

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

Introduction

The largest innovation required in the Internet or in large WAN networks is not to increase the network speed, but to improve the control time of the flow control. Network delay is mainly dependent on flow control, not line speed, and the largest problem in the Internet today is Web access time. Flow control in the Internet today is based on TCP. Operating from the edge of the network, with no knowledge of the state of the network except packet loss, TCP cannot react fast enough for today's network size and speed. TCP's life has already been overextended to 15 years and a major improvement is required in flow control in order to reduce network delay, improve Web access time, and reduce network cost. Explicit rate flow control, currently the standard for ATM, operates from inside the network detecting the maximum rate for each flow that will not congest the network, and sends this information to the sources at near the speed of light. The impact of this new flow control is a 100:1 improvement in the time to control the source, which if implemented in the Internet, would result in the following major improvements:

- Reduction of Web access time by 100:1
- Reduction of network delay variance by 100:1
- Reduction of the switch memory required in the Internet by 100:1
- Near elimination of packet loss in the Internet, improving line utilization by up to 20%

There is no simple fix possible by just improving TCP. The network must participate in order to make any significant improvement in control time. Until new flow control, with the capability of explicit rate, is incorporated into the Internet, Web access will remain extremely slow and the true potential of the Internet will never be realized.

TCP - The Current Internet Flow Control

The Internet Engineering Task Force (IETF) supports TCP/IP. TCP/IP is used in the Internet and many LAN's. It consists of several parts, but the one that manages the rate of data flow into the network is TCP. TCP operates in the following way. It starts sending data at a very slow rate (slow start) and watches the acknowledgments from the destination to see if any data is lost. If none is lost, it speeds up. It keeps speeding up until data is lost at which time it decreases its rate. It keeps decreasing its rate until no data is lost. It then increases its rate again, oscillating like this continually to track the network capacity.

The most critical problem today in the Internet is the slow control-time of the TCP flow control. TCP ships about one second of data into the network before congestion can be stopped. This one-second time caused by the Round Trip Time (RTT) of data packets over the network times the number of steps TCP takes to adjust to the network capacity. The RTT is primarily due to the size and speed of the network since each switch in the connection path adds delay proportional to its speed. As the Internet has grown, the number of routers or switches in the connection path has grown to 15-30 and the trunk speed has increased to 622 Mbps. This results in an RTT of 100-200 ms today, and growing as the Internet grows. So long as TCP operates only in the end-stations its operation cannot be substantially improved. When the router or switch memories in the data path all have sufficient memory, TCP works, although slowly. However, the memory required is the delay-bandwidth product of TCP's cycle time and the full input bandwidth of the router or switch. With smaller memories, the packet loss soars, which not only hurts goodput (successful data transfer rate) but seriously increases all data transfer times.

The worst problem caused by the long control time of TCP is that the average delay through a switch or router is proportional to the control time and the switch speed. As the number of switches and/or their speed increases, the RTT also increases and thus the control time increases. Thus, the slow

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

Web access time we experience today will not only get worse as the Internet grows, it will feed on itself and grow exponentially with network size. Explicit rate flow control not only reduces the control time by 100:1 today, but it eliminates the interdependence of average switch delay and control time.

The IETF has not yet considered revising TCP. In fact no study has been done by the IETF on flow control because everyone seems to believe, "if it worked in the past, it will continue to work". Nothing can be further from the truth in an environment like the Internet where growth is a factor of 5 per year. TCP must be replaced with a new flow control as good as explicit rate flow control as soon as possible. This can be done by using ATM (or Cells In Frames) with explicit rate flow control or by the IETF revising TCP/IP to incorporate explicit rate flow control.

ATM Explicit Rate Flow Control

In May 1997 the ATM Forum adopted as part of its traffic management protocol, TM 4.0, explicit rate flow control as the new specification for Available Bit Rate (ABR) traffic. ABR is intended for ATM data traffic, although the previous mode of operation for data, Uncontrolled Bit Rate (UBR) is also available. ABR has two modes of operation, Explicit Rate for the future, and Explicit Forward Congestion Indication (EFCI) to maintain compatibility with the past. EFCI, sometimes called relative rate, has about the same performance as TCP and thus has almost no added value.

