

D-AIMD——一种改进的 TCP 拥塞控制机制

D-AIMD: An Improved TCP Congestion Control Mechanism

徐昌彪¹ 隆克平³ 杨士中²

(重庆邮电学院移动通信工程研究中心 重庆400065)¹ (重庆大学通信与测控学院 重庆400044)²

(北京邮电大学程控交换技术与通信网国家重点实验室 北京100876)³

Abstract With the rapid growth of rate-based services and wireless applications, improving TCP congestion control has been becoming an important study task in network research fields. This paper first briefly introduces the AIMD (a, b) algorithm, and then presents an improved TCP congestion control (D-AIMD) whose principles and simulation results are discussed in detail. This mechanism can be easily implemented with lower additional overheads and can efficiently improve the networks communication performances.

Keywords TCP, Congestion Control, D-AIMD

0 引言

TCP 是目前 Internet 中广泛采用的一种传输控制协议, 为各主机之间提供可靠按序的传输服务^[1]. TCP 拥塞控制^[1,2]机制是其成功应用于 Internet 的关键所在. TCP 拥塞控制机制主要基于 AIMD(a, b) (Additive-Increase Multiplicative-Decrease) 算法^[3~7], 即每一往返时间, 拥塞窗口值增加 a , 而当网络发生拥塞时, 拥塞窗口降为原先的 $(1-b)$ 倍. 对所有连接采用相同 a, b 的系统, 易知其资源利用率大致为 $(2-b)/2$. 目前, TCP 拥塞控制大多采用的是 $a=1, b=0.5$. 其主要优点是应用能快速地使用网络中的有用资源, 而当网络拥塞程度加大时, 又能急剧地降低其数据发送速率, 快速减轻网络拥塞. 正因为如此, 致使其数据发送速率波动较大, 资源利用率不是很高 (通常只利用了 75% 左右的资源). 这对于那些要求速率波动不太大的应用以及资源比较紧缺的应用环境 (如无线网络) 来说, 是不大适宜的. 另外, TCP 拥塞控制机制的前提是网络拥塞是引起数据丢失的唯一重要原因, 即只要终端检测到有数据丢失, 均认为是网络拥塞所致, 于是调用拥塞控制, 这对于链路质量很好的有线网络而言是适宜的. 然而若将其直接应用于链路质量相对较差的无线环境却是不适宜的, 这主要是因为无线环境中, 网络拥塞不能再看作是数据丢失的唯一重要原因. 随着基于速率的应用业务的增多, 无线应用的越来越广泛, 对 TCP 拥塞控制机制的改进自然成了网络研究领域一个重要的研究课题.

本文所提出的改进的 TCP 拥塞控制机制仍主要基于 AIMD 算法, 针对网络拥塞导致数据丢失的情形, 对有多种原因引起数据丢失的情况, 假定有各种机制让数据源端能知道是哪一种原因所致, 对非拥塞引起数据丢失的情形, 调用其它相应的控制机制, 从而实现差错控制与拥塞控制的有效分离.

1 AIMD 拥塞控制

基于 AIMD(a, b) 算法的拥塞控制机制, 均遵循如下基本思想: 每一往返时间, 拥塞窗口值增加 a , 当网络发生拥塞时, 拥塞窗口降为原先 $(1-b)$ 的倍. 下面就 AIMD 响应函数以及 AIMD(a, b) 与 AIMD(1, 1/2) 拥塞控制间的公平性问题进行

简要的分析.

1.1 AIMD 响应函数

为了探讨 AIMD(a, b) 的响应函数, 拟采用文[3]中所描述的 AIMD 确定模型而非文[4]中所描述的 TCP 随机模型. AIMD 确定模型假定每当拥塞窗口达到 W 个分组时就会出现一个分组的丢失, 如图1所示; 而 TCP 随机模型则假定数据以随机概率 p 丢失, 且考虑了 TCP 重传超时的影响. 不过, AIMD 确定模型对集中讨论 AIMD(a, b) 中加性因子 a 和乘性因子 b 的作用却是很有用的.

