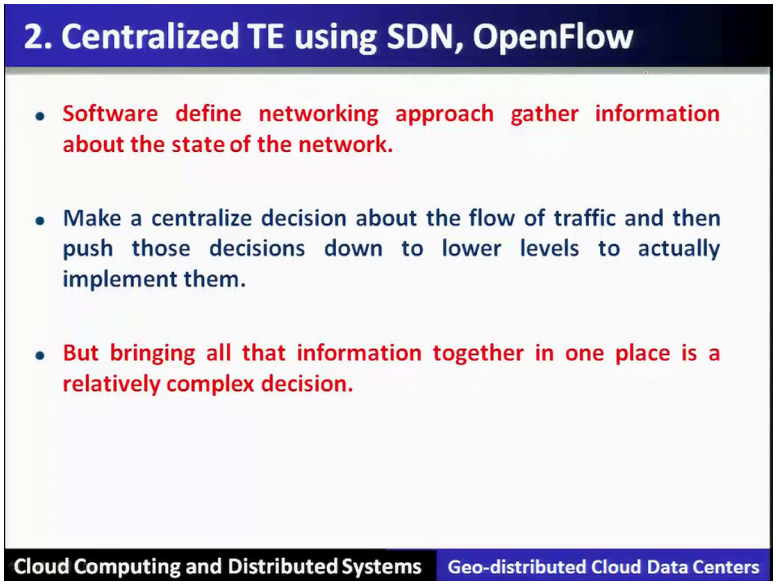
available after being allocated to the services, which are very stringent requirement of the bandwidth at a particular moment of time.

Take the example that in a Google search engine, if there is a query, this particular query will now go to a data center, nearest data center. Now, that nearest data center; does not have the information about that particular query. So, this data center in turn will flood to the inter-data center network over other data centers. Now, this particular traffic, which is now generated to know or to satisfy the query has to be done very efficiently or in a within a particular time. So, this is latency sensitive operation over a wide area network.

So, this operation has to be immediately addressed, so that this particular response can be basically given. So, they are called latency sensitive queries, so which basically are need to be addressed they are inflexible. So, the bandwidth has to be allocated, whereas maybe some other kind of traffic may not be that latency sensitive. So, therefore there is a service diversity across different services, so that is to be leveraged for the bandwidth utilization in the wide area network. So, this is the first way.

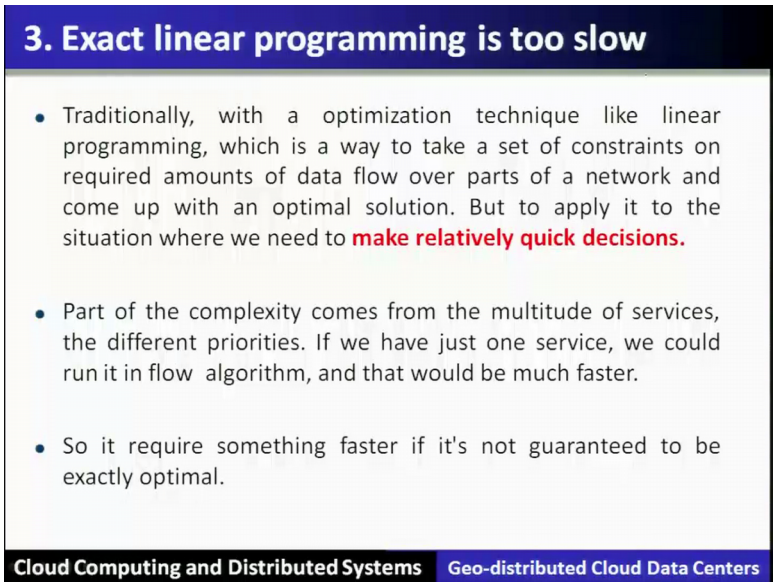
Now, the second is called the tolerance that is the delay. So, once that means, different services will ensure how much tolerance is basically can be basically allowed for a particular application. And based on that different services can be classified, and the bandwidth can be put on the utilization according to the criteria.