Explicit Rate flow control uses Resource Management (RM) cells to carry the control information around the network. The source sends a RM cell at the start of each data burst and then an additional one every 32-128 data cells. In the RM cell there is an Explicit Rate (ER) field which the source sets to its peak rate. The RM cell is passed through the network to the destination, which turns it around and sends it back to the source. On its return, each switch reduces ER, if necessary, to the maximum rate which it can support without congestion. This rate is computed by dividing the available bandwidth by the number of active virtual circuits. Thus, when the RM cell returns to the source, it contains the maximum rate which it can send data at without congesting some switch on the path. The source then adjusts its rate for this VC to ER.

Since the RM cells are not tied to any data and only constitute 1-3% of the total bandwidth, they can be given extremely high priority by the network so that the control information can be returned to the source at near the speed of light. In the Internet today, this would be ten times faster than the data flow rates. Thus, the time to control the source rate can be cut dramatically. This is not possible for TCP since its control depends on data packet loss nor is it possible for EFCI since this depends on marking data packets.

Control Time Analysis of TCP and Explicit Rate

There are three major factors involved in determining the control-time of a flow control loop:

The propagation speed of the control signal. With explicit rate there is the option to give the RM cells priority so that they can move near the speed of light, not the speed of the data flow. With TCP this is impossible because the control signal is the loss of data from the data path. For the Internet this factor results in, on the average, a factor of 10 improvement in control-time.

The oscillation cycle. With explicit rate there is no oscillation and the source adjusts immediately upon receiving the first RM cell. With TCP, the source adjusts over typically 5 round trips times to the network capacity. This results in another factor of 5 improvement in control-time.

The distance which the control must travel. TCP must send data around the whole network to detect loss whereas Explicit Rate only needs to send the RM cell back $\frac{1}{2}$ to $\frac{1}{4}$ of the network round trip distance. This results in at least another factor of 2 improvement in control-time.

