

文章编号: 1001-0920(2004)07-0799-05

TCP 拥塞控制的混杂建模与分析

郑刚¹, 宋永华², 谭民¹

(1. 中国科学院自动化研究所复杂系统与智能科学重点实验室,
北京 100080; 2. Brunel 大学电子与计算机工程系, 伦敦, 英国)

摘要: 互联网中的拥塞控制是一个离散事件与连续变量相互作用的混杂动态过程。假定拥塞事件为一个更新过程序列, 采用含有一个状态的随机混杂自动机模型, 描述了 TCP 拥塞控制中的 A MD 过程。通过分析发送端拥塞窗口的演化行为, 发现拥塞窗口的演化过程与 TCP 连接的吞吐量具有稳定的统计特征, 与初始条件无关。解释了当路由器采用去尾策略时, 引起的流同步现象与分析结果的一致性。最后对分析结果进行了仿真实验。

关键词: 拥塞控制; 混杂系统; 混杂自动机

中图分类号: TN 913 **文献标识码:** A

Hybrid modeling and analysis of TCP congestion control

ZHENG Gang¹, SONG Yong-hua², TAN Min¹

(1. Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China; 2. Department of Electronic and Computer Engineering, Brunel University, London, U.K. Correspondent: ZHENG Gang, E-mail: zhenggang@comp.sys.ia.ac.cn)

Abstract: Based on the assumption that congestion events consist of a renewal process sequence, the TCP A MD congestion control is modeled as one-state stochastic hybrid automaton. By analyzing the evolutionary behavior of congestion window size of the TCP source, it is drawn that the statistic performances of both congestion window size and TCP connections throughput are stable, without respect to their initial conditions. In addition, it is formulated that the analysis results accord with the flow synchronization phenomenon resulting from dropping tail strategy taken in the network routers. The network simulation shows the validity of the method.

Key words: congestion control; hybrid system; hybrid automaton

1 引言

互联网中的拥塞现象主要表现在数据包的丢失和传输延时的增加, 使构成网络连接的端系统处理应用任务的性能下降, 严重影响了网络的稳定性和鲁棒性。拥塞控制是通过对互联网中拥塞现象的检测, 及时调整数据源的发送速率, 减小互联网的拥塞程度, 以保证整个互联网系统的稳定运行。

目前, 在互联网中普遍采用 TCP/IP 分层结构,

TCP 层实现了端到端的可靠数据传输, 除完成差错控制、流量控制外, 还完成拥塞控制的功能。TCP 拥塞控制对保证互联网的稳定性和鲁棒性起着关键的作用。Jacobson^[1] 提出端到端基于窗口的 TCP 拥塞控制算法, 通过检测发送端对接收端返回的 ACK 标志来调整拥塞窗口的大小, 从而改变发送速率。以此为基础, 逐渐形成了 Tahoe, Reno, New Reno, Vegas 和 SACK 等多种拥塞控制机制^[2,3]。与 TCP

收稿日期: 2003-06-27; 修回日期: 2003-09-26

基金项目: 国家基础研究重大项目前期研究专项基金资助项目 (2001CCA 02900); 中国科学院海外杰出学者基金资助项目

作者简介: 郑刚 (1974—), 男, 河南宜阳人, 博士生, 从事混杂系统理论及应用的研究; 谭民 (1962—), 男, 广东新会人, 研究员, 博士生导师, 从事机器人控制、多机器人协调等研究

层拥塞控制策略相适应,在IP层,通过在路由器中采用不同的队列管理策略和数据包调度算法,如先进先出(FIFO)、公平排队(FQ)和主动队列管理(AQM)策略等,有效地改善拥塞控制的效果^[2,3]。

建立TCP拥塞控制模型对于分析网络数据流特征,提高网络服务质量和网络资源的利用率非常重要。TCP拥塞控制建模的主要目标是找到网络吞吐量的表达式,从而了解网络参数对吞吐量的影响。文献[4~7]采用不同的方法,得到吞吐量与丢弃概率的平方根和平均往返时间成反比关系。由于考虑因素的不同,其形式可能略有差异。本文利用混杂系统的建模方法^[8],将互联网中发生的拥塞事件看成一个离散事件序列,虽然具有随机性,但符合一定的统计规律,在相邻的拥塞事件之间,数据包传输速率的变化过程是连续的,整个拥塞控制过程是连续状态与离散事件相互作用的混杂动态过程。文献[9]利用含有两个状态的确定性混杂自动机模型描述TCP拥塞控制机制,也得到了吞吐量与丢弃概率的平方根与往返时间的反比关系,但本文考虑了拥塞事件发生的不确定性,利用含有一个状态的随机混杂自动机模型描述TCP拥塞控制过程,得到了相似的结果,而且更具一般性。

