

无线网络中基于显式拥塞反馈的 TCP 协议跨层设计研究

马金忠 叶 进 张向利

(桂林电子科技大学信息与通信学院 桂林 541004)

摘 要 无线网络的特殊性使得有线网络下 TCP 的拥塞控制方法不能直接应用在无线网络中,因此出现了一些跨层设计方法来优化 TCP 的性能。但是目前跨层优化的实现方法中没有充分考虑与已有 TCP 协议的相互融合,使得其应用受到一定的限制。为此,提出了基于显式拥塞反馈的 TCP 协议跨层设计模型,并详细讨论了其可行性、兼容性和扩展性。该模型对混合网络中的 TCP 协议的研究具有较好的参考价值。

关键词 无线网络,跨层设计,显示拥塞通告,显示控制协议,变结构拥塞控制协议,多级反馈拥塞控制协议

Research on TCP Cross-layer Design Based on Explicit Congestion Feedback in Wireless Networks

MA Jin-zhong YE Jin ZHANG Xiang-li

(College of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract The algorithm of congestion control can not be extended to wireless networks in a straightforward manner because the different characters of wireless links. Subsequently, cross-layer re-design of TCP was brought forward to optimize TCP performance. However, it is less considered about how to syncretize cross-layer methods into existing TCP protocols, which result in limited applications. We proposed a TCP model with cross-layer explicit congestion feedback, whose feasibility, compatibility and expansibility were discussed in detail. This model is worthy referring to researches on TCP protocols in hybrid networks.

Keywords Wireless networks, Cross-layer design, ECN, XCP, VCP, MLCP

1 引言

随着移动设备的大量普及,多种无线网络技术如无线局域网(Wireless LAN)、移动自组网(Ad Hoc)、无线传感器网(Wireless Sensor Network)等也随之出现并被迅速应用,人们开始关注无线网络和传统 Internet 网络互相融合的问题,从网络的体系结构方面提出的跨层设计方法就是其中一种解决方案。从实现的难易程度、花费的代价高低、网络的兼容性和可扩展性等角度综合考虑,无线网络的 TCP 协议设计应注重与现有的有线 TCP 协议相融合,才能使已有的大量成熟的 TCP 速率调控机制得以应用,而跨层设计只是对它们的补充和修正。在这种思想指导下,我们提出了一种基于显式拥塞通告的 TCP 跨层设计模型。

本文第 2 部分比较了无线网络 TCP 协议的跨层设计方法,第 3 部分设计了一种基于跨层显式反馈的 TCP 协议模型,重点对其中的显式反馈机制进行了详细讨论,最后提出了进一步研究的设想。

2 无线网络的跨层设计原理及方法

TCP/IP 的拥塞控制算法多是使用丢包或超时重传作为发现网络拥塞的隐式通告信号。这种机制在有线网络里很好地解决了拥塞问题。而无线网络中无线链路的广播特性和终

端的移动性造成网路性能较差,其中包括传输错误率高,服务质量差,网络动态变化等。在拥塞控制方面无法判别是因为链路性能差,还是因为拥塞而造成的丢包现象。因而传统的具有分层结构的 TCP/IP 协议不能适应有线与无线网络混合的结构。一些研究提出一种通过协议栈的跨层设计来解决这个问题,并提出了许多跨层方案^[1]。

跨层设计是通过在协议栈的各层之间传递特定的信息来协调协议栈各层之间的工作过程,使之与无线网络环境相适应,从而使网络能够满足业务的不同需求。其核心就是使协议栈能够根据无线环境的变化来实现对资源的自适应优化配置,实现协议栈各层之间信息交互。该机制使得信息能够在协议栈的各层之间传递,而不再将信息的传递限定在相邻的两个层之间。

目前跨层设计已经成为无线网络性能优化的重要手段之一。Vivek Raghunathan 等人强调 MAC 与 TCP 的跨层设计才是无线网络拥塞控制的最好出路^[2],因此 Lijun Chen 等人提出了 Flow Contention Graph 和 Contention Matrix 的方法建立无线网络的链路状态模型,并提出将拥塞控制转化成为某种约束条件下资源分配的最优化问题^[3]。此外,传统有线网络中物理层的作用是传送来自高层的数据包,或把接收到的包向高层传送。而目前的先进的信号处理技术可以使物理层发挥更大的作用。冀晓亮等人研究了基于空时编码发射分

