Buffer Reservation for TCP over ATM Networks^{*}

Clarence S. C. Lee, K. F. Cheung, Danny H. K. Tsang
Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
Phone: (852) 2358-7045 Fax: (852) 2358-1485
Email: eetsang@ee.ust.hk

Abstract

Transmission Control Protocol (TCP) has already been a widely used protocol in many networks. Recent studies show that the performance of TCP over ATM networks is severely degraded. In this paper, we propose to apply the buffer reservation technique on TCP traffic over ATM networks. While the ATM block transfer (ABT) method (Boyer, 1992) is to reserve bandwidth for bursts of cells, our proposed buffer reservation scheme is to reserve buffer space for TCP packets. We study the performance achieved by the Delayed Transmission (DT) and the Immediate Transmission (IT) buffer reservation schemes. Simulation results show that under the DT buffer reservation scheme, TCP can achieve better throughput and delay performance compared to the plain TCP and the TCP with Early Packet Discard (EPD) except in the WAN environment with short TCP packet length. In the WAN environment with short TCP packet length, the IT scheme can achieve similar throughput performance but smaller delay compared to the plain TCP and the TCP with EPD.

Keywords

Transport control protocol, TCP, ATM, buffer reservation

^{*}This work is supported by Hongkong Telecom Institute of Information Technology grant HKTIIT 93/94. EG01

1 INTRODUCTION

Since TCP has already been a widely used protocol in many networks to provide best effort service, the performance of TCP on ATM networks is of major interest. Recent studies show that the TCP performance with UBR service on ATM networks is severely degraded due to the fragmentation of TCP packets into ATM cells as required by ATM networks (Romanow, 1995). Usually a TCP packet is segmented into several 53-byte ATM cells, which will then be transmitted in an ATM network. As a result, even only one of the cells in a TCP packet is dropped due to switch buffer overflow, the whole TCP packet cannot be reconstructed at the destination. Therefore, network resources are wasted to deliver the corrupted packet.

The ATM block transfer (ABT), also known as the fast reservation protocol (Boyer, 1992), is to reserve bandwidth for bursts of cells, which are also referred to as blocks. A preceding RM cell is sent to request the bandwidth required for the block. Another RM cell follows at the end of the block to release the reserved resources. Each block negotiates for a change in the bandwidth allocation by either Delayed Transmission (DT) or Immediate Transmission (IT).

In this paper, we propose a buffer reservation technique for TCP traffic over ATM networks. Buffer space in ATM switch for TCP packets is reserved in our proposed reservation scheme while bandwidth is reserved for the ATM blocks in the ABT method. Our proposed reservation scheme is simple to implement because no target reservation bandwidth is to be determined.

In our proposed reservation scheme, before sending TCP data, a pilot packet is sent to reserve the buffer space in the switches along the connection route. Successful buffer reservation allows the entire TCP packet to be transmitted to the destination. It can dramatically reduce the number of corrupted TCP packets that otherwise can waste network resources. In our investigation, both Immediate Transmission (IT) and Delayed Transmission (DT) buffer reservation schemes are compared to the plain TCP and the TCP with Early Packet Discard (EPD) (Romanow, 1995).

2 TCP OVER ATM WITH BUFFER RESERVATION

In conventional TCP/ATM stack, the TCP packets are given to the AAL layer and are converted to ATM cells. Since the TCP/IP is a standardized protocol, we suggest implementing our proposal by adding a buffer reservation module (see Figure 1) between the TCP/IP and the AAL layer. In this way, users can obtain the buffer reservation improvement without changing the original TCP/IP infrastructure. The major functionality of this buffer reservation module is to handle pilot packet for buffer reservation and pass data between the TCP/IP and the AAL layer. The AAL layer is assumed to be able to convert the pilot packet into the ATM cell format and can loop back the pilot packet to its source.

Two modes of buffer reservation are investigated. In the DT mode of buffer reservation, the data are buffered at the source and the pilot packet is sent first. When the pilot packet reaches the destination, it will be looped back to the source to indicate the buffer reservation status. The data cannot be sent until the buffer space of the switches in the path has been reserved. As a result, TCP/IP data units are stored temporarily at the source. In the IT mode, the data are sent immediately after the pilot packet. The buffer reservation is performed on a hop-by-hop basis. When the network is congested, the pilot packet and the data packet are immediately dropped. The TCP will retransmit the data, and the buffer reservation module will resend the pilot packet along with a reduced number of data packets.

Figure 1 TCP on ATM protocol stack with buffer reservation module.

