

基于行为监控的自适应动态信任度测模型

李小勇 桂小林 毛 倩 冷东起

(西安交通大学计算机科学与技术系 西安 710049)

摘 要 大规模分布式系统中的动态信任关系模型本质上是最复杂的社会关系之一,涉及假设、期望、行为和环境等多种因子,很难准确的定量表示和预测.将粗糙集理论和信息熵理论结合起来,应用于开放环境下动态构建基于行为数据监控与分析的信任关系度测(度量与预测)模型.该方法直接从分析传感器监测到的动态数据入手,针对影响信任的多个度测指标进行自适应的数据挖掘与知识发现,从而改变了传统的信任关系建模思路,跳出了传统信任关系建模过程中各种主观假设的束缚,并克服了传统模型对多维数据处理能力不足的问题.实验结果表明,与已有模型相比,新模型能够快速准确地实现开放分布式环境下实体的可信性判别,而且具有良好的行为数据规模的扩展能力.

关键词 信息安全;动态信任模型;粗糙集;信息熵

中图法分类号 TP311

DOI号: 10.3724/SP.J.1016.2009.00664

Adaptive Dynamic Trust Measurement and Prediction Model Based on Behavior Monitoring

LI Xiao-Yong GUI Xiao-Lin MAO Qian LENG Dong-Qi

(Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049)

Abstract In the large-scale distributed systems, trust relationship model is one of the most complex concepts in social relationships, and it also is an abstract psychological cognitive process, involving assumptions, expectations, behavior and the environment, and other factors. So, it is very difficult to quantify and predict trust relationship accurately. In this paper, rough set theory and information entropy theory are combined and applied to the study of distributed dynamic trust measurement and prediction model based on behavior data. The new model works through analysis monitored behavior data by sensors directly, changes the traditional modeling thoughts, brakes away from the fetter of various subjective assumptions in traditional modeling methods, and overcomes the problem of inadequate handling capacity for multi-source behavior data in the traditional trust model. Simulating results shows that the new model can accurately implement trust measurement and prediction process between entities in open and complex distributed environment, and has a better scalable capacity of behavior data.

Keywords information security; dynamic trust model; rough set; information entropy

收稿日期:2008-12-09;最终修改稿收到日期:2009-01-18. 本课题得到国家自然科学基金(60873071)、国家“八六三”高技术研究发展计划项目基金(2008AA01Z410)、教育部新世纪优秀人才计划项目(NCET-05-0829)和陕西省科技计划攻关项目(2007K04-05)资助. 李小勇,男,1975年生,博士研究生,主要研究方向为动态信任管理理论、计算机网络. E-mail: lxyxjtu@163.com. 桂小林(通信作者),男,1966年生,博士,教授,博士生导师,主要研究领域为网络计算、动态信任管理理论. E-mail: xlgui@mail.xjtu.edu.cn. 毛倩,女,1986年生,硕士,主要研究方向为动态信任管理技术. 冷东起,男,1986年生,硕士,主要研究方向为动态信任管理技术.

1 引言

随着以 Internet 为基础平台的、各种大规模的分布式应用(如 P2P、网格计算和普适计算等)的深入研究,系统表现为由多个软件服务组成的动态协作模型.在这种动态和不确定的环境下,传统的安全机制中基于 PKI(Public Key Infrastructure)的静态信任机制不能适应这种需求,因此,针对这些新型应用环境的动态信任管理技术已成为一个研究热点^[1-3].动态信任管理是在原有网络安全技术的基础上增加行为可信的安全新方法,强化了对网络状态的动态处理,为实施智能自适应的网络安全和服务质量控制提供策略基础.研究动态信任管理技术对确保互联网的可靠运行、资源的安全共享和可信利用,具有重要的现实意义:首先,该技术是近几年才发展起来的,仍然属于探索性研究课题,对很多相关理论和技术性问题都没有达成共识,仍缺乏系统明确的方法论指导,还无法完全解决互联网发展过程中对于信任关系快速和准确的度测的需求.其次,信任关系的合理量化不但是“信任管理”理论的基础性工作和必须首先解决的核心科学问题,也是可信软件^[3]、可信网络^[4]等新型可信计算领域的基础性研究课题.因此,针对复杂开放环境下网络节点行为的多样性、动态性和协同性,特别是数据与控制同时动态变化的新特征,系统而深入地开展动态信任关系量化机理的研究,具有广阔的应用前景.

目前的代表性工作研究了多种开放系统中的动态信任关系,并使用不同的数学方法和数学工具,建立了信任关系模型,文献[5]开发了一个具有稳健性和伸缩性的 P2P 声誉系统 Power-Trust,该系统利用幂次法则收集本地节点反馈并将这些反馈聚合起来,生成全局声誉;文献[6]提出了多 Agent 系统中实体之间基于模糊逻辑的动态信任模型,使用模糊逻辑的推理理论建立信任关系度量和预测的推理规则,具有从大量输入数据中自学习以获取评估规则的能力;文献[7]提出了一种基于半环(Semi-ring)代数理论的信任模型,将信任问题定义为一个有向图 $G(V, E)$ 的路径问题,在信任链的建立过程中能够较准确地地区分诚实的实体和恶意的实体;文献[8]提出了基于概率论的信任模型,使用该方法,反馈信任度的计算使用多级多路径的信任链方式,能够较准确地反映全局的信任度.文献[9]的主要贡献是将风险作为决策要素引入信任模型,建立了由直接风

险和间接反馈相耦合的信任评估机制.

现有的成果有效地推动了相关研究的发展,极大地丰富了人们对信任度测基本问题的认识,但也呈现出一些不足:(1) 现有模型一般都是建立在经典的基于状态演绎的概率统计基础之上的,人们为建立演绎模型不得不做出各种主观的假设,这样,模型的自适应性必然会受到影响;(2) 由于一种有效的度测方法必然具有多维(multi-dimensional)的决策属性(指标),而目前模型由于采用的数学工具本身的限制而无法处理多维的度测指标(快速响应性是动态信任管理技术的基本需求,对于一个时空开销较大的计算方法是没实用价值的),表现出对多维数据处理能力不足的问题.目前看来,如果总是根据各种的主观假设或者简单的上下文信息建立新的模型,只会给软件开发人员对模型应用不一致性的处理带来更大的困惑.因此,改变传统建模思路,采用新的观点、方法和新的数学工具来研究信任推理过程,摆脱传统建模方法主观假设的束缚,不受其制约,才有可能为建立较普适的、具有快速响应能力和动态自适应能力的信任度测模型提供可能.

将粗糙集理论和信息熵理论结合起来,应用于开放环境下动态构建基于行为数据监控的动态信任度测模型.该方法直接从分析传感器监测到的行为数据入手,针对影响信任的多个度测指标进行自适应的数据挖掘与知识发现,从而改变了传统的建模思路,跳出了传统信任关系建模过程中各种主观假设的束缚,并克服了传统模型对多维数据处理能力不足的问题.实验结果表明,与已有模型相比,新模型能够快速准确地实现开放分布式环境下实体的可信性判别,而且具有良好的行为数据规模的扩展能力.

