文章编号: 1001-0920(2006)11-1244-05

基于时空混沌系统构造Hash函数

刘光杰, 单 梁, 孙金生, 戴跃伟, 王执铨
(南京理工大学 自动化学院, 南京 210094)

摘 要: 提出一种基于时空混沌系统的单向Hash函数构造方法. 该方法通过使用单向耦合映射格子和基于迭代Logistic映射的初始状态生成函数实现明文和密钥信息的混淆和扩散, 并基于密码块连接方式产生任意长度明文的128位Hash值. 理论分析和实验表明, 提出的Hash函数满足Hash函数所要求的单向性、初值以及密钥敏感性和抗碰撞性等安全性能要求.

关键词: 时空混沌; 耦合映像格子; Hash函数
中图分类号: TP273 **文献标识码**: A

Construction of Hash Function Based on Spatiotemporal Chaotic Systems

LIU Guang-jie, SHAN Liang, SUN Jin-sheng, DAI Yue-wei, WANG Zhi-quan
(Institute of Automation, Nanjing University of Science and Technology, Nanjing 210094, China. Correspondent: LIU Guang-jie, E-mail: guangj_liu@yahoo.com.cn)

Abstract: A one-way Hash function is constructed based on spatiotemporal chaotic systems. The one-way coupled map lattices (OCMLs) and initial condition generation function based on iterative Logistic map are used to realized the data confusion and diffusion between plain-text and key information. And the cipher block chaining mode is used to generate the 128-bit Hash value for plaint-text with arbitrary length. Theoretical analysis and experimental results indicate that the proposed algorithm can satisfy the required security performance, such as one-way, initial value and key sensitivity and collision resistance.

Key words: Spatiotemporal chaos; Coupled map lattices; Hash function

1 引 言

Hash函数又称为单向散列函数, 它在现代密码学中起着非常重要的作用, 可作为文件的唯一表示而用于内容标识和认证. 随着互联网和电子商务以及数字文档等应用的不断兴起和广泛应用, 对Hash函数的要求越来越高, 而密码分析的手段也不断地提高. 传统的Hash函数如MD2, MD4, MD5, SHA等[1], 为获得最终的Hash值, 需要进行大量复杂的异或和位操作运算, 效率不高且不够安全. 最近, 王小云等人成功破解了MD5, 以及SHA-1[2]等过去被认为足够安全的Hash函数, 因此更为安全高效的Hash函数的研究是非常必要的.

混沌密码学是现代密码学中一个比较重要的分支, 混沌由于其本身具有的丰富的非线性复杂性在混沌块密码[3]以及伪随机数发生器[4,5]方面都取得了重要的应用. 最近, 基于混沌的Hash函数的研究也取得了一定的进展. Yi[6]提出了一种基于混沌帐篷映射的Hash函数构造方法; Xiao等人[7]研究了基于具有可变参数的分段线性混沌映射的Hash函数构造方法, 基于Chebyshev混沌映射的构造方法[8], 和基于混沌神经网络的构造方法[9]; 王小敏

收稿日期: 2005-09-05; **修回日期**: 2005-12-12
基金项目: 国家自然科学基金项目(60374066); 江苏省自然科学基金项目(BK2004132); 教育部博士点基金项目(20020288025).
作者简介: 刘光杰(1980—), 男, 江苏徐州人, 博士生, 从事多媒体信息安全的研究; 王执铨(1939—), 男, 武汉人, 教授, 博士生导师, 从事动态大系统控制、混沌控制等研究.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

等[10]也提出了一种基于广义混沌映射切换的单向Hash 函数; 李红达等人提出了基于符合混沌动力系统的构造方法[11]. 这些研究均是基于低维的混沌动力系统, 设计基于块密码的Hash 函数

本文考虑时空混沌系统在时空域所具有的复杂非线性动力学特性, 设计了基于单向耦合映射格子(OCML)的单向Hash 函数构造方法 OCML 和初始状态生成函数共同用于实现明文和密钥信息的混淆和扩散, 并基于密码块连接实现任意长的明文消息到128 位的Hash 值的Hash 函数, 理论和实验显示了该方法具有很好的密钥和明文敏感性, 且具有一定的抗碰撞性

