Applied Mechanics and Materials
ISSN: 1662-7482, Vols. 687-691, pp 2979-2983
doi:10.4028/www.scientific.net/AMM.687-691.2979

Submitted: 2014-09-26
Accepted: 2014-09-27
Online: 2014-11-27

TCP proxy with data redundancy elimination and compressing technology

Lili Yu[1, a], Hongyuan Chen[1,b]

[1]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, People's Republic of China

[a]Yulili1990825@163.com, [b]chy@mail.njust.edu.cn

Keywords: TCP protocol; WAN accelebration; TCP flow optimization; data redundancy elimination and compressing technology

Abstract. TCP protocol is a connection-oriented transmission protocol with mechanisims such as sliding window,data acknowledgement,slow start and congestion avoidance to ensure reliability[1].But these machanisims bring low efficiency because of delay and loss.Thus,we need to design an improved TCP protocol to achieve the purpose of WAN acceleration.This paper mainly introduces two kinds of WAN accelebration technology of TCP flow optimization and data cache technology.TCP flow optimization uses TCP transparent proxy technology based on an improved TCP protocol to solve the problem of large delay and high packet loss rate.The improved TCP protocol includes slow start improved,large receive buff, the improvement algorithm of congestion avoidance,etc.Besides,data redundancy elimination and compression technology[2] are also used in TCP data interation to reduce link load and improve transmission efficiency.

Introduction

In WAN,due to its characteristics of large delay,high packet loss rate,TCP transmission efficiency is not high.So we should do some corresponding optimizations.Since the concept of WAN acceleration[3] was first suggested in 2006,it has been a hot research field over these years. How to use some software or modify communication protocol to improve the network performance and maximize the use of existing cyber source,without changing the existing wide area network topology structure and network interconnection[4] ,has great research value and wide market.

In order to improve the performance of TCP protocol,transmission flow optimization(TFO) using TCP transparent technology that based on slow start improved,large receive buff and the improvement algorithm for congestion avoidance is proposed to solve the problems of loss and delay that brought by the traditional TCP.

This paper describes the improvementation of TCP proxy technology based on the improved TCP protocol.Traditional TCP protocol uses the transmission mode of end-to-end interaction respond and confirmation[5].Althrough this can ensure the reliability of data transmission,it brings large delay in WAN.By using the mode of local-end respond and developing a private TCP protocol[6] combined with data redundancy elimination and compression technology,TCP proxies can not only reduce the delay,but also eliminate a lot of unnecessary transmission of duplicate data.Thus the efficiency of data transmisson can be greatly improved.

Transmission flow optimization

TCP proxies are deployed in the two gatewaies located between LAN and WAN.The deployment status is shown in Figure 1.

When a certain client somewhere in a LAN makes TCP connection with the server in another LAN, TCP segments sending to the server will go through the two TCP proxies on both sides of WAN. Figure 2 shows the data transmission progress of the two TCP proxies between the client and the server.

To SYN segments,TCP proxy rebuilds it by filling in with the WAN optimization option(It will be introduced later),then it sends the new one to the other end.When receiving data in order,proxy does

not wait for ACKs sent by the next proxy or the receiver,but sends a pseudo ACK to the prior proxy or the sender immediately.For FINs,proxy also uses local-end mode.

Besides,in order to meet the needs of WAN,TCP proxies make some optimizations based on the original TCP mechanisms.The following several aspects of optimizations are considered:

1. Slow start improved

The congestion window always starts from one MSS when the traditional TCP startups.And after receiving ACKs from the other end,it grows exponentially.To accelebrate this process,TCP proxy optimizals the slow start mechanism by changing the initial value of congestion window starting from one to two.Therefore,Compared with the previous,it just takes half the time to achieve the same value of congestion window.

2. Large receive buff

Window size field in TCP header,which can only be 64K,limits the maximum receive window of traditional TCP.In condition of high bandwidth and high latency in WAN,even if there is surplus bandwidth,the sender can not send more than 64K data to the receiver until receiving ACKs.In order to improve the transmission performance of TCP,proxy uses the window scale option to increase the receive window from 64K to 512K or even more.The improvement can be seen in Figure 3.