2 动态信任决策问题的形式化描述

开放系统中的动态信任管理是指在对信任关系进行建模和管理时,综合考察影响可信性的多种因素(特别是行为数据),针对信任关系的多个属性进行有侧重点的建模;在信任关系评估中,动态地收集相关的主观因子和客观数据的变化,以一种快速及时的方式实现对信任度的计算、管理和决策的相应调整.因此,根据数学模型建立的运算规则,在时间和观测到的证据上下文的触发下自适应地进行信任度的重新计算(度测),是动态信任决策的核心工作^[1-2].

图 1 为本文基于行为数据监控的开放系统信任

管理系统结构。网络实体、网络服务、支撑软件和网络资源等网络实体之间的协同依赖于它们各自之间建立的信任关系,而在这些网络实体进行交互和协同的过程中,行为监控模块对实体行为进行监控,监控到的行为数据作为信任度量与预测的证据,而信任度测的结果可以作为实体获得网络服务的凭证。例如,当某实体需要访问网络计算环境时,首先要经过认证授权,认证授权模块会调用信任度量与预测模块来对实体的信任度进行度量及预测。在度量与预测的过程中,信任度量与预测模块使用行为数据库中的信息,这些信息是通过行为监控模块对网络实体和网络计算环境中的资源进行实时监控而收集到的。认证授权模块得到实体的信任度之后,通过查找策略库,按照匹配的策略给用户分配相应的权限和 QoS(Quality of Service)。在访问的过程中,环境监控模块持续对网络实体和网络计算资源进行监控并将相应信息存入行为数据库,信任度量与预测模块通过对上下文信息数据库的信息进行分析,发现并预测可能出现的意外情况,实时地通知认证授权模块,认证授权模块及时查找策略库并修改授予实体的权限及 QoS。

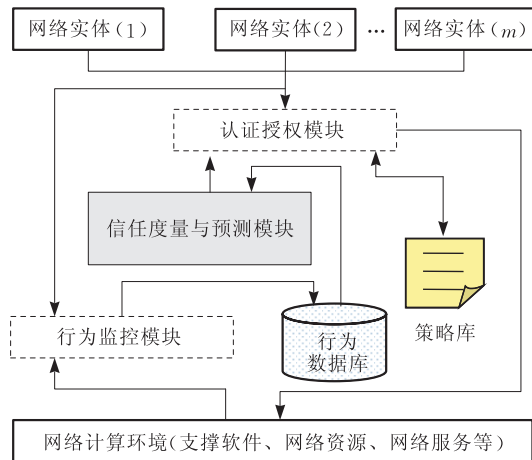


图 1 基于行为监控的开放系统信任管理系统结构

定义 1. 设 P_1, P_2, \dots, P_N 表示分布式系统中发生交互行为的 N 个网络实体(network entities),网络实体可以是任何网络中的元素,可以是一个用户,也可以是一个软件服务(资源),还可以是一个网络设备或者数据集。称集合 $\Omega = \{P_1, P_2, \dots, P_N\}$ 为实体域,则在任意系统 Ω 中,都存在集合 $Q, Q \subset \Omega$,且必定存在着元素 $o, o \in \Omega$,使得 $\forall q \in Q$,都有操作使得 $q \rightarrow o$,则 Q 称为主域(truster domain), o 的集合称为客体域(trustee domain),记为 O 。本文模型中,服务请求者(Service Requester, SR)属于 O ,

而服务提供者(Service Provider, SP)属于 S 。

定义 2. 设 $\exists P_i \in Q$ 度量 $\exists P_j \in O$ 的信任程度有 m 项测量指标 I_1, I_2, \dots, I_m ,其测量值分别表示为 $\xi_1, \xi_2, \dots, \xi_m$,其中每一个元素 $I_i (1 \leq i \leq m)$ 称为一个决策属性(Decision Factor, DF)。设 ϖ_i 表示第 i 个 DF 相对于其它 DF 的重要性程度,并且满足

$$0 \leq \varpi_i \leq 1, \sum_{i=1}^m \varpi_i = 1 \quad (1)$$

定义 3. 设 $\Gamma(P_i, P_j)$ 表示实体 $P_i \in Q$ 对实体 $P_j \in O$ 的总体信任评价,称为总体信任度(Overall Trust Degree, OTD),OTD 是基于信任度的访问控制决策的依据。令

$$\Gamma(P_i, P_j) = \sum_{k=1}^m \varpi_k \times \xi_k \quad (2)$$

定义 4. 设总体可信度 $\Gamma(P_i, P_j)$ 有 p 个等级划分 c_1, c_2, \dots, c_p ,其中 $0 \leq c_p \leq 1 (1 \leq p \leq m)$ 。度测等级空间记作 U ,表示为 $U = \{c_1, c_2, \dots, c_p\}$ 。若度测等级空间 U 具有如下性质: $c_i \cap c_j = \emptyset (i \neq j)$,且 $c_1 < c_2 < \dots < c_p$,即 c_{k+1} 比 c_k 强,则称 U 为一个有序分割类。

定义 5. 设 $\exists P_i \in Q$ 可提供 k 个级别的服务 $S = \{s_1, s_2, \dots, s_k\}$,且 S 是一个有序分割类,则 S 和 $\Gamma(P_i, P_j)$ 之间的映射函数 Ψ 定义为

$$\Psi(\Gamma(P_i, P_j)) = \begin{cases} s_k, & c_k < \Gamma(P_i, P_j) \leq 1 \\ s_{k-1}, & c_{k-1} \leq \Gamma(P_i, P_j) < c_k \\ \dots & \\ s_2, & c_1 \leq \Gamma(P_i, P_j) < c_2 \\ s_1, & 0 \leq \Gamma(P_i, P_j) < c_1 \end{cases} \quad (3)$$

其中分界点 $c_1, c_2, \dots, c_k \in [0, 1]$ 由定义 4 确定, $\exists P_i \in Q$ 向 $\exists P_j \in O$ 请求某种质量的服务时,首先要根据 P_j 的信任级别决定它所能得到的服务质量,这样既可以分级对不同的实体提供不同的服务,也有利于降低系统可能存在的风险。

例如,某 FTP 服务提供了 3 个等级的服务质量, $S = \{s_1, s_2, s_3\}$, s_1 表示拒绝服务, s_2 表示只读, s_3 表示既可以读也可以写。相应的决策等级空间设定为 $U = \{c_1, c_2, c_3\} = \{0, 0.2, 0.5\}$,则服务决策函数可表示为

$$\Psi(\Gamma(P_i, P_j)) = \begin{cases} s_3, & 0.5 < \Gamma(P_i, P_j) \leq 1 \\ s_2, & 0.2 \leq \Gamma(P_i, P_j) \leq 0.5 \\ s_1, & 0 \leq \Gamma(P_i, P_j) < 0.2 \end{cases}$$

若 $\Gamma(P_i, P_j) = 0.19$,则根据决策函数 Ψ ,决策过程为 $\Psi(\Gamma(P_i, P_j)) = \Psi(0.19) = s_1 =$ 拒绝服务。

3 行为数据的获取与预处理

行为数据是指可直接根据软硬件检测获得的用来定量评估网路实体 OTD 的基础数值^[10]. 获得全面可信、划分粒度适中、满足应用需求的行为数据是动态信任度测的基础. 网络实体行为的各种数据就蕴含在包括各种应用协议报文的巨大网络流量中. 数据获取要全面、实时、真实可靠、尽量不影响网络的正常流量. 在获得数据后要进行“清理”, 即剔除冗余的、无效的数据, 将无序的、杂乱的数据整理成有序的、完备的证据, 并进行规范化表示, 为基于行为数据的动态信任评估奠定坚实的基础.

