教育资源网格中的一种信任评估模型*)

董晓华[1,2] 李 季[1] 吴中福[1]

(重庆大学计算机学院 重庆 400044)[1] (重庆大学经济与工商管理学院 重庆 400044)[2]

摘 要 在开放的数据网格中,动态反映资源可信度是一个主要安全问题。本文提出了一种新的信任模型,动态评估资源的信任值,从而提高教育资源网格的安全性和可扩展性。该模型根据实体访问资源的推荐信任值、资源的被访问频率以及资源所在域的信任度综合评估资源信任度。通过和目前流行的针对行为的信任评估模型的实验对比,本文提出的信任模型具有更低的时间复杂度,信任度的评估也更可靠。实验结果表明,该信任模型可作为一种有效的手段,不但可以为数据网格中对资源的信任决策提供支持,防止恶意资源破坏的扩散,同时还能提高资源检索的可靠度。

关键词 教育资源,数据网格,信任模型,安全

A Trust Evaluating Model in Education Resources Grid

DONG Xiao-Hua[1,2] LI Ji[1] WU Zhong-Fu[1]

(Department of Computer Science and Technology, Chongqing University, Chongqing 400044)[1]

(College of Economics & Business Administration, Chongqing University, Chongqing 400044)[2]

Abstract In the opening data grid, reflecting resources dependability dynamically is a basic problem. In this paper a new trust model is presented to enhance security and extensibility of education resource grid by evaluating resources' trust dynamically. For evaluating resources' trust, the trust model combines entity's commendatory trust, resource access frequency and the current domain trust of the resource. Compared with other current trust models, our experiment results show that the trust model proposed has lower time complexity and better dependability. Furthermore, experiment results also show our trust model is an effective method in that it not only provides measures for decision-making of resource trust, which helps to limit the bad effects of malicious resources in data grid, but also increases retrieval accuracy of resources.

Keywords Education resource, Data grid, Trust model, Security

1 引言

教育资源包括视频、音频和文本等内容,网格环境方便了人们对这些资源的访问。在教育资源网格中,用户可根据资源的特征描述进行资源定位与发现。在搜索出来的资源中有大量的相关资源,其中有些资源是需要的资源,而有些资源不是必需的,我们可把有用资源称为较高可信度资源,而其它资源可称为较低可信度资源或不可信资源。同样地,在安全性方面,会存在个别网格节点恶性发布病毒、木马程序,破坏网格资源,这些资源也称为不可信资源。针对这种情况,我们应如何来动态确定哪些是可信资源,哪些是不可信资源呢?

目前,在 P2P 环境中的信任模型研究较多,如基于 PKI[1]的信任模型、基于局部推荐的信任模型[2]、基于角色的信任模型[3]、基于 Bayesian 的信任模型[4]和全局可信度模型[5],这些信任模型都在是 P2P 环境下的各节点没有 QoS 认证的信任模型,不能较好地应用于网格环境。

在网格环境中,信任模型主要针对实体行为的信任而建立的,还没有针对网格中的资源进行动态度量的信任模型。如 Beth 信任度模型[6]、Rahman 信任度模型[7]、Azzedin 信任模型[8~10]等都是针对实体行为的信任模型。在这些基于行为的信任模型中,行为实体的信任度基本上都由直接经验和间接经验得出的。如 Beth 模型引入了经验的概念来表述和度量信任关系,并给出了由经验推荐所引出的信任度推导和综合计算;Rahman 模型将实体间信任关系定义为直接信任关系和推荐信任关系,并通过推荐的形式来传递经验信息;Azzedin 模型根据实体所在的管理域之间的直接信任和推荐信任关系来计算。针对网格资源文[11]提出了一种基于赔偿的网格资源交易模型,把信号博弈理论应用于网格资源可靠性辩识,但这种模型具有主观性,并且对常用资源的信任评估没有激励机制。

在本文中,根据教育资源网格的特点,提出了一种针对教育资源网格的资源信任评估模型—EMFRT(Evaluating Model For Resource Trust)。在该模型中,在每一个域的主节点中通过建立信任代理,管理域内资源的信任度表和域的信任度;通过整合实体访问资源时对资源的信任度评价、资源的被访问频率以及资源所在域的信任度,综合得出评估资源信任值。