到稿日期:2008-04-16 本文受桂科青 0832084,桂科自 0832246 资助。

马金忠 讲师,主要研究方向为计算机应用,E-mail:damojian@guet.edu.cn;叶 进 副教授,主要研究方向为计算机网络;张向利 教授,主要研究方向为计算机网络。

集 OFDM 系统的跨层传输技术。该技术结合了传统物理层自适应调制和链路层自动请求重发,利用误包率计算门限并在每个子载波上进行自适应调制,在系统给定延时和误包率约束的基础上最大程度地提高频谱利用率^[4]。

目前研究人员从不同的角度提出了很多的不同的实现方法,表 1 将它们总结为以下几类^[1,5,6]。

分组头信息传递法:这种方法将跨层信息存储在 IP 分组的扩展头中进行传递。由于 IP 分组通常只被逐层地处理,这使得这种信息传递方式构成一个自上到下的管道,设计容易,扩展性好。

联合设计法:无线网络中,具有服务质量(QoS)约束的网络应用往往需要将协议栈中多层协议模块联合设计才能达到满意的效果,这种设计方法用信令来传输跨层信息。互联网控制信息协议(ICMP)就属于这种方法。

合并相邻层法:无线网络中某些相邻层的联系紧密,信息交互频繁,可以考虑将这几层进行“合并”,使得它们成为一个新的“超层”。该超层保持了原来层的功能,模糊了原来层的界限。TCP 网络协议栈中,物理层与 MAC 层联系非常紧密,G. Dimic 等人建议把这两层合并设计,使之更有利于无线网络的优化^[7]。

添加分层法:针对特定的网络环境,在当前的协议栈中添加中间层,用来解决特定的问题。Singh A 等提出,在传输层和网络层里添加一个子层 ATCP,但不修改标准 TCP 协议,仍然使用网络层反馈技术,来解决 ad hoc 网络中因为终端移动性和链路状况差而造成的丢包识别问题^[8]。

表 1 跨层设计方法比较

设计方法	分组头信息传递法	联合设计法	合并相邻层法	添加分层法
传递方式	IP 头部	信令传输	模糊层间接口	中间层
实现代价	易	难	难	较难
扩展性	好	较好	差	一般
兼容性	好	差	一般	一般

从表 1 中我们可以看出,从传递方式上看,利用 IP 头部信息是直接的,并且不需要更改目前的协议栈,其它 3 种方法都不同程度地更改了协议栈结构。从实现代价上看,分组头信息传递法是最容易实现的,其它 3 种都较难实现。从扩展性上看,分组头信息传递法最好,合并相邻层最差,其它两者次之。从兼容性上看,分组头信息传递法最能和目前的协议兼容,其它 3 种方法因为要改变协议栈结构,所以表现出的兼容性都较差。