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

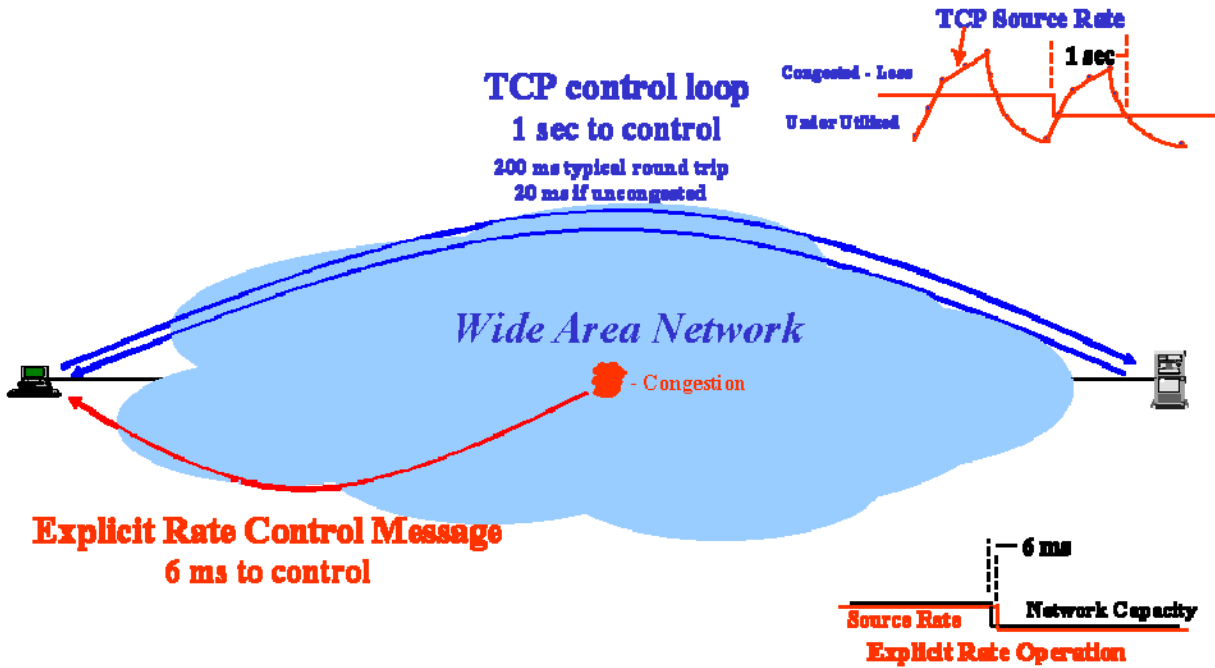


Figure 1. Comparison of TCP and Explicit Rate Time to Control

Overall these factors result in a factor of 100 in the control-time reduction from TCP to Explicit Rate. Comparison of Control-Times of Flow Control Options

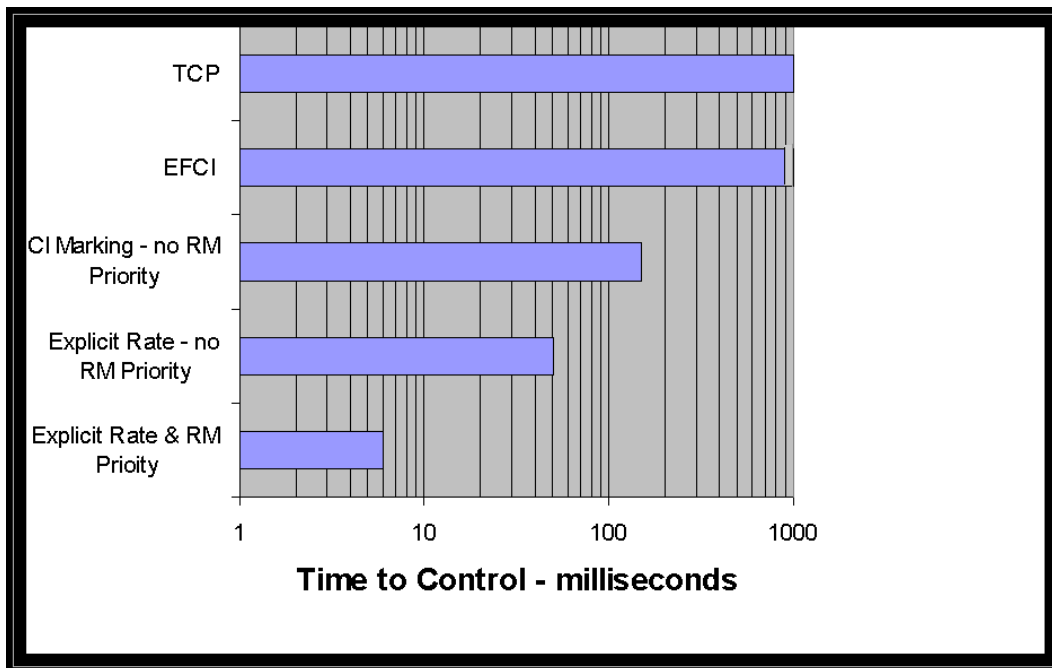


Figure 2. Control-times for Flow Control Options

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

The graph in figure 2 shows the control-time for TCP and several ATM ABR options: EFCI, CI Marking, and Explicit Rate. CI marking is not done today since to do it fairly is just as hard as Explicit Rate, but if done is somewhat better than EFCI marking. Explicit Rate is shown with and without RM cell priority, a critical feature that reduces the time by a factor of 10.

Impact of Control-Time Reduction from Explicit Rate For the Network Operator

The network operator can gain some of the advantages of Explicit Rate, even without extending Explicit Rate end-to-end between users. If a network or sub-network is operated with Explicit Rate and the Explicit Rate is terminated at the edge of the net, then network delay, data loss, switch memory, and transmission wastage can all be reduced for that network or subnet.

Network Delay Explicit Rate reduces the size of the data queues in each switch in the subnet by about 100:1. This is because the control time with TCP is 100 times longer than Explicit Rate and the network must buffer 100 times more data to maintain control. Also, with TCP each switch waits for its memory to fill up to a threshold before tossing out cells or packets whereas Explicit Rate switches compute the correct rate per VC even with an empty buffer, further improving the delay reduction. However, when the Explicit Rate network is not end-to-end, and is only a subnet, then the delay reduction is reduced since the first switch converting to Explicit Rate has the same delay as a TCP router. Thus, the effect in a subnet will result in a delay reduction of about 10:1 if there are 10 nodes in the path.