目前可用于获取行为的方法有^[11]: (1) 利用网络流量检测与分析工具, 例如 Bandwidthd, 可以获得每个网关的各种协议的详细 IP 流量, 查看网络状态, 如数据包的传输和接收速率等. (2) 利用已有的入侵检测系统, 例如, RealSecur, 可以获得访问次数、操作失败次数和时延. (3) 利用审计跟踪系统产生的系统事件记录和实体行为记录, 包括系统日志、审计记录、应用程序日志、网络管理日志截获的用户数据包以及相应的操作记录. (4) 专门的数据采集工具, 如 Cisco 的 NetFlow Monitor, NetScout 公司的 NetScout 网络性能管理产品, 可以实时获得网络的带宽利用率, 不同用户对带宽的占用等. (5) 自开发的软硬件系统. 尽管如此, 仅凭现有技术还有很多证据不能获得, 如何获得全面的证据也是动态信任管理技术的重要内容之一. 对于目前检测不到的证据可以采用其它的方法进行研究, 例如, 可以根据以往用户的相关证据进行推理和预测, 也可以与专家的经验判断相结合来确定未知的证据值.

本文设计和部署了两种类型的软件传感器用来获取系统中实体的行为数据^[12]: (1) 监测传感器 (monitoring sensors). 负责采集软件和计算资源交互过程中的常用行为参数, 例如, 网络带宽利用率、内存和 CPU 利用率和应用行为隐患 (包括端口、系统调用等可能潜在的入侵行为和恶意行为) 等; (2) 计算传感器 (calculating sensors). 从该处获取作业执行的成功率、错误修复率、资源站点自防御的能力以及平均无故障时间 (MTBF) 等. 所有可能获取的这些行为数据分别可划分到不同的信任属性之中 (图 2). 遵循行为属性选取的专业性和可操作性原则, 对于不同的应用环境应该从不同的角度去抽取影响信任度测的行为数据, 例如 P2P 文件共享系统和网格计算系统可能关注的行为证据有很大的差

异, P2P 文件共享系统更关心节点的非法操作带来的安全隐患, 所以应尽可能多地收集安全相关的行为数据, 而网格计算系统在任务调度时, 主要关注资源的可用属性, 所以应尽可能多地收集可用性相关的数据.

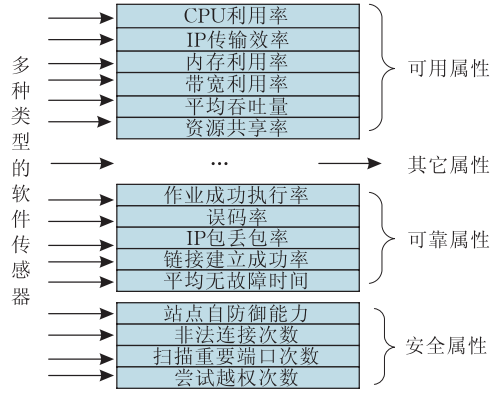


图 2 开放系统中基于行为数据的多维信任度测指标

通过软件传感器获得的行为数据, 其表现形式为在一定范围内的具体值 (物理量纲值: dimension data) 或者百分比数据 (percentage data), 例如平均无故障时间和扫描重要端口次数, 都是一个在某一范围内的具体值, 平均无故障时间是沿正向递增的 (positive-increasing), 即越大越好, 扫描重要端口次数是沿正向递减的 (positive-decreasing), 即越小越好. 由于行为监测数据的表示多样性, 为了便于融合计算, 需要把数据表示进行规范化, 即把它们全部表示为在 $[0, 1]$ 区间沿正向递增的无量纲值, 这样不仅便于数值融合计算而且也与网络实体信任度测值的范围和方向相一致.

将软件传感器获得的数据按照时间序列进行排列, 则在某个时间戳 n 共有 n 组需要处理的行为数据, 这 n 组数据中的每一组数据称为一个样本. 这样, 共有待处理的 n 个样本 $X = \{x_1, x_2, \dots, x_n\}$, 每个样本的属性集合表示为 $x_j = \{x_{j1}, x_{j2}, \dots, x_{jm}\}$, 可用阶特征矩阵:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

表示某一层的关系数据模型. 规范化后的矩阵为 $B = (b_{ij})_{m \times n}$, 称 B 为评判矩阵, 则

$$b_{ij} = \begin{cases} x_{ij} & (a) \\ 1 - x_{ij} & (b) \\ (x_{ij} - r_{\min}^j) / (r_{\max}^j - r_{\min}^j) & (c) \\ (r_{\max}^j - x_{ij}) / (r_{\max}^j - r_{\min}^j) & (d) \end{cases} \quad (4)$$

其中, $r_{\max}^j = \max_{1 \leq j \leq m} \{x_{ij}\}$, $r_{\max}^j = \min_{1 \leq j \leq m} \{x_{ij}\}$. (a) 表示 x_{ij} 是正向递增百分比; (b) 表示 x_{ij} 是正向递减百分比; (c) 表示 x_{ij} 是正向递增量纲值; (d) 表示 x_{ij} 是正向递减量纲值. 通过规范化处理, 所有的行为证据都可以转换为 $[0, 1]$ 范围的正向递增值. 这样每个行为证据的值越大, 该证据对实体的 OTD 的贡献也越大.

4 自适应的动态信任度测方法

如前所述, 现有的信任度量与预测模型一般都是建立在经典的基于状态演绎的概率统计基础之上的^[8], 例如: D-S 信任度测模型和 Bayesian 信任度测方法等. 人们为建立演绎模型不得不做出一些主观的假设, 以便将可信性动态变化这个复杂的物理过程进行简化, 从而最终建立度测模型, 这样, 模型的准确度必然会受到影响. 模糊数学的发展深化了人们对可信性模糊现象的认识^[6], 但它只是概率方法的合理补充, 且建立评估规则过于复杂, 要花大量时间培训, 这就降低了它的实用性. 另外, 由于一种有效的可信性度测方法必然具有多维决策属性(度测指标)^[10], 而目前模型在计算可信程度时只考虑少数指标(甚至在有些文献中只考虑单一的指标), 或者由于数学模型本身的限制而无法处理较多的指标(例如, 基于模糊逻辑的方法中^[6], 当输入的度测指标较多时, 就会产生所谓的“规则爆炸”问题, 规则的获取将需要巨大的时空开销). 由此导致模型对一些重要的环境和行为上下文考虑不够, 在网络日益呈现复杂化、网络攻击形式也日益多样化和难以预测的情况下, 模型就不能很好地刻画可信关系的不确定性, 表现出对多源上下文处理能力不足的问题.

粗糙集和信息熵理论分别在金融管理、医疗诊断、遥感遥测、组织管理决策、故障诊断和分布式入侵检测等诸多领域展开广泛而深入的研究与应用, 而在动态可信性度测领域的研究与应用, 到目前为止尚未见到相关的报道. 鉴于此, 本文进行了相关问题的探索, 将粗糙集和信息熵理论结合起来, 应用于在开放的分布式环境下构建基于行为数据挖掘的动态信任度测模型. 新模型直接从分析传感器监测到的行为数据入手, 跳出了现有信任关系建模过程中的各种统计分布假设的束缚, 并可克服现有模型对多源行为数据处理能力不足的问题.

