基于经验和信誉的MAS信任评价的学习机制

刘哲元 慕德俊 王晓伶
(西北工业大学自动化学院 西安 710072)

摘 要 为合理利用MAS中存在的经验和信誉两种信任评价资源,准确评价合作agent,提出了根据活动因子的学习结果动态评价MAS的学习机制。采用该机制,MAS根据活动因子的取值赋予不同信任资源以不同权值,动态计算可信度,评价合作目标,使得MAS取得的总体报酬最优。仿真结果验证了学习机制的有效性。

关键词 多智能体系统,经验模型,信誉模型,动态学习机制
中图法分类号 TP393 **文献标识码** A

Learning Mechanism of Trust Evaluation of MAS Based on Experience and Reputation

LIU Zhe-yuan MU De-jun WANG Xiao-ling
(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract To accurately evaluate cooperation agents, it's necessary to use the experience-based and reputation-based models existed in MAS. Therefore, a learning mechanism of MAS according to the learning result of active parameter was proposed. MAS using this mechanism gives different trust model the corresponding weight, dynamically calculates the reliability and evaluates the cooperation objects to make the optimal rewards. Simulation shows the efficiency of the learning mechanism.

Keywords MAS, Experience-based model, Reputation-based model, Dynamic learning mechanism

MAS(Multi-Agent System,多智能体系统)中agent在协作过程中会交互使用对方的资源、信息、服务,为了避免这些资源被不可信任的或恶意的agent访问及使用,保障交互任务顺利完成,agent在协作交互前须对合作者进行信任评价,以此判断合作者是否可靠到可以与之进行交互。信任可以从两个方面评价:一是agent间的交互直接经验;二是由对agent过去行为的评价报告所组成的信誉(如见证人信誉、认证信誉等)。在目前的MAS研究中,利用这两方面的信任信息派生出了两种信任评价模型:基于经验的信任模型[1,2]和基于信誉的信任模型[3-5]。基于经验的信任模型利用agent之间交互的历史信息推断合作者的可信度,适用于小规模且有大量的重复性交互的MAS。采用这种模型agent可以准确地推断出合作者的可信度,但是需要大量的交互历史信息,并且交互协作的发起方可能会被恶意agent攻击。而基于信誉的信任模型,是从第三方的信誉提供者处获取关于目标agent的信誉信息,推断目标的可信度,适用于大规模且交互协作不多的MAS。但是由于信誉信息来源于第三方的提供者,提供者的可靠度、提供信息的准确度直接影响着信任评价的准确性。

在许多环境中需要采用经验模型和信誉模型相结合的混合模型来推断合作者的可信度[6]。就如何在混合模型中决策信任方式,文献[6]提出了采用经验和信誉两种资源综合评价合作者的方法,通过赋予经验模型和信誉模型一定的权值,在评价过程中通过选择适当的权值以使MAS协作报酬最大化。但是该方法调整权值的过程是一个在预定义参数集合中选择某一个特定值的过程,没有根据系统环境信息对具体的权值进行修改及调整,因此当前选择的权值未必是当前状态下评价最优的值。

本文提出了根据活动因子ψ动态信任评价的学习机制。采用该机制,MAS根据ψ的取值赋予不同信任资源(经验或者信任)不同权值,动态计算可信度;并且在协作过程中通过对ψ的学习,使得MAS取得的总体报酬最优。实验表明,采用该机制,MAS在取得目标奖励的同时,可以有效地对目标agent进行评价。

1 基于经验和信誉的信任评价

随着时间的推移和协作交互的不断进行,用MAS来评价信任的环境特征、合作者的信誉等都在不断变化。为了使agent在协作中取得最大的报酬,agent在协作过程中需要不断调整其信任决策方式。

1.1 评价方法

在经验模型中,agent a选择信任一个合作者b,a需付出P_r的代价,a选择信任b后,b返回给a值为P_e的回报,如果$P_e-P_r>0$,则认为b是值得信任的,反之认为不值得信任。但在进行评价时,a并不知道P_e的大小,a通过对P_e的估计来决定是否信任b,估计值为$Q(T)$[6]。若$Q(T)>0$,选择信

到稿日期:2009-09-21 返修日期:2009-12-18
刘哲元(1982-),男,博士生,主要研究方向为网络信息安全,E-mail:f64897011@gmail.com;慕德俊(1966-),男,教授,博士生导师,主要研究方向为网络信息安全、并行计算;王晓伶(1980-),男,博士生,主要研究方向为网络信息安全、多智能体系统。