2　时空混沌系统构造Hash 函数的可行性

单向Hash 函数$H(M)$ 作用于任意长度的明文消息M, 它返回一个具有固定长度的Hash 值h. 即Hash 函数应该有压缩特性以及单向特性, Hash 函数的单向性可描述为:

给定M, 很容易计算h;

给定h, 根据$H(M)=h$ 计算M 很难;

给定M, 要找到另一消息M'并满足$H(M)=H(M')$ 很难

此外, 在一些其他的应用背景要求下, 仅具有单向性是不够的, 还需要抗碰撞性, 即要找出两个随机的消息M 和M'满足$H(M)=H(M')$, 在计算上很困难

时空混沌系统是一个无穷维的动力学系统, 是一个分布系统而不是集总系统 时空复杂行为虽然表现多样, 但却有着共同的特性 首先, 它们一般处于开放的远离平衡的系统中; 其次, 存在着非线性的相互作用, 且发展过程是不可逆的; 再次, 系统对初值变化和参数改变具有很强的敏感性 和一般的低维度混沌系统相比, 时空混沌系统具有更好的数据混淆和扩散特性, 因此也被用于设计流密码[12] 和块密码系统[13]. 时空混沌具有的不可逆性(单向性)、良好的混淆和扩散特性以及对密钥的敏感性, 使得其在理论上可以用于设计性能较好的Hash 函数

3　基于时空混沌系统的Hash 函数

3.1　单向耦合映像格子

单向耦合映像格子系统(OCML), 是指一个有限维且具有周期边界条件的格子系统 任意格子n 在i 时刻的状态变量为$x_n(i)$, 它的局部更新规则由下面的映射给出:

$$x_n(i)=F(x_n(i-1),x_{n-1}(i-1)),$$

或

$$x_n(i)=F(x_n(i-1),x_{n-1}(i+1)),\qquad(1)$$

其中F 是关于它的变量的连续函数 本文研究的是一种格子映射为分段线性混沌映射(PLCM)[14] 的单向耦合映像格子模型, 即

$$\begin{aligned}x_n(i+1)&=(1-\epsilon)f(x_n(i))+\\&\quad\epsilon f(x_{n-1}(i)),\\n&=1,2,\cdots,N,\end{aligned}\qquad(2)$$

且满足周期边界条件

$$x_0(i)=x_N(i),\ \forall i\in Z.\qquad(3)$$

其中: ϵ 为耦合参数, 本文取$\epsilon=2/3$; 映射f 即分段线性混沌映射, 可描述为

$$X(k+1)=f(X(k),Q)=\begin{cases}X(k)/Q, & 0\le X(k)<Q;\\(X(k)-Q)/(0.5-Q), & Q\le X(k)<0.5;\\(1-Q-X(k))/(0.5-Q), & 0.5\le X(k)<1-Q;\\(1-X(k))/Q, & 1-Q\le X(k)\le 1.\end{cases}\qquad(4)$$

这里Q 为满足$Q\in(0,0.5)$ 的控制参数, 本文取$Q=0.25$.

3.2　基于密码块连接的Hash 函数构造

图1 给出了一个基于16 维的单向耦合映像格子的Hash 单元, 该单元的运算可由下面的方程描述:

$$h=H(P,\text{key}),\ X^0=G(\text{key}),\ X^1=T(X^0,P),\ h=C(X^1).\qquad(5)$$

其中: P 为128 位的明文数据, key 为128 位的密钥, X^0 为耦合映像格子的初始状态值, X^1 为经过OCML 迭代演化后的状态值, h 为128 位的Hash 值

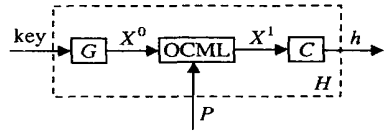

图1　Hash 函数的密码块结构

这里G 是一个初始状态生成函数, 用以产生OCML 的初始格子状态值$X^0=[a_1,a_2,\cdots,a_{16}]$. 为此, 首先将128 位密钥按先后顺序分解成4 个32 位的整数: V_1,V_2,V_3,V_4; 然后通过除2^{32} 量化至$[0,1]$ 之间的小数: v_1,v_2,v_3,v_4; 并基于混沌Logistic 映射g: $y=4x(1-x)$, 以式(6) 产生16 个格子的初始值

$$\begin{aligned}a_{4k-3}&=g^T(v_k),\ a_{4k-2}=g^T(a_{4k-3}),\\a_{4k-1}&=g^T(a_{4k-2}),\ a_{4k}=g^T(a_{4k-1}),\\k&=1,2,3,4\end{aligned}\qquad(6)$$

其中: g^T 表示Logistic 映射迭代T 次, 本文取$T=16$.

