一种按需确认的TCP协议改进算法

易发胜 夏梦芹 李巧勤 曾家智
(电子科技大学计算机学院 成都 610054)

摘 要 本文提出了一种基于按需确认的TCP协议改进算法,该算法规定接收方根据发送方要求来对数据报文进行确认。同传统TCP相比,大大减少了ACK的数量,并消除了延迟ACK对传输效率的影响。仿真试验结果表明,该算法减少了TCP协议对网络系统的开销要求,并保持良好的传输效率。

关键词 TCP,滑动窗口,ACK

TCP-DBA: A TCP Improved Algorithm with Demand-based Acknowledgement

YI Fa-Sheng XIA Meng-Qin LI Qiao-Qin ZENG Jia-Zhi
(Department of Computer Science, UEST of China, Chengdu 610054)

Abstract This paper puts forward a TCP improved algorithm with demand-based acknowledgement, which ask that the receiver send an acknowledgement according to the sender's request. Compared to traditionary TCP, the algorithm reduces greatly the number of ACK, and eliminates delayed ACK's bad influence on efficiency of the data transportion. The simulation results demonstrate that the algorithm reduces the request to network sources in TCP protocol and keeps nice transporting efficiency.

Keywords TCP, Slide window, ACK

1 引言

因特网中的IP协议在网络层采用数据报方式工作,对分组报文只是尽力转发,并不保证其可靠到达目的地,也没有处理网络拥塞。为了达到可靠传输数据的目的,在传输层使用TCP协议,不仅实现了数据的可靠传输,还具有流量控制和拥塞控制功能。在TCP协议中,接收方每收到一个数据报文,都要发送一个ACK(确认),发送方根据收到的ACK时间及其反馈的信息,可以判断网络的拥塞情况、接收方的缓冲区大小以及数据报文的接收情况,并进行相应处理从而实现可靠传输、流量控制和拥塞控制。

标准的TCP协议要求接收方每收到一个数据报文,就必须发送一个确认给发送方。这给TCP协议的实现带来一些问题。首先,大量确认报文增加了协议处理的复杂性,耗费了很多CPU时间;其次,大量确认报文也加剧了网络拥塞的可能,降低了网络的性能。特别地,对于半双工网络来说(如以太网、无线局域网等)大量确认报文将大大降低有效的网络数据吞吐量。虽然确认可以捎带在接收方向发送方传输的数据报文中,但是这样的情况毕竟是有限的。为了提高TCP协议性能,针对TCP中的滑动窗口协议做了很多有效的改进。如TCP Tahoe[1]、TCP Reno[2]、TCP SACK[3]、TCP Vegas[4]等,这些研究对TCP的性能进行了很多有益的改进,但这些改进主要集中在如何减少TCP报文重传方面。

在RFC1122[5]中,提出了延迟ACK策略来减少ACK的数量。接收方收到一个报文以后,并不立即发送ACK,而是等待0~200ms,这期间若有回送数据报文就捎带ACK,如果收到两个连续的数据报文或者等待超时则发送一个独立ACK。延迟ACK的应用虽然减少了ACK的数量,但是不及时的确认有时会影响发送方的发送效率[6],特别是对于RTT比较短的局域网来说尤其明显。此外,延迟ACK也影响RTT的正确估计,从而对TCP的整体性能带来一定影响。针对无线局域网中TCP表现不佳的现状,文[7]提出了自适应延迟ACK的设想,有效减少了ACK的数量,但是没有细致考虑这种方案对TCP其它性能的影响。

本文通过对TCP滑动窗口协议和网络传输环境的分析,提出了一种根据发送方要求进行确认的TCP协议改进算法。仿真结果表明,该算法有效地减少ACK的数量,更好地改善了TCP的性能。

2 TCP协议中的确认研究

2.1 确认的作用

在TCP滑动窗口协议中,发送方根据接收方的确认才知道发送的数据报文是否被接收方收到。同时,还要根据确认到来的时间来判断网络是否发生拥塞,然后控制发送流量来实现拥塞控制。因此,发送方需要不断接收来自接收方的确认以进行相应的发送处理操作。