3. The improvement algorithm of congestion avoidance

Reno algorithm is mainly optimized for LAN.When congestion occurs, the send window will be reduced by half immediately, and slowly increased when receiving ACKs.Compared with Reno algorithm,BIC algorithm searches for the most suitable send window for the network based on binary algorithm.It can achieve high efficiency in WAN transmission.

Proxies based on the improved TCP protocol including aspects above fully consider the characteristics of high latency,high packet loss rate and lagre bandwidth of WAN,which are more suitable for WAN transmission.

The working principle of TCP proxy with data cache technology

In this section,we will show how proxy works when the client makes TCP connection with the server.

First,proxy creats a listen TCP in local in order to support the service of transmission flow optimization(TFO).The WAN optimization option should be negotiated during the connection establishment between the two proxies on both sides of WAN.We show it in Figure 4.

The version and ID field in it inform the other end of local CPU,and the status field can help do data redundancy elimination.The ability field includes TFO|DRE|LZ.

Figure 2 and 5 show the connection establishment.When TCP proxy A receives an SYN segment from LAN,it first creates a Lan TCP in local,then another TCP(Wan TCP) related to the LAN one can be created.Wan TCP builds a new SYN' segment filled in with the WAN optimization option and sends it to TCP proxy B.The SYN segment and SYN' segment have the same initial sequence.When TCP proxy B receives the SYN' segment from WAN,it will also create a Wan TCP and a Lan TCP.Lan TCP sends an SYN without WAN optimization option to the server and waits for the expected SYN+ACK.When Lan TCP receives SYN+ACK or the last ACK from LAN,the related Wan TCP will send them to the other end after filling in with optimization option.If Wan TCP receives SYN+ACK or the last ACK from WAN,the related Lan TCP will send them to the other end without the optimization option.The two Wan TCPs do WAN optimization option negotiation during the connection establishment.

TCP proxy uses data redundancy elimination and compression technology to deal with data segments.Data redundancy elimination inspects TCP traffic and identifies patterns within the message.Once patterns have been identified,redundant patterns can be safely replaced by small signatures,thus minimizing bandwidth consumption significantly.And LZ compression has a connection-oriented compression history to further minimize the amount of bandwidth consumed by a TCP connection.

We have refered that proxy uses local-end mode while dealing with data segments and FIN segments.When Lan TCP receives an FIN,it sends back an FIN+ACK of the received sequence

number plus one and notifies Wan TCP to do connection termination with the other end.The connection termination is the same with traditional TCP,so we do not introduce it in detail.

As shown in Figure 6,both of the two TCPs have three queues:Retransmission Queue,Send Queue,Out-of-order Queue.

1.When receiving data segment from LAN,Lan TCP first checks the legitimacy of the segment.By dealing with its acknowledge sequence number,the TCP will delete data that has been acknowledged in its Retransmission Queue.

2.On one hand,if the sequence number of received data is in order,Lan TCP will send ACKs back and put them into Wan TCP's Send Queue.

3.These data should be compressed and de-duplicated before Wan TCP sends them to the server.To meet the need of retransmission,Retransmission Queue is used to deposite data which has been sent but not acknowledged.

4.On the other hand, data out of order will be put into Out-of-order Queue in local.

Figure 7 shows the data processing when receiving data from WAN Side.

1.Data must be decompressed immediately when Wan TCP receives it from WAN.Data in the Retransmission Queue should be deleted according to the acknowledge sequence number.

2.Wan TCP will send ACKs back and put data into Lan TCP's Send Queue in order.

3.Lan TCP sends data in its Send Queue and stores data that has been sent but not be acknowledged in its Retransmission Queue.

4.Data out of order also needs to be put into local Out-of-order Queue.

Evaluations

In this section,we evaluate the performance of WAN optimization which combines transmission flow optimization with data redundancy elimination and compression technology,and compares it with the traditional TCP.We use simware to simulate gatewaies.

1. Transmission performance over long distance

Firs, we choose three files with different sizes as followed:File A,13.2M.File B,16.3M File C,10.0M.. Then, we set an envirnment of long distance between two simwares,and let PC download files of different sizes from the server. Table1 records the download time for two times.