(Refer Slide Time: 38:28)



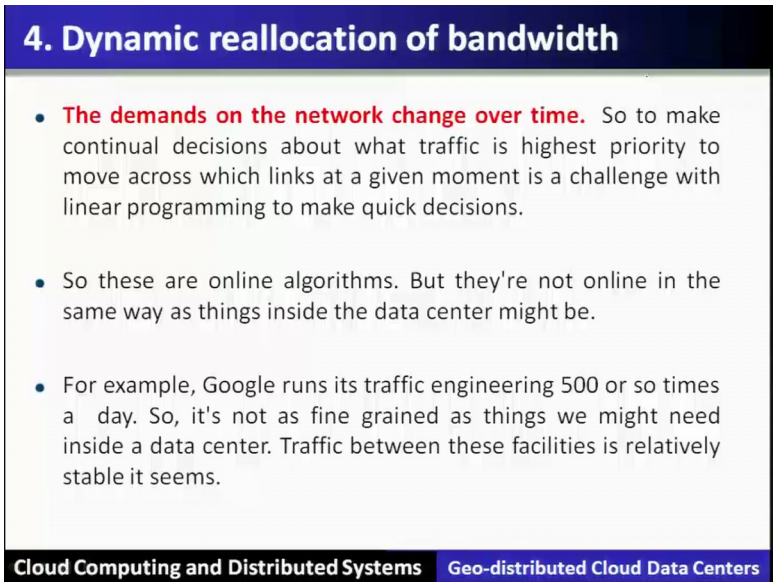
Now, second one is called centralized traffic engineering using software defined networks. Now, here the software defined networking approach will gather the information about the state of the network. So, there will be a centralized decision about the flow of traffic, and then push these decisions down to the lower level actually, implement them done. But, bringing all this information in one place is a complex decision, why because in its distributed system.

(Refer Slide Time: 39:03)



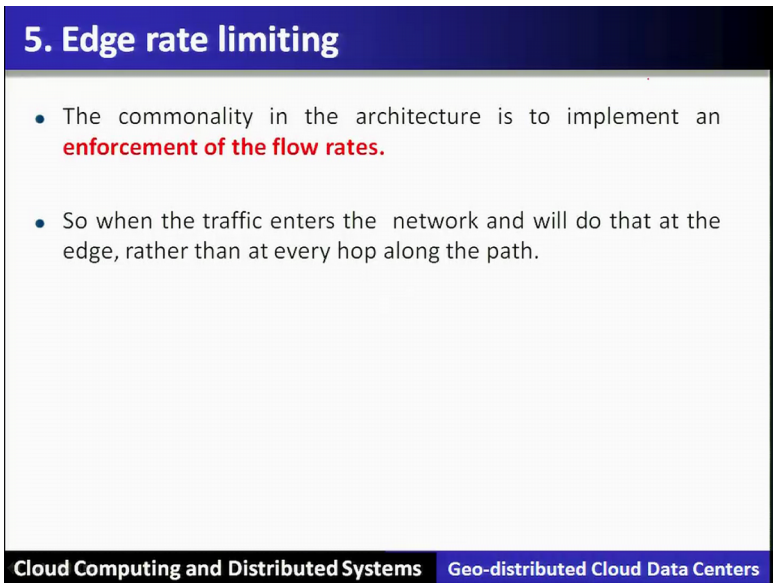
Now, third one is called exact linear programming will be here, what is slow for example, if the information is collected and optimization with all the constraint is applied through the linear programming, it requires time for the optimization problem to be solved. So, here the situation is that it has to make a quick decisions. So, the part of the complexity comes from the multitude of the services, with different priorities. So, if we have just one service, we could run it in a simpler method. But, the scenario is quite complex, it requires a different kind of different schemes, that means a complicated optimization techniques are required. However, it is required something faster, if it is not guaranteed to be exactly optimal to ensure to make the quick decision.

(Refer Slide Time: 40:05)