在不同的时候,发送方对确认的要求不一样。在慢启动或重传恢复阶段,发送方需要尽快收到确认,使发送速率尽快恢复到正常水平。在拥塞避免阶段,只要在超时限到来之前收到ACK,滑动窗口正常工作。在这个时候,发送方实际上并不需要每个确认都可靠到达,因TCP是累计确认,后面的ACK会一起确认前面的报文。

根据上述分析,发送方可以知道在什么情况下需要接收方尽快确认,而在有些情况下,可以延迟确认。因此可以考虑由发送方控制接收方发送必要的确认,从而在不影响TCP传输效率的情况下,减少ACK报文数量。我们称这种根据需

*)国家自然科学基金资助项目,编号:69871005。易发胜 讲师,博士生,主要研究方向:新型网络体系结构,QoS。夏梦芹 博士生,主要研究方向:新型网络体系结构。李巧勤 博士生,主要研究方向:新型网络体系结构。曾家智 教授,博导,主要研究方向:计算机网络与通信。

要进行确认的方式为按需确认(DBA),相应的 TCP 协议为 TCP-DBA。

2.2 TCP 传输环境和确认

在 TCP 滑动窗口协议中,由于下层的 IP 是数据报方式,因此有些报文可能失序,而且数据报文或 ACK 报文都有可能丢失。

实际上同一个应用的数据报文失序可能性很小。根据研究网络在一段时间内的状态是相对稳定的[8]。为了尽快让发送方知道报文丢失的情况,提高传输效率,可以假定收到不连续序号的报文就认为有报文丢失。

因此接收方在正常情况下,按需确认;但是若收到一个失序报文或者收到报文前面序号不连续,都主动立即发送确认让发送方及时了解接受情况。为防止确认丢失的影响,在丢失报文收到以前,对每个收到的报文都需要确认,而不管报文是否有发送方的确认控制信息。

2.3 延迟确认的影响

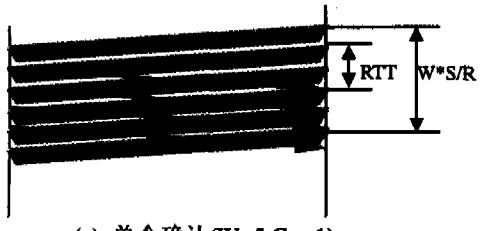
TCP 发送方依靠收到确认来释放已经发送报文所占用的缓冲区。同时从发送报文到收到确认的时间 RTT 是发送方判断网络拥塞的依据[4]。因此延迟确认将可能导致 TCP 保持一定的发送速率的情况下需要更多的发送缓冲区,或者影响发送方对拥塞的错误判断,导致不必要的数据报文重发。

延迟 ACK 的实现完全是接收方随机的延迟处理,目的是通过捎带 ACK 减少独立 ACK 的数量,这存在一些副作用。RFC2525[6]分析了延迟 ACK 对 TCP 的各种不利影响。

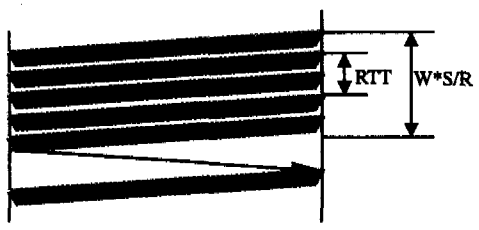
在本文基于发送方控制按需确认的方案中,接收方依据发送方的要求立即给出确认,既减少了 ACK 的数量,又不影响对拥塞控制的判断。不过,在发送窗口一定的情况下,如果发送了太多的报文没有得到及时确认,将会导致传输吞吐量的下降。因此,发送方在什么时候要求接收方必须进行确认需要进行认真分析。

3 按需确认的算法分析

设发送窗口的当前大小为 W 个报文段,每个报文最大传输量是 S bit,本地链路带宽 R bps,往返时延 RTT 的值为 RTTs。还假设发送方每发送 G 个报文段之后要求接收方确认,在正常传输情况下分别采用连续 ACK 和 DBA 的效果如图 1 所示。



(a) 单个确认(W=5,G = 1)



(b) 按需确认(W=5,G = 5)