我们认为,无线网络跨层设计应当充分考虑和目前有线网络的兼容性,并具有良好的扩展性能,实现容易。综合以上考虑,目前分组头信息传递法是无线网络中跨层设计的最佳选择。

3 基于跨层显式反馈的 TCP 协议设计模型

鉴于 TCP/IP 体系结构在传统的有线网络中的主导地位,从经济角度考虑,无线网络中的跨层设计应该注重和目前体系结构的兼容性并具有可扩展性。

这样可以使已有的大量成熟的 TCP 速率调控机制得以应用,而把无线网络的跨层设计作为一种延伸和补充。从上述的分析中,我们知道分组头信息传递法是目前的首选。在

先前工作的基础上^[9],我们设计出了融合无线网络跨层设计的 TCP 协议模型,如图 1 所示。

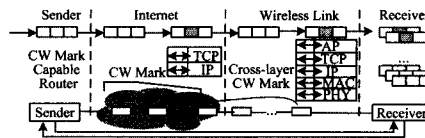


图 1 融合无线网络跨层设计的 TCP 协议模型

这个协议模型包括发端(Sender)、收端(Receiver)、有线网络(Internet)、无线网络(Wireless Link)4个部分。CW (Congestion Warn) Mark 是拥塞警告标记,CW Mark Capable Router 是具有显式拥塞标记能力的路由器。当拥塞发生的时候,该路由器在数据包里添加信息,告诉收端链路处于拥塞状态。收端将该信息添加到返回的 ACK 信号里面,当发端接收到 ACK 信号后,知道网络处于拥塞状态,进行相应的速率控制。在目前的网络中也有不支持这个标记的路由器,为了兼容,可以再设置一位是否支持 CW Mark 的位填充到 IP 包头部,这一位被首先判断,如果支持就使用上面描述的方法;如果不支持,后续的处理与传统网络没有区别。无线链路部分采用基于反馈机制的跨层技术,使得信息可以在协议栈各层之间互相传递。其中 AP, TCP, IP, MAC, PHY 分别代表应用层、传输层、网络层、数据链路层和物理层,左右的双向箭头线指信息可以进入该层,也可以由该层提供给其他层来实现优化控制,中间的直线指参数信息是添加在数据包的头部。

这个协议模型在无线链路部分通过使用分组头信息传递法的跨层设计机制获得网络的性能参数,然后传递给相应的层进拥塞度量,根据度量结果由源端或者中间路由器采取一定的拥塞控制算法。有线网络部分可以不变,也可以使用具有 CW Mark 的路由器。这样的协议模型使得有线和无线网络部分都可以发挥最好的性能,从而使整个网络得到更好的吞吐量、更高的线路利用率等,其体系结构能很好地和目前的网络有效兼容,并具有很好的扩展性。

以上模型中我们采用了基于显式反馈的机制。该机制在无线网络协议栈之间跨层传递网络参数,由中间节点向源端显式地反馈网络拥塞,从而采取相应的速率控制。具体的显式反馈机制目前有以下几种:

(1) ECN (Explicit Congestion Notification), 即显示拥塞通告^[10]

ECN 是 S. Floyd 等提出的一种和 AQM (Active Queue Management) 机制结合使用的方法。在执行 AQM 机制的路由器中,当拥塞发生并且队列没有溢出时,对数据包按照一定的概率进行拥塞标记,接收端收到拥塞标记包后将此拥塞信息通过确认包传送到发送端,于是发送端激发拥塞控制机制。ECN 工作原理如图 2 所示。

实际的做法是把 IP 包头部的服务类型 (Type Of Service) 字段中的第 7 位定义为 CE (Congestion Experienced), 用来表明网络中是否发生了拥塞,这一位由路由器来填充。当路由器通过 AQM 发现队列长度超过了平均队列长度,那么在转发出的报文中,将把 CE 位填充为“1”,用来通知接收端网络中发生了拥塞。接收端收到该报文后,通过 ECN-ECHO 来反馈关于网络拥塞的信息给发送端,发送端可以据此判断网络的状态并进行相应的处理。它只是在现有的拥塞控制策略的基础上,在报文头部定义了警告位,由路由器发起,经过

接收端间接地告知发送端网络拥塞的发生。因此 ECN 有效地防止了报文的丢失和避免了不必要的重传,从而提高了网络的吞吐量。

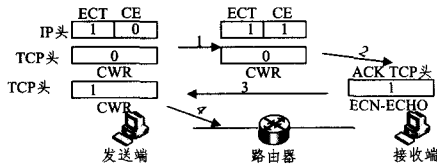


图2 ECN工作原理

(2)XCP (eXplicit Control Protocol),即显式控制协议^[11]

TCP协议的拥塞控制机制都是基于加性增加乘法减少(AIMD)算法,因此该协议在高带宽时延乘积(BDP)的网络中表现出来的性能很差。根据TCP存在的问题和高带宽时延乘积(BDP)网络的特点,Dina Katabi等提出了一种新的Internet拥塞控制协议,该协议称为XCP。它是一种联合端系统和路由器共同协作地使用RTT和队列长度来反馈的跨层设计协议,将拥塞控制从带宽分配策略中解耦。具体实现中,它在IP头部添加了如图2所示的信息。