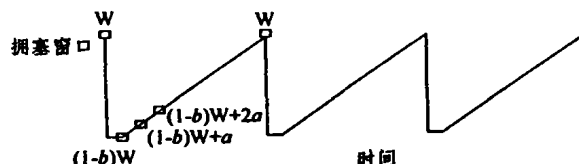


图1 稳态时基于 AIMD(a, b) 的拥塞窗口

对于 AIMD 确定模型, 定义拥塞周期为相邻两次拥塞所经历的时间. 这样, 拥塞周期 T 为 $(\frac{b}{a}W + 1)t_{RTT}$, 其中 t_{RTT} 为往返时间. 一个拥塞周期内所发送的分组总数 S 为:

$$S = (1-b)W + [(1-b)W + a] + \dots + [(1-b)W + \frac{bW}{a}a] \\ = (\frac{b}{a}W + 1)\frac{2-b}{2}W \quad (1)$$

于是一个拥塞周期内平均的数据发送速率 R 为:

$$R = \frac{S}{T} = \frac{2-b}{2R}W \text{ packets/s} \quad (2)$$

另外, 由于 AIMD 确定模型假定一个拥塞周期会丢失一个分组, 因此, 分组的丢失率 p 为:

$$p = \frac{1}{S} = \frac{2a}{b(2-b)W^2 + a(2-b)W} \quad (3)$$

上式可近似为:

$$p \approx 2a/b(2-b)W^2 \quad (4)$$

由式(4)可得:

$$W \approx \sqrt{2a/b(2-b)p} \quad (5)$$

对 AIMD 确定模型, 定义 AIMD 响应函数 $R(a, b, t_{RTT}, p)$ 或 R 为 AIMD 流的稳态数据发送速率. 于是由式(3)和式

徐昌彪 讲师, 博士生, 研究方向为计算机网络通信. 隆克平 教授, 博士后, 硕士生导师, 研究方向为宽带通信网性能分析、IPQoS 结构及其实现机制、TCP/IP 协议改进及其性能分析. 杨士中 工程院院士, 教授, 博导, 长期从事通信技术的研究.

(2)得^[6,7]:

$$R = \frac{pa(-2+b) + \sqrt{p(-2+b)a(pba-8b-2pa)}}{4pbt_{RTT}} \quad (6)$$

或将式(5)代入式(2)得如下近似式^[6,7]:

$$R \approx \sqrt{2-b} \sqrt{a} / \sqrt{2bt_{RTT}} \sqrt{p} \quad (7)$$

1.2 AIMD(a,b)与AIMD(1,1/2)拥塞控制的公平性

根据式(7)可得基于AIMD(1,1/2)的TCP拥塞控制的响应函数:

$$R(1,1/2,t_{RTT},p) \approx \sqrt{1.5/t_{RTT}} \sqrt{p} \quad (8)$$

为了使基于AIMD(a,b)的拥塞控制与基于AIMD(1,1/2)的TCP拥塞控制之间能够公平地竞争网络资源,则要求它们在相同的条件下具有相同的数据发送速率,也就是响应函数相同,即 $R(a,b,t_{RTT},p) = R(1,1/2,t_{RTT},p)$,于是由近似式(7)和式(8)可得a,b应大至满足以下关系式^[6,7]:

$$a = 3b/(2-b) \quad (9)$$

这样,满足式(9)的基于AIMD(a,b)的拥塞控制(如AIMD(1/5,1/8))就能够基本上与基于AIMD(1,1/2)的TCP拥塞控制公平地竞争网络资源。

下面我们从基于AIMD确定模型的图2来简要说明式(9)的几何含义。由图2知,稳态时TCP拥塞控制的平均数据发送速率为 $3W/4$ packets/s,要使基于AIMD(a,b)的拥塞控制在相同的条件下能够与TCP拥塞控制公平地竞争网络资源,显然要求其平均数据发送速率也为 $3W/4$ packets/s,并且其拥塞周期、往返时间、拥塞丢包率等分别与TCP拥塞控制下的拥塞周期、往返时间、拥塞丢包率一样。这样,由图2也就易得式(9)的结果。

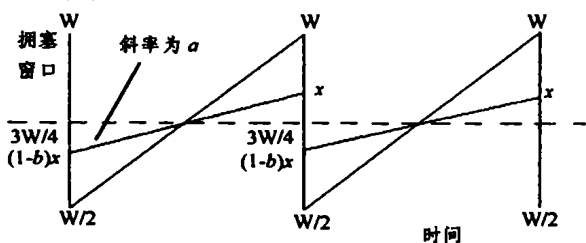


图2 稳态时基于AIMD(a,b)的拥塞控制与基于AIMD(1,1/2)的TCP拥塞控制公平竞争网络资源时的拥塞窗口

2 改进的TCP拥塞控制机制

基于AIMD(a,b)的拥塞控制算法,当a,b取不同的值时即可得到不同的拥塞控制机制,显然若取较小的加性因子a和乘性因子b时,可大幅度降低数据发送速率的波动性,充分利用网络资源(如b=1/8时,资源利用率高达93.75%左右)。然而较小的加性因子a和乘性因子b却不能使应用快速地使用网络中的有用资源,且当网络拥塞程度加大时,与TCP相比,数据的发送速率下降较慢,不能快速地减轻网络拥塞。基于上述原因,我们提出了一种改进的TCP拥塞控制机制。