图 1 按需确认的效果

图中一个窗口的数据报文全部发送完毕的时间是 $W*S/R$,当 $W*S/R > RTT+S/R$ 的时候,如果使用连续确认,滑动窗口可以实现连续发送;如果使用 $G = W$ 的成组确认,则每发送一个窗口内的数据,将会等待一个时间 RTT 才能继续发送。虽然可以增加 W 来提高传输效率,但是 W 的增加受到接受方流量控制、网络拥塞控制和发送方资源限制的影响。因此如果要考虑传输效率,保持连续发送,则应该选择一个合理的 G 值。下面分三种情况来讨论这个问题。

(1)当 $W*S/R > RTT+S/R$ 时,如果这时要保持连续发送,必须满足

$$G*S/R+RTT \leqslant W*S/R \quad (1)$$

所以 $G \leqslant W - RTT * R/S$ (2)

设 a 为 $(W-1)*S/R$ 和 RTT 的比值

则 $a = (W-1) * S / (R*RTT)$ (3)

在本情况中 $a > 1$。根据式(3)知道

$$RTT = (W-1) * S / (R*a) \quad (4)$$

将(4)代入公式(2),得到

$$G \leqslant [W - (W-1)/a] \quad (5)$$

根据 $W*S/R > RTT+S/R$,在 $RTT = W*S/R - S/R$ 时,即 $a=1$ 时 G 得到最小值 1;

(2)当 $W*S/R < RTT+S/R$ 时,这时无论哪种确认方式发送完一个窗口的数据报文,都需要等待一段时间 T 才能开始发送下一个窗口的数据。很容易知道,$T = RTT+G*S/R - W*S/R$,在 $G= 1$ 的时候,T 最小。

在改进算法中设定这个阶段的 a 值范围为 $0.1 < a < 1$,让 G 随 a 按比例变化,可以得到:

$$G=[(1-a)(W-1)/0.9 +1] \quad (6)$$

(3)当 T 很大,$W*S/R << RTT+S/R$ 时。这时 $a < 0.1$,无论连续确认还是按需确认都需要等待相当于 RTT 的时间。在改进算法中采用 $G = W$ 的确认要求对整体性能影响不大。这主要体现在广域网上的数据传输。比如本地是 100Mbps 的以太网接入,RTT 等于 100ms 的时候,如果发送窗口为 8K 字节,$W*S/R$ 小于 1ms,大大小于 RTT 的值。

上面三种情况主要考虑了保持高效 TCP 数据吞吐量的情况下,如何选择按需确认的时机,但是在 TCP 的不同阶段和不同使用场合,对确认还需要区别对待。

在 TCP 的慢启动阶段,每收到一个确认,将倍增拥塞窗口的大小,使 TCP 快速达到最大发送速率。这时需要接收方每收到数据报文就给出确认才好。

而像无线局域网这样的半双工网络,发送一个单独的确认报文会影响整个传输性能,要求确认报文越少越好,这时可以让发送方发送完一个窗口的数据后再要求接收方确认[7]。

4 TCP-DBA 改进算法

为了提高 TCP 传输效率和保持原 TCP 协议的兼容性,基于 DBA 的改进算法根据 TCP 处于的不同阶段和不同场合来决定使用的 G 值。

在慢启动阶段令 G=1,在拥塞避免阶段则使用前面的分析,发送方根据 RTT 和 W 值计算 G,或者直接配置 $G=W$(无线局域网场合)。然后依次发送 G 个报文,发送最后一个报文的时候,给报文设置要求要求确认标记(为了不修改 TCP 帧格式,使用 TCP 首部的一位保留域作为标记),发送完 G 个报文,立即启动定时器。接收方依次接收各种数据报文,检测到要求确认的数据报文时,发出一个确认。发送方收

到后计算 RTT 的值，并根据通知窗口和拥塞窗口计算下一次 G 的值，开始发送下一组报文。如果发送窗口待发报文数量小于 G，在发送发送窗口最后一个数据报文的时候设置强制确认标记。