单向耦合映像格子的演化方程T 可描述为

$$\begin{aligned}x_{4k+1}(i+1)&=(1-\epsilon)f(x_{4k+1}(i))+\\&\quad\epsilon f[(x_{4k}(i)+p_k)\bmod 1],\end{aligned}$$

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$x_{4k+2}(i+1) = (1-\epsilon)f(x_{4k+2}(i)) + \epsilon f(x_{4k+1}(i)),$$
$$x_{4k+3}(i+1) = (1-\epsilon)f(x_{4k+3}(i)) + \epsilon f(x_{4k+2}(i)),$$
$$x_{4k+4}(i+1) = (1-\epsilon)f(x_{4k+4}(i)) + \epsilon f(x_{4k+3}(i)),$$
$$k = 0, 1, 2, 4;\ i = 0, \cdots, 11. \tag{7}$$

其中: p_1, p_2, p_3, p_4 分别是 128 位明文数据对应的 4 个 32 位整数除 2^{32} 量化得到的[0, 1] 区间上的小数, $x_i(0) = a_i, i = 1, \cdots, 16$, 且有边界条件 $x_0(i) = x_{16}(i)$. 经过 24 轮演化迭代后, 得

$$X^1 = [x_1(24), x_2(24), \cdots, x_{16}(24)] = [b_1, b_2, \cdots, b_{16}],$$

其中 $b_1, b_2, \cdots, b_{16}$ 均为[0, 1] 上的小数

图 1 中的 C 为压缩函数, 对 $X^1 = [b_1, b_2, \cdots, b_{16}]$, 首先乘 2^{32} 并取整得到 $B_i = \mathrm{int}[2^{32} \times b_i], i = 1, \cdots, 16$. 明文 P 对应的 128 位 Hash 值 h 为

$$c_1 = B_1 \oplus B_2 \oplus B_3 \oplus B_4,$$
$$c_2 = B_5 \oplus B_6 \oplus B_7 \oplus B_8,$$
$$c_3 = B_9 \oplus B_{10} \oplus B_{11} \oplus B_{12},$$
$$c_4 = B_{13} \oplus B_{14} \oplus B_{15} \oplus B_{16},$$
$$h = c_1 \odot c_2 \odot c_3 \odot c_4. \tag{8}$$

这里: c_1, c_2, c_3, c_4 均为 32 位二进制整数, $\odot$ 表示位连接, $\oplus$ 表示位异或. 图 1 所示的 Hash 单元可将 128 位的明文数据映射为 128 位的 Hash 值. 通过引入密码块链接模式(CBC), 可对具任意长度的明文数据产生 128 位的 Hash 值, 密码块链接模式如图 2 所示

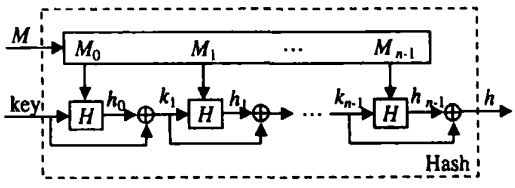

图 2 CBC Hash 函数模型

任意长的明文数据首先要进行位填充为 M, 保证 M 的长度为 128 的倍数. 然后将 M 分成 n 个 128 位的子明文块, 分别为 $M_0, M_1, \cdots, M_{n-1}$. 整个基于 CBC 的 Hash 函数可描述为

$$h_i = H(k_i, M_i) \oplus k_i,\ k_0 = \mathrm{key},\ h = h_{n-1} \oplus k_{n-1}. \tag{9}$$

在该模型中, 上一个环节输出的 Hash 值和上一个环节的密钥进行异或作为本环节的 128 密钥, 这一过程保证了不同明文数据块之间信息的更好的混淆和扩散

4 安全性分析

4.1 文本数据的 Hash 结果

根据本文提出的方法, 选择密钥"EDFC04F13CA8A2A0634CD57C4953F2D2", 分别计算了下面 5 种情况下文本的 Hash 值:

1) Secure communications based on chaos synchronization have attracted much attention in both fields of practical communication and chaos application since the pioneering work of Pecora and Carroll[1-8]. However, in the nonlinear science community as well as in the communication community the opinions on chaos communications have changed considerably during the past decade;

2) 将上述文字中的第一个"in"改为"on";

3) 将"communications"改为"communication";

4) 在文本最后增加一个空格;

5) 将十六进制密钥"EDFC04F13CA8A2A0634CD57C4953F2D2", 改为"EDFC04F13CA8A2A0634CD57C4953F2D1".