从图3看出,相对于ECN来讲,XCP机制使用IP头部的多位信息,而不是一位。在上述IP包结构中,XCP首先让发送方探测网络中的拥塞节点,并将其拥塞的程度通过H-feedback反馈回来,然后发送方根据反馈时延调整发送速率。实验表明XCP比TCP在高带宽时延乘积(BDP)网络环境中有着更好的性能,在效率性、公平性、灵活性以及稳定性方面表现都非常出色。甚至是在传统的网络环境中,XCP也能提高网络带宽的利用率,减小队列长度,使丢包的情况降至极低^[12]。但该方法使用多比特来反馈拥塞信息,对现有协议栈改动较大,甚至要重新设计协议栈,花费代价较高。

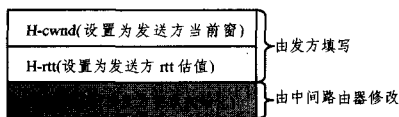


图3 XCP使用的拥塞头部

(3)VCP(Variable-structure congestion Control Protocol)即变结构拥塞控制协议^[13]

当网络的带宽或者时延增大时,TCP协议的性能严重下降,表现最突出的是网络瓶颈处带宽利用率很低。Yong Xia等提出的变结构拥塞控制协议(VCP)可有效地解决上述问题。它使用队列速率和链路使用率来反馈到发送端进行跨层设计。图4是VCP协议的工作过程。

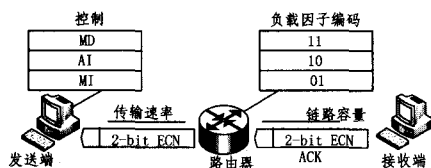


图4 VCP协议工作过程

在VCP协议中,每个路由器每隔一定时间计算一个“负载因子” p ,用来反映网络的拥塞程度,并将其编码放入每个经过数据包的两位ECN位中。源主机根据数据包中的负载因子,执行不同的拥塞控制算法,分别提高带宽利用率和公平

性。与其他的显式拥塞控制协议相比,VCP仅使用了两位比特,但是仍获得了较好的性能。

该协议具体工作过程为:每个路由器每200毫秒计算负载因子的值,源主机发送的数据包经过路由器时,路由器将该值编码放入数据包的两位ECN中。当数据包经过的下个路由器有更高的负载因子时,将会用这个高的负载因子替换原来的,否则不变。这样得到的将是整个路径中最拥塞处的,即瓶颈路由器的负载因子。将该值划分为Low-load, High-load, Overload三个区间,用2-bit ECN位编码后由ACK包从接收方送回发送方;最后根据收到的负载因子的等级,发送方执行下列不同的算法之一:MI, AI, MD。当负载因子为Low-load时,执行MI; $cwnd(t+rtt) = cwnd(t) \times (1-\xi)$,用于效率控制器;当负载因子为High-load时,执行AI; $cwnd(t+rtt) = cwnd(t) + \alpha$;当负载因子为Overload时,执行MD; $cwnd(t + \delta t) = cwnd(t) \times \beta$ 。其中, $cwnd$ 为源主机的拥塞窗口值, rtt 为源主机估计的数据流往返时间值,这里 rtt 是200ms, $\delta t \rightarrow 0$, $\alpha = 1.0$, $\beta = 0.875$, $\xi = 0.0625$ 。

模拟仿真证明该协议在高时延带宽积(BDP)网络中具有带宽利用率高,瓶颈队列持续时间低,可忽略的拥塞包丢失,较合理的公平性等优点^[13]。

(4)MLCP(Multi-Level feedback congestion Control Protocol),即多级反馈拥塞控制协议^[14]

VCP使用了“负载因子”作为拥塞的信号,并用2bit来显示反馈到发送端。在高速带宽时延乘积(BDP)网络中,VCP的速率收敛不是最优化的,而且通过模拟,对于不同的RTT流,VCP显示出了较差的公平性。因此Ihan Ayyub Aazi等提出了MLCP协议。该协议使用4比特对负载因子进行编码,然后显示反馈到发送端。其中拥塞窗口调整过程为:当负载因子低于80%时,执行MI; $cwnd(t+rtt) = cwnd(t) \times (1 + \xi(\sigma))$,其中 $\xi(\sigma) = \kappa(1-\sigma)/\sigma$, σ 是负载因子, $\kappa = 0.35$ 。当负载因子高于80%时,系统获得了高使用率,发送端使用AI-II-MD规则来控制公平性。发送端,应用AI直到 σ 达到95%之后使用II。当系统转到过载区域(负载因子大于100%),应用MD来控制。3种算法如下所示:

$$\text{AI: } cwnd(t+rtt) = cwnd(t) + \alpha$$

$$\text{II: } cwnd(t+rtt) = cwnd(t) + \alpha / (cwnd(t))^{1/2}$$

$$\text{MD: } cwnd(t+\delta t) = cwnd(t) \times \beta(\sigma)$$

其中: tp 是往返时间, $rtt = tp$, $\delta t \rightarrow 0$, $\alpha = 5.0$, $0 < \beta(\sigma) < 1$ 。每一个 tp 间隔执行一次MD。通过模拟可知这种算法在高带宽时延乘积(BDP)网络中获得了接近0的丢包率,维护了较低的路由队列长度,达到了接近最优的公平性和效率。

表2对这4种用于跨层设计的显式反馈机制进行了小结,它们都可以应用到图1所示的TCP模型中。

表2 4种显式拥塞反馈算法的比较

方法	反馈依据	反馈结果	使用标识位数	处理
ECN	Queue length	Congestion Warn (拥塞标记)	1位	0: 正常响应 1: 启动拥塞控制
XCP	RTT Queue length	H-Feedback (拥塞程度)	16位	$Cwnd = \max(Cwnd + H\text{-feedback}; s)$ S为一个包的大小
VCP	Queue Rate Link Utilization	Load Factor (负载因子)	2位	00(Lowload); MI 01(Highload); AI 10(Overload); MD

MLCP	Queue Rate Link Utilization	Load Factor (负载因子)	4 位	<80%:MI <95%:AI <100%:II >100%:MD
------	--------------------------------	-----------------------	-----	--

参考文献

综上,已有的显式拥塞反馈具备了一些成熟的拥塞控制方法,但它们所反馈的拥塞测度都来自路由器,不能体现无线链路上的拥塞情况。我们提出了跨层设计的 TCP 模型,它能够将这些显式拥塞反馈机制延伸到底层,增强了 TCP 协议中拥塞感知的范围。最重要的是,由于该模型中 TCP 响应的各层反馈是“或”的关系,即使某个节点不支持某层的反馈,也不会影响其他层反馈发生作用,因此这种跨层显式反馈具备较好的兼容性和可扩展性。

结束语 本文首先研究了无线网络 TCP 协议中的跨层设计问题,然后在充分考虑与有线网络已有的拥塞控制机制相融合的基础上,提出了一种跨层的显式拥塞通告方法以及相应的 TCP 协议设计模型,并且详细分析了几种具体的实施方案。

对于该设计模型的进一步研究包括:底层拥塞测度的分析和提取;显式反馈机制的跨层扩展和实现。我们已经实现了基于 MAC 层重传次数的跨层 ECN 机制,并取得了预期的效果^[15]。接下来需要从这两个角度进行更加广泛的研究和应用,尤其需要将本文提出的模型扩展到已有的 TCP 协议中,考察跨层的显式拥塞反馈与原有的拥塞控制之间的相互作用。

总之,传统的网络体系结构协议栈不同层之间相对独立,便于标准化和配置,而跨层设计结构主要通过增加协议栈中不同协议层的直接通信,共享有效信息来提高网络性能,需要协议栈内部不同层之间的相互作用。这两种结构的互联互通面临着巨大的挑战,因此设计具有弹性的松散耦合的跨层设计结构,在保持传统网络各层相对独立性的同时,优化网络的性能是未来的发展方向。