一些意外情况下，及时有效地确认可以改进滑动窗口的效率。在改进的算法中，除了正常的按需确认，在如下几种情况下，接收方都需要发送确认信息：

· 接收方收到不连续的报文，都要立即发送一个 SACK。这样实现选择性重传的功能，并实现快速重发。

· 如果接收方向对方发送数据报文，可以捎带 ACK，可以让发送方尽快释放发送窗口占用的缓冲区空间。

· 接受方收到重复报文（窗口外报文）要求强制 ACK。这主要是因确认丢失造成重发，需要根据滑动窗口规定进行确认。

如果路径出现拥塞，可能导致重发定时器超时，发送方需要依次对没有收到确认的报文进行重发。并设置强制确认标记。

4.1 发送窗口的改进算法

发送方首先计算出 G 的值依次发送；收到确认时进行判断给出相应处理；如果超时则进行超时处理。算法如下：

1. 根据 RTT、W 和待发报文数量计算 G 值
2. 依次发送各个报文，对最后一个报文设置要求确认标记。
   启动定时器
3. if 收到确认
   判断确认类型
   释放已经确认的缓冲区
   计算 RTT 的值
   if（属于 SACK）
   重发确认报文以后的数据报文
   else
   计算 W，取出下一组待发数据继续步骤 1；
4. if 定时器时间到
   依次重发窗口下届帧，启动定时器设置要求确认标记

4.2 接收窗口的改进算法

接收窗口正常情况下，只确认要求确认的报文。但是如果收到窗外报文、收到重复报文、收到不连续的报文、或者有数据送到对方可以捎带确认，都必须发送确认。

1. 收到数据报文
2. if 有待发数据
   捎带确认；
   返回；
3. if 有要求确认标记：
   if 是下界报文设置新的等待序号
   改变接收窗口的大小
   发送 ACK（等待序号）
   返回
4. if 报文序号在接收窗口内：
   if 是下界报文
   设置新的等待序号
   改变接收窗口的大小
   else //重复的或者失序的
   发送 ACK（等待序号）
   返回
5. if 报文序号接收窗口外：//重复报文
   发送 ACK（等待序号）
   返回

5 改进算法仿真分析

使用 NS2 作为仿真工具，对改进的滑动窗口协议算法进行了分析。NS2 是 Berkeley 大学开发的仿真平台。仿真采用的网络拓扑结构如图 2 所示。

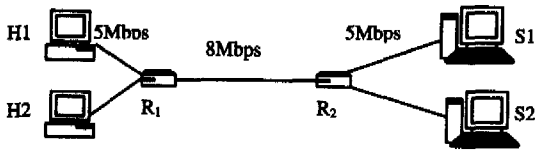


图 2 仿真试验网络拓扑

在图 2 中，共有 4 台主机和 2 台路由器。主机 H1 和 H2 通过 5Mbps 的链路和路由器 R_1 相连，服务器 S1 和 S2 通过 5Mbps 的链路和路由器 R2 相连。路由器 R_1 和 R_2 之间通过一个带宽为 8Mbps 的链路相连。路由器使用 FIFO 和 Drop-Tail 队列管理算法，缓冲区的最大队列长度为 45 个报文段大小，接收方的通知窗口大小设为 20 个报文段。最大数据包长度分别设为 1500 字节。

这里同时传输两个 ftp 流来进行仿真对比。在 H1 到 S1 之间使用标准的 TCP Reno，采用延迟 ACK。而在 H2 到 S2 之间使用在 TCP Reno 基础上利用 GACK 改进滑动窗口协议的 TCP 连接。R_1 和 R_2 之间为瓶颈带宽，在仿真过程中，设定此链路迟延分别为 10ms、50ms、250ms，H1 和 S1 的延迟 ACK（DACK）的值分别设为 0ms、50ms、200ms，在不同的情况下测试从服务器 S1、S2 分别发送 1M 字节文件到 H1 和 H2 的传输时间和 H1、H2 产生 ACK 的数量，由于每轮测试 DBA 均参与，DBA 的值是三次的平均值。另外，由于不同路线的数据流完成时间可能不一样，我们规定先传输完的数据流继续发送随机数据，保持类似的背景流量，但是随机数据的 ACK 数量不算在内。最后得到的数据如表 1 所示。