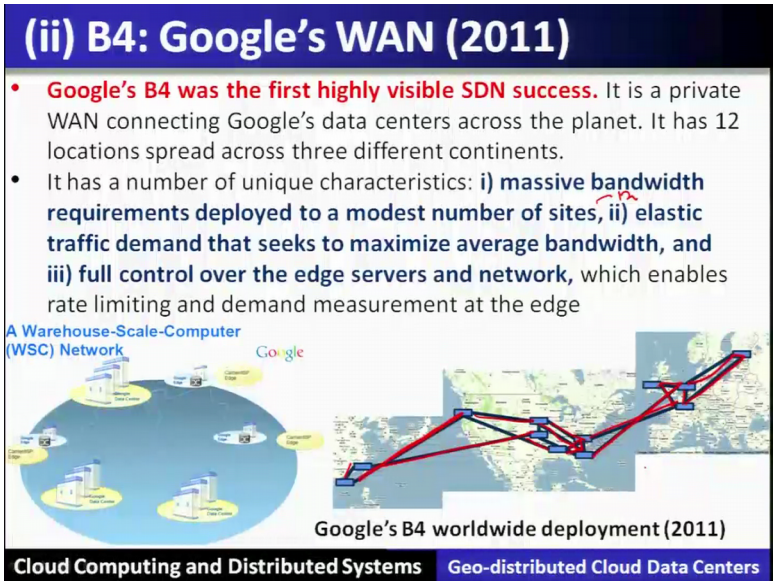
Now, the fourth one is called dynamic reallocation the bandwidth. The demands from the network changed over the time. So, to make the continual decisions about the traffic is the highest priority to move across, which link at a given moment is a challenge with the linear programming to make this particular decisions. These are online algorithms. But, with the data center they are not online they are not online in terms of fine grain. But, they are doing this traffic engineering 500 times in a particular day, and it is not the fine grained as the things inside the data center.

(Refer Slide Time: 40:47)



Now, fifth one is called edge rate limiting. The commonality in the architecture is to implement an enforcement of flow rates. So, when the traffic enters the network and will do that at the edge only, rather than at every half along the network; as, we have seen an MPLS that ingress and egress there are two routers, which has to deal with all the flow or a traffic engineering.

(Refer Slide Time: 41:16)

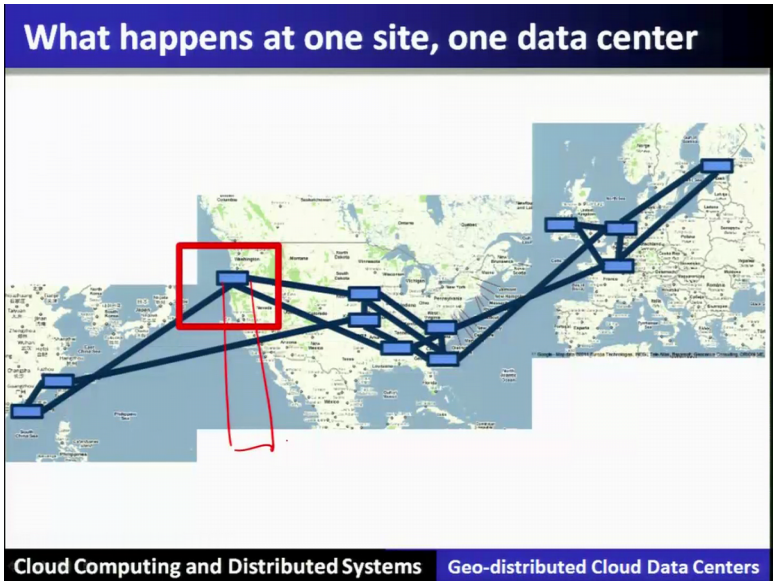


Now, let us see how these aspects are covered in Google's wide area network that is B4, as far as 2011 figure is concerned. So, Google's B4 was first highly visible software defined networking, which is applied to the wide area networks. So, it is a private wide area network connecting Google's data center across the planet, which is spread over twelve locations and in three different continents. It has number of unique characteristics first is the massive bandwidth requirements deployed to the modest number of sites, so that means, here how many sides are there the 12 different sites. So, lot of traffic here is seen across 12 different sites.

Now elastic traffic demand that seeks to maximize the average bandwidth and full control over the edge servers and the network; which enables the rate limiting and demand measurements at the edge; so, here you can see in the picture, there are 12 different locations, and they are connected with the high bandwidth links. There are few edges, which will basically see lot of traffic movement across inter-data center networks.