上面 5 种情况计算得到的 Hash 值如下:

1) 67ACA29EF99F62CC04B565515041D1E7;

2) 2C9DB1A2FC5846E8A57EBE85E0E7C9D1;

3) D3206951455DADA96DB8D9B9607A4893;

4) 58839A26E384A38A12572755CC2770B8;

5) EFC5D927BF9E593FCD9A395939F58084.

从上面的仿真结果可以看出, 本文的 Hash 函数对初值的变化相当敏感, 即使是很小的明文改变也会导致最终得到的 Hash 值发生很大的变化, 此外 Hash 函数对密钥的变化也相当敏感

4.2 单性性分析

从第 3 章的叙述可以看出, 从明文消息 M 和密钥 K 计算 Hash 值是非常容易的. 由式(6), 耦合格子的初值状态的生成函数是不可逆的混沌 Logistic 映射, 耦合映像格子中的映射(4) 也是不可逆的分段线性函数, 因此根据最终的 Hash 值计算明文 M 和密钥 K 是非常困难的. 若在密钥未知情况下, 进行穷尽搜索攻击, 对一个仅有 10 比特的明文消息来说, 也要在 2^{138} 的穷举空间中进行尝试

4.3 明文和密钥安全性分析

在本文所提算法中, 基于 Logistic 映射的初始状态生成函数和基于分段线性映射的 16 维耦合映像格子, 实现了对明文信息和密钥信息的混合, 这一混合操作实现了密码编码所必须的混淆和扩散. 这种良好的混淆和扩散作用, 保证了所构造的 Hash 函数对统计攻击的安全性. 理论上混淆和扩散特性越

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

好, 最终所得到的Hash 值对密钥和明文的敏感性越强, 因此本文就所构造的Hash 函数对明文和密钥的敏感性进行了测试

对一个二进制表示的 128 位Hash 值而言, 其每个位置的值非 1 即 0, 因此理想的敏感性应保证任何明文或者密钥的轻微改变将导致 Hash 比特发生 50% 的变化概率. 对一个具有 1 024 个比特的明文消息每次改变其一位上的值, 即将第 i 个比特的"0"("1") 改为"1"("0"), 计算改变后的明文消息的 Hash 值 h_i, 然后将其和原始消息的Hash 值 h_0 进行比较, 并计算 h_i 和 h_0 二进制表示的 Hamming 距离 $D(h_i, h_0)$, 最终得到Hash 比特变化率

$$r(i) = \frac{D(h_0, h_i)}{128} \times 100\%. \qquad (10)$$

图 3 给出了比特率的分布情况, 从图中可以看出, 比特变化率非常接近理想的 50% (64 比特), 说明所构造的Hash 函数具有非常好且稳定的明文敏感性. 明文消息的任一微小变换都会导致Hash 值发生较大的变化, 保证了当攻击者可进行选择明文攻击时, 根据已知的明文 - 密文对很难伪造和推导出其他的明文 - 密文对.

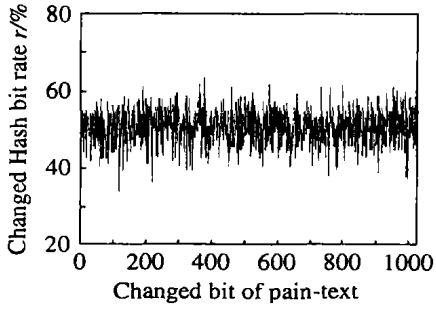

图 3　明文敏感性分析

在本文提出的Hash 函数中, 128 位密钥通过基于迭代Logistic 映射的生成函数产生 16 个(0, 1) 之间的小数作为耦合映像格子的初始状态, 因此密钥空间的大小为 2^{128}, 保证了密钥对任何的密钥穷举是安全的. 通过下面的实验可以看到Hash 值对密钥也是相当敏感的. 对 128 位的密钥"ED FC 04F1 3CA 8A 2A 0634CD 57C 4953F 2D 2"和 4.1 中的文本明文, 每次改变密钥中的一位, 即将第 i 个比特的"0"("1") 改为"1"("0"), 计算对应的Hash 值, 并计算如式(10) 的Hash 比特变化率. 图 4 给出了Hash 比特变化率的分布情况