2 TCP拥塞控制的随机混杂自动机模型

TCP拥塞控制主要是指调整发送端拥塞窗口的大小,它决定了发送端的数据传输速率。发送端根据来自接收端的ACK标志判断网络是否发生拥塞,从而对拥塞窗口的大小进行相应的调整,间接改变发送端的数据传输速率,减轻网络的拥塞程度。TCP拥塞控制是一种端到端的基于窗口调整的控制机制。

TCP拥塞控制机制有多种形式,本文针对常见的Reno机制展开讨论。拥塞控制分为慢启动、拥塞避免、快速重传和快速恢复4个阶段。在慢启动阶段,发送端每收到一个ACK标志,拥塞窗口增加1个数据包,拥塞窗口大小在慢启动阶段成指数增加。当超过慢启动阈值,进入拥塞避免阶段,发送端仅在收到对应前一次发送数据包的全部ACK标志后,拥塞窗口增加1个数据包大小,在此阶段拥塞窗口的增加是线性的,又称和式增加。如果发送端收到3个相同的ACK标志,则认为数据丢失发生拥塞,进入快速重传阶段。慢启动阈值减半,拥塞窗口变为慢启动阈值加上3个数据包大小,重传丢失数据包。随后进入快速恢复阶段,每收到一个ACK标志,拥塞窗口加1,直到收到新的ACK标志,表明发送端收到重

传数据包,快速恢复阶段结束,重新回到拥塞避免阶段。此时拥塞窗口变为慢启动阈值,即相当于拥塞事件发生时拥塞窗口的1/2,实现了积式减少。如果超过规定的重传定时器计时时间,仍未收到新的ACK标志,则拥塞窗口恢复一个数据包大小,重新进入慢启动阶段。这就是整个TCP拥塞控制过程,详细描述参见文献[2,10]。

慢启动阶段持续时间较短,在长时间网络连接中,其影响相比拥塞避免阶段很小,而快速重传之后的快速恢复阶段对网络连接吞吐量的影响也不明显。因此本文建立的模型主要考虑对拥塞控制起核心作用的“和式增加和积式减少(A MD)”过程,认为在发送端收到3个重复的ACK标志时,数据丢失发生拥塞,拥塞窗口减半,重传数据包,开始新的线性增加。这里不考虑超时对网络吞吐量的影响,关于超时的影响参见文献[5~7]。整个AMD拥塞控制过程为一个混杂过程,拥塞事件构成一个离散事件序列,在拥塞事件之间,如果不考虑接收端通告窗口的影响,则拥塞窗口是连续的线性增加,拥塞事件的发生使拥塞窗口的大小发生减半的跳变。本文采用含有一个状态的随机混杂自动机模型描述此过程。

定义1 如图1所示,随机混杂自动机H由如下七元组构成:

$$H = (Q, X, \text{Init}, f, \text{Inv}, G, R). \quad (1)$$

其中:Q表示离散状态集, $Q = \{q\}, q \in Q = \{0\}$,该系统具有一个离散状态;X表示连续状态集, $x = (W, \tau) \in X = \mathbb{R}^2$,其中W为拥塞窗口尺寸,τ为系统位于当前离散状态的持续时间;初始条件: $\text{Init} = \{0\} \times \{x \in \mathbb{R}^2 \mid W = w_0, \tau > 0\} \subseteq \mathbb{R}^2$;连续变量动态特征: $\dot{x} = f(x) = (\alpha, 1), \alpha = \text{Const}$;状态保持条件: $\text{Inv}(q) = \{\tau < T\} \subseteq X, T$ 为随机变量,符合某种概率分布 $P\{T = t\} = F(t), F(t) \in [0, 1]$;跳变条件: $G(q, q) = \{t \geq T\} \subseteq X$;重置关系: $R(q, q, x) = \{x \in \mathbb{R}^2 \mid W = W^- / 2, \tau = 0\}$ 。