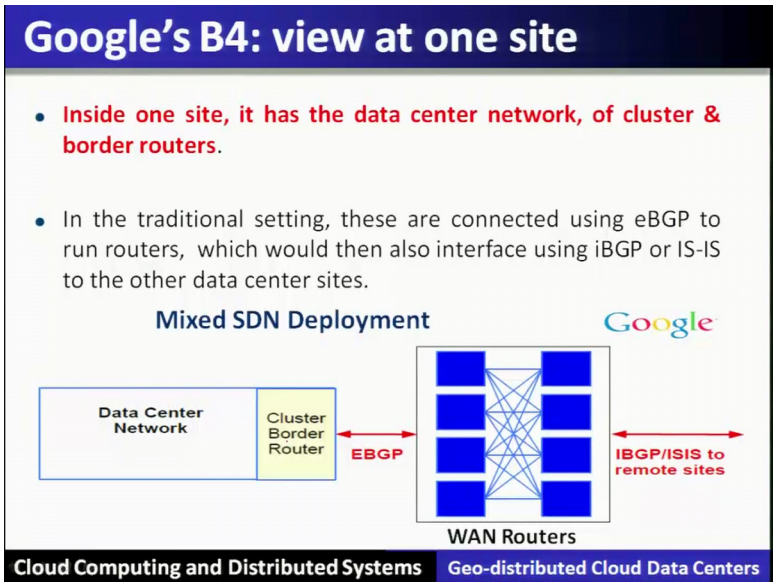
And this is all owned by the Google, so let us see how they have exploited this use of optimizing the wide area network.

(Refer Slide Time: 43:05)



Now, let us see what happens at inside, one data center. And that will be basically without laws of generality at all other data centers will manage.

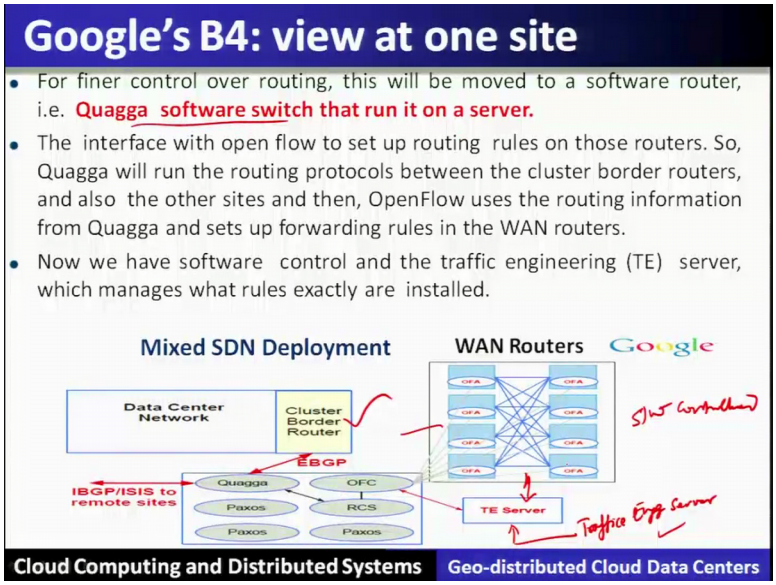
(Refer Slide Time: 43:19)



So, inside one data center nothing but a cluster and the border routers or a cluster border routers. So, this is a cluster border router in one data center, which is connected to the wide area network using EBGP. So, cluster border router using EGP is connected to the

wide area network routers, and which intern using IBGP and IS-IS, it is connected to the remote sites. So, in the traditional settings these are connected using EBGP to run the routers, which would then interface with IBGP or IS-IS for other data centers.

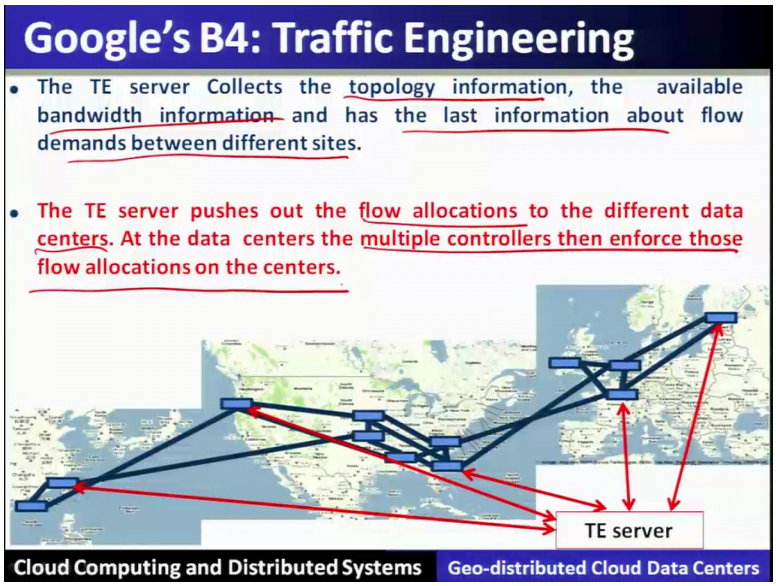
(Refer Slide Time: 44:09)



So, for final control over the routing, this will be move to the software router, meaning to say that this particular WAN router will be now the software controlled router. So, the software is basically pulled at one place and that is called a traffic engineering server, which will contains the entire logic. And so this particular software is are the Quagga software switch that runs on this particular server.