Table 1. Download time of three files in different conditions

	First Time			Second Time		
	A(s)	B(s)	C(s)	A(s)	B(s)	C(s)
WAN TCP	22.2	32.0	19.8	29.2	35.0	20.2
WAN TFO	7.0	5.3	6.9	13.7	7.6	6.6
LAN TCP	6.2	4.9	5.7	7.2	5.5	5.9

Table 1 compares transmission flow optimization with traditional TCP both in WAN and in LAN.It is seen easily that the download speed in WAN trasmission using trasmission flow optimization can be nearly three times of traditional TCP,and nearly reachs the transmission speed in LAN.So that we can come to an conclusion that WAN optimization can effectively solve the problem of lage delay brought by WAN transmission.

2. Analysis under congestion

BIC algorithm provides a larger initial value of congestion window.When congestion occurs,it uses binary algorithm to search for the send window to make full use of the bandwidth in WAN.Figure 8 shows the comparation between transmission flow optimization and tradition TCP.

3. Data compressing ratio

We estimate data compressing ratio by observing data bytes before and after DRE/LZ.It is tested by five groups of data.

From Table 2,we can see that using data redundancy elimination and compression technology can nearly save 80% bandwidth.

4. Network stablity

The result also shows that WAN optimization device has a high stablity in WAN transimssion.

Table 2. Data compressing ratio

Data Size	Before DRE/LZ (Bytes)	After DRE/LZ (Bytes)	Saved (%)
A	4538,318	897,875	81
B	305,001	100,192	67
C	790,652	203,400	74
D	2053,366	275,343	87
E	217,970	66,093	70

Conclusions

This paper mainly describes how to optimizal TCP transmission in WAN.By considering its characteristics of large delay,high packet loss rate and large throughput,we combine transmission flow optimization with data redundancy elimination and compression technology.TCP proxy technology based on the improved TCP protocol is the core technology in WAN optimization.And data redundancy elimination and compression technology improve transmission efficiency further.

The method of TCP transparent proxy is tested in laboratory that provides a virtual environment of WAN.Three parameters,which are transmission rate,compressing rate and the stablity,are observed during the test.The result of test comes to an conclusion that TCP transparent proxy with data redundancy elimination reconciles the contradiction between processing time and transmission delay.

Acknowledgement

This research is supported by Natural Science Foundation of Jiangsu Province(#BK20130773). Part of work has been supported by H3C Project of TCP Transmission Optimization Development. We would like to thank Hubaotao for insightful discussion and suggestions.

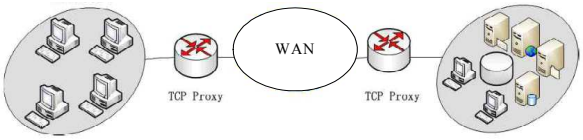

Fig 1. The deployment status of TCP proxy

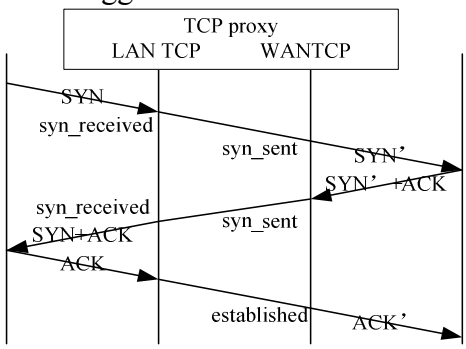

Fig 5. The connection establishment

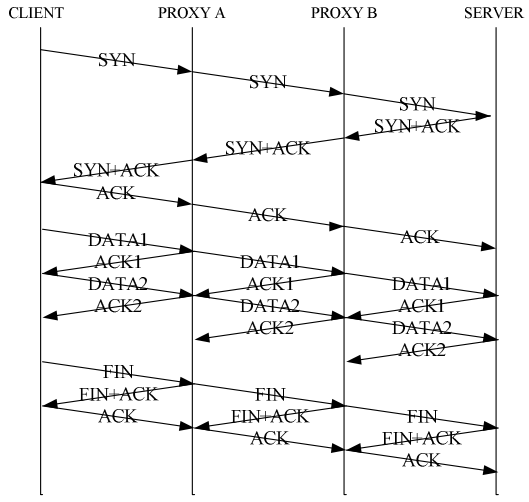

Fig 2. Progress of data transmission through TCP proxies between the client and the server