任 b；否则，选择不信任。

在信誉模型中，agent 从不同的信誉提供者获取 $Q_i(T)$（i 表示信誉提供者序号），通过对 $Q_i(T)$ 求平均值来计算目标 agent 的可信度。

在混合模型中，引入参数 ψ，通过 ψ 赋予经验模型、信誉模型以不同的权值，利用 $Q(T)$ 和 $Q_i(T)$ 共同进行信任评价。ψ 越小，信任评价越依赖于双方之间的交互经验，ψ 越大越依赖于第三方提供的信誉值。即当 ψ 为 0 时，仅使用经验模型评价目标的可信度；当 ψ 为 1 时，仅使用信誉模型评价。混合模型中，可信度 $Q'(T)$ 的求解公式如下：

$$Q'(T)=(1-\psi)Q(T)+\psi\left(\frac{\sum_{i\in I}Q_i(T)}{|I|}\right) \tag{1}$$

式中，I 表示信誉提供者的集合，$\psi\in[0,1]$。如果 $Q'(T)>0$，则 a 选择信任合作者 b，并将 $Q(T)$ 更新为 P_e-P_r。

在协作过程中，MAS 存在诸多因素（如 agent 的数量、交互频率、信誉准确度等）影响对 agent 信任的评价，所以在混合模型中需要针对不同时段、不同环境对 ψ 进行动态调整并计算 $Q'(T)$。

1.2 活动因子 ψ 的动态学习过程

混合模型在信任评价过程中动态调整 ψ 的过程，是根据环境参数动态增加或者减小当前参数 ψ 的过程。首先在预定义的参数值集合 $S=\{\psi_1,\psi_2,\cdots,\psi_n\}$ 中选取一个参数作为活动因子，评价协作 agent，计算协作取得的报酬；同时计算采用其他参数使 agent 协作取得的报酬，如果在 S 中存在参数 ψ_i，使得当前的活动因子 ψ 小于 ψ_i，则表示混合模型中信任评价趋于 ψ 逐渐增加的趋势，应适当增加 ψ（见图 1(a)），否则减小 ψ（见图 1(b)）。

$$f_{Inc}(t)=1-e^{-t} \tag{2}$$

$$\psi_{new}=f_{Inc}(t_{\max}+\Delta t) \tag{3}$$

式中，$t_{\max}$ 为：

$$t_{\max}=-\ln(1-\psi_{\max}) \tag{4}$$

$$f_{Dec}(t)=e^{-t} \tag{5}$$

$$\psi_{new}=f_{Dec}(t_{\max}+\Delta t) \tag{6}$$

式中，$t_{\max}$ 为：

$$t_{\max}=-\ln(\psi_{\max}) \tag{7}$$

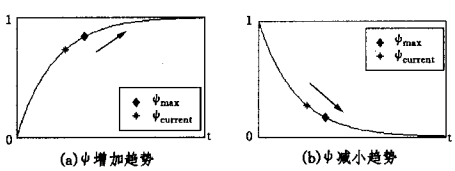

图 1 活动因子 ψ 的动态调整

经过调整得到新的 ψ 后，将之添加到集合 S 中，用于下一次的学习过程。ψ 的动态学习算法如下：

1）为预定义 ψ 值的集合 $S=\{\psi_1,\psi_2,\cdots,\psi_n\}$ 创建表，表中 ψ_i 一一对应于 $Q(\psi_i)$，$Q(\psi_i)$ 表示当选择 ψ_i 时 agent 获取报酬的估计值；