So, the interface with the open flow to setup the routing rules on these routers. So, this is an open flow routing which will; so the Quagga will run the routing protocols between the cluster border routers and also other sites. And open flow uses the routing information from Quagga and sets of the forwarding rule in the wide area network routers. Now, we have a software control and the traffic engineering server, which manages what exactly are installed over here.

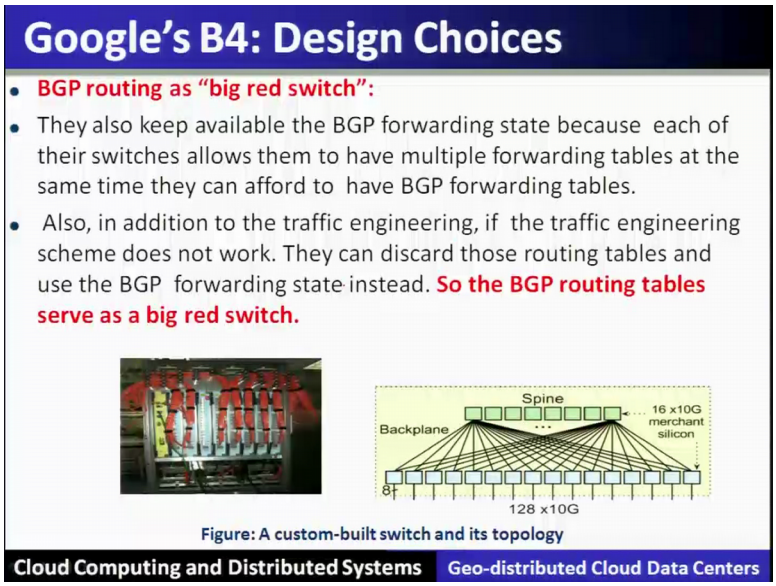
(Refer Slide Time: 45:41)



So, let us see the details of the traffic engineering. So, traffic engineering server collects the topology information, available the available bandwidth and the last information about the flow demands between different sites. So, there are three different kind of information, which traffic engineering will collect from all different locations, they are the topology informations and available bandwidth information that was already done in MPLS. But, besides that the last information about the flow demands between different sites that also is collected.

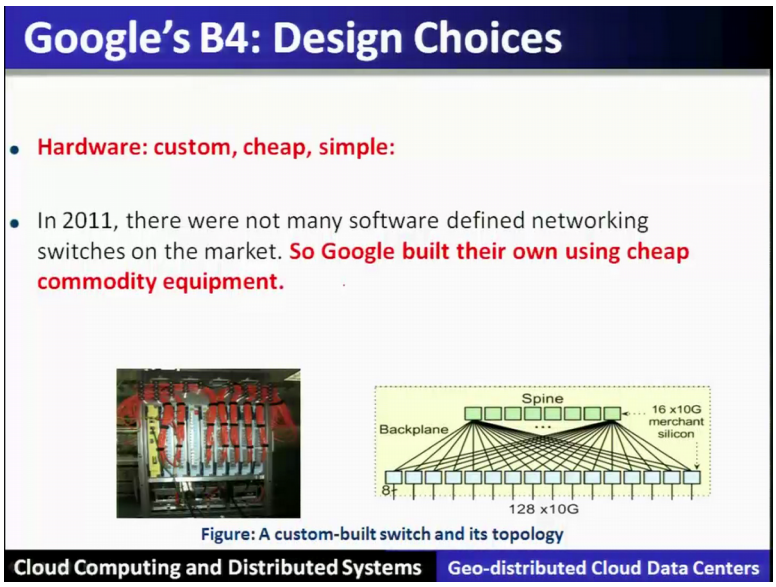
So, the traffic engineering server pushes out the flow allocation to different data centers, that means, after collecting it will compute the flow allocation, these flow elevations will be pushed back to the different data centers. So, at the data centers these multiple controller then force these flow allocation to the centers.

