一个含有信任度的反馈式信任管理模型的设计*)

耿秀华 韩 臻 杨 波
(北京交通大学计算机与信息技术学院 北京 100044)

摘 要 提出了一种含有反馈的信任管理模型,将信任度结合到信任管理中,模型具有自主学习反馈的功能,能够将自己及别人的经验融入系统,使系统在学习中不断得到修正完善。通用、灵活、自动学习是本模型的特点。

关键词 信任管理,信任度,反馈,证书

Design of a Trust Management Model with Degrees of Trust and Feedback

GENG Xiu-Hua HAN Zhen YANG Bo
(College of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044)

Abstract In this paper, a trust management model with feedback is presented, in which degrees of trust are combined. Direct Experience and indirect experience from others can return the model so that it improves itself continuously. The model has the characteristic of generality, flexibility and auto-studying.

Keywords Trust management, Degrees of trust, Feedback, Credential

1 引言

在网络飞速发展的今天,许多网络服务需要做出授权判断,而作为典型的大规模分布式网络,Internet 比以往的集中式网络拥有更多的实体,这些实体之间有些互相并不了解,也没有一个大家可以依赖的权威中心,因此,许多基于传统软件系统形式的安全技术和手段,尤其是安全授权机制,如访问控制列表 ACL(Access Control List)和一些公钥证书体系(X.509,PGP)等不再能够满足 Internet 的安全需求,在这种情况下,信任管理理论应运而生。信任管理的思想最早由 Blaze、Feignbaum 和 Lacy 提出[1],是一种统一地说明和解释安全策略、证书和信任关系的方法,能够灵活跨越多个管理域,准确实现身份识别。它能在整个系统中保持统一的安全策略,同时还具有很强的表达能力和可扩展性,很好地适应了分布式系统的需要,是一种独立于应用的、更适合在分布式网络中使用的访问控制体系。信任管理引擎回答这样一个问题:"证书集合 C 能够证明某个请求 r 与本地安全策略 P 一致吗?"。信任管理引擎是一个独立的系统组件,将(r, C, P)作为输入数据,而输出则是一个判断,即"是否能证明输入数据与策略一致",如果不能证明,则可能还会输出一些附加信息,提示用户应如何处理。

信任是对一个实体具有在特定环境下可信、安全、可靠地行动的能力的坚定信心。信任是一个涉及很广的话题,包括信任的建立、信任管理和安全性相关的问题。在网络与分布式计算安全环境中,信任与信任关系起着非常重要的作用,是基础的和早期决策的部分之一。从这个意义上理解,Blaze 等人提出的信任只是一种理性信任,但信任除理性信任(也称客观信任)外还有感性信任(也称主观信任)。例如,用户信任自己的防火墙系统,也信任网络上的另一个用户,信任前者是因为认为它能抵抗恶意的攻击,是理性信任,后者的信任是相信对方不会有恶意的行为,是带有感性的信任。

在理性信任模型中,实体 A 信任实体 B 是指 A 相信 B 在特定环境下会以一定的方式执行或不执行某项活动,其特点为:

(1)信任是精确的、客观的,要么信任,要么不信任。

(2)信任与活动没有直接的关系,活动的执行与否不会影响信任的情况。

(3)活动多次成功执行或不成功执行也不能影响信任的情况。

理性信任模型中一般实体之间存在固定的关系。在感性信任模型中,信任是表征一个实体(网络中可以推荐信任和接受推荐的对象)执行特定活动的主观可能性程度,该特定活动在事先是不可监控的(与能否被我们监控无关),并且被它自己的活动所影响。它包括三个特性:

(1)信任是主观的;

(2)信任受活动的影响;

(3)信任程度随双方活动的结果而不断修正。

感性信任模型放弃了实体间的固定关系,认为是一种经验的体现,对信任进行量化或者分等级,成为近来研究热点。在不同的感性信任模型中,信任程度的划分和计算也不同。

既然存在两种信任,那么如何将理性信任和感性信任有机地结合起来,建立一个符合现实世界思维习惯的信任管理模型,使信任管理更具有人性化,能较好地反映出 Internet 环境的动态性和不确定性,便是一种有实用价值的思路。

但现有的信任管理系统普遍存在以下问题:

1)信任关系描述的简单性。目前的大部分信任管理系统[2~4]的本质是使用一种精确的、理性的方式来描述和处理复杂的信任关系,实体要么可信,要么不可信。但实际上信任有时是一种主观的行为,是非理性的,是一种经验的体现,取决于经验并随着客体行为的变化不断修正,不仅要有具体的内容,还应有程度的划分。信任关系是动态变化的,会随着时间而改变。文[5]中虽然提出了一种将信任度结合到信任管