2 相关术语

为了更好的论述和理解 EMFRT,我们给出一些定义和形式化描述。

定义 1 主体(subject) 指需要访问网格资源的实体。

*)基金项目:国家发改委基金项目(CNGI-04-15-3A)。**董晓华** 讲师,博士生,研究方向为网格技术、信息系统。**李 季** 讲师,博士,研究方向为网格技术。**吴中福** 教授,博导,主要研究方向为远程教育、网络技术等。

主体可以是网格用户，或者是代表用户的进程。我们用 s 表示主体。

定义 2 客体(object) 指受保护的各种资源。客体可以是 CPU、数据、程序等，用 o 表示。教育资源网格中的视频、音频和文本等教育资源是本文进行信任评估的资源(resource)，用 r 表示，显然 $r\in R, R\subseteq O$。

定义 3 访问(access) 指主体对客体的操作。这些操作包括对 CPU 的调用，数据的添加、修改和删除，程序的运行，文件的创建、读写等。访问操作是一个行为，我们用 Access 表示，Access(s, r)就表示主体 s 对数据、程序等资源 r 的一次访问。

定义 4 自治域(autonomous domain) 指具有相同的管理策略、安全策略的一组网格实体的集合。根据网格实体所属组织的不同以及地理位置的不同，可把网格划分为若干个独立的自治域，自治域之间通过网络连接，如图 1。在本文中，自治域简称为域，用 D 表示。$D_1, D_2, \cdots, D_i, \cdots, D_N$ 表示网格系统中的各个不同的自治域。

定义 5 信任(trust) 一般是指网格实体之间在交互中所能体现的可靠性、诚信度和提供服务的能力。在本文中用 Trust(s,r)表示 Access(s, r)后实体 s 对 r 的推荐信任值，T_D 表示域 D 的信任值，T_r 表示资源 r 的信任值。

信任具有时间性、动态性、应用性、不可传递性和主观性等特点[12]。即严格意义上的信任应该是在特定时间、针对一定应用环境以及根据交互行为动态变化的。

3 教育资源信任评估模型(EMFRT)

在教育资源网格中，不同域的主节点通过网络相连接。每个域中，域主节点连接多个网格节点，网格节点的下层还可以继续连接网格节点。这样，教育资源网格中的每个域就形成了以域主节点为根节点的层次型结构，如图 1。

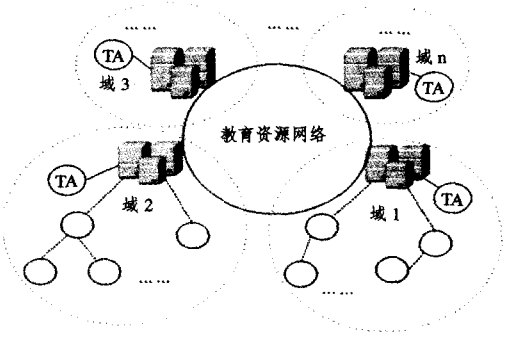

图 1 教育资源网格的结构模型

在每个域中都设置了一个信任代理[13,14](trust agent, TA)，主要完成：

1)更新网格区域中的信任表，包括资源信任度、域信任度；

2)允许实体和客体自由出入网格域并继承其信任特征，并初始设定新加入域或网格节点的信任值；

3)应用时间衰减函数来反映网格域中的信任随时间而变化的关系。

在此网格模型中，网格实体(s)访问网格资源(r)时，资源由每个域内的网格节点提供，根据网格实体所在域的不同可分为域内访问和域间访问。域内访问和域间访问的定义如下：