注1 状态保持时间是拥塞窗口发生跳变的时间间隔,即拥塞事件间隔,因为每次跳变发生后,都

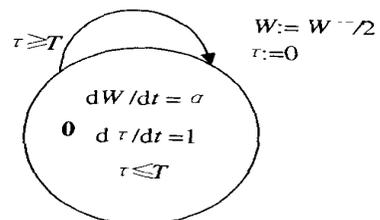


图1 TCP拥塞控制随机混杂自动机模型

间间隔,即拥塞事件间隔,因为每次跳变发生后,都

有 $\tau = 0$, 故可假定系统每次的状态保持时间是独立且同分布的, 即拥塞事件间隔是独立同分布的, 拥塞事件序列为一个更新过程序列

3 模型分析

设第 n 次拥塞事件发生的前一时刻拥塞窗口的大小为 W_n , 第 $n - 1$ 次拥塞事件与第 n 次拥塞事件发生的时间间隔为 T_n , $\{T_n\}$ 为更新过程序列, 期望 $E[T_n] = \int_0^\infty t dF(t) = \lambda^{-1}$, 方差 $D[T_n] = d$, 自相关函数 $R(k) = E[T_n T_{n-k}] = E[T_n]E[T_{n-k}] = \lambda^{-2}$, $k > 0, R(0) = E[T_n^2] = \lambda^{-2} + D(T_n) = \lambda^{-2} + d$. 令 α 表示拥塞窗口的增加斜率, 有如下随机线性差分方程表达式:

$$W_{n+1} = W_n/2 + \alpha T_n \quad (2)$$

引理 1^[7] 对随机线性差分方程 $X_{n+1} = vX_n + \alpha S_n$ (v 和 α 为常数, $v < 1$), 若 $\{S_n\}$ 为平稳遍历的随机过程序列, 则该方程有唯一平稳解 $X_n^* = \alpha \sum_{k=0}^\infty v^k S_{n-1-k}$, 且无论其初始状态如何, W_n 几乎处处收敛于该解, 即满足 $\lim_{n \rightarrow \infty} |X_n - X_n^*| = 0, a.s.$

利用引理 1, 因为 $\{T_n\}$ 为更新过程序列, 满足平稳遍历的条件, 故 W_n 几乎处处收敛于 W_n^* , 即有

$$W_n^* = \alpha \sum_{k=0}^\infty \left(\frac{1}{2}\right)^k T_{n-1-k} \quad (3)$$

由式 (3) 得其期望形式

$$E[W_n^*] = E\left[\alpha \sum_{k=0}^\infty \left(\frac{1}{2}\right)^k T_{n-1-k}\right] = \alpha \sum_{k=0}^\infty \left(\frac{1}{2}\right)^k E[T_{n-1-k}] = \alpha \lambda \quad (4)$$

拥塞事件的随机性导致拥塞窗口演化过程也具有随机性, 下面对拥塞窗口变化的统计特征进行分析

命题 1 拥塞窗口的均值为 $\bar{W} = \frac{3\alpha\lambda}{2} + \frac{\alpha d}{2\lambda}$

证明 用 $\tau_k (k = 1, 2, \dots)$ 表示第 k 次拥塞事件发生时刻, τ_0 表示初始时刻, $\tau_k - \tau_{k-1} = T_k$. 根据 $\{T_n\}$ 的平稳遍历性和 $\{W_n\}$ 的收敛性, 拥塞窗口的均值为