*)国家自然科学基金资助项目(60573043)。耿秀华 副教授,博士生,研究方向:信息安全;韩 臻 教授,博导;杨 波 教授,博导。

理系统的框架，但是并没有说明一致性验证器和信任度评估该如何协调得出最终结论，而且这种模型也只是针对软件服务协同的，应用领域有局限性。

2)经验反馈问题。虽然有人也提出了基于信任度的信任管理模型[5]，但该模型没有考虑到经验反馈问题，只是单纯地引用别人积累的经验，而没有将自己的经验融入系统，从而使"经验"成为无源之水，无本之木，在实际应用中缺乏可操作性。

在 Internet 中，实体能不能接受另一个彼此之间并不熟悉的实体(不存在证书链)的访问？如果能行的话，访问依据是什么？本次访问的结果能不能记录下来作为评判该实体在 Internet 中信誉的一个输入？这些便是本文要讨论和解决的问题。本文首先介绍了几个相关的基本概念，之后提出了一种含有信任度的反馈式信任管理模型，将信任度结合到信任管理系统中，分析了其中的几个关键性问题，最后总结并指出了有待于进一步研究深入的一些问题。

2 含有信任度的反馈式信任管理模型框架

2.1 相关概念

一致性验证器：在信任管理系统中用来检验用户证书和用户间的信任关系是否符合本地策略的机制，是信任管理系统中的核心部分。

经验：信任评估主体对客体的观察结果以及第三方提供的该客体的观察结果，前者称为直接经验，后者称为间接经验。

信任度：主体对客体的信任程度，本文信任度值域范围为[0,1]。

2.2 含有信任度的反馈式信任管理模型框架

Blaze 等人提出的信任管理系统，其本质是使用一种精确的、理性的方式来描述和处理复杂的信任关系，实体要么可信，要么不可信，非黑即白，不存在中间状态，而此后的许多学者认为信任是非理性的，是一种经验的体现，不仅有具体的内容还应有程度的划分，并提出了一些基于此观点的信任评估模型。

与以往的基于信任度的信任管理框架不同的是，含有信任度的反馈式信任管理模型框架将理性信任和感性信任有机综合在一起，资源请求者若想访问某资源，在符合资源拥有者本地策略的前提下，或者能够收集到足够的证书证明自己具备访问的资格，或者具有一定的信任度。而对于特定的资源可能要求满足上述两个条件，这些均和本地的安全策略有关。

虽然有信任管理框架[5]也将一致性验证器和信任度评估结合起来构成信任管理引擎，但文中并未说明其具体的结合方式及工作原理，且该框架只是针对软件服务的，应用范围有局限性，更重要的是该框架没有经验反馈机制，只是利用现有的经验推导得出一个信任度，而没有将系统本次访问的经验累积到系统，使系统在经验的基础上得到不断的修正完善。如图 1 所示，含有信任度的反馈式信任管理模型从功能上来讲可分为两大系统：综合判断系统和评估反馈系统，其中，综合判断系统将感性经验和理性经验集中起来并结合本地安全策略对资源请求者是否具备访问资格做出最终判断。评估反馈系统是本模型一个重要的组成部分，负责对访问者本次的行为进行综合评判并做出相应反馈。

综合判断系统共有四个输入：请求描述、本地安全策略、一致性验证器的验证结果以及信任度评估系统的评估结果。一致性验证器用来处理理性经验，即和证书相关的部分，资源请求者将请求及收集到的相关证书送交一致性验证器，一致性验证器首先要验证证书的签名，然后到每张证书的证书颁发处去核实验证资源请求者提交的证书有没有被撤消，是否仍然合法有效，在过滤掉回收证书的基础上，如果能在本地 policy 和资源请求者之间找到一条信任链，那么便认为请求操作是可接收的，若找不到这样的信任链，则认为不能确定请求操作可否接收(还可有要求请求者重新举证)。最后一致性验证器将验证结果 LV(1 同意访问、0 未知待定)输出给综合判断系统。