2）选定一个 ψ_i 作为当前使用活动信任因子 $\psi_{current}$，即 $\psi_{current}=\psi_i$；

3）使用 $\psi_{current}$ 计算 $Q'(T)$，根据 $Q'(T)$ 判断是否信任目标 agent；

4）同时对其他的 ψ_i，计算相应的 $Q(\psi_i)$；

5）if $Q'(T)\leqslant 0$，不信任合作者，则双方无交互事务发生，跳转到 3），等待下一次评价过程；

6）else，选择信任，agent 双方进行事务交互，观察 P_e；

7）对每个 ψ_i，计算协作交互取得的报酬 P_e-P_r，并用 P_e-P_r 更新 $Q(\psi_i)$；

8）选择 $Q(\psi_i)$ 值最大的 ψ_i，即 $\psi_{\max}$；

9）if $\psi_{\max}>\psi_{current}$，需要将活动因子增大，使用式(3)计算得到新的活动因子 ψ_{new}；

10）else，需要将活动因子减小，使用式(6)计算得到新的活动因子 ψ_{new}；

11）采用 ψ_{new} 替换集合 S 中对应于 $\psi_{current}$ 的参数 ψ_i；

12）在下一次交互中，选择 $\psi_{\max}$ 作为活动因子，$\psi_{current}=\psi_{\max}$；

13）重复 3）。

在 MAS 协作过程中，一方面交互经验在不断积累，同时信誉的准确性和可靠性也在不断变化，通过 ψ 的动态学习过程，调整信任评价方式可以准确地评价协作 agent。

1.3 动态学习算法分析

1.3.1 ψ_{new} 的选择

在动态学习中，ψ_{new} 的选择是对 $\psi_{\max}$ 增加或者减小的过程，受 Δt，ΔE，ΔR 的影响，其中 ΔE，ΔR 分别表示 Δt 时间内 $Q(T)$ 的差值及 $E(Q_i(T))$ 的差值。综合 Δt，ΔE，ΔR 3 个因素：

$$\text{当}|\Delta R|<|\Delta E|\text{时，}\Delta t=\left(\frac{|\Delta R_i-\Delta R_{i-1}|}{|\Delta E_i-\Delta E_{i-1}|}\right)\Delta t \tag{8}$$

$$\text{反之，}\Delta t=\left(\frac{|\Delta E_i-\Delta E_{i-1}|}{|\Delta R_i-\Delta R_{i-1}|}\right)\Delta t \tag{9}$$

对于活动因子 ψ 的调整，首先取得 $\psi_{\max}$，根据 $\psi_{\max}$ 得到 $t_{\max}$，然后采用经式(8)、式(9)得到的 Δt 值计算而得 ψ_{new}。由于在计算 ψ_{new} 过程中综合考虑了影响活动因子选择的几个因素，ψ 的调整幅度因环境不同而不同，避免了在调整过程中活动因子值因调整过大或者过小而发生的反复震荡的情况。

1.3.2 抗干扰能力

在开放分布式网络中，恶意节点或者通过频繁地提交不诚实反馈，或者在累积一定的反馈可信度之后开始提交不诚实的反馈来进行破坏。对于这种信誉值不断变化的 agent，之前的经验模型或者信誉模型并不能很好地发现恶意的欺诈或者攻击。采用本文动态学习机制，结合直接经验和间接信誉两种资源：当系统中有恶意节点频繁提交不诚实的反馈，混合模型会逐步减小 ψ 值，主要依赖于采用直接经验对之评价；当恶意 agent 在累积到一定的可信度之后开始提交不诚实信任反馈时，混合模型首先会逐步增大 ψ 值，此时倾向于采用第三方的信誉对之评价，之后在获取到足够的关于该 agent 恶意欺诈的直接经验后，则会逐步减小 ψ 值。混合模型由于在交互过程中调整评价方式，因此可以相对及时、准确地评价该 agent 是否可信。

另外，在同陌生人 agent 时，交互初期双方并没有交互经验可以借鉴，而经过一定的交互后双方可能发展为熟人。对陌生人的评价是一个根据 ψ 调整评价方式的过程，ψ 随着对陌生人的熟悉度、可靠度的变化而改变。

1.3.3 复杂性分析

动态学习算法的时间复杂度为 $O(|S|\times n)$，$|S|$ 表示集合 S 的大小，n 表示 n 次动态学习过程，其中 $|S|<<n$。S 是 ψ_i 的集合，在一次动态学习过程中 S 的大小仅决定了计算

$Q(\psi_i)$的次数、ψ_{max}的比较次数，其复杂度为 O(|S|)。ψ_{new}值是根据ψ_{max}得到t_{max}，Δt之后的一次计算结果，其复杂度为 O(1)。集合 S 的更新过程是一个使用ψ_{new}替换$\psi_{current}$的过程，其复杂度为 O(1)。