$$\bar{W} = E[W(t)] = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t W(u) du = \lim_{n \rightarrow \infty} \frac{E\left[\sum_{k=0}^n \int_{\tau_k}^{\tau_{k+1}} \left(\frac{1}{2}W_k + \alpha t\right) dt\right]}{E[T_k]} = \frac{3\alpha\lambda}{2} + \frac{\alpha d}{2\lambda}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{E\left[\sum_{k=1}^n \left(\frac{1}{2}W_k T_k + \frac{1}{2}\alpha T_k^2\right)\right]}{n\lambda} &= \\ \lim_{n \rightarrow \infty} \frac{E\left[\sum_{k=1}^n \left\{ \frac{1}{2}\alpha \sum_{j=0}^{k-1} \left(\frac{1}{2}\right)^j T_{k-1-j} T_k + \frac{1}{2}\alpha E[T_k^2] \right\}\right]}{n\lambda} &= \\ \lim_{n \rightarrow \infty} \frac{E\left[\sum_{k=1}^n \left\{ \alpha \sum_{j=0}^{k-1} \left(\frac{1}{2}\right)^{j+1} T_{k-1-j} T_k \right\} + \frac{1}{2}n\alpha R(0)\right]}{n\lambda} &= \\ \lim_{n \rightarrow \infty} \frac{\alpha \sum_{k=1}^n \left[\alpha \sum_{j=0}^{k-1} \left(\frac{1}{2}\right)^{j+1} R(j+1) \right] + \frac{1}{2}n\alpha R(0)}{n\lambda} &= \\ \lim_{n \rightarrow \infty} \frac{n \left[\alpha \sum_{j=0}^\infty \left(\frac{1}{2}\right)^{j+1} R(j+1) \right] + \frac{1}{2}n\alpha R(0)}{n\lambda} &= \\ \frac{\alpha R(j+1) \sum_{j=0}^\infty \left(\frac{1}{2}\right)^{j+1} + \frac{1}{2}\alpha R(0)}{\lambda} &= \\ \frac{\alpha\lambda^2 + \frac{1}{2}\alpha(\lambda^2 + d)}{\lambda} = \frac{3\alpha\lambda}{2} + \frac{\alpha d}{2\lambda} \end{aligned}$$

若平均往返时间为 RTT , $X(t)$ 为数据包的传输速率, 由命题 1 可得, 在该 TCP 连接另一端——目的节点接收数据包的平均速率, 即为 TCP 连接的平均吞吐量

$$\bar{X} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X(u) du = \frac{\bar{W}}{RTT} = \frac{1}{RTT} \left(\frac{3\alpha\lambda}{2} + \frac{\alpha d}{2\lambda} \right) \quad (5)$$

拥塞的发生是由数据包丢失引起的, 由于拥塞发生的随机性, 数据包丢弃概率(拥塞发生概率)也是衡量网络拥塞程度的一个重要指标, 丢弃概率即为丢失数据包量与共发送数据包量的比值. 在 AIMD 拥塞控制过程中, 一旦拥塞发生, 拥塞窗口减半, 重传数据包, 故丢弃概率实际上又表现为拥塞次数与发送数据包量之比

定义 2 丢弃概率

$$p = \lim_{t \rightarrow \infty} \frac{t/\lambda}{\int_0^t X(u) du} = \frac{1}{\lambda \bar{X}} \quad (6)$$

TCP/IP 协议中规定接收端收到 $b (b = 1 \text{ 或 } 2)$ 个数据包后, 返回给发送端一个 ACK 标志, 则拥塞窗口的线性增加斜率为 $\alpha = \frac{1}{bRTT}$.

综合上述分析, 加入丢弃概率 p , 可得到平均吞吐量的另一种表示形式

命题 2 平均吞吐量

$$\bar{X} = \frac{1}{RTT \sqrt{bp}} \sqrt{\frac{3}{2} + \frac{d}{2\lambda^2}} \quad (7)$$

证明

$$\begin{aligned} \bar{X} &= \frac{1}{RTT} \left(\frac{3\alpha\lambda}{2} + \frac{\alpha d}{2\lambda} \right) = \frac{1}{RTT} \alpha \left(\frac{3\lambda}{2} + \frac{d}{2\lambda} \right) = \\ &= \frac{\lambda}{bRTT^2} \left(\frac{3}{2} + \frac{d}{2\lambda^2} \right) = \frac{1}{bRTT^2 p \bar{X}} \left(\frac{3}{2} + \frac{d}{2\lambda^2} \right) \Rightarrow \\ \bar{X} &= \frac{1}{RTT \sqrt{bp}} \sqrt{\frac{3}{2} + \frac{d}{2\lambda^2}} \end{aligned}$$

相应地, 拥塞窗口的均值可表示为另一种形式, 即命题 3

命题 3 拥塞窗口均值为

$$\bar{W} = \frac{1}{\sqrt{bp}} \sqrt{\frac{3}{2} + \frac{d}{2\lambda^2}} \quad (8)$$

经过上述分析, 可得到如下结论:

1) 本文的随机混杂自动机模型建立在假设拥塞事件为一个更新过程序列的基础上, 它要求相邻的拥塞事件间隔是独立同分布的。在具有这种特征的网络连接中, 拥塞窗口变化的均值以及网络连接的平均吞吐量仅与拥塞间隔的概率分布有关, 与网络连接的初始条件无关, 由此表现的数据包丢弃概率也是固定的

2) 根据式(4)的结果和 w_n 的收敛性, 当多个发送端与接收端建立多个TCP连接时, 如果相邻的拥塞间隔相互独立且具有相同的概率分布, 则发送端的拥塞窗口随时间的推移, 趋于相同的演化过程

对于互联网中采用去尾排队策略的路由器, 在队列满时, 丢弃后继数据包, 拥塞间隔可认为是独立同分布的。如果有多个TCP数据流流向该路由器, 当拥塞发生时, 各个TCP数据源同时将拥塞窗口减半, 并开始线性增加。由上述分析结果可知, 不同TCP数据流拥塞窗口变化的均值相同, 且与初始条件无关。根据引理1中 w_n 的收敛性, 可推断拥塞窗口趋于相同的演化过程, 这与实际中由“去尾”策略引起的“流同步”现象是一致的。如果采用RED队列管理策略^[11], 根据路由器中队列长度的不同, 具有不同的丢弃概率, 拥塞间隔则不再同分布, 不同数据流具有不同的演化过程, 将避免流同步现象的发生

若拥塞事件为Poisson序列, 即拥塞发生间隔服从指数分布, 则有 $d/\lambda^2 = 1$ 。由命题2, 得 $\bar{X} =$

$$\frac{1}{RTT} \sqrt{\frac{2}{bp}}, \text{ 平均吞吐量与丢弃概率 } p \text{ 的平方根成反}$$

比关系, 其形式与文献[4~7]的结果相似

4 模型验证

利用NS-2网络仿真软件, 模拟TCP拥塞控制过程, 验证上述分析结果。采用图2所示的哑铃型网络拓扑结构, 节点 S_1, \dots, S_n 为TCP数据源, 节点 N_1 为路由节点, 采用“去尾”排队策略, 节点 N_2 为目的节点。源节点与路由节点之间建立TCP连接, 传输链路为10MB的高速链路, 而路由节点到目的节点为瓶颈链路, 传输带宽1.5MB。

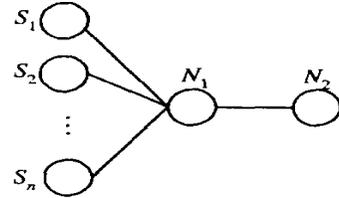


图2 网络仿真拓扑结构

图3给出了数据源节点为3个时拥塞窗口的演化过程, 图4为节点 N_2 吞吐量的变化过程, 并且指出3个数据源开始发送数据的时间是随机的, 并不同时。

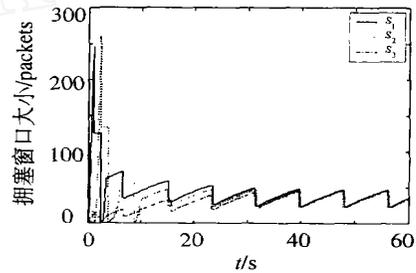


图3 源节点拥塞窗口的演化过程

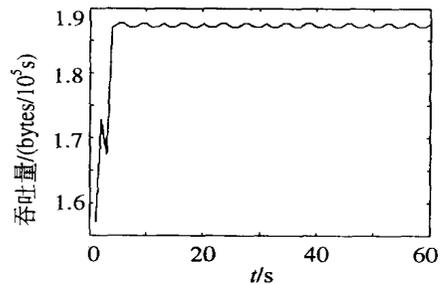


图4 N_2 吞吐量的变化过程

从仿真结果可以看出, 尽管3个数据源经历慢启动阶段的先后顺序不同, 具有不同的初始条件, 但随着时间的推移, 拥塞窗口同时增大或减小, 趋于相同的演化过程, 与初始条件无关。从节点1吞吐量的变化过程看, 在经历了初始阶段的迅速增加后, 稳定在某个量附近。这与前面的分析结果是一致的