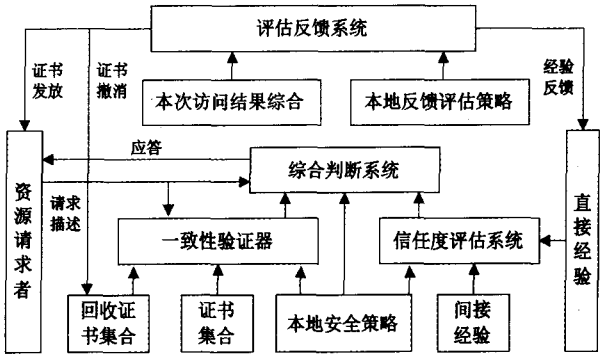

图 1 含有信任度的反馈式信任管理模型框架

信任度评估系统则用来处理感性经验，它有三个输入：本地安全策略、直接经验和间接经验。直接经验和间接经验是对实体进行信任度综合评估的两种途径，引入间接经验是因为在网络复杂的环境下，当对一个实体信任度进行评估时，评估者自己可能对其一无所知，而且即使在相对熟悉的情况下，获得的信息越全，越有助于评估。信任度评估系统依据经验信息和本地安全策略对资源请求者进行评估，得出其信任度 $DT(0\leqslant DT\leqslant 1)$。$DT$ 和 LV 一起作为综合判断系统的输入，综合判断系统参照本地安全策略，如资源的敏感性等因素作出最后决择 RT 并将 RT 返给资源请求者，$RT=f(ACTION, LV, DT, POLICY)$(1 同意访问、0 拒绝访问)，其中 $ACTION$ 是请求动作描述。例如，若某资源为极敏感资源，其相关的安全策略可以制定为 $RT=LV$ OR $DT\geqslant 0.95$，甚至可以是 $RT=LV$ AND $DT\geqslant 0.95$，而对于一般性资源其相关的安全策略可以制定为 $RT=LV$ OR $DT\geqslant 0.60$，依此类推。这样在访问者和资源请求者之间不存在证书链的前提之下，访问者也有可能凭借其在网络中良好的信任度而获得对资源的访问权。

评估反馈系统是本模型另一个重要的组成部分，是本模型区别于其它模型的关键所在，其输入为本次访问结果综合和本地反馈评估策略，依据评判结果的不同，有三种不同形式的反馈：其一，认为访问者 R 完全可信，可将证书 C 直接颁发给访问者 R 并将 R 在直接经验中的信任度提升为 1，这样在下次需要访问时，R 可免去收集证书的过程，只提交 C 即可，因此可以大大提高 R 的访问速度。其二，认为访问者 R 完全不可信，撤消 R 的证书并将 R 在直接经验中的信任度降低为 0。其三，前两种都是极端情况，更多时候访问者 R 介于完全可信和完全不可信之间，这时可依据实际情况在直接经验中修正对 R 的信任度。对信任度的修改不仅是本实体将来判断 R 是否值得信任的重要依据，也为网络中其它实体对 R 的评估提供了有价值的参考。使网络真正成为一个"人人为我，我为人人"的有机整体。

(下转第 138 页)

问题:一是条件较强,很多偏好信息无法获取;其次,大多数攻击检测模型只能检测对某一种攻击模型所产生的攻击,但实际中企业不知道其竞争对手将采用何种攻击模型,所以今后必须提高攻击检测模型的普遍适应性,使之能检测出各种类型的攻击。

(3)综合 Web 挖掘等其它方法,提高攻击检测的准确性与成功率。

目前所有推荐安全研究所涉及的用户资料都是基于用户评价信息。在实际中,可采用 Web 挖掘方法,通过结合对访问路径分析,获取用户的偏好模式,同时通过对所有用户 Web 日志分析,发现假用户,提高攻击检测的准确性与成功率。

(4)切实以推荐模型的稳定性和健壮性为研究目标,提高推荐系统的准确性。

攻击模型和攻击检测模型是电子商务推荐安全研究的基础,电子商务推荐安全研究要求不但要能检测假用户,还要根据检测到的假用户以及假用户所用的攻击模型,提出推荐模型的改进办法,真正在存在攻击的情况下仍能做出较为准确的推荐,提高推荐模型的稳定性和健壮性。

参 考 文 献

1 Resnick, Varian. Recommender systems [J]. Communications of the ACM, 1997, 40 (3): 56～58

2 Schafer J B, Konstan J, Riedl J. Recommender Systems in E-Commerce [C]. In: EC '99 Proceedings of the First ACM Conference on Electronic Commerce, Denver, CO, 1999. 158～166

3 余力,刘鲁. 电子商务个性化推荐研究综述. 计算机集成制造系统,2004,10(10):1306～1313

4 Herlocker J, Konstan J, Tervin L G, et al. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, 2004, 22(1): 5～53

5 赵亮,胡乃静,张守志. 个性化推荐算法设计[J]. 计算机研究与发展,2002,39(8):986～991

6 Ben J, Konstan J A, John R. E - commerce recommendation applications [R]. University of Minnesota, 2001

7 Billsus D, Pazzani M. Learning Collaborative Information Filters [C]. In: Proceedings of the International Conference on Machine Learning (Madison WI, July 1998), Morgan Kaufmann Publishers