2.1 Buffer Reservation Module

The buffer reservation module can be implemented at the source as shown in Figure 2. It consists of a sending buffer, a receiving buffer and a pilot packet processing unit (PPPU). In the sending buffer, a sliding window algorithm (Jacobson, 1988) same as the TCP with the slow start and congestion avoidance phases is implemented to make advice on the reservation buffer size. The pilot packet should contain three fields: pilot packet indication bit (PI) is used to indicate a pilot packet, buffer space reserved (BUF_RES) is used to specify the amount of data buffer to be reserved, and reservation successful indication bit (RES_SUCC) is used to indicate whether the reservation is successful.

Figure 2 Buffer reservation module implementation.

The PPPU needs to keep a record of the amount of data that have successful buffer reservation. After all the data that have successful buffer reservation are sent and acknowledged, another pilot packet will be sent for the next window of data. If the reservation is unsuccessful, the transmission window will be reduced down to one segment size. After a backoff delay, another pilot packet with the desired buffer space equal to one segment size will be sent.

A time-out timer is also implemented in the buffer reservation module. Whenever a transmission time-out is triggered, the packet is put in the retransmission list. The retransmission packet also requires a pilot packet to reserve its buffer space. When packets wait in the reservation module, the TCP time-out timer in the upper TCP layer may also be triggered for packet retransmission if the packets are not acknowledged. When these retransmission packets are received by the buffer reservation module, the duplicate packets will be dropped if the same packets are still in the sending buffer. On the other hand, if the same packets have already been transmitted but without acknowledgment, these packets will be put into the retransmission list.

The pilot packet is sent with the reservation successful indication bit (RES_SUCC) set to one. When the pilot packet traverses along the connection path, the switches at which the reservation is unsuccessful will reset the RES_SUCC bit. When a return pilot packet is received, the RES_SUCC bit is checked. When the RES_SUCC bit is equal to one, the reserved buffer size in the BYTE_RES field is successfully reserved by all the switches in the connection path. In the DT mode, this indicates that the block of data can be sent. In the IT mode, this indicates that the complete block of data has been successfully transmitted. If RES_SUCC=0, the reservation is unsuccessful. At the destination, non-pilot packets will be passed to the upper TCP/IP layer as normal data packets.

2.2 Buffer Reservation ATM Switch

In Figure 3, the block diagram of the buffer reservation ATM switch is shown. One forward direction output queue is shown in detail while others can be implemented similarly. An unreserved buffer size (UBS) memory is implemented with an initial value equal to the maximum buffer size of the output queue. When an ATM cell enters (leaves) the output queue, the UBS is increased (decreased) by one cell. The switch keeps a record of the buffer size that is not reserved. To ensure efficient buffer reservation, the pilot packet has a higher priority such that it will be put in front of the non-pilot ATM cells in the output queue. Note that in the IT mode, data with unsuccessful reservation are dropped at the switch.

Figure 3 Buffer reservation ATM switch.

When a pilot packet is switched to an output queue, the `BYTES_RES` field is compared with the `UBS` memory. If `UBS` is larger than the desired reservation value, the reservation will be granted and the `UBS` memory is updated by subtracting the value of `BYTES_RES`. Otherwise, the `RES_SUCC` field will be set to zero. In our simulation setup, the pilot packets will only be looped back at the destination. The switch will then check the `RES_SUCC` field of the backward pilot packets and release the previously reserved buffer space appropriately.

Alternatively, the unsuccessful pilot packets can be looped back at the switch. If the `UBS` is smaller than the request reservation size, the pilot packet will be looped back to the source in the backward direction. The switch can then check for the `RES_SUCC` field of the backward pilot packets. In this way, the switch can release the buffer space faster than the destination loop back mechanism. The performance

can be increased. However, the switch must be implemented with the functionality to loop back the pilot packet when necessary. In particular, if the pilot packet cannot be switched to the next switch in the forward direction, the pilot packet is looped back to inform the previous switches to release their buffer space.

3 PERFORMANCE EVALUATION IN LAN SCENARIO

The simulation setup in Figure 4 models an ATM network supporting TCP. Acknowledgment and pilot packets are returned to the source in the return path with no background traffic. 20 TCP sources are implemented in the LAN scenario to model a situation with a moderate number of connections. The Maximal Transmission Units (MTUs) of the TCP packets are chosen to be 512 and 9180 bytes to model the effect of long and short packet lengths for the simulations of the throughput and delay performance. The TCP sources are modeled by on-off source sources with mean burst length equal to 1000 and 20000 bytes respectively for the cases of MTU=512 bytes and MTU=9180 bytes. The off period is set to different values to achieve different loading conditions. The transmission speed of the TCP sources and the switch capacity are 150 Mbps. The link propagation delay between the source/destination end system and the ATM switch is taken to be 0.5 μ s, which is equal to 100 m in distance. The link propagation delay in the ATM backbone is set to 0.05 ms which is equal to 10 km in distance. The simulation time is set to 2 s, which is reasonably long for the transfer of 30 MB of data. In our investigation, network components are implemented with 4.3-Tahoe BSD TCP version with window size equal to 160 kbytes, while the EPD threshold is set at half of the switch maximum occupancy as suggested in (Romanow, 1995).