从图 4 可看出, Hash 值的比特变化率接近理想的 50%, 因此具有较好的密钥敏感性保证了对统计分析的安全

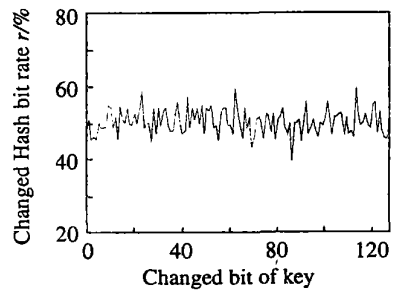

图 4　密钥敏感性分析

4.4　抗生日攻击和碰撞攻击分析

对生日攻击而言, Hash 值的比特长度决定了密码系统的安全性. 对本文 Hash 函数而言, 128 位 Hash 值长度意味着 2^{64} 的攻击难度. 这个数量级的攻击难度对一般应用来说是足够的.

Hash 函数的抗碰撞性是指找到任意两个不同明文具有同样的Hash 值在计算上是不可行的. 下面的实验对本文提出的基于时空混沌的 Hash 函数(SCH) 的抗碰撞性进行了初步的测试. 首先选择 Hash 值的前 8 个和 16 个比特作为 128 位Hash 值的摘要比特 d, 对应地取二进制明文消息的比特长度也分别为 8 和 16, 这样可使得Hash 函数的原像空间等于像空间. 设像空间中具有 k 个原像点的像点数目为 $N(k)$. 从抗碰撞性的要求考虑, $N(1)$ 越大, 发生碰撞的概率越小. 因此从 $N(k)$ 的分布情况, 可观察到Hash 函数的抗碰撞性能, 记 $n(k)$ 为

$$n(k) = \frac{N(k)}{\sum_{k=0}^{K} N(k)}. \qquad (11)$$

其中: K 为发生最大碰撞的数值, 对一般 Hash 函数而言 K 不会超过 12, 这里记 $K = 15$.

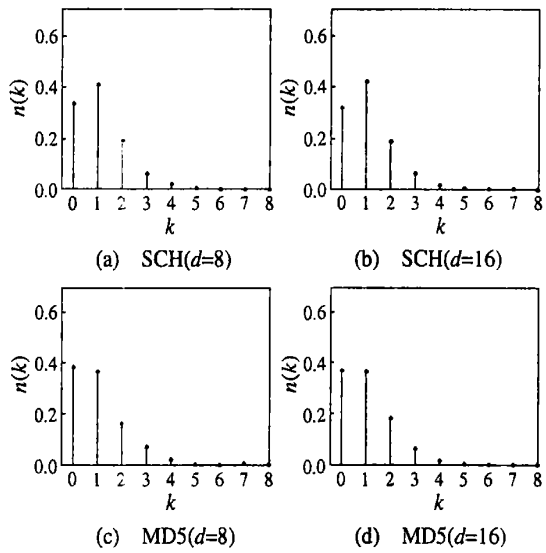

图 5　SCH 和MD5 的 k - n(k) 对比

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

图 5(a) 和 5(b) 分别给出了 $d = 8, 16$ 时 SCH 的 $n(k)$ 分布情况. 将同样的实验用于MD5, 图 5(c) 和 5(d) 分别给出了 $d = 8, 16$ 时MD5 的 $k - n(k)$ 分布情况. 从图中可以看出, 在摘要比特 8 和 16 情况下, SCH 的抗碰撞性均好于传统的MD5.