(Refer Slide Time: 46:49)



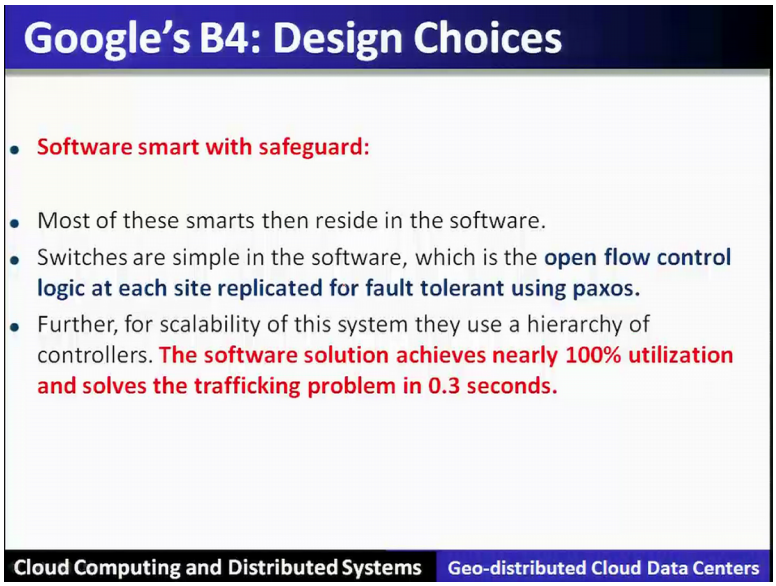
Now, if we see that the entire process of doing, this is nothing but it looks like a big switch, which does this kind of BGP routing across all the data centers, that is the wide area network implementation using software; software for wide area network.

(Refer Slide Time: 47:08)



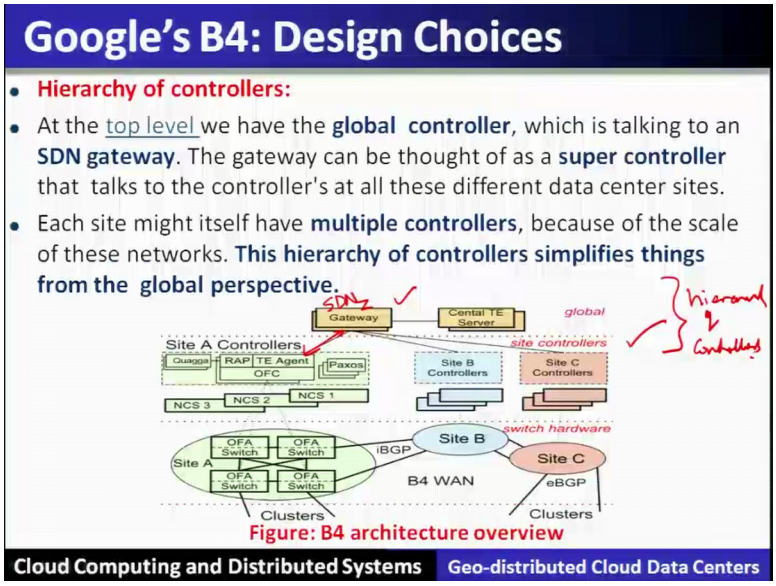
Now, here we can see that doing this will not required to have to do this will require a cheap commodity equipments.

(Refer Slide Time: 47:24)



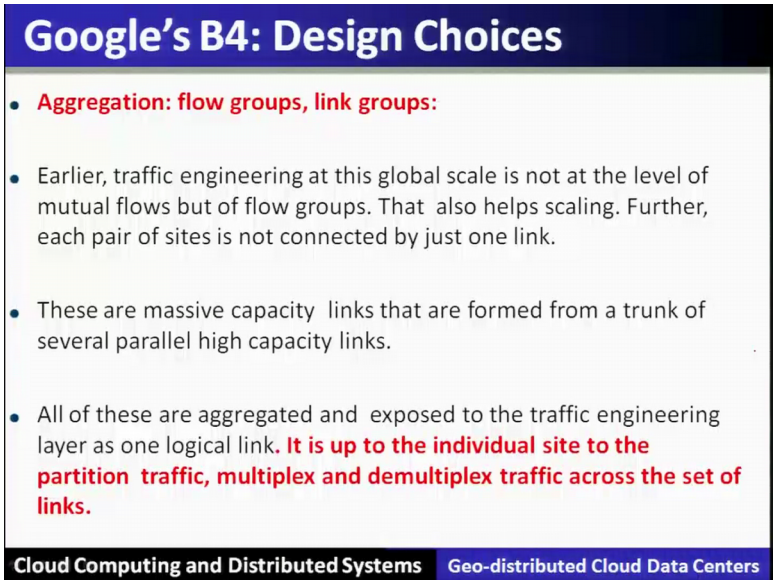
Because, most of the things are done in the software that means, software all the controls are done through the software. So, the switches are or simple, which are nothing but and open flow logic at each site replicated for fault tolerant using paxos. Further, the scalability of the system is ensured by hierarchy of controllers and the software solution achieves 100 percent utilization and solves the traffic problem traffic engineering problem in 0.3 seconds.