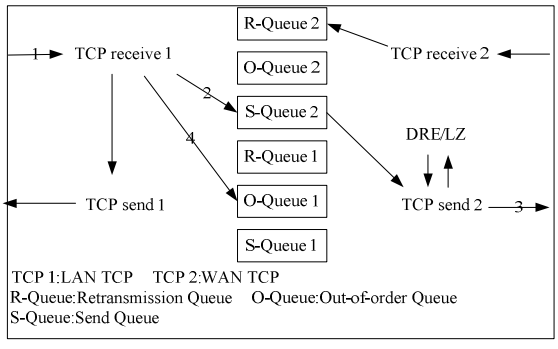

Fig 6. Processing When Receiving Data from LAN Side

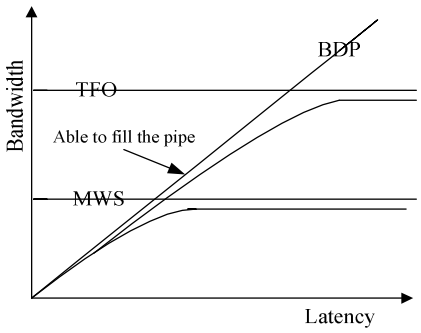

Fig 3. Receive window of TFO

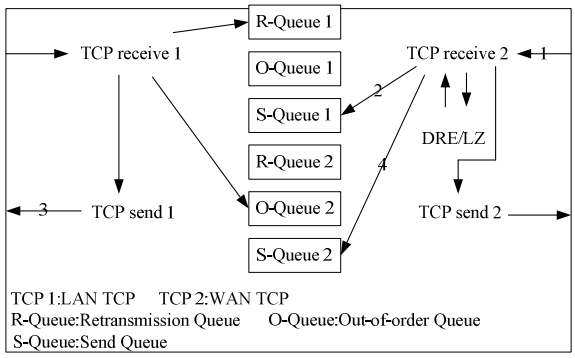

Fig 7. Processing When Receiving Data from WAN Side

Kind=34	Len=12	version	ID	status	ability
1 byte	1 byte	2 byte	6 byte	1 byte	1 byte

Fig 4. WAN optimization option

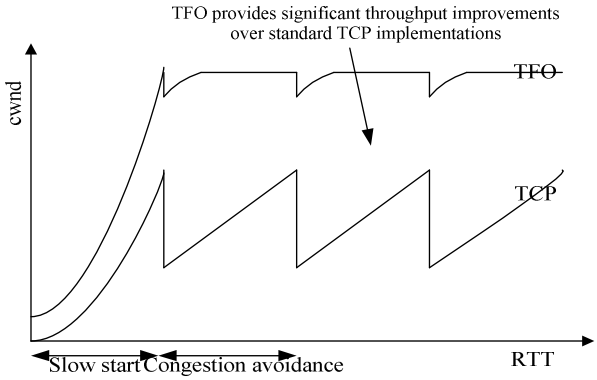

Fig 8. Comparing TFO and TCP under congestion

Reference

[1] W. Richard Stevens. 2000,4. *TCP/IP Illustrated.* Machinery Industry Press.

[2] Yun ting MAO. 2012. Research and Implemetation of data compression algorithm for WAN[D]. Nanjing: Nanjing University of Posts and Telecommunications.

[3] Yongping ZENG, Jianxin WANG. 2010.6. *WAN acceleration technology and Network deployment overview[J]*. Telecom: 32-34.

[4] Jian xin WANG, Na PENG. 2009,5. *Study on the WAN acceleration technology [J]*.

[5] G. Xylomenos and G. C. Polyzos, Mar.1999, *"TCP and UDP Performance over Wireless LAN"*, Proc. IEEE INFOCOM'99, pp. 439-446.

[6] V. Jacobson, R. Braden, and D. Borman. *Tcp extensions for high performance.* Technical Report RFC 1323, IETF, 1992.