2.1 基本思想

本文所提出的改进的TCP拥塞控制仍基于AIMD(a,b)算法,所不同的是在一个连接的通信过程中a,b并不唯一。其基本思想是将较大a,b下的拥塞控制机制与较小a,b下的拥塞控制机制有机地结合起来,将较大a,b下的拥塞控制机制用于系统暂态,以实现应用快速地使用网络中的有用资源,以及在网络拥塞程度加重时,快速减轻网络拥塞;将较小a,b下的拥塞控制机制用于系统稳态,达到速率波动性小,拥塞次数

少,资源利用率高的目的。另外,由于此改进的TCP拥塞控制机制在a,b的选取上考虑了与TCP拥塞控制机制的公平性,因此,当它与TCP一起使用时,能与TCP之间实现资源的公平竞争。显然,此机制的实现基于双重AIMD算法,于是可简称为D-AIMD(Dual AIMD)。

2.2 状态转移图

基于上述基本思想,可得D-AIMD的状态转移图,如图3所示。此拥塞控制机制有两种工作状态:基于AIMD(a1,b1)的暂态和基于AIMD(a2,b2)的稳态,a1,b1以及a2,b2的选取满足式(9)以确保资源的公平竞争。也就是当系统处于暂态时,调用基于AIMD(a1,b1)的拥塞控制机制;而当系统处于稳态时,则调用基于AIMD(a2,b2)的拥塞控制机制。可以用条件1来判定系统是否进入了稳态,用条件2来判定系统是否进入了暂态。

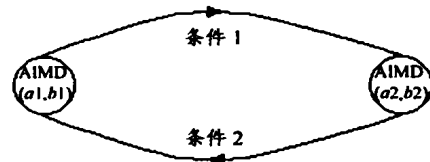


图3 改进的TCP拥塞控制状态转移图(a1>a2,b1>b2)

2.3 状态转移条件

由图3知,条件1和条件2即为D-AIMD的状态转移条件,它们是系统用以确定是调用基于AIMD(a1,b1)的拥塞控制机制,还是调用基于AIMD(a2,b2)的拥塞控制机制的主要依据。根据暂态与稳态之间的区别,我们认为可通过检测数据源端平均数据发送速率的波动性来判定系统是处于暂态,还是稳态,这一点在数据源端是容易办到的。如果平均数据发送速率波动较大,说明系统处于暂态;反之,如果数据的平均发送速率波动性较小,可认为系统基本上已进入了稳态。

2.4 具体实现

对链路质量较好的有线网络,与目前的TCP拥塞控制机制一样,D-AIMD仍通过重传定时器超时和数据源端收到3个重复应答来检测数据的丢失;而对链路质量相对较差的无线网络,由于拥塞不再是数据丢失的唯一重要原因,因此,应采用其它判定数据丢失的机制,以确保拥塞控制机制仅当网络拥塞时调用。另外,此机制也借用目前TCP拥塞控制中的慢起动机制,也就是当连接刚刚建立初期或重传定时器超时时,调用慢起动算法。这样,D-AIMD的具体实现过程如下:

(1)连接建立初期或重传定时器超时,采用慢起动算法,直到拥塞窗口达到慢起动门限,之后至(2)。

(2)令 $a=a1, b=b1$,调用AIMD(a,b)算法,检测此状态下数据的平均发送速率,直到平均数据发送速率的波动性小于某一门限为止,之后至(3)。

(3)令 $a=a2, b=b2$,调用AIMD(a,b)算法,检测此状态下数据的平均发送速率,直到平均数据发送速率的波动性大于某一门限,若为重传定时器超时转至(1),否则转至(2)。

关于a1,b1和a2,b2的选取,在遵循本改进思想的基础上,可灵活考虑。一般来说a1,b1可尽量取大一点的值,a2,b2尽量取小一点的值。对现有的系统,可取 $a1=1, b1=1/2, a2, b2$ 取一对较小的值即可,于是可沿用目前的TCP拥塞控制机制与基于AIMD(a2,b2)的拥塞控制机制一起实现端到端的拥塞控制。