定义 6 $\exists Access(s, r)$，若 $s, r\in D_i$，则称 s 对 r 的访问为域内访问。

定义 7 $\exists Access(s, r)$，若 $s\in D_i, r\in D_j$，则称 s 对 r 的访问为域间访问。

对于任何一次访问 $Access(s, r)$，信任代理 TA 的执行过程如下：

1)返回 r 的信任值 T_r；

2)接收 s 对 r 的推荐信任 $Trust(s,r)$；

3)更新信任表。

为了叙述方便，在这里我们假定当前待进行信任评估的资源为 r'，$r'\in D'$。

3.1 域内访问

对于域内访问，影响资源 r' 的信任值的因素包括三方面：实体 s 访问资源 r' 的历史推荐值、资源信任度的时间衰减函数 $c_{r'}$ 以及资源的活动因子 $e_{r'}$，计算方式为：

$$T_{r'}=(1+e_{r'})\times T'_{r'}\times c_{r'} \tag{1}$$

其中，

$$T'_{r'}=\frac{\sum_{i=1}^{n} Trust(s_i, r')}{n} \tag{2}$$

表示当前资源 r' 被访问时其历史推荐值的平均值。

$e_{r'}$ 为活动因子，表示资源 r' 在近期某一设定时间内被访问次数的比重。因为某些资源的影响因子(推荐信任度)小，但经常被使用到，所以在此增加了 e_r，从而增加资源 r' 的信任度。其值由近期某一设定时间内资源 r' 被访问次数除以当前域内所有资源 r_k 被访问的次数：

$$e_{r'}=\sum_{i=1}^{n} Access(s_i, r')/\sum_{j=1,k=1}^{j=m,k=l} Access(s_j, r_k) \tag{3}$$

$c_{r'}$ 为资源信任度的衰减函数。资源的信任度是与时间有关的，如资源 r' 是之前发布的过期资源，而最近发布的资源是在资源 r' 之上进行更新的新资源，在访问资源时，新资源应该具有被访问的更大的可能性，而以前发布的资源因为知识陈旧而应该具有较小的信任度。因此，在综合评估资源的信任度时，增加时间衰减函数 $c_{r'}$，令 Δt 为信任评估时间与 r' 上次被访问时间之差，则 $c_{r'}$ 是 Δt 的一元递减函数，如式(4)，λ 可根据具体应用情况设置不同的值，本文中用如下公式表示：

$$c_{r'}=f(\Delta t)=(1+\Delta t\cdot\lambda)^{-1} \tag{4}$$

3.2 域间访问

对于域间访问，影响资源信任度的因素除了上述域内访问时考虑的因素外，还需考虑域自身的信任度。如某些域中节点经常提供不可信或不可靠资源，则此域的信任度将会随着降低，域信任度的高低也将影响域内资源的可信度。因此，域间访问时资源信任的评估模型可表示为：

$$T_{r'}=\alpha\cdot T_{r'}+\beta\cdot T_{D'} \tag{5}$$

其中，α、β 为资源信任度和域信任度的权值，$\alpha+\beta=1$，$\alpha, \beta\in[0,1]$，α 和 β 的值由资源 r' 的访问频率确定，如式(6)：

$$\alpha=\sum_{i=1}^{n} Access(s_i, r')/\left(\sum_{j=1,k=1}^{j=m,k=l} Access(s_j, r_k)+\sum_{i=1}^{n} Access(s_i, r')\right)\beta=1-\alpha \tag{6}$$

α 的值为当域内资源 r' 被访问次数除以当域内所有资源访问次数的平均数与资源 r' 被访问次数之和。因此，如果资源 r' 被访问次数大于平均数，则 $\alpha>0.5$；否则 $\alpha\leqslant 0.5$。

$T'_{r'}$ 为资源 r' 域内访问的信任度，其计算公式为式(1)。

$T_{D'}$ 为当前域的信任度，影响因素包括域间访问时资源信任平均值、域活动因子 $e_{D'}$ 及域信任度时间衰减函数 $c_{D'}$ 三

方面决定：

$$T_{D'}=(1+e_{D'})\cdot\sum_{j=1,k=1}^{j=m,k=l}Access(s_j,r_k)\cdot c_{D'}$$
$$r_j\in D',s_i\notin D' \quad (7)$$

式(7)中，s_i 和 r_j 属于不同的域，$e_{D'}$ 和 $c_{D'}$ 值的设定基本和域内访问的 $e_{r'}$ 和 $c_{r'}$ 相同，不同的是 $e_{D'}$ 表示当前域 D' 的访问次数占整个网格系统中域访问次数的比重，$c_{D'}$ 表示当前域 D' 的信任评估时间与上次域内节点被其它域访问时间衰减函数。