4.1 信任度测知识表达系统的构建

在粗糙集理论中, “知识”被认为是一种分类的

能力, 集合是根据“知识”划分的, 粗糙集理论把用于划分集合的知识嵌入集合本身, 从而扩展了经典的集合论. 设有待处理的 n 个样本(历史证据集或者行为属性集)组成的集合 $X = \{x_1, x_2, \dots, x_n\}$, 为了对样本的获取时间戳进行标识, 我们对 X 中的样本按照实体之间交互的时间顺序进行排列, x_1 表示离现在较久的一次交互样本, x_n 表示离现在最近的一次交互样本, 每个样本拥有 m 个输入行为数据(图 1), 某个样本的属性值用集合可表示为 $x_j = \{x_{j1}, x_{j2}, \dots, x_{jm}\}$, 若将某个样本的属性数据视为研究对象的一条信息, 则可定义论域为 $X = \{x_1, x_2, \dots, x_n\}$. 由拥有 m 个属性的 n 个样本的数据所构成的二维信息表就是动态信任度测的知识表达系统.

由论域、条件属性和决策属性构成的知识表达系统可表示为 $T = (X, A, C, D, f)$, 其中, $C = \cup \{x_{ij}\} (1 \leq i \leq n, 1 \leq j \leq m)$ 为条件属性, $D = \{d_1, d_2, \dots, d_n\}$ 为决策属性, $A = C \cup D, C \cap D = \emptyset, f: X \times (C \cup D)$ 为综合评判函数, 通常将具有条件属性 C 和决策属性 D 的知识表达系统称为系统决策表(System Decision Table, SDT), SDT 中的每一行表示一种决策规则. X 中对象根据决策规则的不同被划分到不同决策类中. 由于 X 中的样本是按照交互发生的时间顺序排列的, 因此, 我们称此时 $T = (X, A, C, D, f)$ 为建立的基于时间戳的可信系统决策表(图 3).

论域X	条件属性				决策属性	
	x_1	x_2	...	x_m	OTD	
时间 序 列 ↓	x_1	x_{11}	x_{12}	...	x_{1m}	d_1
	x_2	x_{21}	x_{22}	...	x_{2m}	d_2

	x_j	x_{j1}	x_{j2}	...	x_{jm}	d_j

	x_n	x_{n1}	x_{n2}	...	x_{nm}	d_n

图 3 基于时间序列的信任系统决策表

4.2 分类知识获取算法(CKAA)

由式(2)和(4)可以看出, OTD 融合计算的关键问题是如何合理地分配各级属性(就是规范化后的行为证据)的分类权重 ω_i 的值, 进而完成对行为数据的量化和 OTD 的计算. 对于包含条件属性 C 和决策属性 D 的知识表达系统 $T = (X, A, C, D, f)$, 在建立系统决策表后, 利用决策表对拥有全部属性的论域中对象进行分类, 并逐次删减某一属性后重新分类, 再结合粗糙集理论中的属性依赖度和重要度

就可计算出各决策属性的分类权重. 但现在的关键问题是, 由于先验知识的不足, 在没有得到各决策属性的分类权重之前, 我们还无法使用式(2)给出的综合函数进行融合计算, 也就无法确定决策属性 D 的值, 当无决策属性时, 也就无法利用粗糙集中的属性依赖度和重要度对论域中的对象进行分类. 针对这一问题, 本文利用模糊聚类的方法获取相关证据(行为数据)的分类知识, 然后使用信息熵理论中的互信息量来确定各指标的权重, 使得结果具有相对的客观性, 而且各权重可以根据网络数据的动态变化进行动态的调整, 具有良好的自适应性.

算法 1. 分类知识获取算法 (Classification Knowledge Acquisition Algorithm, CKAA).

1. 输入 $T' = (X, C, f)$, 其中 $X = \{x_1, x_2, \dots, x_n\}$, $C = \bigcup \{x_{ij} \mid (1 \leq i \leq n, 1 \leq j \leq m)\}$;
2. 根据式(4)求解评判矩阵 $B = (b_{ij})_{m \times n}$;
3. 计算出表征被分类对象间相似程度的相似系数 δ_{ij} , 从而建立 X 上的相似关系矩阵, 由于各属性对样本的相似性影响不一致, 为了更好地反映各属性对相似性的影响程度, 本文采用几何平均最小法计算模糊相似矩阵 \tilde{R} :

$$\delta_{ij} = \frac{\sum_{k=1}^m \min(b_{ik}, b_{jk})}{\sum_{k=1}^m \sqrt{b_{ik} \times b_{jk}}} \quad (5)$$

4. 计算相似矩阵的等价闭包矩阵 $e(\tilde{R})$, 步 3 建立起来的相似关系一般只满足自反性和对称性, 不满足传递性, 所以还需要 \tilde{R} 改造成等价闭包矩阵, 用平方法求 $e(\tilde{R})$:

$$\tilde{R} \rightarrow \tilde{R}^2 \rightarrow (\tilde{R}^2)^2 \rightarrow \dots \rightarrow \tilde{R}^{2k} = \tilde{R}^{2k-1} \quad (6)$$

则 $e(\tilde{R}) = \tilde{R}^{2k-1}$;

在得到等价闭包矩阵 $e(\tilde{R})$ 之后, 我们可以根据矩阵系数确定若干个分类的置信水平 $\gamma_k (k=1, 2, \dots, \lambda)$;

5. 分别在各个不同的置信水平 $\gamma_k (k=1, 2, \dots, \lambda)$ 上, 以置信区间的左区间值为阈值, 根据全部属性的等价闭包矩阵将论域 X 进行等价类的划分, 此时, 共可以得到 λ 个分类结果. 设在置信水平 $\gamma_k (k=1, 2, \dots, \lambda)$ 上将 X 划分为 r 个等价类, 记为 $X/R_{\gamma_k} = \{U_1, U_2, \dots, U_i, \dots, U_r\}$;

6. 依次从全部属性中删除某个属性, 重复步 3、步 4, 分别计算等价闭包矩阵. 得到删除某一属性的等价闭包矩阵之后, 在相同置信水平 $\gamma_k (k=1, 2, \dots, \lambda)$ 上, 将论域 X 划分为 v 个等价类, 记为 $X/V_{\gamma_k} = \{U'_1, U'_2, \dots, U'_j, \dots, U'_v\}$;

7. 输出知识 $R_{\gamma_k}, V_{\gamma_k} (k=1, 2, \dots, \lambda)$;

8. 结束.

此时, $R_{\gamma_k}, V_{\gamma_k}$ 称为在 $T = (X, A, C, D, f)$ 上导出的两种知识. 据此, 可分析和计算各属性对分类的影响和两种知识的互信息量, 进而确定各属性所包含的信息量与分类权重.

4.3 分类权重计算方法(CWCA)

信息熵在事件发生之前, 它是结果不确定性的

量度, 在事件发生之后, 它是我们从该事件中所得到信息的量度(信息量). 因此, 事件的信息熵, 是一个事件的不确定性或信息量的量度, 也可以理解为包含在这个事件本身中的关于它自己的信息. 文献[13]讨论了知识粗糙性与信息熵之间的关系, 证明了在无决策信息系统中, 知识约简在信息和代数两种不同表示下是等价的, 从而从信息论角度刻画了粗糙集理论的本质. 下面在文献[13]的基础上, 给出基于信息熵理论中互信息量的分类权重计算方法.