Data Loss Explicit Rate can be configured to achieve as low a cell loss ratio as desired. This is done by adjusting the Transient Buffer Exposure (TBE) parameter during signaling. Typically, it would be likely that the network operator would reduce the data loss to 10-12. Since TCP typically has loss rates of 10⁻¹ to 10⁻² this is a dramatic improvement. One major effect of this lower loss rate is that delay is dramatically reduced since retransmissions take lots of time.

Switch Memory Size Even though there tends to be statistical smoothing of the data transients at a switch for both techniques, the reduction of control time is directly reflected in reduced buffer size. This is critical because high-speed ATM switches cannot afford the same memory as the older software switches and routers had. A 100 Mbps router often has 12 Megabytes of buffer storage or about 1 second of buffering at full input rate. A 10 Gbps ATM switch typically has about 10 ms of storage at full input rate and could not afford to increase it to 1 second since the cost would be unreasonably high. Explicit rate flow control makes it possible for an ATM switch to use only 10 ms of memory and still support near loss-less data transmission.

The reduction of switch memory requirements by 100:1 results in a switch cost reduction of about 7:1. Today, the price spread between ATM switches and routers is more like 10:1 but that is partly due to memory and partly due to the greater cost of routing compared to switching.

Line Utilization Line utilization can be controlled to be much better with a faster control-time. In simulation examples (see Explicit Rate Flow Control) the improvement was 6-13%. However, there is a total tradeoff possible between utilization and queuing delay. One can fill memory with data to insure high line utilization but this increases the delay and buffer size required. There are some Internet sites which run their lines at 60% utilization to reduce delay and data loss. These sites could greatly increase utilization with Explicit Rate. Alternatively, other sites run their lines at 90% load with high delay and data loss. These sites could reduce their delay, data loss and memory requirements with

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

Explicit Rate. In general, a 10% improvement in line utilization will result from Explicit Rate in parallel with the 100:1 memory savings

For the End User

Once Explicit Rate has been established from end-to-end the user can gain other advantages. To gain these, TCP must be avoided or modified. To avoid TCP one can run a native ATM application. But to gain the broadest benefit, TCP should be modified so that when a VC has been established end-to-end with Explicit Rate, then TCP does not use slow start and lets the data flow be controlled by ER. If the VC path includes a router without ER, then TCP should continue to use slow start. This change to TCP is minor and a revised version could easily be downloaded if the user wants the benefits.

World Wide Web Access Today, TCP is used for all WWW activity. Each page access starts a new connection with TCP in slow start at about 2,000 bits/second. Since the typical page (not including graphics) is small – perhaps 5,000 bytes, then TCP typically takes 10 seconds to retrieve a page. If no loss is detected the rate might be increased, but most page access is done at slow start today. The graphics, being a larger block, may move TCP out of slow start, but even so many seconds are lost up front getting TCP up to speed. Thus slow start is the largest negative impact on Internet performance today.

The reason for slow start is that TCP has no knowledge of the network capacity until it has operated for several seconds, testing the network operating point. It would be catastrophic to the network if TCP were to start faster without any ability to determine the network status faster. The network would be inundated. However, with a flow control that could determine network capacity faster, the start rate could be made proportionately faster. This is because:

Start Rate = Buffer Size / Control-Time

Thus, since explicit rate is 100 times faster than TCP, the start rate could be increased to at least 200 Kilobits/second which would result in WWW page access in 0.1 second rather than 10 seconds. Such a change would be a major increase in performance for the WWW.

Reduction in Delay Variation Explicit rate switches have 100 times less delay variance the IP switches, down from 3 sec. to 30 ms. This dramatically improves the consistency of the service for all data applications and also improves the utility of the data service for highly interactive activities.

Cells in Frames (CIF) to Extend Explicit Rate End-to-End

Implementing Explicit Rate Flow Control in the core of the Internet will improve overall bandwidth utilization and performance significantly. However, the benefits of flow control extend only as far as the RM cell can travel. If it is too difficult or expensive to expand ATM from the ATM-ER core to the desktop, Cells In Frames technology can extend the ATM protocol including ABR-ER across legacy frame based media, to the desktop.