- [1] 俞一帆. 第四代移动通信系统的跨层设计研究. 博士论文. 北京:北京邮电大学电信工程学院,2006
- [2] Raghunathan V, Kumar P R. A Counterexample in Congestion Control of Wireless Networks. *Performance Evaluation*, 2007, 64(5):399-418
- [3] Chen Lijun, Low S H, Doyle J C. Joint Congestion Control and Media Access Control Design for Ad Hoc Wireless Network. Infocom, 2005
- [4] 冀晓亮, 衡伟. 结合自适应调制和自动请求重发的无线跨层设计 // 2006 北京地区高校研究生学术交流会——通信与信息技术会议论文集(上). 北京, 2006
- [5] 卢先领, 孙亚民, 周灵, 等. Ad Hoc 无线网络跨层设计综述. *计算机科学*, 2007, 34(10)
- [6] Srivastava V, Motani M. Cross-Layer Design: A Survey and the Road Ahead. *IEEE Communications Magazine*. December 2005: 112-119
- [7] Dimic G, Sidiropoulos N D, Zhang R. Medium Access Control - Physical Cross-Layer Design // *IEEE Sig. Proc.* 2004, 21(5):40-50
- [8] Singh A K, Iyer S. ATCP: Improving TCP performance over mobile wireless environments. *Mobile and Wireless Communications Network // 2002 of the 4th International Workshop*. 2002: 239-243
- [9] 叶进, 王建新. 混合网络中一种基于拥塞概率预测的 TCP 协议. 小型微型计算机系统(已录用)
- [10] Floyd S. TCP and Explicit Congestion Notification. <http://www.icir.org/floyd/ecn.html>
- [11] Katabi D, Hanley M, Rohrs C. Congestion Control for High Bandwidth-Delay Product Networks. SIGCOMM, 2002
- [12] 陶小梅. XCP 协议的改进研究. 硕士论文. 广西:广西师范大学, 2005
- [13] Xia Yong, Subramanian L, et al. One More Bit Is Enough. SIGCOMM, 2005
- [14] Aazi I A, et al. On the Design of Load Factor based Congestion Control Protocols for Next Generation Networks. to be appeared in *IEEE Infocom 2008*
- [15] 叶进, 王建新. 融合 MAC 层拥塞通告的 TCP 协议设计. 高技术通讯(已录用)

(上接第 81 页)

- [6] Wang L, Chen Z P, Jiang X H. Ringed Petersen spheres connected hypercube interconnection networks[C] // 10th IEEE International Conference on Engineering of Complex Computer Systems. June 2005:127-131
- [7] Abachi H, Walker A-J. Network Expandability and Cost Analysis of Torus, Hypercube and Tree Multi-Processor Systems[C] // Proceedings of the 28th Southeastern Symposium on System Theory (SSST'96). 1996:426-430
- [8] Ould-Khaoua M. On the Optimal Network for Multicomputers: Torus or Hypercube? [C] // Proceedings of the 4th International Euro-Par Conference on Parallel Processing. 1998:889-992
- [9] Shahabi A, Honarmand N, Navabi Z. Programmable Routing Tables for Degradable Torus-Based Networks on Chips[C] // IEEE International Symposium on Circuits and Systems. May 2007:1065-1068
- [10] Lusala A K, Manet P, Rousseau B, et al. NoC Implementation in FPGA using Torus Topology[C] // International Conference on Field Programmable Logic and Applications. Aug. 2007:778-781
- [11] Wu Chang, Li Yubai, Chai Song. Design and simulation of a torus structure and route algorithm for network on chip[C] // 7th

- International Conference on ASIC. Oct. 2007:1289-292
- [12] Yeh C H, Behrooz P. Routing and embeddings in cyclic Petersen network; An efficient extension of the Petersen graph[C] // Higaki H, Page TW, eds. Proc. of the Int'l Conf. on Parallel Processing. Fukushima: IEEE Computer Society, 1999:58-65
- [13] Ohring S, Das S K. Folded Petersen cube networks; New competitors for the hypercubes[J]. *IEEE Trans. on Parallel and Distributed Systems*, 1996, 7(2):151-68
- [14] Das S K, Banerjee A K. Hyper Petersen network; Yet another hypercube-like topology[C] // Siegel H J, ed. Proc. of the 4th Symp. On the Frontiers of Massively Parallel Computation. Virginia: IEEE Computer Society, 1992:70-77
- [15] 刘方爱, 刘志勇, 乔香珍. 一种实用的互联网络拓扑结构 RP(k) 及路由算法[J]. *中国科学(E 辑)*, 2002, 32(3):380-385
- [16] 刘方爱, 刘志勇, 乔香珍. 一类层次环网络的构造及路由算法. *计算机学报*[J], 2002, 25(12):1397-404
- [17] 王雷, 林亚平, 夏巍. 双环 Petersen 图互联网络及其路由算法研究[J]. *软件学报*, 2006, 17(5):1115-1123
- [18] 王雷, 林亚平. 基于超立方体环连接的 Petersen 图互连网络研究[J]. *计算机学报*, 2007, 9(6):339-343