4 实验结果及分析

4.1 资源检索算法描述

本模型中，在资源的检索过程中增加了信任值检索条件，因此检索算法可简单描述为：

①给定待检索资源的信任阈值 t'；

②根据元数据描述找出一系列合适的潜在资源 $R'\{r_1,r_2,\cdots,r_n\}$，检索资源按信任值降序排列，并满足 $Tr_i\geqslant t'$，$r_i\in R'$；

③根据一定的筛选算法从中选择一个(些) 合适资源 R''，$R''\subseteq R'$；

④根据用户对资源的使用情况对资源进行信任评价并按照式(1)或式(5)修改资源信任值。

4.2 实验结果及分析

在本研究小组的 CNGI 综合实验室中，搭建了 CROWN 网格平台，在平台上建立了教育资源网格，并设置有 500 个域网络，每一个域中包括了 200 个节点。我们收集整理了 5000 种不同的资源，每种资源按照如图 2 设定的资源属性复制 6 个不同的版本，6 个资源复制采用相同的索引词元数据描述，但采用不同的标志，然后把这些资源随机分布在网格节点上，资源类型的含义见表 1。同时，我们设定实体 s 对资源 r 的一次访问 access 称为一次事务(Transaction)，所有实体访问都是可信行为，因为每种资源的描述采用相同的索引词元数据，并且信任度的初始设定采用相同的策略，因此每个资源的 6 种不同的属性版本都可能被检索。

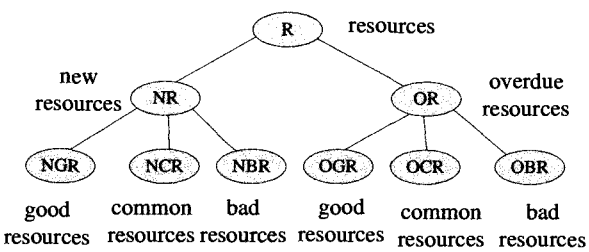

图 2 资源的组织形式

表 1 资源类型含义对应表

资源类别	简称	说 明
new resources	NR	新资源，包括新发布资源及对老资源的更新
overdue resource	OR	旧资源或过期资源
good resources	GR	好的资源，具有较高的应用价值
common resources	CR	普通资源，具有较低的应用价值
bad resources	BR	坏资源，没有应用价值或者对网格资源具有破坏性

在实验中，我们首先在教育资源网格不加入任何信任评估模型进行了实验，实验结果如表 2。随着访问次数的增加，各种资源被访问频率基本保持不变，对网格系统有破坏作用的资源以及一些过期资源仍然具有相当高的访问率。

表 2 未加入信任评估模型各资源的访问频率

Transactions	NGR	NCR	NBR	OGR	OCR	OBR
300	51	53	50	51	48	47
600	103	99	95	109	103	91
900	154	153	150	143	151	149
1500	245	256	257	248	245	249
5000	825	879	819	847	853	777
10000	1716	1763	1626	1728	1705	1462

然后，我们在网格中加入信任模型进行对比实验：一种信任模型是对现有流行的网格实体行为信任评估模型进行改造，用于对资源的信任评估，这种信任模型的信任度是以直接信任和推荐信任通过加权计算而综合得出；另一种就是本文提出的资源信任评估模型。在资源发现时增加了信任度阈值 σ，当某资源 r' 的综合信任值 $Tr'\geqslant\sigma$ 时，才能被检索，检索结果根据资源的 Tr 进行排序。

为了在实验中能方便地进行信任度评价，我们对图 2 中的资源分类模式进行了信任度的等级量化，如表 3 所示。

表 3 资源信任度量化标度

资源名称	标度
NGR	8-10
NCR	4-6
NBR	0-1
OGR	5-7
OCR	1-3
OBR	0-1

在对比实验中，我们分别对资源检索的时间效率和资源的访问频率进行了统计。在资源检索过程中，当资源访问次数较少时，传统模型时间复杂度和 EMFRT 的时间复杂度相近，随着资源访问次数的增加，传统模式的时间复杂度比本文提出的信任模型时间复杂度增加得更快，如图 3。

