DARPATech, DARPA's 25th Systems and Technology Symposium
August 9, 2007
Anaheim, California
Teleprompter Script for Dr. Timothy Gibson, Program Manager, Strategic Technology Office – Networks Presentations

DARPA Moving & Storage: We Do the Heavy Lifting

» **TIM GIBSON:**
If you want the network as a whole to do more,
the network --
as a piece of technology -- must do more work.

In the past, this has meant increasing the amount of work done by the network routing devices, particularly those closer to the network core.

Devices at or near the network's edge have been unaffected.

Over time, the cost of the core devices has risen, ostensibly because they are doing more work.

Conversely, the cost of devices at or near the edge has fallen while performance and capability has risen.

We can continue down this path, but I suggest that we more evenly distribute the work across the network.

Before I tell you where
we can go from here,
let's spend a little time looking at how we got where we are.

The network protocols we use today to pass data, IP, TCP, and their ilk are essentially unchanged from when they were designed for the ARPANET forty years ago.

Suffice it to say that any new PDA or MP3 player has more memory than the largest 1960’s computer.

CPUs were also anemic, with processor speeds normally below one megahertz until the late 1970s.

An unspoken -- but key -- factor behind the design of the ARPANET protocols was the need to limit memory and processing requirements.

For example, today’s networks would be more secure if we had an agent that brokered connections between devices.

Any brokering device needs an enormous amount of memory to track pending connections,
and while memory is cheap now it wasn’t then.

So while brokered connections were seriously discussed during the TCP/IP design period, the cost of the brokering devices was so high that the idea was discarded, leaving us with today’s insecure system.

The key to making tomorrow’s networks better:
faster, cheaper, and more reliable is to use the things we have now that weren’t available forty years ago:
cheap and abundant memory, cheap and abundant processing power.

Now this sounds all well and good,
but is it really possible?

I think it is.

In the Control Plane project the challenge was very simple:
given abundant processing power and memory at the edge,
can you provide edge devices with just a little information about the network and increase data throughput?

The goal was to increase speeds by a factor of ten, while remaining compatible with the Internet Protocol,
versions 4 and 6.

This was an ambitious goal when the collective efforts of the entire network researching community have made TCP/IP five times faster in the last forty years.

So what happened?

At the end of the first phase, 18 months, we had improved performance over twenty times.

We had a new router with lots of memory that handles packets differently than conventional routers.

This different approach to handling packets allows the new routers to be made from field-programmable gate arrays, that is, FPGAs.

The result is a high-speed 50 Gbps router that is five times cheaper and uses 80% less power than conventional designs.

How is all this possible?

Because we are using the tools we have today to rebuild the foundations we built forty years ago.

“What else can be changed?”

Building on the idea that we can make things faster by exploiting cheap memory and knowledge of the networks, we need to look at how packets are routed on the networks.

Currently, packets are routed using a variety of protocols.

OSPF, short for Open Shortest Path First,
routes packets over short to medium distances.

Now think about that name and what it means: Open Shortest Path First.

It does not say best path, or fastest path,
or perhaps a path with the some characteristics,
just the shortest path.

While OSPF was fine forty years ago when the average ARPANET communications speed was 56,000 bits per second, today's network speeds have increased to the billions of bits per second.

You would think we should be a little more sophisticated by now.

Well we aren't --
but we can be.

We should develop routing protocols that allow computers to choose the best path for their traffic.

To enable this we need to expand the current route discovery and route dissemination protocols, as well as the network data protocols,
to regularly use multiple paths if they are available.

Letting computers transmit data on multiple paths means other changes.

Because we currently only send data on one path we always receive packets in the order they are sent.

If a packet is missing, clearly it was lost -- and it must be retransmitted.

But if you transmit on multiple paths, packets will regularly arrive out of order, confusing everything.

So the receiving computer has to decide when packets are lost and when they are just slow.

This requires more memory but that's okay because memory is cheap.

While improving local routing protocols is
a fairly hard problem,
wide area routing poses a more difficult problem.

Wide area routing is done by a different set of protocols, chiefly the Border Gateway Protocol or BGP.

BGP does some route discovery ala OSPF,
but also uses policies to decide where to send things.

Policies can be as simple as preferring to send traffic via a specific carrier because they charge less.

Developing new methods for route discovery for wide area networks will be difficult because of the installed legacy base,
but it is possible.

Given the pricing changes for memory and processing in the last forty years, we should not only be making networks faster and more reliable, the equipment should also cost less.

The last thing I want to talk about is interconnecting an optically switched backbone with metropolitan level IP networks.

I am not talking about an optically routed network, but an optically switched network.

The DARPA CORONET program is developing one of these networks.

It is a fully wave division multiplexed network with a capacity of 100 terabits per second.

It is also an extremely dynamic optical network, requiring 100 milliseconds to setup or teardown a link.

While we don’t have one of these networks now,
we will in a few years.

The element I want to focus on is the device that interconnects the two networks.

It will talk IP on one side and fiber|wavelength pairings on the other.

The problem for this device is that existing routers don’t really know how to get packets from one place to another.

They only know the general direction they need to shove a packet to have it eventually arrive at its destination.

On the other hand, an optical|IP interconnection device must know the exact fiber|wavelength pair to reach every other interconnection device in the world.

This is an immense amount of information to discover and keep current, and this workload will make the interconnection device very expensive.

However, if we modify the addressing scheme we can make the device's job easier and have it cost less.

Suppose we assign addresses by some type of hierarchy.

For example, we can use a geographical hierarchy.

In the example on the screen, there is a machine at DARPA with a geographic name trying to connect to another geographically named machine in the UK.

I admit these names are a little long, but we can have the people type in the first part just like they do now and have the computers add the rest.

To decide which fiber|wavelength pairing to place a packet, the optical|IP interconnection device only needs look at the geographical address from right to left.

As soon as it gets to a field that is not the same as its own address,
it knows now to send it
to that place.

In the example,
the destination address of Europe is different from America,
so the interconnection device knows it needs to send the packet to Europe.

This approach requires the interconnection device to know much less information.

It only needs to know some or all of the paths that lead to Europe because all of them eventually get to the United Kingdom.

Overall, it just needs to know how to get to the other continents,
Mexico and Canada,
the other states, and the other towns in Virginia.

Clearly, that is easier than keeping track of the exact paths to everywhere on the planet.

With this scheme you merely place a packet on the next available path going in the general direction and it eventually arrives, which is what we do now.

So by modifying the addressing scheme the interconnection devices can be simpler.

They can also be cheaper and faster -- even though they are doing more work.

This type of routing -- which is probabilistic -- lets packets go different directions and arrive out of order -- but we know how to solve that problem.

Memory.

We place the burden for reassembling and properly ordering the data on the end user devices with plenty of memory and processing power.

The core does less,

the edges do more,
the work is more evenly distributed, and the overall network is faster, better,
and cheaper.

Faster, more reliable, and cheaper.

At the beginning of this speech I said that the key to making tomorrow's networks --
better is to use the things we have now that weren't available forty years ago: memory and processing power, both of which are now cheap and abundant.

We are making some changes with these new resources.

The performance and reliability improvements we are seeing are,
quite frankly, very nice.

However, there is much more we need to do,
to make the Internet
we know today
ready for tomorrow.

For this new work we need your help --
and more importantly -- your ideas.

I thank you for your time and look forward to hearing from you.

I will be followed by
Dr. Larry Stotts.