算法 2. 分类权重计算算法 (Classification Weight Calculated Algorithm, CWCA).

1. 输入知识 $R_{\gamma_k}, V_{\gamma_k} (k=1, 2, \dots, \lambda)$;
2. 计算 $R_{\gamma_k}, V_{\gamma_k}$ 在 X 的子集组成的 σ 代数上的概率分布:

$$[R_{\gamma_k} : p] = \begin{bmatrix} X_1 & X_2 & \dots & X_r \\ p(X_1) & p(X_2) & \dots & p(X_r) \end{bmatrix} \quad (7)$$

其中 $p(X_i) = \frac{|X_i|}{|X|}$, $i=1, 2, \dots, r$;

$$[V_{\gamma_k} : p] = \begin{bmatrix} X'_1 & X'_2 & \dots & X'_v \\ p(X'_1) & p(X'_2) & \dots & p(X'_v) \end{bmatrix} \quad (8)$$

其中 $p(X'_i) = \frac{|X'_i|}{|X|}$, $i=1, 2, \dots, v$;

3. 计算 R_{γ_k} 的初始熵 $H(R_{\gamma_k})$:

$$H(R_{\gamma_k}) = - \sum_{i=1}^r p(X_i) \log(p(X_i)) \quad (9)$$

4. 计算 R_{γ_k} 相对于 V_{γ_k} 的条件熵 $H(R_{\gamma_k} | V_{\gamma_k})$:

$$H(R_{\gamma_k} | V_{\gamma_k}) = - \sum_{i=1}^r p(X_i) \sum_{j=1}^v p(X'_j | X_i) \log(p(X'_j | X_i)) \quad (10)$$

其中,

$$p(X'_j | X_i) = |X'_j \cap X_i| / |X_i|, \quad i=1, 2, \dots, r; \quad j=1, 2, \dots, v;$$

5. 根据步 4、步 5 的计算结果可进一步计算两种知识之间的平均交互信息量(简称为互信息量) $I_{\gamma_k}(R_{\gamma_k} | V_{\gamma_k})$:

$$I_{\gamma_k}(R_{\gamma_k} | V_{\gamma_k}) = H(R_{\gamma_k}) - H(R_{\gamma_k} | V_{\gamma_k}) \quad (11)$$

6. 互信息量反映知识 R_{γ_k} 从知识 V_{γ_k} 上获取的信息量, 某一属性所含信息量可表示为

$$K_i = \frac{1}{\lambda} \sum_{k=1}^{\lambda} \gamma_k I_{\gamma_k}(R_{\gamma_k} | V_{\gamma_k}), \quad i=1, 2, \dots, m \quad (12)$$

7. 将各因素所含信息量相对大小归一化处理确定权重分配

$$\omega_i = \frac{K_i}{\sum_{i=1}^m K_i}, \quad i=1, 2, \dots, m \quad (13)$$

8. 结束.

由算法 2 的计算步骤 6、7 容易判断, 得到的分类权重值 $\omega_i (1 \leq i \leq m)$ 满足归一性与非负性的要求. 由此, 我们可以使用算法 1、2, 求得各行为数据

的分类权重,然后由式(2)求得服务请求者(SR)的 OTD.

5 实验与性能分析

性能评测的主要指标是模型能够在两个网络实体之间准确快速地建立起信任关系. 因此性能的评估主要从 2 个方面进行考查:(1)模型的准确性,检查所提出的度测模型与算法是否能提供准确和一致的可信性判别;(2)输入行为数据规模的可扩展性,随着对动态信任属性内涵的深入研究,可能会不断有新认识和新定义的行为属性加入,所以模型应该具有较好的行为数据规模的可扩展性.

5.1 实验方法

由于受许多不确定性因素的影响,不可避免地存在预测误差(forecast error). 准确度是指预测值与真值之间的符合程度,准确度的高低常以误差的大小来衡量. 即误差越小,准确度越高;误差越大,准确度越低. 度测模型的最终目标是实现无偏差的预测,如果一个预测值被认为是准确的,只要在特定的上下文和时间内,对于给定的网络实体的预测信任值与实际的信任值小于一个特定精确度误差值即可. 设 \bar{d}_n 为时刻 n 对于下一个时刻 $(n+1)$ 的预测信任值, d_{n+1} 为时刻 $(n+1)$ 的实际信任值. 下面给出实验中用到的两个衡量算法预测准确性的指标^[14]:

(1) 平均绝对偏差 (Mean Absolute Deviation,

MAD)

$$MAD = \frac{\sum_{t=1}^n e_t}{z}$$

(14)

其中 e_n 为时刻 n 的预测误差, $e_n = d_{n+1} - \bar{d}_n$, z 为实验进行的总次数. MAD 用来衡量整个预测周期内每一次预测值与实际值的绝对误差(不分正负,只考虑偏差量)的平均值. MAD 能较好地反映预测的准确度,但它不容易衡量预测的无偏性.

(2) 平均绝对误差百分比 (Mean Absolute Percentage Error, MAPE)

$$MAPE = \frac{1}{z} \sum_{t=1}^n \left| \frac{e_t}{d_{n+1}} \right| (\times 100\%)$$

(15)

其中 e_n 为时刻 n 的预测误差, $e_n = d_{n+1} - \bar{d}_n$, z 为实验进行的总次数.

实验共采用 1000 组数据样本,这些实测的训练样本来源于西安交通大学开发的基于校园网的网格实验系统 Grid-WADER2.0. 每一个样本主要包括图 1 给出的多个行为证据(数据). 作为参照,比较了本文模型 RTM(Rough Trust Model)中 F-I 方法和 FTM^[6] (Fussy Trust Model)中基于模糊集的方法的性能. 实验使用自开发的原型系统和 matlab 程序相结合的方法完成,原型系统实现了本文介绍的相关算法的全部功能,matlab 程序主要用来对计算结果进行性能分析,图 4 为原型系统的主控界面. 具体的实验方法如下:

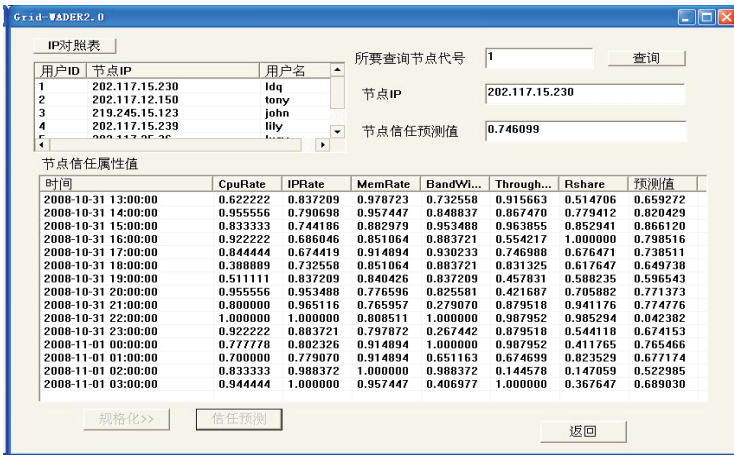


图 4 原型系统的主界面

(1) 在 1000 组数据样本中随机选择一个时间戳 $n(n < 1000)$, 则 n 对应一组样本观测值 $x_n = \{x_{n1}, x_{n2}, \dots, x_{nm}\}$, 那么实验中可使用(输入)的样本总数就是 n 个, 这 n 个数据(称为训练样本)可以构成一个知识表达系统 $T = \{X_n = \{x_1, x_2, \dots, x_n\}$,