3 仿真结果分析

在对D-AIMD的仿真实验中,暂态沿用基于AIMD(1,

1/2)的 TCP 拥塞控制,稳态采用基于 AIMD(1/5,1/8)的拥塞控制。

3.1 仿真拓扑及相关假设

D-AIMD 的提出主要基于以下三个目标,一是期望系统暂态时,应用能快速地使用网络中的有用资源,网络拥塞程度加重时,又能快速减轻网络拥塞;二是期望系统稳态时,数据发送速率波动性小,拥塞次数少,资源利用率高;三是当它与 TCP 一起使用时,可以与 TCP 之间实现资源的公平竞争。因此,在仿真实验中,我们选用了有名的单瓶颈(哑铃)拓扑作为我们的仿真拓扑,如图4所示。

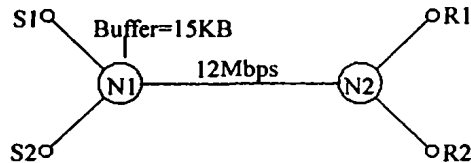


图4 单瓶颈(哑铃)拓扑

图4中 S1、S2 为数据发送端, R1、R2 为数据接收端,可创建两个 TCP 连接: S1-N1-N2-R1 与 S2-N1-N2-R2。并假设:

(1)忽略 S1、S2 到 N1 以及 N2 到 R1、R2 的时延,端到端的时延以 N1 至 N2 的时延为主,往返时间约为 100ms;

(2)拥塞只发生在 N1 节点处, N1 处用于缓存 S1 或 S2 发来的数据的缓冲大小为 15kB;

(3)不考虑重传超时时的影响;

(4)N1 至 N2 链路上的带宽为 12Mbps;

(5)主要考虑三种数据流: TCP 流、基于 AIMD(1/5,1/8)的数据流(以下简称为 AIMD(1/5,1/8)流)、基于 D-AIMD 的数据流(以下简称为 D-AIMD 流),它们有着相同大小的数据帧(1500bytes)。

3.2 单数据流的仿真

单数据流的仿真采用的是 TCP 流和 D-AIMD 流,使用 S1-N1-N2-R1 的连接,仿真结果如图5、图6所示。从图中可以看出:(1)由于暂态时 D-AIMD 流采用的是 TCP 拥塞控制机制,因此它在有用资源的利用速度上大致与 TCP 流一样;(2)稳定状态下, TCP 流平均拥塞窗口约 80packets, D-AIMD 流平均拥塞窗口约 95packets,显然, D-AIMD 流比 TCP 流在资源的利用率上要高;(3)在相同的通信量的情况下, D-AIMD 流比 TCP 流发生的拥塞次数要少;(4)稳态时 D-AIMD 流比 TCP 流数据发送速率的波动性要小。