(Refer Slide Time: 47:58)



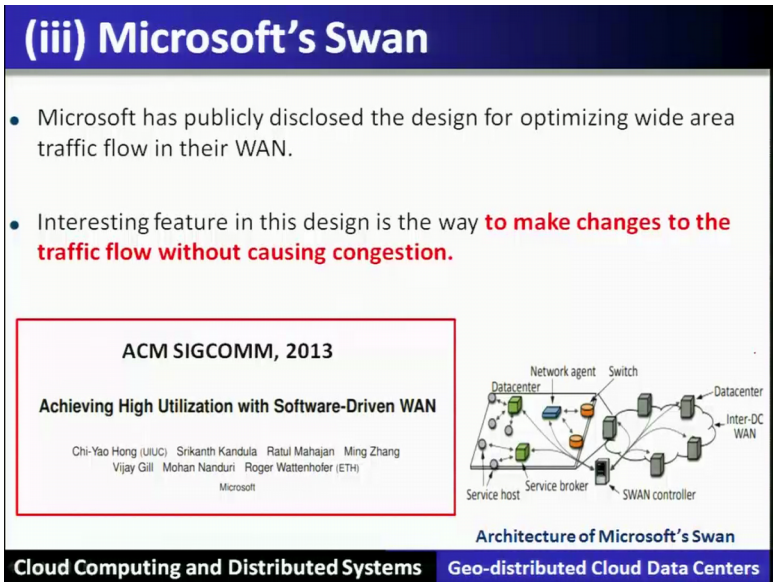
Let us see the hierarchy of controller means that at the top level, we have the global controllers. So, all these are the global controllers for different sites, which is talking to an SDN gateway. So, they are talking to the SDN gateway. So, the gateway can be thought of a super controller that talks to the controllers at all different sites of the data center so, this forms and hierarchy of controllers. Now, each site might itself have multiple control, because of the scale of the network. This hierarchy controller simplifies the things from global perspective.

(Refer Slide Time: 48:54)



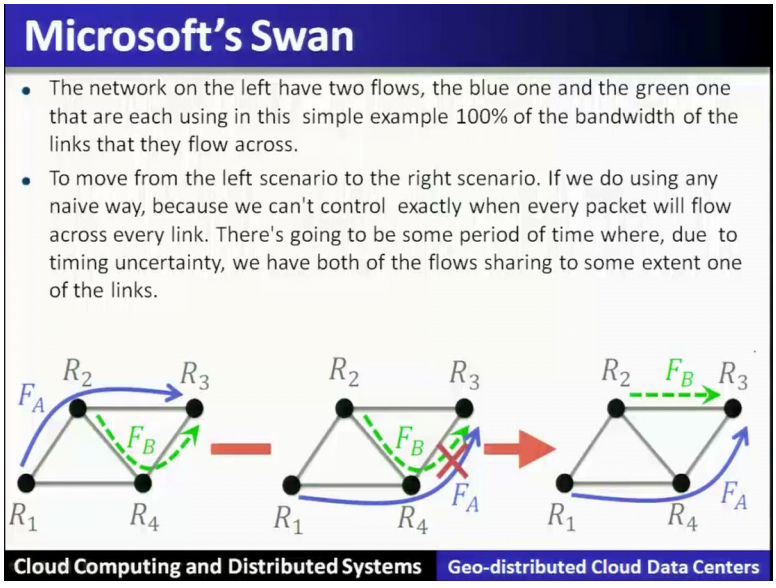
Now, then this particular hierarchy of controller will do the flow groups will form the flow groups and also will ensure the link groups. So, the traffic engineering will be confined or will focusing on the flow groups not at a particular flow. So, the traffic engineering will be done at this globally scale and it is not at the level of mutual flows, but of the flow groups. So, these massive capacity link that are formed from the trunks of several high capacity links are now going to be utilized using this particular flow groups.