$C, f\}$;

(2) 根据算法 1、算法 2 和式(2)计算得到的 n 个时刻的 OTD (d_1, d_2, \dots, d_n) 的预测值作为信任的输出值, 也就是 \bar{d}_n (注意, 在实际应用环境中, 在下次交互行为没有发生之前, $X_n = \{x_1, x_2, \dots,$

x_n }是已知数据,而 x_{n+1} 还不知道);

(3)将下一个时间戳 $n+1$ 时刻的样本观测值 $x_{n+1}=\{x_{n+1,1},x_{n+1,2},\cdots,x_{n+1,m}\}$ 加入训练样本集,那么实验中可使用(输入)的样本总数就是 $n+1$ 个,根据算法 1、2 计算得到的时刻 $n+1$ 的 DTD 作为信任的实际输出值,也就是 d_{n+1} ;

(4)重复步(1)~(3)共进行 z 次实验(本文中取 $z=10$),然后利用式(14)、(15)进行性能评估.

5.2 实验结果分析

(1)准确性评估

首先观察在不同输入的训练样本情况下两种模型的 MAD 的比较结果(图 5),实验中输入的训练样本总数 $20<n<200$,共进行 10 次相关计算.由于 MAD 用来衡量模型测定结果对平均值的偏离程度,它的值越趋近于 0,预测的结果准确度越高,从图 5 的计算结果可以看出,当输入训练样本数目较少时,RTM 的 MAD 要明显优于 FTM,例如当 $n=20$ 时,RTM 的 $MAD=0.130945$,而 FTM 的 $MAD=0.171945$,RTM 要高出 FTM 将近 5 个百分点.从图中还可以看出,输入样本数目大于 40 时,RTM 的 MAD 的拟合曲线的变化情况要比 FTM 的 MAD 的拟合曲线变化情况平缓,且 RTM 的 MAD 比 FTM 平均低 3%,说明 RTM 需要较少的训练样本就能获得比较好的准确性;而当输入样本数目较多时,从图 5 中可以看出,RTM 和 FTM 两者的 MAD 逐渐接近,并且变化逐渐地平缓,说明当输入样本数目增加到一定程度之后,再显著增加样本的数目,并不能再明显地提高系统的性能,所以,在一个实际的应用系统中,可以根据系统的计算能力适当选择训练样本的数目,也就是训练样本数目的输入不宜过多,这样可以有效提高系统的快速响应性,增强系统的动态适应能力.

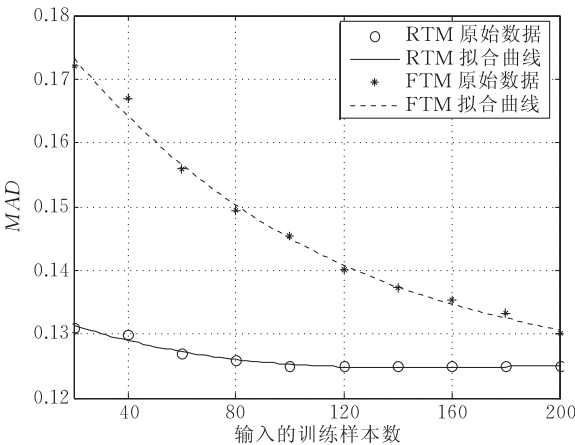


图 5 不同训练样本情况下 MAD 的比较

MAD 用来衡量整个预测周期内每一次预测值与实际值的绝对误差(不分正负,只考虑偏差量)的平均值. MAD 能较好地反映预测的准确度,但它不容易衡量预测的无偏性.通过指标 MAPE 可以反映出度测模型的无偏性,该值也是越小越好,较小的 MAPE 值,说明度测模型具有较好的无偏的预测准确性.图 6 为不同输入训练样本情况下两种模型的 MAPE 的比较结果,实验中输入的训练样本总数 $20<n<200$,共进行 10 次相关计算.从图 6 所示的整个实验周期中,RTM 的 MAPE 平均为 12.56,而 FTM 的 MAPE 平均为 14.75,可以看出,实验结果与图 5 的结果基本吻合,RTM 模型具有更好的预测准确性和无偏性.

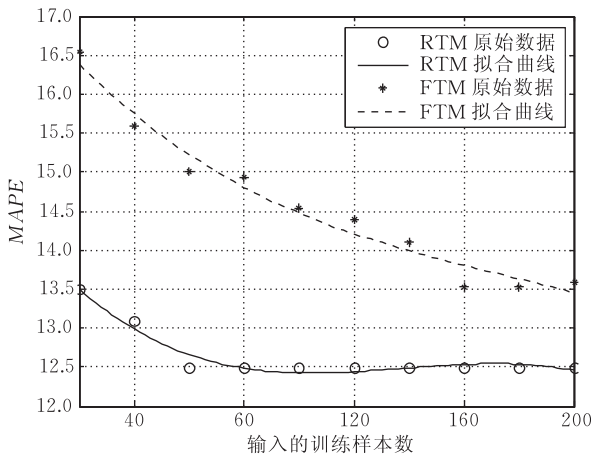


图 6 不同训练样本情况下 MAPE 的比较

(2)可扩展性评估

由于一种有效的度测方法必然具有多维(multi-dimensional)的决策属性(指标),而目前模型由于采用的数学工具本身的限制而无法处理多维的度测指标(例如,基于模糊逻辑的方法中,当输入的度测指标较多时,就会产生所谓的“规则爆炸”问题,规则的获取将需要巨大的时空开销,而快速响应性是动态信任管理技术的基本需求,对于一个时空开销较大的计算方法是没实用价值的),表现出对多维数据处理能力不足的问题.随着对动态信任属性内涵的深入研究,可能会不断有新认识和新定义的行为属性加入,所以模型应该具有较好的数据规模的可扩展性.

图 7 显示了使用不同的输入训练样本数目情况下 RTM 和 FTM 融合计算时间(Fusion Computing Time,FCT)的比较结果.从图 7 可以看出,在输入训练样本数目较少时,RTM 和 FTM 的 FCT 比较接近,但是随着输入样本数目的线性增加,RTM 的

FCT 拟合曲线变化仍然比较平缓,而 *FTM* 的 *FCT* 拟合曲线增长较快,说明当输入样本数目较多时,*RTM* 比 *FTM* 需要更少的计算时间,而且输入样本的数目不会显著增加系统的 *FCT*,反映出 *RTM* 比 *FTM* 具有更好的模型的快速响应性,而对于一个实际的分布式应用系统来说,系统必须能够及时发现与处理恶意的实体行为,所以快速响应性是衡量一个系统是否具有较好的动态适应能力的主要指标.从实验结果可以看出,*RTM* 比 *FTM* 具有更好的快速响应性,所以,*RTM* 比 *FTM* 具有更好的模型的动态适应能力. *RTM* 出现这种较好的快速响应性的主要原因是其主要算法采用自适应的动态评估方法,它的 *F-I* 算法具有直接从软件传感器检测到的输入数据快速获取分类知识的能力,而不需要其它的辅助决策手段(模糊逻辑的方法中计算规则的获取需要主观的定义隶属函数).