2 实验与评估

2.1 实验环境描述

实验中通过设置不同 ψ 值，比较采用不同信任模型时 MAS 取得的报酬，即当 $\psi=0$ 时采用经验模型、$\psi=1$ 时采用信誉模型、文献[6]的混合模型以及采用本文 ψ 动态学习机制的混合模型(后称为改进后的模型)。

实验参数的设定如表 1 所列：P_r 表示 agent 付出的代价，P_e 表示 agent 收到的报酬。为实验不同环境下混合模型评价信任的准确性，P_e 值服从 $N(\mu,\sigma_1^2)$，通过设定不同的 P_e，模拟系统中不同的协作回报。为模拟信誉提供者提供信誉的准确度，使信誉准确度的分布服从 $N(\mu,\sigma_2^2)$，σ_2 表示信誉提供者提供信誉的偏移度，σ_2 越小表示偏移越小，提供的信誉越准确。

表 1 实验环境参数设定

参数名称	参数值
P_r	10
μ	{8,10,12}
σ_1	10
σ_2	{0.1,10}
ψ	{0,1.0,0.5,动态 ψ}

2.2 实验结果分析

本文实验分为 3 部分：

(1)P_e 服从 $N(8,10)$，$N(12,10)$环境下信任模型之间的比较

图 2、图 3 分别是 P_e 在 $N(8,10)$，$N(12,10)$分布下，σ_2 分别为 0.1 和 10 情况下，不同信任模型取得的报酬比较。当 μ 等于 8 时，目标 agent 返回给评价 agent 的 P_e 小于 P_r，系统中 agent 不信任协作目标，不愿同其他 agent 协作，此时改进后的模型收益同于文献[6]的混合模型。当 μ 等于 10 时，P_e 等于 P_r，即 agent 协作取得的总体平均报酬为 0；当 μ 等于 12 时，P_e 大于 P_r，此时采用本文改进后的机制的混合模型在处理过程中的交互初期取得的报酬有了明显的提高，并且经过一定的交互后由于经验的积累，这种提高逐渐趋向于平稳，取得的报酬趋于文献[6]中的混合模型。

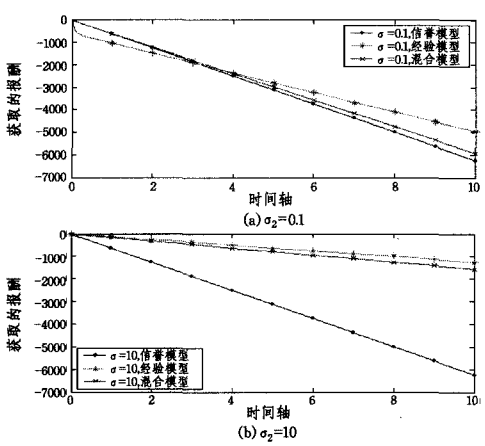

图 2 μ 服从 $N(8,10)$下模型之间的比较

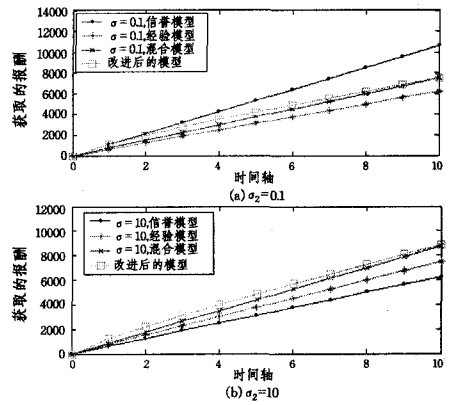

图 3 μ 服从 $N(12,10)$下模型之间的比较

从图 2、图 3 可以看出，当 σ_2 为 0.1 时，由于信誉提供者提供的信誉度基本可信，采用 ψ 的混合模型趋向于信誉模型；当 σ_2 为 10 时，则趋向于经验模型。在 MAS 交互初期，由于系统中没有足够的历史交互经验可以分析、利用，这时混合模型偏向于信誉机制，而一旦采集到足够的历史数据后，由于经验模型快速、准确等特点，则趋向于采用经验模型。

(2)不同环境下 ψ 学习过程的比较

图 4 是 MAS 采用改进后的学习机制，取不同 μ 值，$\sigma 2$ 为 0.1 或 10 时，ψ 动态学习的曲线图。当 σ_2 为 0.1，μ 为 8 时，由于 P_e 小于 P_r，开始的一段时间内 ψ 比较大，说明此时信任的评价主要依赖于信誉模型；而 μ 为 12 时则很稳定地依赖于信誉模型。当 σ_2 为 10 时，由于信誉提供的不准确性，混合模型的信任评价主要倾向依赖于经验模型。