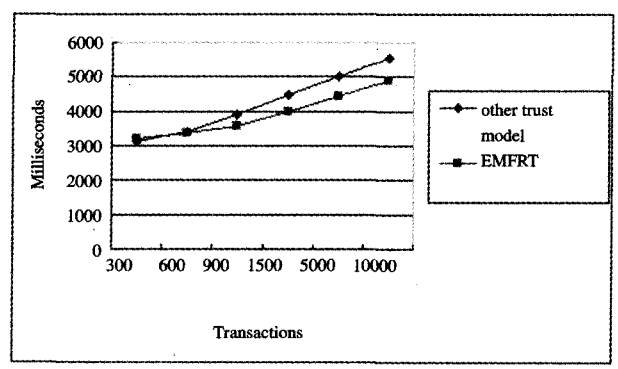

图 3 时间复杂度对比图

同时，通过加入两种信任模型的教育资源网格的对比实验发现，随着访问次数的增加，在 RETM 中经常被访问的资源，其访问频率增加较快，相应地，低可信或者不可信资源的访问频率比传统信任模型要低，如表 4、表 5。特别地，某些价值不高但经常使用的普通资源，也能很好地提高其访问频率，如表 4 中的 NCR，尽管其信任度量化值比 OGR 低，但因其使

用频率高，因此其访问频率比 OGR 略高。

表 4 EMFRT 模型下的各资源访问频率

Transactions	NGR	NCR	NBR	OGR	OCR	OBR
300	91	73	11	76	42	7
600	211	154	13	156	57	9
900	342	235	14	232	67	10
1500	689	362	14	349	76	10
5000	2614	1151	14	1044	167	10
10000	6225	1846	14	1673	232	10

表 5 传统信任模型下的各资源访问频率

Transactions	NGR	NCR	NBR	OGR	OCR	OBR
300	85	74	13	77	42	9
600	185	154	16	159	73	13
900	304	229	18	237	97	15
1500	583	372	20	392	116	17
5000	2189	1225	23	1287	257	19
10000	5535	1989	23	2129	305	19

结束语 网格环境的动态、多组织的特性引入了新的具有挑战性的安全问题，信任模型对安全信息进行度量和评估，能较好地反映出网格环境的动态性和不确定性，因此与实际的应用策略相结合是一种服务于网格系统安全的有意义和有实用价值的途径。本文提出了针对教育资源管理的一种新的信任评估模型，根据实体访问资源的推荐信任值、资源的活动因子、资源信任度的时间衰减函数以及资源所在域的信任度进行资源信任度综合评估。通过对信任评估模型的量化实验表明，该模型可作为一种有效的手段，既可以有效地防止恶意资源对数据网格中资源破坏的扩散，又可以在资源定位时增加定位的准确度，且为数据网格中为资源的信任决策提供支持。同时，该信任模型同样适用于对实体行为的信任评价。将来进一步的工作是对本文提出的信任模型进行进一步完善，如根据不同实体对相同资源的信任推荐的差异性建立模型，以及研究策略防止一些实体的恶意评价。

参 考 文 献

1 Altman J. PKI Security for JXTA overlay network: [Technical Report, TR-I2-03-06]. Palo Alto: Sun Microsystem, 2003

2 Cornelli F. Choosing reputable servents in a P2P network. In: Lassner D, ed. Proc. of the 11th Int'l World Wide Web Conf. Hawii: ACM Press, 2002. 441～449

3 Khambatti M, Dasgupta P, Ryu K D. A role-based trust model for Peer-to-Peer Communities and dynamic coalitions. In: Cole J L, Wolthusen S D, eds. Proc. of the 2nd IEEE Int'l Information Assurance Workshop. New York: IEEE Press, 2004. 141～154

4 Wang Y, Vassileva J. Bayesian network trust model in peer-to-peer networks. In: Moro G, ed. Proc. of the 2nd Int'l Workshop on Agents and Peer-to-Peer Computing. Berlin: Springer-Verlag, 2004. 23～34