3.3 多数据流的仿真

多数据流的仿真采用的是:两个 D-AIMD 流、两个

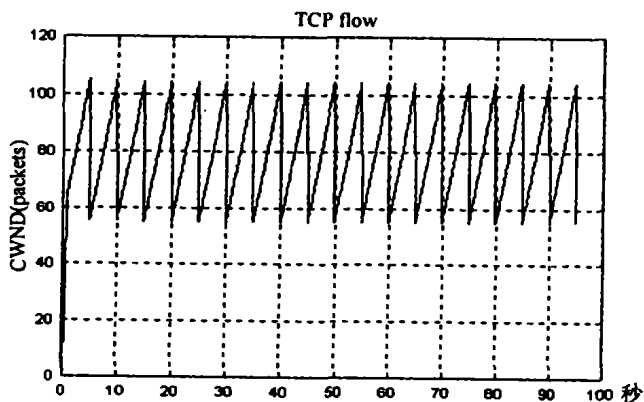


图5 TCP 流拥塞窗口

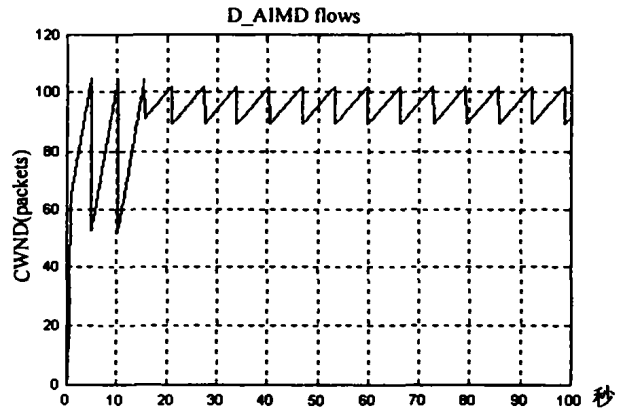


图6 D-AIMD 流拥塞窗口

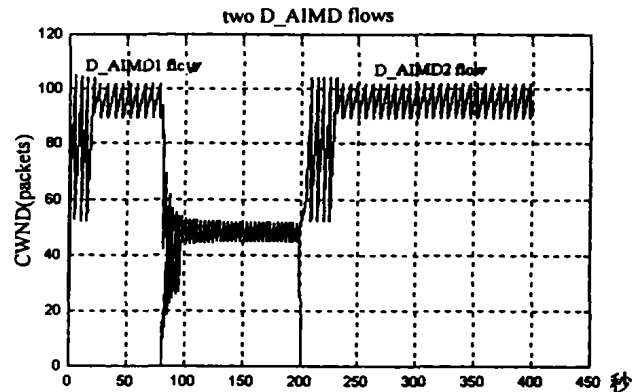


图7 两个 D-AIMD 流的拥塞窗口

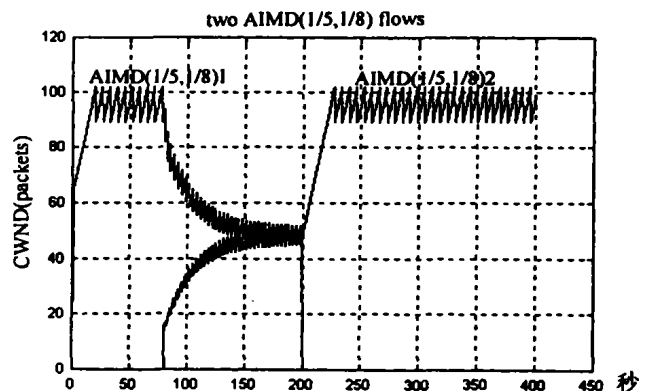


图8 两个 AIMD(1/5,1/8)流的拥塞窗口

AIMD(1/5,1/8)流、两个 TCP 流、一个 TCP 流与一个 D-AIMD 流。仿真结果分别如图7、图8、图9、图10所示。

图7、图8分别为两个 D-AIMD 流、两个 AIMD(1/5,1/8)流竞争网络资源的仿真结果,其中,第一个连接先建立,80s 时第二个连接加入,200s 后第一个连接断开。从图中易知:(1)两个 D-AIMD 流能够实现资源的公平竞争,当一个新的 D-AIMD 流加入,造成网络拥塞加重时,另一个 D-AIMD 流也会迅速降低其数据发送速率,并很快稳定;而当一个 D-AIMD 流断开后,另一个 D-AIMD 流也能快速使用有用的网络资源;(2)两个 AIMD(1/5,1/8)流能够实现资源的公平竞争,但当一个新的 AIMD(1/5,1/8)流加入,造成网络拥塞加重时,另一个 AIMD(1/5,1/8)流不会快速地降低其数据发送速率,系统达到新的稳态较慢;而当一个 AIMD(1/5,1/8)流

(下转封四)

(上接第 97 页)