5 结 论

本文提出一种基于时空混沌系统的Hash 函数构造方法. 通过基于分段线性函数的单向时空耦合映像格子和基于迭代Logistic 映射的初始状态, 生成函数来实现明文信息和密钥信息的混淆和扩散, 并基于密码块连接模式实现任意长度明文消息到 128 位摘要之间的单向Hash 函数. 相关实验表明, 该Hash 函数具有很好的明文和密钥敏感性, 以及足够大的密钥空间和较好的抗碰撞性. Hash 函数在安全性能上的这些优点使其适合于数据签名和认证等诸多领域.

参考文献(References)

[1] Vanstone S A, Menezes A J, Oorschot P C. *Handbook of Applied Cryptography* [M]. CRC Press, 1996.

[2] Wang X Y, Yu H B. How to Break MD5 and Other Hash Functions [A]. *Advances in Cryptology - Eurocrypt* [C]. LNCS 3494, 2005: 19-35.

[3] Pareek N K, Patidar V, Sud K K. Discrete Chaotic Cryptography Using External Key [J]. *Physics Letters A*, 2003, 309(1-2): 75-82.

[4] Stojanovski T, Kocarev L. Chaos-based Random Number Generators — Part I: Analysis [J]. *IEEE Trans on Circuits and Systems I — Foundamental Theory and Application*, 2001, 48(3): 281-288.

[5] Stojanovski T, Kocarev L. Chaos-based Random Number Generators— Part II: Practical Realization [J]. *IEEE Trans on Circuits and Systems I — Foundamental Theory and Application*, 2001, 48(3): 382-385.

[6] Yi X. Hash Function Based on Chaotic Tent Maps [J]. *IEEE Trans on Circuits and Systems — II: Express Briefs*, 2005, 52(6): 354-357.

[7] Xiao D, Liao X F, Deng S J. One-way Hash Function Construction Based on the Chaotic Map with Changeable-parameter [J]. *Chaos Solitions and Fractals*, 2005, 24(1): 65-71.

[8] Xiao D, Liao X F, Tang G P, et al. Using Chebyshev Chaotic Map to Construct Infinite Length Hash Chains [A]. *Int Symposium on Circuits and Systems* [C]. Vancouver: IEEE, 2004: 11-12.

[9] Xiao D, Liao X F. A Combined Hash and Encryption Scheme by Chaotic Neural Metwork [A]. *Int Symposium on Neural Network* [C]. Dalian: Springer, 2004: 633-638.

[10] 王小敏, 张家树, 张文芳. 基于广义混沌映射切换的单向Hash 函数构造[J]. ***物理学报***, 2003, 52(11): 2737-2742.
(Wang X M, Zhang J S, Zhang W F. One Way Hash Function Construction Based on the Extended Chaotic Maps Swith [J]. *Acta Physica Sinica*, 2003, 52(11): 2737-3742.)

[11] 李红达, 冯登国. 复合离散动力系统与Hash 函数[J]. ***计算机学报***, 2003, 26(4): 460-464.
(Li H D, Feng D G. Composite Nonlinear Discrete Chaotic Dynamical Systems and Keyed Hash Functions [J]. *Chinese J of Computers*, 2003, 26(4): 460-464.)

[12] 单梁, 李军, 王执铨. 时空混沌序列在语音保密通信中的应用[J]. ***东南大学学报***, 2004, 34(增): 20-23.
(Shan L, Li J, Wang Z Q. Speech Secure Communications Using Spatiotemporal Chaotic Sequence [J]. *J of Southeast University*, 2004, 34(S): 20-23.)

[13] Wang S H, Kuang J Y, Li J H, et al. Chaos-based Communications in a Large Community [J]. *Physical Review E*, 2002, 66(6): 1-4.

[14] Papadimitriou S, Bountis T, Mavroudi S, et al. Probabilistic Symmetric Encryption Scheme for Very Fast Secure Communication Based on Chaotic Systems of Difference Equations [J]. *Int J on Bifurcation and Chaos*, 2001, 11(12): 3107-3115.

(上接第 1243 页)

[12] Massoulie L. Stability of Distributed Congestion Control with HeteHogeneous Feedback Delays [J]. *IEEE Trans on Automatic Control*, 2002, 47(6): 495-902.

[13] Vinnicombe G. *On the Stability of End-to-end Congestion Control for the Internet* [R]. CUED/F-INFENG/TR. 398. http://www-control.eng.cam.ac.uk/gv/internet, 2001-11-17.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net