3.1 Throughput Performance

As shown in Figure 5, the effective throughput of the simulation model is plotted against different network loading while the switch buffer size is fixed to 160 kbytes. The effective throughput is defined as the number of correct TCP packets that travel through the ATM network in an elapsed time. Retransmitted, loss or duplicate packets are excluded from the calculation. The effective throughput is measured as a percentage of maximum network capacity. An effective throughput of 100% is achieved when the ATM link is never idle and no packet is corrupted and retransmitted. The loading of the network is measured by the amount of data generated by the sources.

Figure 4 Simulation scenario for ATM LAN.

Figure 5 Effective throughput vs. loading (160 kbytes switch buffer size):
(left) MTU=9180 Bytes, (right) MTU=512 Bytes.

When the loading becomes larger, throughput degradation is observed for the plain TCP. The major reason for the degradation is the transmission of partially dropped TCP packets that lead to a large number of retransmissions and wastage of bandwidth resource.

The TCP with EPD shows an improvement with respect to the plain TCP; however, throughput degradation is still observed. The number of corrupted packets is reduced since the whole packet is either transmitted or dropped. However, the utilization of buffer is limited due to the EPD dropping threshold.

For the DT buffer reservation scheme, almost no throughput degradation is observed with short and long TCP packet lengths, because every packet has buffer space reserved in advance. The whole TCP packet can be served without being partially dropped and nearly no retransmission is required.

For the IT scheme, the performance is similar to that of TCP with EPD because both schemes will either admit or drop the whole TCP packet at the switch.

With short TCP packet length, both TCP with EPD and plain TCP perform similarly with small throughput degradation. Since the TCP packet length is short, the loss of resources due to corrupted packets is small compared to that of long TCP packet length.

We also investigate the throughput performance with different switch buffer sizes and the results are shown in Figure 6. In all the schemes investigated, larger switch buffer size can always achieve higher effective throughput. With both long and short TCP packet lengths, the DT buffer reservation scheme achieves the highest effective throughput among all the schemes.

Figure 6 Effective throughput vs. switch buffer size (loading=1.1):
(left) MTU=9180 Bytes, (right) MTU=512 Bytes

3.2 Delay Performance

The mean packet transmission delay is investigated. It is defined as the time between the packet generated at the source and the packet correctly received at the destination. The mean packet transmission delay with long and short TCP packet lengths are plotted in Figure 7.

From the results, the mean packet transmission delay becomes longer when the loading increases. It is because the packets are easier to be corrupted and the number of retransmissions is larger at heavy load. With long TCP packet length, the mean packet transmission delays of the DT scheme, the IT scheme and TCP with EPD are smaller than that of plain TCP. The DT buffer reservation scheme can still achieve small mean packet transmission delay because of the short propagation delay time. Packets do not have to wait for a long time to receive the buffer reservation notification.

With short TCP packet length, the loss of resources due to segmentation is reduced. As a result, TCP with EPD and the IT scheme behave similarly compared

Figure 7 Mean packet transmission delay vs. loading (160 kbytes switch buffer size): (left) MTU=9180 Bytes (right) MTU=512 Bytes

to the plain TCP. At high network loading, the mean packet transmission delay of the DT scheme is smaller than that of the others. It is because few retransmissions are required in the DT scheme. As a result, data packets wait at the source only due to network congestion but not due to frequent retransmission time-out. The network can transmit data more efficiently in the DT scheme.

4 PERFORMANCE EVALUATION IN WAN SCENARIO

To study the buffer reservation schemes over the ATM WAN, we perform similar investigation as in the LAN case. As shown in the WAN simulation scenario (Figure 8), the distance between the ATM switches is set to 1000 km, which is equal to 5 ms propagation delay time. 100 TCP sources are used in this simulation to model a large number of TCP connections in this ATM network.