表 1

链路迟延 / 确认类型	10ms		50ms		250ms	
	Tp(s)	N_ack	Tp(s)	N_ack	Tp(s)	N_ack
DBA(平均)	2.63	82	4.43	230	11.46	58
0msDACK	2.54	690	4.12	694	10.97	687
50msDACK	2.83	353	4.89	369	11.58	386
200msDACK	2.97	348	4.56	358	11.45	350

表中的 Tp 表示 1M 字节数据传输所需的秒数，N-ack 表示监测到的 ACK 的个数（包含捎带 ACK）。从表 1 中可以发现，在链路延迟比较小的时候，延迟 ACK 设定的延迟时间越小，数据传输所需要的时间越短，表示吞吐量更大，但是产生的 ACK 报文的数量迅速上升；在链路延迟比较大时，数据传输所需要的时间相差不大，但是不同的延迟 ACK 的 ACK 报文的数量保持类似的比例。但无论什么情况下，DBA 所需要的数据传输时间都较低，并保持很低的 ACK 报文数量，达到了很好的设计效果。

这主要是因为，延迟 ACK 设定的延迟时间越大，发送的报文越难得到及时的确认，因此吞吐量会有所降低，需要的 ACK 数量呈减少趋势。不过 DACK 设定的延迟从 50ms 改变到 200ms，ACK 的数量减少不多，这主要是连续到达的报

文强制产生独立的确认数量差不多(延迟 ACK 规定,收到两个连续的数据报文必须确认)。另外,由于瓶颈带宽的存在,网络会产生报文丢失现象,加上捎带确认的作用,使 DBA 的 ACK 数量比理论计算的要多。

为了消除改进算法可能抢占带宽资源的影响,设定 R_1 和 R_2 之间带宽为 12Mbps,这样消除了瓶颈带宽的影响,重新重复了上述试验,得到表 2 所示的数据。

表 2

链路迟延 / 确认类型	10ms		50ms		250ms	
	Tp(s)	N_ack	Tp(s)	N_ack	Tp(s)	N_ack
DBA	1.91	43	4.23	206	10.81	38
0msDACK	1.86	668	4.02	667	10.44	667
50msDACK	2.25	345	4.45	357	11.38	356
200msDACK	2.45	346	4.52	349	11.25	363

表 2 的数据体现了链路带宽足够,传输报文不会丢失的情况下的对比数据。表 2 具有和表 1 相似的统计特性。对比表 1 的数据发现,在链路状态比较好的时候,DBA 的 ACK 报文数量进一步减少。同时 DBA 和其它延迟 ACK 的传输时间保持类似的比例,表明改进的算法和其它 TCP 是兼容的。

瓶颈带宽仅在链路延迟比较小时对数据传输率有影响。当链路延迟比较大的时候,如果要充分利用带宽,不但需要增加发送窗口大小,也需要增加接收窗口大小,从而让发送窗口不受小接收窗口的限制。

结论 本文通过对 TCP 滑动窗口协议的分析,发现在接收方进行确认的时候,或者因为延迟确认影响了数据传送效率,或者会增加确认报文的数量,给系统带来较重的负担。经过认真分析目前 TCP 的应用环境,提出一种基于按需确认的 TCP 协议改进算法。仿真分析发现,该算法在大量减少 TCP ACK 数量的同时,还保持良好的数据传输效率,提高了整体网络性能,并能同传统 TCP 协议友好共存。

参 考 文 献

1 Van Jacobson. Congestion avoidance and control. ACM Computer Communication Review, 1988, 18(4): 314~329

2 Stevens W. TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms. RFC 2001, 1997

3 Matthew M, Jamshid M, Sally F, et al. TCP selective acknowledgement option. RFC 2018, 1996

4 Brakmo L S, Peterson L L. TCP vegas: End-to-end congestion avoidance on a global Internet. IEEE Journal on Selected Areas in Communications, 1995, 13(8): 1465~1480

5 Braden R. Requirements for Internet Hosts – Communication Layers. STD 3, RFC 1122, October 1989

6 Paxson V, Allman M, Dawson S, et al. Known TCP implementation problems. RFC2525, Internet Engineering Task Force, Mar. 1999