CIF is ATM with variable length packets on the lines and trunks. The CIF Alliance has specified a protocol which allows ATM to be embedded into various frame based legacy protocols (Ethernet & Token Ring), using only one ATM header for up to 31 cells from the same virtual circuit in a packet. The specification of CIF over PPP and Sonet is underway. A significant feature of CIF is that ATM can be transported to workstations without changing the legacy NIC card because the necessary processing is done in simple downloaded software "Shim" on the workstation. Figure 3 shows the placement of the Shim in the Windows software stack.

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

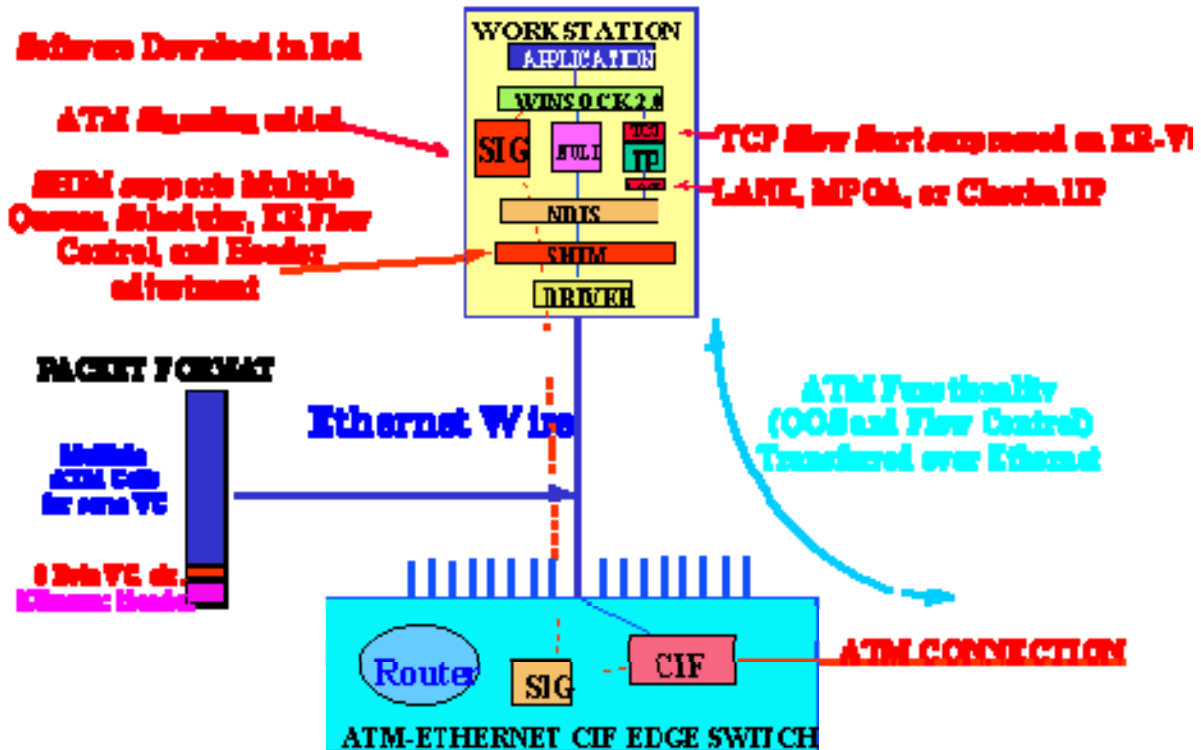


Figure 3. CIF Workstation and CIF Switch Configuration

CIF also permits the connection of workstations and servers with about the same expense as standard Ethernet. CIF is described in another paper, "Cells In Frames", and the specification is available "ATM Forum 96-1104, Request for Coordination of Cells In Frames Specification, August 1996".