8 余力,刘鲁,罗掌华. 我国电子商务推荐策略的比较分析. 系统工程理论与实践,2004-08

9 Robin B. Hybrid Recommender Systems: Survey and Experiments [R]. Department of Information Systems and Decision Sciences, California State University, Fullerton

10 Ben J, Konstan J A, John R. E-Commerce Recommendation Applications [R]. University of Minnesota, 2001

11 Burke R, Mobasher B, Zabicki R, et al. Identifying attack models for secure recommendation. In: Beyond Personalization: A Workshop on the Next Generation of Recommender Systems, San Diego, California, 2005

12 Yu Li, Liu Lu, Li Xuefeng. A Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-content Recommendation in E-Commerce. International Journal of Expert System with Application, Jan. 2005, 28: 67～77

13 Burke R, Mobasher B, Zabicki R, et al. Analysis and Detection of Segment-Focused Attacks Against Collaborative Recommendation

14 Burke R, Mobasher B, Bhaumik R. Limited knowledge shilling attacks in collaborative filtering systems. In: Proceedings of the 3rd IJCAI Workshop in Intelligent Techniques for Personalization, Edinburgh, Scotland, 2005

15 Mobasher B, Burke R, Bhaumik R, et al. Effective attack models for shilling item-based collaborative filtering systems. In: Proceedings of the 2005 WebKDDWorkshop, held in conjuction with ACM SIGKDD 2005, Chicago, Illinois, 2005

16 O'Mahony M, Hurley N, Kushmerick N, et al. Collaborative recommendation: A robustness analysis. ACM Transactions on Internet Technology, 2004, 4(4): 344～377

17 Lam S, Reidl J. Shilling recommender systems for fun and profit. In: Proceedings of the 13th International WWW Conference, New York, 2004

18 Burke R, Mobasher B, Zabicki R, et al. Segment-based Injection Attacks Against Collaborative Filtering Recommender Systems-icdm05

19 Resnick P, Iacovou N, Suchak M, et al. Grouplens: an open architecture for collaborative filtering of netnews. In: CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work, ACM Press, 1994. 175～186

20 Burke R, Mobasher B, Zabicki R, et al. Detecting Profile Injection Attacks in Collaborative Recommender Systems-cec06

21 O'Mahony M, Hurley N, Kushmerick N, et al. Collaborative recommendation: A robustness analysis. ACM Transactions on Internet Technology, 2004, 4(4): 344～377

22 Chirita P A, Nejdl W Zamfir C. Preventing shilling attacks in online recommender systems. In: WIDM '05: Proceedings of the 7th annual ACM international workshop on Web information and data management, New York, NY, USA, ACM Press, 2005. 67～74

(上接第 82 页)

由于本地安全策略、信任度计算策略、反馈评估策略等均可由本地用户根据实际情况具体制定,因此该模型具有灵活、通用、自动学习的特点。

结束语 本文设计了一个含有信任度的反馈式信任管理模型,该模型在对实体进行信任评估时,综合考虑了感性信任和理性信任两方面的因素,同时该模型还具有自主学习反馈的功能,能够根据访问者在访问过程中的不同表现做出不同的反应,将每次访问的经验反馈至系统,使系统在学习中不断得到修正和完善。但本文所提出的模型只是一个初步的框架,有许多工作尚需进一步深入研究,例如如何合理有效地制定本地安全策略,如何对本次访问结果进行综合评估,如何根据不同类型的访问来确定评估发生的时间等都需要细化和完善。

参 考 文 献

1 Blaze M, Feigenbaum J, Lacy J. Decentralized Trust Management. In: Proc. of the 17th Symposium on Security and Privacy, pages164-173. IEEE Computer Society Press, Los Alamitos, 1996. 164～173

2 Blaze M, Feigenbaum J, Strauss M. Compliance-checking in the PolicyMaker trust management system. In: Proceedings of Second International Conference on Financial Cryptography (FC'98), volume 1465 of Lecture Notes in Computer Science, Springer, 1998. 254～274

3 Blaze M, Feigenbaum J, Ioannidis J, Keromytis A. The KeyNote Trust-Management System. Version 2. Internet RFC 2704, September 1999

4 Chu Y, Feigenbaum J, LaMacchia B, Resnick P, Strauss M. REFEREE: Trust management for web applications. World Wide Web Journal, 1997(2): 706～734

5 徐锋,曹春,等. 一个软件服务协同中的信任管理框架设计. 计算机科学,2003,30(6)