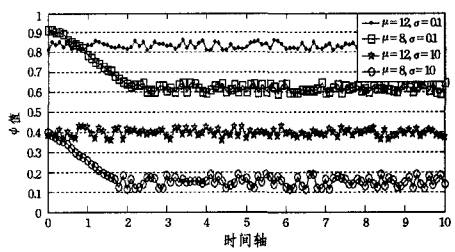

图 4 不同 μ 情况下 ψ 学习曲线图

(3)对一个陌生人 agent 的评估过程仿真

图 5 表示对陌生人 agent 的评价过程，在系统开始阶段由于缺少同陌生人 s 的交互经验，评价 agent 对陌生人处于不可信状态，即评价值为 −1；在之后的交互中，由于交互历史的积累，agent 逐渐开始信任陌生人 s，即信任值大于零；然后由于 s 给出虚假评价及时间权重值减小，评价 agent 对 s 的信任度减小，直到不信任状态(评价值小于 0)；最后随着 s 提供的信息的可靠性的增加，s 又逐渐恢复到可信任状态。对陌生人的评价是一个 ψ 随着陌生人可靠度、信誉准确度变化而调整的过程。

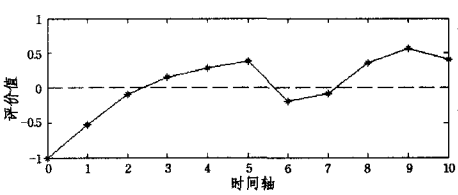

图 5 对陌生人 agent 评价过程

结束语 MAS 中对协作目标的信任评价直接关系到系

统资源的安全性保护及交互任务能否实现，为能准确评价 agent，本文提出了根据活动因子 ψ 动态评价合作目标的学习机制。该机制根据 ψ 的取值赋予不同信任资源不同权值，并且在协作过程中根据系统特征及交互历史对 ψ 进行学习，动态调整信任评价方式。仿真实验表明，采用该机制在取得目标奖励的同时，可以有效地对目标 agent 进行评价，是一种有效的评价方法。

参 考 文 献

[1] Banerjee B, Mukherjee R, Sen S. Learning Mutual Trust[C]// Proc. of the Workshop on Deception, Fraud and Trust in Agent Societies at AGENTS-00. 2000: 9-14

[2] Littman M, Stone P. Leading Best-Response Strategies in Repeated Games[C]// Proc. of the Workshop on Economic Agents, Models, and Mechanisms, at IJCAI-01. Seattle, Washington, 2001

[3] Yolum P, Singh M P. Self-Organizing Referral Networks: A Process View of Trust and Authority. Engineering Self-organizing Systems[J]. Lecture Notes in Artificial Intelligence, 2003: 195-211

[4] Huynh T D, Jennings N R, Shadbolt N R. Certified Reputation: How an Agent Can Trust a Stranger[J]. AAMAS06, 2006: 1217-1224

[5] Huynh T D, Jennings N R, Shadbolt N R. An integrated trust and reputation model for open multi-agent systems[J]. Journal of AAMAS06, 2006, 13(2): 119-154

[6] Fullam K K, Barber K S. Dynamically Learning Sources of Trust Information: Experience vs. Reputation[J]. AAMAS07, 2007: 1062-1069

[7] 王少杰，郑雪峰，等. 基于个体经验的信任模型[J]. 通信学报，2008，4(29)：130-135

[8] 陶海军，王亚东，等. 基于熟人联盟及扩充合同网协议的多智能体协商模型[J]. 计算机研究与发展，2006，43(7)：1115-1160

[9] 王平，张自力. 多 agent 系统中信任的动态性处理[J]. 计算机科学，2005，32(3)：182-185

(上接第 106 页)

4.2 TCP 报文的主动重传

文献[9]提出了一种基于主动重传的切换策略，即在移动 IPv4 通信中进行 TCP 传输时，当注册完成后，为了避免 TCP 进入超时等待，由移动节点主动重发 TCP 数据报文。然而，这种方法需要对每一个移动通信节点做出改进，这不太现实。特别是在网络移动通信中，进出移动网络的节点的活动可以是随机的，如飞机上的笔记本使用者到达目的地后或许就不再使用无线网络进行通信。