断开后,另一个 AIMD(1/5,1/8)流使用有用网络资源的速度也较慢。

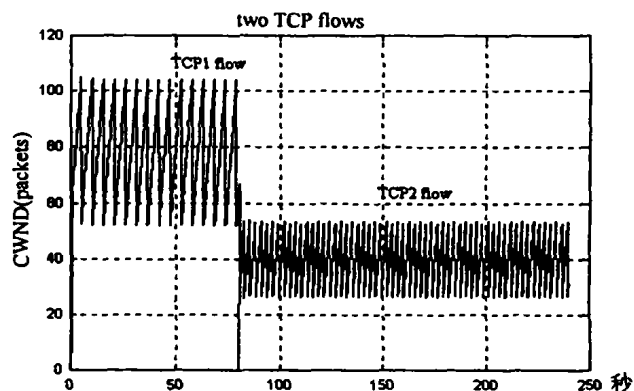


图 9 两个 TCP 流的拥塞窗口

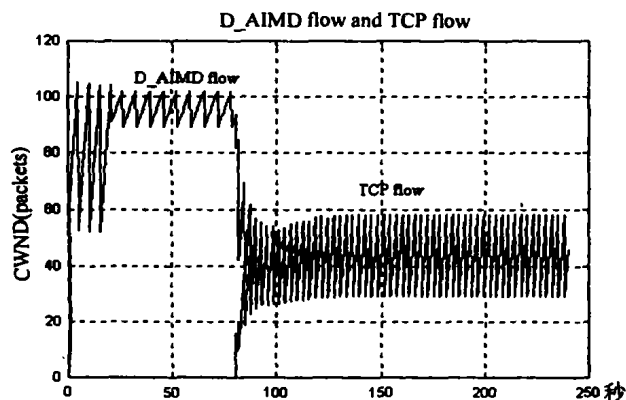


图 10 D AIMD 流和 TCP 流的拥塞窗口

图 9、图 10 分别为两个 TCP 流、一个 TCP 流与一个 D-AIMD 流竞争网络资源的仿真结果,其中,第一个连接先建立,80s 时第二个连接加入。从图中易知:(1)D-AIMD 流与 TCP 流之间能够实现网络资源的公平竞争;(2)D-AIMD 流的加入能够提高网络资源的利用率,譬如:稳态下,图 9 中两个 TCP 流其平均的拥塞窗口约 40packets,而图 10 中 D-AIMD 流与 TCP 流平均的拥塞窗口约 43packets。

结论 随着基于速率的应用业务的增多,无线应用的越来越广泛,对 TCP 拥塞控制机制的改进已成了网络研究领域一个重要的研究课题。本文所提出的改进的 TCP 拥塞控制(D-AIMD)仍基于 AIMD(a, b)算法,所不同的是在一个连接的通信过程中 a, b 并不唯一。其基本思想是将较大 a, b 下的拥塞控制机制与较小 a, b 下的拥塞控制机制有机地结合起来,将较大 a, b 下的拥塞控制机制用于系统暂态,将较小 a, b 下的拥塞控制机制用于系统稳态。通过仿真可看出, D-AIMD 具有以下优点:(1)系统暂态时,应用能快速地使用网络中的有用资源,网络拥塞程度加重时,又能快速减轻网络拥塞;(2)系统稳态时,数据发送速率波动性小,拥塞次数少,资源利用率高;(3)当它与 TCP 一起使用时,不仅可以与 TCP 之间实现资源的公平竞争,还可以提高网络资源的利用率;(4)实现简单,系统额外开销小。

参考文献

- 1 Postel J B. Transmission Control Protocol. Internet Engineering Task Force, Sep. 1981, RFC793
- 2 Jacobson V. Congestion Avoidance and Control. In: Proc. ACM SIGCOMM, Aug. 1988. 314~329
- 3 Floyd S. Connections with Multiple Congested Gateways in Packet-Switched Networks Part 1: One-Way Traffic. ACM Computer Communication Review, 1991, 21(5): 30~47. URL <http://www-nrg.ee.lbl.gov/nrg-papers.html/>
- 4 Padhye J, Firoiu V, Towsley D, Kurose J. Modeling TCP Throughput: A Simple Model and its Empirical Validation. SIGCOMM Symposium on Communications Architectures and Protocols, Aug. 1998. URL <http://www.acm.org/sigcomm/sigcomm98/tp/abs-25.html>
- 5 Chiu D-M, Jain R. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. Computer Networks and ISDN Systems, 1989. 1~14
- 6 Widmer J. Equation-based Congestion Control. Feb. 2000. Diploma Thesis, URL <http://www.aciri.org/tfrc/>
- 7 Floyd S, et al. Equation-Based Congestion Control for Unicast Applications: the Extended Version. March 2000; [ICSI Technical Report TR-00-03]. URL <http://www.aciri.org/tfrc/>

计算机科学

(1974 年 1 月创刊)

第 29 卷第 4 期 (月刊)

2002 年 4 月 25 日出版

中国标准刊号: ISSN 1002-137X
CN50-1075/TP

定价: 18.00 元 国外定价: 5 美元

邮发代号: 78-68

发行范围: 国内外公开

主管单位: 国家科学技术部

主办单位: 国家科技部西南信息中心

编辑出版: 《计算机科学》杂志社

重庆市渝中区胜利路 132 号 邮政编码: 400013

电话: (023) 63500828 E-mail: jsjcx@swic.ac.cn

社长: 牟炳林

主编: 朱宗元

印刷者: 重庆科情印务有限公司

总发行处: 重庆市邮政局

订购处: 全国各地邮政局

国外总发行: 中国国际图书贸易总公司 (北京 399 信箱)

国外代号: 6210-MO