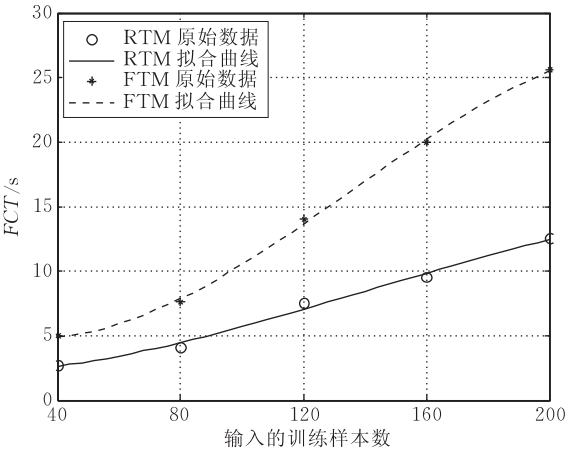


图 7 不同训练样本情况下 *FCT* 的比较

图 8 所示的是在不同的输入行为数据(证据)数目情况下 *RTM* 和 *FTM* 的 *FCT* 的比较结果. 在实验中,训练样本的总数为 80 条,分别比较输入行为证据数目为 3,6,9,12,15 的情况下,分别计算 *RTM* 和 *FTM* 二者的 *FCT*. 从图 8 可以看出,随着证据指标数目的增加,*FTM* 的 *FCT* 迅速增加,其拟合曲线呈现出指数增长趋势,而 *RTM* 的 *FCT* 拟合曲线变化比较平缓,呈现线性缓慢增长的趋势. 实验结果说明,当证据指标增加到一定数目时,*FTM* 模型融合计算过程需要较大的系统时间开销,而 *RTM* 模型对证据指标数目不是特别敏感,说明 *RTM* 具有比 *FTM* 更好的输入行为证据数目的可扩展性. 与 *FTM* 相比,*RTM* 出现这种快速响应性的原因也是十分明显的:在 *RTM* 中,主要的运算为矩阵的算术运算,而没有复杂的迭代计算,所以系统的计算复杂

性较低,而 *FTM* 系统中,当输入变量个数(行为证据数目)较多时,为建立完备的规则库,造成描述数据变得较大,使规则的提取需要大量的时空开销,例如信任评估需要 9 个输入参量,每个参量隶属函数标记数目分别为 3,3,3,4,4,4,5,5 和 5,为了建立输入样本与规则的完全匹配,则 *FTM* 需要建立起完备的规则库应当包含 $3^3 \times 4^3 \times 5^3 = 216000$ 条数据,显然建立这么大的规则库需要比较多的系统开销.

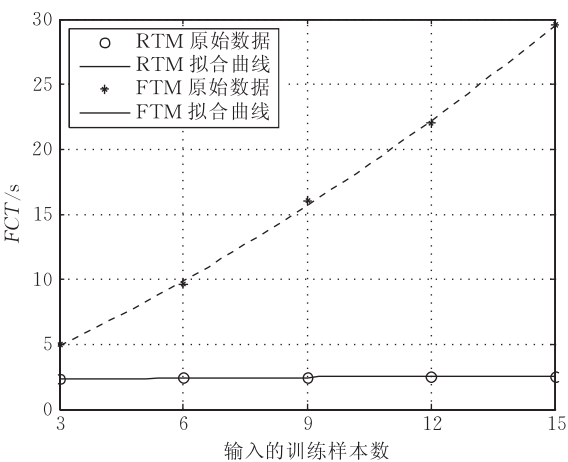


图 8 不同的行为数据情况下 *FCT* 的比较

5.3 进一步讨论

(1) 动态自适应性问题

传统信任模型在总体信任程度融合计算时,大多采用专家意见法或平均加权法等主观的融合计算方法,致使预测结果带有较大的主观成分,影响了可信决策的科学性,而且缺少灵活性,一旦决策属性的权重通过主观方式确定,将在实际应用中很难由系统动态地去调整它,致使预测模型缺少自适应能力. 粗糙集和信息熵理论的结合可以直接从分析传感器监测到的动态数据入手,针对影响信任的多个度测指标进行自适应的数据挖掘与知识发现,完全不需要任何的先验知识与主观的假设,因此具有更好的实际应用价值. 综合图 5 和图 6 的实验结果,可以看出,本文的信任度测方法能提供准确和无偏的信任关系度测,也说明本文模型具有较好的动态自适应性.

(2) 计算的效率问题

动态信任管理强调动态地收集相关的主观因素和客观证据的变化,以一种及时的方式实现对网络实体信任度测、管理和决策,并对网络实体的信任度进行动态更新与演化. 由此可见,信任关系度量与预测模型的快速响应性是开放分布式环境信任管理的

基本需求之一。和模糊逻辑理论类似,当输入的训练样本数据较多时,利用粗糙集和信息熵相关的算法进行总体信任度融合计算也需要较多的时空开销,实验结果见图 7。然而,从社会学的角度来看,信任关系本质上是最复杂的社会关系之一,具有不确定性、不对称性、部分传递性和时空衰减性等一系列复杂的动态属性,是一个抽象的心理“认知”过程,按照人类对信任的认知习惯,信任关系是一种随时间变化而动态衰减的量,也就是隔的时间越久,以前的信任值对现在信任预测的贡献越小。根据这一基本认知,我们在计算总体信任度时,也没有必要对很久以前的数据作为输入的训练样本,而仅仅需要将最近若干个时间戳内采集到的数据作为输入的训练样本,这样,不但可以提高算法的执行效率,也更加符合人类社会对信任关系的心理认知习惯。

6 结论与下一步工作

动态信任管理技术是近几年才发展起来的,仍然属于前瞻性的研究课题:从近几年国内外关于这一问题的研究进展来看,当前工作对很多相关理论和技术性问题都没有达成共识,仍缺乏系统明确的方法论指导,还无法完全解决互联网发展过程中对于信任关系快速和准确的度量与预测的需求;从目前研究的发展趋势来看,动态信任关系的度量模型与预测技术不但是解决大规模开放分布式网络安全问题的基础性工作和必须首先解决的核心科学问题,也是近年来可信网络、可信软件等新型可信计算领域的基础性研究课题。

粗糙集和信息熵理论分别在金融管理、医疗诊断、遥感遥测、故障诊断和分布式入侵检测等诸多领域得到广泛而深入的研究与应用^[13-16],而在分布式动态可信性度测领域的研究与应用,到目前为止尚未见到相关的报道。鉴于此,本文进行了相关问题的探索,将粗糙集理论和信息熵理论结合起来,应用于开放环境下动态构建基于行为数据监控与分析的信任关系度测(度量与预测)模型。该方法直接从分析传感器监测到的动态数据入手,针对影响信任的多个度测指标进行自适应的数据挖掘与知识发现,从而改变了传统的信任关系建模思路,跳出了传统信任关系建模过程中各种主观假设的束缚,并克服了传统模型对多维数据处理能力不足的问题。实验结果表明,与已有模型相比,新模型能够快速准确地实现开放分布式环境下实体的可信性判别,而且具有