基于文献[9]的设想，本文提出可以通过移动路由器 MR 来触发主动重传。在移动网络的移动路由器 MR 上，增加一个缓存空间，对应于移动网络内的不同的节点，用于存取移动节点当前发送的数据包。切换完成以后，MR 将该数据包发送给通信对端，从而触发了 TCP 的切换过程。如果移动节点是 TCP 连接的发送方，切换完成后，MR 缓存将当前存储的 TCP 数据包重发一遍给通信对端；如果移动节点作为 TCP 的接收方，切换完成后，MR 缓存将当前存储的 ACK 数据包重复发送 3 遍，使得通信对端采取快速重传机制。

在 NS 中，需要在 mip-reg. cc 中添加相关函数并更改代理广播周期等参数，还需要更改 agent. h, tcp. cc, tcp-sink. cc, ns-default. tcl 和 ns-mip. tcl 中的相关函数。需要注意的是，并不是移动网络在数据链路层的每次切换都会触发 TCP 的重传，需要区分移动网络所切换的两个子网是否归属于同一个代理。

4.3 与原有切换策略的性能对比

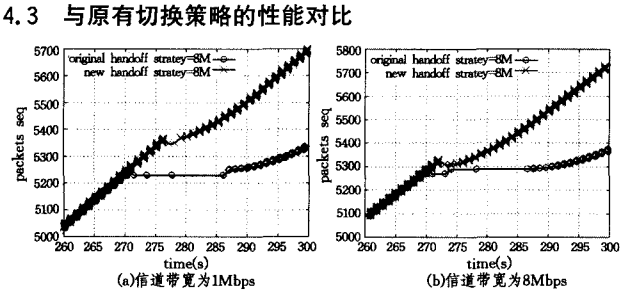

图 8 改进的切换策略与原有切换策略的对比

针对上述切换策略，在 NS 中将基于 NEMO 的原有切换策略和新的切换策略进行仿真对比。如图 8 所示，空心圆点代表使用原有切换策略时的 TCP 连接的包序号，叉点代表使用新的切换策略时的 TCP 连接的包序号，改变系统的带宽分别设置为 1Mbps 和 8Mbps。可以看出，使用新的切换策略后，切换开始的时间推后，而切换结束的时间提前，从而大大缩短了 TCP 传输过程中由切换引起的中断时间。

结束语 由于卫星通信具有大范围和全球无缝覆盖，采用卫星通信网来向这些移动平台提供对地面网络连接的延伸，是一种比较经济、可行的解决方案。本文首先建立了基于卫星通信的网络移动通信系统模型并在 NS2 软件中实现了该模型，然后基于该模型对 TCP 的性能以及切换对其性能的影响进行了仿真，最后通过分析卫星通信的特点提出了一种改进的切换策略，通过计算机仿真证明了该策略可以大大缩短切换引起的 TCP 传输中断时间。

参 考 文 献

[1] 张更新，李江华，钱宗峰. 基于卫星移动通信的网络移动研究[J]. 电信快报，2005(3)：14-17

[2] Devarapalli V, Wakikawa R, Petrescu A, et al. Network Mobility (NEMO) Basic Support Protocol[S]. IETF, RFC 3963. 2005

[3] Ernst T, Lach H Y. Network Mobility Support Terminology [S]. IETF, RFC 4885. 2007

[4] Ernst T. Network Mobility Support Goals and Requirements [S]. IETF, RFC 4886. 2007

[5] Thubert P, Wakikawa R, Devarapalli V. Network Mobility Home Network Models[S]. IETF, RFC 4887. 2007

[6] Ng C, Thubert P, Watari M, et al. Network Mobility Route Optimization Problem Statement[S]. IETF, RFC 4888. 2007

[7] James F K, Keith W R. Computer Networking: A Top-Down Approach Featuring the Internet [M]. Pearson Education Inc., 2004: 159-190

[8] Stevens W. TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms[S]. IETF, RFC 2001. 1997

[9] 赵阿群. 移动支持协议切换性能研究[J]. 软件学报，2005，16(4)：587-594

[10] 叶敏华，张惠民，刘雨. Mobile IP 切换时的 TCP/UDP 性能分析[J]. 电信交换，2001(4)：1-6