5 Kamvar S D, Schlosser M T, EigenRep: Reputation management in P2P networks. In: Lawrence S, ed. Proc. of the 12th Int'l World Wide Web Conf. Budapest: ACM Press, 2004. 123～134

6 Beth T, Borcherding M, Klein B. Valuation of trust in open system, In D. Collmann, Editor, Computer Security, ESORICS'94, volume 875 of Lecture Notes in Computer Science, Springer Verlag, Berlin, 1994. 3～18

7 Abdul-Rahman A, Hailers S. A distributed trust model. In: Proceeding of the 1997 New Security Paradigms Workshop, Cumbia, UK, ACM Press, 1997. 48～60

8 Azzedin F, Maheswaran M. Integrating Trust into Grid Resource Management Systems. In: Proceedings of the International Conference on Parallel Processing, 2002. 47～54

9 Azzedin F, Maheswaran M. Evolving and Managing Trust in Grid Computing Systems. In: Proceedings of the IEEE Canadian Conference on Electrical&Computer Engineering, 2002. 1424～1429

10 Azzedin F, Maheswaran M. Towards Trust-Aware Resource Management in Grid Computing Systems. In: Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, 2002. 452～452

11 Li M S, Yang S B. A grid resource transaction model based on compensation. Journal of Software, 2006, 17(3): 472～480

13 Abdul-Ranman A, Hailes S. Supporting Trust in Virtual Communities. In: Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000. 908～918

14 Dyson J R D, Griffiths N E, Lim Choi Keung. Trusting Agents for Grid Computing. IEEE International Conference on Systems, Man and Cybernetics, 2004. 3187～3192

15 Azzedin F, Maheswaran M. A Trust Brokering System and Its Application to Resource Management in Pubic-Resource Grids. In: Proceedings of the 18th International Parallel and Distributed Processing Symposium, 2004. 289～298

（上接第 43 页）

从图 1 可以看出，当多播组数量增加时，三种算法所产生的代价也随之增加，但其中 BDCMA 增加的幅度最小，即在相同多播组的条件下，它的代价最小。从图 2 和图 3 可以看出，当时延约束增加时，三种算法寻径成功率和平均消息开销也随之而增加。但相比较而言，BDCMA 比 SPT 和 QMRP 的寻径成功率要高。这主要是由于 SPT 采用单路径搜索，QMPR 在不满足 QoS 要求时就自动转换到多路径，而 BDCMA 则采用动态链接而形成的结果。

结束语 笔者重点讨论了具有多 QoS 约束的多播路由问题，主要涉及到带宽、时延和代价等特征值，提出了一种具有多 QoS 约束的动态多播路由算法 BDCMA，该算法避免了以往算法中大部分的盲目路径搜索，它仅仅要求网络的局部状态信息，而不要求维护全局网络状态信息。BDCMA 算法可通过分布式方式搜索到多条可行的路径分支，且能选择较优路径将新成员连入多播树，从而有效减少多播树生成的开销，使之有效的分配网络资源来满足不同的 QoS 需求。仿真试验结果表明，BDCMA 为多约束 QoS 多播路由技术提供了一种新的有效途径，能适用于各种网络规模和群组模型，可扩展性良好，具有较好的应用前景。

参 考 文 献

1 许红梅，许毅. 多 QoS 约束的动态多播路由算法. 交通与计算机，2006(1)：70～72

2 李道奇，许毅. 基于时延及带宽约束的多播路由算法. 武汉理工大学学报-信息与管理工程版，2005(4)：72～76

3 李腊元，李春林. 多 QoS 约束的多播路由协议[J]. 软件学报，2004，15(2)：286～291

4 石坚，董天临，石瑛. 基于 QoS 的动态组播路由算法. 通信学报，2001，8(8)：14～22

5 Waxman B W. Routing of Multipoint Connections [J]. IEEE Journal on Selected Areas in Communication. 1988, 6(9): 1617～1622

6 高茜，李勇，罗军舟. 一种新的 QoS 约束的多播路由协议[J]. 计算机学报，2003. 1441～1449