7 Singh A K, Kankipati K. TCP-ADA : TCP with Adaptive Delayed Acknowledgement for Mobile Ad Hoc Networks. WCNC 2004 / IEEE Communications Society, 2004. 1685~1690

8 Paxson V. Measurements and analysis of end-to-end Internet dynamics: [Ph. D dissertation]. UC Berkeley, 1996

(上接第 31 页)

27 Bahl P, Padmanabhan V. RADAR: An in-building RF-based user location and tracking system. In: Proc of Infocom, 2000, 2: 775~584

28 Meguerdichian S, Slijepcevic S, Karayan V, et al. Localized algorithms in wireless ad-hoc networks: Location discovery and sensor exposure. Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing, 2001

29 Wylie M P, Holtzman J. The non-line of sight problem in mobile location estimation. Proc IEEE International Conference on Universal Personal Communications, 1996, 2: 827~831

30 Kleinrock L, Silvester J. Optimum transmission radii for packet radio networks or why six is a magic number. IEEE National Telecommunications Conference, Birmingham, Alabama, 1978

31 Nagpal R. Organizing a global coordinate system from local information on an amorphous computer: [Technical Report 1666]. MIT AI Lab, 1999

32 Capkun S, Hamdi M, Hubaux J-P. GPS-free positioning in mobile ad-hoc networks. Cluster Comput, 2002, 5(2): 157~167

33 Benbadis F, Friedman T, de Amorim M D, et al. GPS-Free-Free Positioning System for Wireless Sensor Networks. IFIP International Conference on Wireless and Optical Communications Networks (WOCN), Dubai, United Arab Emirates, March 2005

34 Niculescu D, Nath B. DV based positioning in ad hoc networks. Journal of Telecommunication Systems, 2003, 22(1/4): 267~280

35 Savarese C, Langendoen K, Rabaey J. Robust Positioning Algorithms for Distributed Ad-Hoc Wireless Sensor Networks. In: Proc. Usenix Annual Technical Conference, Monterey, CA, June 2002. 317~328

36 Bulusu N, Heidemann J, Estrin D. GPS-less Low Cost Outdoor Localization for Very Small Devices. IEEE Personal Communications Magazine, Special Issue on Smart Spaces and Environments, 2000, 7(5): 28~34

37 Howard A, Mataric M, Sukhatme G S. Localization for Mobile Robot Teams Using Maximum Likelihood Estimation. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Lausanne, Switzerland, October 2002

38 Niculescu D, Nath B. Error characteristics of ad hoc positioning systems (APS). Proc 5th ACM MobiHoc, Tokyo, May 2004

39 Robinson D P, Marshall I W. An Iterative Approach to Locating Simple Devices in an ad-hoc Network. London Communications Symposium, 2002

40 Benson J P, Sreenan C J. High-Precision Ad-Hoc Indoor Positioning in Challenging Industrial Environments. Proceedings of the 1st Workshop on Positioning, Navigation and Communication (WPNC2004), Hanover, Germany, March 2004

41 Patwari N, O'Dea R J, Wang Y. Relative location in wireless networks. In: IEEE VTC, May 2001, 2: 1149~1153

42 Cong L, Zhuang W. Hybrid TDOA/AOA mobile user location in wideband CDMA systems. IEEE International Conference on Third Generation Wireless Communications, June 2000

43 Krishnamachari B, Wicker S, Bejar R. Phase transition phenomena in wireless ad-hoc networks. GLOBECOM, San Antonio, TX, 2001

44 Bulusu N, Estrin D, Heidemann J. Tradeoffs in location support systems: The case for quality-expressive location models for applications. Proc of the Ubicomp 2001 Workshop on Location Modeling for Applications, Atlanta, 2001

45 Avvides A, Park H, Srivastava MB. The bits and flops of the N-hop multilateration primitive for node localization problems. Proc of the 1st ACM Int'l Workshop on Wireless Sensor Networks and Applications, Atlanta: ACM Press, 2002

46 Bergamo P, Mazzini G. Localization in sensor networks with fading and mobility. IEEE PIMRC, 2002. Lisbon, Portugal, September 2002

47 Sundaram N, Ramanathan P. Connectivity based location estimation scheme for wireless ad hoc networks. Proceedings of Globecom, 2002, 1: 143~147