4.1 Throughput Performance

As shown in Figure 9, the maximum effective throughput is approximately 0.7, which is smaller than the results in the LAN case (effective throughput ≈ 1). The major reason is the relatively long propagation delay time compared to the packet transmission time. After a window of TCP packets are transmitted, the sources can transmit data again when either the acknowledgments return or the retransmission time-out is triggered. Usually, the retransmission time-out is longer than the time for the acknowledgment to return if the data are not lost. As a result, the link idle time is longer in the WAN scenario, which leads to a lower effective throughput. With long TCP packet length, all throughput performance is similar to that of the LAN scenario. The DT scheme does not suffer from throughput degradation while all the other schemes show throughput degradation due to packet corruption and a larger number of retransmissions.

Figure 8 Simulation scenario for ATM WAN.

With short TCP packet length, the DT buffer reservation scheme achieves a smaller effective throughput than the other three schemes. This is because the link idle time in the case of short TCP packet length is longer compared to that in the case of long TCP packet length.

In Figure 10, the effective throughput is plotted against the switch buffer size. In all the schemes, larger switch buffer size always results in higher effective throughput. With long TCP packet length, the DT scheme achieves higher effective throughput than the others except for very small buffer size. For very small buffer size, the DT buffer reservation scheme is not able to reserve the required buffer for data packets effectively. Most of the buffer space is often allocated for the ATM cells in transit. Hence, the incoming pilot packets at the switch cannot easily reserve buffer space successfully. The IT buffer reservation scheme and the TCP with EPD perform similarly with respect to the switch buffer size.

Figure 9 Effective throughput vs. loading (160 kbytes buffer size):
(left) MTU=9180 Bytes, (right) MTU=512 Bytes.

With short TCP packet length, the DT buffer reservation scheme achieves smaller effective throughput than the other schemes. This is due to the longer link idle time when the sources wait for the return of the pilot packets.

Figure 10 Effective throughput vs. switch buffer size (loading=1.1): (left) MTU=9180 Bytes, (right) MTU=512 Bytes

4.2 Delay Performance

The mean packet transmission delay with long and short TCP packet lengths are plotted in Figure 11. For the DT buffer reservation scheme, TCP packets can be sent after the pilot packet is returned with successful reservation indication. This round-trip delay time is significant compared to the packet transmission time. As a result, at small loading the mean packet transmission delay of the DT buffer reservation scheme is longer than that of the other schemes.

With long TCP packet length, the mean packet transmission delay of the DT scheme increases slower than that of the other schemes as the loading increases. This is because buffer overflow and packet drop rarely happen in the DT scheme.

Figure 11 Mean packet admission delay vs. loading (160 kbytes buffer size): (left) MTU=9180 Bytes, (right) MTU=512 Bytes

Hence, retransmission that must wait for time-out trigger is very unusual in the DT buffer reservation scheme. Note that at high loading, the DT and the IT schemes perform similarly.

With short TCP packet length, the mean packet transmission delay of the DT scheme is longer than that of the other schemes. Because of shorter TCP packet length, packet corruption in TCP with EPD and plain TCP does not lead to a significant increase in the mean packet transmission delay.

5 CONCLUSIONS

We proposed to apply the buffer reservation technique for TCP on ATM networks. It can support packet-based protocols such as TCP over ATM networks with higher performance. The DT and IT buffer reservation schemes were investigated. We compared the performance of the two buffer reservation schemes to that of plain TCP and TCP with EPD by simulations. In the LAN environment, results show that the DT buffer reservation scheme out-performs the plain TCP and the TCP with EPD. It can achieve higher throughput and similar delay performance compared to the plain TCP and the TCP with EPD. The IT buffer reservation scheme performs similarly compared to the TCP with EPD. In the WAN environment, the DT buffer reservation scheme can achieve better performance with long TCP packet length. However, for short TCP packet length the DT buffer reservation scheme achieves a lower effective throughput due to the longer propagation delay time. On the other hand, the IT buffer reservation can achieve similar throughput performance and smaller delay compared to the other schemes in the WAN environment.

The major disadvantage of the proposed buffer reservation scheme is the modification of TCP over ATM stack, which requires sending a pilot packet before sending the TCP packets. Therefore, we suggest that the current TCP systems can achieve the buffer reservation improvement by simply adding the buffer reservation module and the corresponding ATM adaptation system. In this way, the TCP infrastructure does not have to be rebuilt.

REFERENCES

- Boyer P. (1992) Reservation Principle with Applications to the ATM Traffic Control, in *Computer Networks and ISDN Systems*, **24**, 321-34.
- Jacobson V. (1988) Congestion Avoidance and Control, in *Proceedings of SIGCOMM 88*, 314-329.
- Romanow A. and Floyd S. (1995) Dynamics of TCP Traffic over ATM Networks, in *IEEE Journal on Selected Areas in Comm.*, **13**, 633-41.