5 结 语

本文利用随机混杂自动机模型, 分析了TCP层

上的AMD拥塞控制过程,其理论分析结果与仿真结果是一致的。在网络中,还存在许多具有混杂特征的网络现象,如何利用混杂系统的方法进行有效地描述和分析,解决一些复杂的问题,是进一步的研究方向。

参考文献(References):

- [1] Jacobson V. Congestion avoidance and control[J]. *Proc ACM SIGCOMM 88 in ACM Computer Communication Review*, 1988, 18(4): 314-329.
- [2] Stallings W. *High-speed Networks and Internets: Performance and Quality of Service*[M]. 2nd edition. New York: Prentice Hall, 2002.
- [3] 罗万明, 林闯, 阎保平. TCP/IP 拥塞控制研究[J]. *计算机学报*, 2001, 24(1): 1-18.
(Luo W M, Lin C, Yan B P. A survey of congestion control in the internet[J]. *Chinese J of Computers*, 2001, 24(1): 1-18.)
- [4] Mathis M, Semke J, Mahdavi J, et al. The macroscopic behavior of the TCP congestion avoidance algorithm[J]. *ACM Computer Communication Review*, 1997, 27(3): 67-82.
- [5] Padhye J, Firoiu V, Towsley D, et al. Modeling TCP throughput: A simple model and its empirical validation[A]. *Proc ACM SIGCOMM 98*[C]. Vancouver, 1998. 303-314.
- [6] Misra V, Gong W B, Towsley D. Stochastic differential equation modeling and analysis of TCP window size behavior[A]. *Proc Performance 99*[C]. Istanbul, 1999. 553-562.
- [7] Altman E, Avrachenkov K, Barakat C. A stochastic model of TCP/IP with stationary random losses[A]. *Proc ACM SIGCOMM 00*[C]. Stockholm, 2000. 231-242.
- [8] Schaff A, Schumacher H. *An Introduction to Hybrid Dynamical Systems*[M]. Berlin: Springer-Verlag, 2000.
- [9] Hespanha J, Bohacek S, Obraczka K, et al. Hybrid modeling of TCP congestion control[A]. *Hybrid Systems: Computation and Control*[C]. Berlin: Springer-Verlag, 2001. 291-304.
- [10] Floyd S. *Congestion Control Principles*[M]. RFC 2914, IETF, 2000.
- [11] Floyd S, Jacobson V. Random early detection gateways for congestion avoidance[J]. *IEEE/ACM Transactions on Networking*, 1993, 1(4): 397-413.

第 16 届中国控制与决策学术年会在黄山召开

本刊讯 2004 中国控制与决策学术年会(16thCDC)于5月29日至6月1日在安徽省黄山市召开。来自国内部分高等院校和科研院所的108余名代表参加了本届年会。

5月29日上午举行了大会开幕式。CDC年会程序委员会副主任委员郭治教授致开幕辞,东北大学信息科学与工程学院副院长高宪文教授致欢迎辞,中国海洋大学博士生导师唐功友教授代表参会人员讲话,中国兵工学会自动控制专业委员会、中国航空学会自动控制分会等单位分别发来贺信和贺电。

开幕式后举行了大会学术报告会,本届年会邀请6位著名专家学者作了专题学术报告。报告人及报告题目分别是:南京理工大学郭治教授:待机控制;西安交通大学万百五教授:基于数据挖掘技术的工业生产产品质量控制;香港中文大学黄捷教授:非线性输出调节的理论与应用;北京航空航天大学贾英民教授:车辆系统的鲁棒解耦控制;上海交通大学李少远教授:大范围工况系统的优化控制;东北大学徐心和教授:机器人控制。这些报告从不同方面分别介绍了不同前沿学科的最新研究成果和未来发展方向,受到与会代表的热烈欢迎和高度评价。

会议期间发行了《2004 中国控制与决策学术年会论文集》,本届年会论文集共收入各方面论文281篇。在分组交流中,代表们宣读了各自的论文,并且回答了其他代表提出的问题,整个交流活动进行得热烈而有序,充满了浓厚的学术气氛。通过广泛交流和深入探讨,本届年会共评选出19篇优秀学术论文。

第16届中国控制与决策学术年会由《控制与决策》编辑委员会、中国航空学会自动控制分会、中国自动化学会应用专业委员会、中国运筹学会决策理论及应用专业委员会、中国兵工学会自动控制专业委员会联合主办,东北大学信息科学与工程学院具体承办。