appeared in 2003

Internet moves into high gear

S.Ananthanarayanan

Scientists at Caltech have developed ‘fast TCP’, that could radically speed up downloads on the Internet, says S.Ananthanarayanan.

This would mean better communication and more products, like movies on line, becoming available. All this may load the Internet with more traffic, but fast TCP would also result in better use of the same resources.

How regular Internet works

On the Internet, fantastic data gets transferred between a large number of users and at the same time, over the same network. This is possible because a pair of points communicate over all possible paths between each other, even while all other pairs of points on the network are using the same paths for their communication as well. This, in turn, is achieved through a scheme called the Transmission Control Protocol or ‘TCP’.

In communication via telephone or telex, a single line has to be reserved for each channel and whole messages are conveyed, completely, over this line. In TCP, a single path is not reserved and the message is also not conveyed in one piece. The message is ‘chopped’ up into pieces of around 1500 ‘bytes’, or pieces of about 100 words. Each of these pieces is packed into a ‘packet’, along with a ‘header’, which states what the packet contains, which terminal on the system sent it and to which terminal it is addressed and also the serial number of the packet being sent out.

These packets are not sent down a particular line that goes all the way to the destination, but are just ‘pushed on to’ the network for the TCP methods to help them get to where they are addressed.

When the packets get to the destination, through different routes and generally not in the same order, the destination computer arranges them by the ‘serial numbers’ in the ‘header’ and reconstructs the original message. This is the way a message uses all possible paths to get from one point to another, which is generally more efficient than expecting just one line to carry all the information, especially if the line were to serve others as well.

What if there is an error?

TCP takes care of errors by asking the destination to send back an ‘acknowledgement’ every time a packet is received. The acknowledgement is not just that the packet was received, but also that some of the details of the packet, as described in the ‘header’, have been checked and found correct. In this way, the sending computer waits for some time for the acknowledgements and if some of them do not come, it sends those packets again. If the failures persist, the transmission may be abandoned, due to a ‘bad line’.

But before just abandoning the message, the sending computer tries to accommodate to the poor communication channel by sending packets with longer gaps between successive ones. That is, by sending the message ‘slower’ than at first.

High traffic is more errors

When a large number of users are competing for paths in the network, packets need to follow longer routes and also need to ‘wait’ between legs of transmission. This causes slower speeds and also more errors. As we have seen, more errors slow down the speed of transmission. Transfer rates can then slow down to a ‘crawl’.

In ‘Fast TCP’ the sending computer uses the same infrastructure of the Internet to draw a ‘larger’ picture of the routes and roadblocks, something like a motorist who scans the road a kilometer ahead, instead of just till the car going in front. The computer takes into account the time gap between a packet and its acknowledgement, and arrives at an ‘optimum’ transmission speed for the network. As before, packets that are not acknowledged are sent again, and so on, for the integrity of the messages, but the ‘knee jerk’ reaction of just slowing down is replaced by a more intelligent strategy.

What it calls for is different software, and some hardware, on the sending computer. Trials have given speeds four times as fast.