ABR-Explicit Rate vs. UBR Simulations

There have been numerous simulations that apparently show that Explicit Rate ABR is no better than TCP over UBR. There are several flaws in these simulations which help explain why something 100 times worse is being made to look similar:

The simulations all presume that TCP is used with BOTH ABR and UBR. This is assumed because routers without explicit rate were assumed to be in the path and people were not sure how TCP flow control would be eliminated when ABR was used. It is simple however, to remove TCP flow control by running ATM native mode, running UDP, or by modifying TCP. Either way the bad effects of TCP can be eliminated. But in all the simulations this was not done and routers without explicit rate were assumed to be in the path in addition to ATM explicit rate switches. Thus, TCP still was allowed to oscillate back and forth between slow start and major data loss. All the loss was typically in the routers. It is like comparing a leaky hose with a good hose by comparing them in all cases with a leaky hose in series with the good hose. Explicit rate must be end-to-end before a reasonable comparison can be made.

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

The simulations all presume no RM cell priority. Thus, a factor of 10 is lost up front. Data on a loaded network typically travels at 10% of the speed of light. This 10:1 data slow down is the actual experience on the Internet.

The test cases were not 1000% overloaded without flow control as in the real Internet. Thus, the severity of the problem with UBR was not demonstrated. It would be a disaster to really use UBR with its small buffers in the Internet with TCP unless it was carefully configured to never be overloaded. Given these flaws, the simulations are meaningless. The comparison is easy to make without simulation. One only has to evaluate time to control as done above, and the results follow fall out without a computer.

Explicit rate flow control is only useful to the user if the RM (Resource Management) cells flow end-to-end and all the congestion points arrange to mark the RM cells with the correct rate. If a current day router or bridge is in the path, the RM cells will be delayed behind data and no marking will be done. This will not have a major effect if the device is always significantly under-loaded, but if it ever becomes overloaded then TCP will need to be operated on top of the explicit rate flow control. Further, both end-stations must participate in the explicit rate flow control in order for it to work. If one or both do not support explicit rate, TCP will be required and all the user benefit will be lost.

Should ATM UBR (or UBR+) be used in the Internet?

The analysis above shows that when TCP is used for flow control, the data buffers must be large enough to support on the order of one second of data from all the input lines. If the buffer is much smaller than this, the large peak demands required by binary rate flow control will cause major data losses with high probability. Since ATM-UBR switches only have 10-40 milliseconds of data storage they clearly are destined to have major data loss if used in the Internet. The only workable option to use ATM-UBR in a wide area Internet environment is to configure the network such that the trunks are always significantly underloaded (like 50% load). However, due to the fact that the trunks cost more than the switches in an ATM network this option is only attractive if no other alternative is available.

Recently UBR+ has been discussed at the ATM Forum. UBR+ is UBR with a minimum rate available. However, UBR+ still has no flow control and is expected to be controlled with TCP the same as UBR. The minimum rate has no effect on any of the discussion above, and UBR+ would fail just as badly as UBR in Internet usage.

Conclusion

Flow control is the most critical factor in the performance of a data network. Since the load always seems to grow quickly to vastly exceed the available capacity, there must be some sort of traffic cop to control the load or the network will have to discard most of the incoming data and then retransmission only makes the problem worse. For the past 15 years the Internet flow control has been TCP. It was designed to work only between the end-stations so it did not need to concern itself with the network.

Today this is an unnecessary and extremely harmful restriction. The network can easily assist in determining and signaling its maximum capacity and the end-station can use this information to greatly reduce the control-time, the time from load determination to load adjustment. The ATM Forum has developed such a flow control protocol, Explicit Rate, and the completed specification was published in May 1996. Explicit Rate operates 100 times faster than TCP to control the rate of the traffic sources. This is near optimal given the limit of the speed of light.

Flow Control - Explicit Rate vs TCP

100 Times Faster

Dr. Lawrence G. Roberts

The 100 times improvement in the control-time which Explicit Rate provides over TCP provides several major benefits. For the network operator, it reduces the cost of the switch by 7:1, eliminates most all data loss, and improves trunk utilization. For the user, it reduces the WWW access time from 10 seconds to 0.1 seconds, reduces the delay variance on the network from 3 sec to 30 milliseconds, and eliminates the need for almost all data retransmission.