(Refer Slide Time: 49:46)



Now, here we will see another approach, which is called Swan approach in the Microsoft. How, they are going to utilize the bandwidth using software defined wide area network. Here is important feature is that to make the changes to the traffic flow without causing the condition.

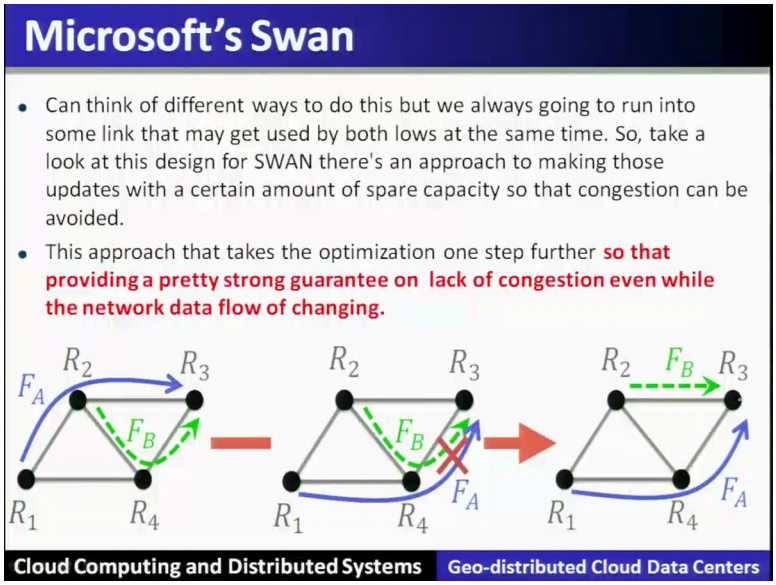
(Refer Slide Time: 50:07)



So, let us see the broad idea over here is that there are two different flows F A and F B, which are shown with the green color. They might be sharing some of the common elements of the routing network. Now, instead of that we do not know at what time they

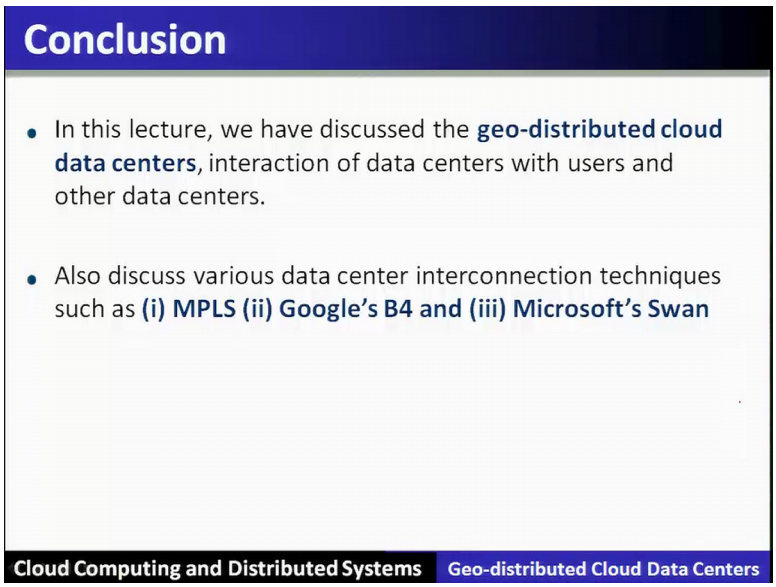
will be sharing the bandwidth. So, instead of that if these flows are routed through two different disjoint paths, then they may be independent, and can share the bandwidth as much as they can without.

(Refer Slide Time: 50:53)



So, doing this kind of global flow control, it is also possible to reduce without means without congestion, it is going to solve the bandwidth utilization issue.

(Refer Slide Time: 51:13)



Conclusion, in this lecture, we have discussed geo-distributed cloud data center; that is the interaction of data center to realizing to ensure the services of the applications to the

users. We have also covered the data center interconnection such as the traditional approaches used that is called MPLS and also the newer approaches, which is in the form of Google's B4 and Microsoft's Swan.

Thank you.