良好的行为数据规模的扩展能力。

下一步的工作重点是:结合人类的心理认知过程,进一步研究信任关系的内涵,尤其是动态信任关系的相关性质、信任的表述和度量的合理性,这对信任关系的建模是非常重要的,也是信任关系建模的基础;对本文模型做进一步的完善,并结合其他学科的知识,如认知学习理论等,继续探索适合描述动态信任关系的普适的度测模型。

参 考 文 献

- [1] Li Xiao-Yong, Gui Xiao-Lin. Research on dynamic trust model in large-scale distributed environment. *Journal of Software*, 2007, 18(6): 1510-1521(in Chinese)
(李小勇, 桂小林. 大规模分布式环境下动态信任模型研究. *软件学报*, 2007, 18(6): 1510-1521)
- [2] Ji M, Orgun M. Trust management and trust theory revision. *IEEE Transactions on Systems, Man and Cybernetics*, 2006, 36(3): 451-460
- [3] Wang Hui-Ming, Tang Yang-Bin et al. Trusted mechanism of Internet software. *Science in China, Series E*, 2006, 36(10): 1156-1169(in Chinese)
(王怀民, 唐扬斌等. 互联网软件的可信机理. *中国科学, E 辑*, 2006, 36(10): 1156-1169)
- [4] Lin Chuang, Peng Xue-Hai. Research on trustworthy networks. *Chinese Journal of Computers*, 2005, 28(5): 751-758(in Chinese)
(林闯, 彭雪海. 可信网络研究. *计算机学报*, 2005, 28(5): 751-758)
- [5] Zhou Rong-fang, Hwang Kai. Power-Trust: A robust and scalable reputation system for trusted Peer-to-Peer computing. *IEEE Transactions on Parallel and Distributed Systems*, 2007, 18(4): 460-473
- [6] Stefan S, Robert S. Fuzzy trust evaluation and credibility development in multi-agent systems. *Applied Soft Computing*, 2007, 7(2): 492-505
- [7] Theodorakopoulos G, Baras J S. On trust models and trust evaluation metrics for ad-hoc networks. *IEEE Journal on Selected Areas in Communications*, 2006, 24(2): 318-328
- [8] Sun Y, Yu W, Han Z, Liu K J R. Information theoretic framework of trust modeling and evaluation for ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 2006, 24(2): 305-319
- [9] Liang Zheng-Qiang, Shi Wei-Song. Enforcing cooperative resource sharing in untrusted Peer-to-Peer environments. *Journal of Mobile Networks and Applications-Springer*, 2005, 10(6): 771-783
- [10] Tian Li-Qin, Lin Chuang. Kind of quantitative evaluation of user behavior trust using AHP. *Journal of Computational Information Systems*, 2007, 3(4): 1329-1334
- [11] Ji Tie-Guo, Tian Li-Qin, Hu Zhi-Xing et al. AHP-based user behavior evaluation method in trustworthy network. *Computer Engineering and Applications*, 2007, 43(19): 123-126 (in Chinese)

(冀铁果, 田立勤, 胡志兴等. 可信网络中一种基于 AHP 的用户行为评估方法. 计算机工程与应用, 2007, 43(19): 123-126)

- [12] Zhu Quan-Xin, Gui Xiao-Lin. Study on the method of active deployment of soft-sensors in grid monitoring system. *Journal of Chinese Computer Systems*, 2007, 28(9): 1630-1636 (in Chinese)
(朱全鑫, 桂小林. 面向网格监控的软件传感器的主动部署方法研究. 小型微型计算机系统, 2007, 28(9): 1630-1636)
- [13] Wang Guo-Ying. Decision table reduction based on conditional information entropy. *Chinese Journal of Computers*, 2002,

25(7): 759-766(in Chinese)

- (王国胤. 基于条件信息熵的决策表约简. 计算机学报, 2002, 25(7): 759-766)
- [14] Gardner Everette S, Jr. Automatic monitoring of forecast errors. *Journal of Forecasting*, 2006, 2(1): 1-21
- [15] Pawlak Z. Rough set theory and its applications to data analysis. *Cybernetics and Systems: An International Journal*, 1998, 29(7): 661-688
- [16] Yulmetyev R M, Emelyanova N A, Gafarov F M. Dynamical Shannon entropy and information tsallis entropy in complex systems. *Physica A*, 2004, 341(11): 649-676



LI Xiao-Yong, born in 1975, Ph. D. candidate. His current research interests include trusted network and dynamic trust management.

GUI Xiao-Lin, born in 1966, Ph. D., professor, Ph. D. supervisor. His research interests include grid computing, cloud computing and dynamic trust management.

MAO Qian, born in 1986, M. S. candidate. Her research interests focus on dynamic trust management.

LENG Dong-Qi, born in 1986, M. S. candidate. His research interests focus on dynamic trust management.

Background

With the widespread applications of large-scale open environments, such as Grid computing, Ubiquitous computing, P2P computing, Ad hoc networks, etc., the technology of dynamic trust management has become a significant requirement from a network security's point of view, and trust evaluating and predicting mechanism has become a determining factor for any given service's success. But the dynamic nature of trust creates the biggest challenge in measuring trust value and predicting trust relationship amongst peers. In recent years, many of state-of-the-art trust models have been proposed, and some of them are very innovative and elaborate, but most of the studies still have some limitations: (1) Many current trust models use simple or one-sided trust decision factors to quantify and predict trustworthiness of service providers or requesters, which may lead to inaccurate or unfair outcome of trust decision. The authors think that when trust relationship between peers cannot be fairly defined, it is unstable, and difficult to manage and predict. (2) In many of previous studies, the subjective assigning method to weights of trust decision factors cannot reflect trust decision scientific and reasonable, and may lead to misjudgment of trust decision result.

Focusing on these problems, in this paper, rough set theory and information entropy theory, are combined and applied to the study of distributed dynamic trust measurement and prediction model based on behavior data. Firstly, a new trusted decision-making method based on historical evidences window is proposed, which not only can reduce the risk and improve system efficiency, but also can solve trust measure-

ment and prediction problem when the direct behavior evidences are insufficient. Then, this paper focuses on trust measurement and prediction model based on behavior evidences: (1) Using the concept of rough set knowledge expression system, trust decision table is set up based on timestamp; (2) Using fuzzy aggregation methods, the new model categories the history evidence records (domain) composed of multi-source monitoring data under different confidence level and obtains relevant knowledge. (3) It uses information entropy theory to determine the classification weight of trust attributes (indicators), and finally implements fusion computing of overall trust degree. The new model works through analysis monitored behavior data by sensors directly, changes the traditional modeling thoughts, brakes away from the fetter of various subjective assumptions in traditional modeling methods, and overcomes the problem of inadequate handling capacity for multi-source behavior data in the traditional trust model. Simulating results shows that the new model can accurately implement trust measurement and prediction process between entities in open and complex distributed environment, and has a better scalable capacity of behavior data.

This work is supported by the National Nature Science Foundation of China (No. 60873071); the National High Technology Research and Development Program (863 Program) of China (No. 2008AA01Z410); Program for New Century Excellent Talents in University of China (NCET No. 05-0829); Scientific and Technological Project in Shaanxi Province, China (No. 2007K04-05).