It is critical that ATM with Explicit Rate be introduced into the Internet to reduce cost and reduce delays. To do this, it first must be introduced into the Internet backbone, which will reduce its cost. Then it can be extended to the users through the use of Cells In Frames (CIF). Using CIF, an ISP can offer standard IP service to all users, but for those users who download the CIF software, the full benefits of ATM protocol can become available immediately. Thus, users can individually select to gain the benefits of Explicit Rate without requiring a flash cut to a new flow control throughout the network. . Figure 4 shows how the Internet could be upgraded in the core with Explicit Rate and then at some ISP's with CIF to get the benefits to the users.

Explicit Rate in the Internet

Explicit Rate Flow Control & QoS End-to-End Fast Web Access, Voice and Video available

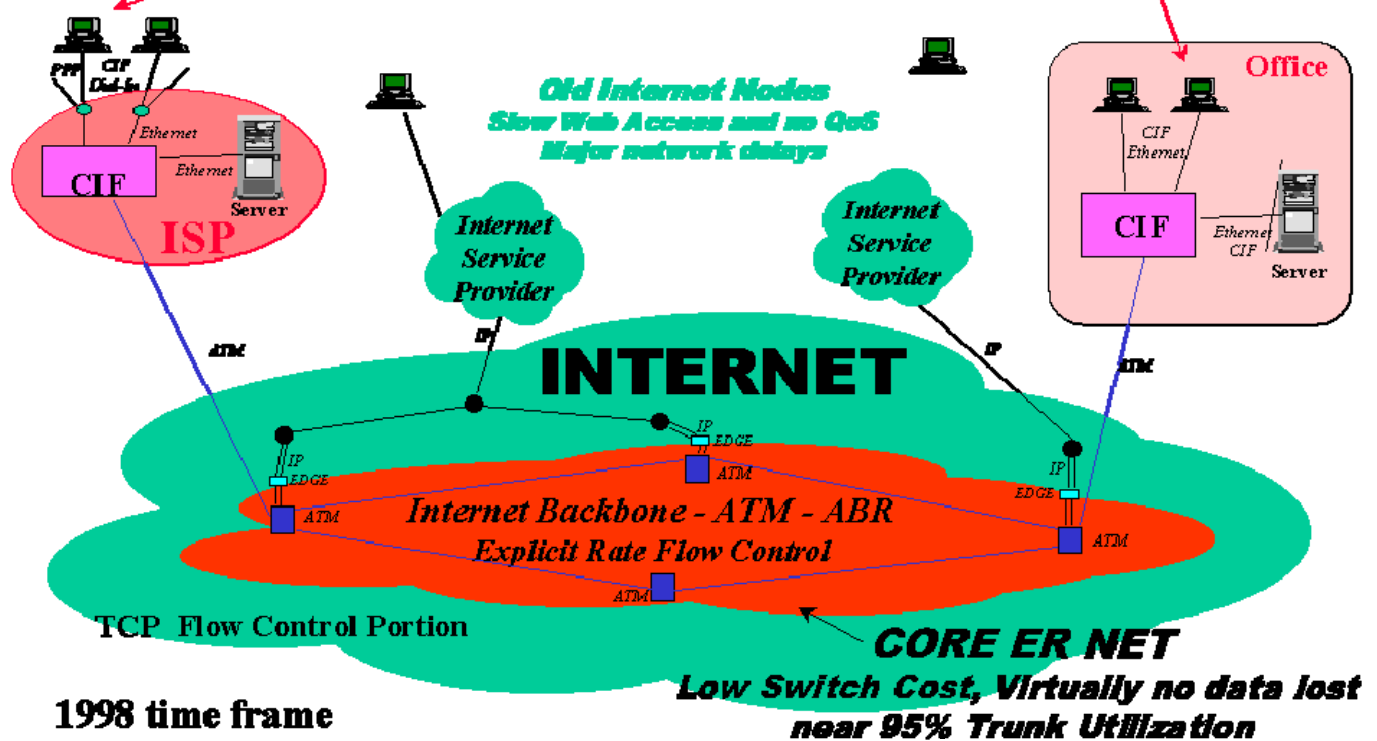


Figure 4. Upgrade Path for the Internet to Support Explicit Rate