云计算环境下用户行为的认证与预测①

李 良[1], 田立勤[1,2], 李君建[1]

[1](青海师范大学 计算机学院, 西宁 810008)
[2](华北科技学院 计算机学院, 北京 101601)

摘 要: 云计算环境下, 传统的身份认证技术表现出一定的缺陷, 为了遏制不可信用户的入侵行为, 本文结合传统的身份认证和行为认证, 论述了云计算环境下的用户行为认证机制. 建立了用户行为认证集, 论述了整个机制的实现过程. 建立了预测用户行为认证等级的贝叶斯网络模型, 并结合历史和实时用户行为信息实现对用户行为认证等级的预测. 通过实例分析论证了预测模型的有效性. 论文理论分析表明该研究对增强用户认证, 有效遏制不可信用户的入侵行为具有重要的理论和实际指导意义.

关键词: 云计算; 用户行为认证; 用户行为认证集; 贝叶斯网络; 用户行为认证等级预测

Authentication and Prediction of User Behavior in Cloud Computing

LI Liang[1], TIAN Li-Qin[1,2], LI Jun-Jian[1]

[1](School of computer Science and Technology, Qinghai Normal University, Xining 810008, China)
[2](School of computer Science and Technology, North China Institute of Science and Technology, Beijing 101601, China)

Abstract: The traditional authentication technology showed some deficiencies in cloud computing, in order to decrease intrusion of untrustworthy user, this paper discusses a scheme for user behavior authentication in cloud computing environment which combines traditional authentication and behavior authentication. The paper presents the creation of the user behavior authentication set. The paper presents the realization process of the whole mechanism. The paper proposes Bayesian network model to predict the level of user behavior authentication based on the combination of the evidence of past transaction and real-time user behavior information. We give an example to demonstrate the effectiveness of the prediction model. This research is of theoretically and practically significant for enhancing user authentication and decreasing intrusion of untrustworthy user effectively.

Key words: cloud computing; user behavior authentication; user behavior authentication set; Bayesian network; user behavior authentication level prediction

1 引言

云计算是一种可以通过网络连接, 便携并且按需访问的可配置共享资源池的服务, 计算资源将以最小的管理和交互代价快速提供给用户. 根据云计算所提供服务类别的不同, 云计算的服务模式可以分为软件即服务(Software as a Service, SaaS)、平台即服务(Platform as a Service, PaaS)和基础设施即服务(Infrastructure as a Service, IaaS)[1]三种模式.

云计算用户与服务提供者的基本架构如图 1 所示, 共分为五个层次, 从上到下依次分为资源提供层、云服务提供层、信息传输层、行业服务提供层、端用户层. 云服务提供者利用资源提供层提供的资源和自己的技术(如虚拟化技术等)最终整合成可向用户提供的云服务(如 SaaS, PaaS, IaaS 等), 并通过信息传输层向用户提供这些服务.

随着云计算服务的广泛使用, 云计算的安全信任问题愈发凸显. 在云计算中, 由于用户直接使用和操作云服务提供者提供的软件、操作系统、甚至是编程

① 基金项目:国家自然科学基金(61163050, 61472137);中央高校基本科研业务(3142014125,3142015022,3142013098)
收稿时间:2015-09-29;收到修改稿时间:2015-11-11 [doi:10.15888/j.cnki.csa.005169]

环境和网络基础性设施, 因此用户对云资源的软硬件的影响和破坏远比目前用户利用因特网进行资源共享要严重的多. 特别是具有合法用户身份的主观破坏行为, 比如竞争者、黑客和对立者等, 例如: 对于 PaaS 服务来说, 由于它可以让用户将自己创建的某类应用程序部署到服务端运行, 并且允许用户对应用程序及其计算环境配置进行控制, 因此, 恶意用户可能会提交一段恶意代码, 这段代码可能恶意抢占 CPU 时间、内存空间和其它资源, 也可能会攻击其它用户, 甚至可能会攻击提供运行环境的底层平台, 因此用户行为是否可信, 以及如何对用户行为可信进行评估是云计算研究的一个重要内容.

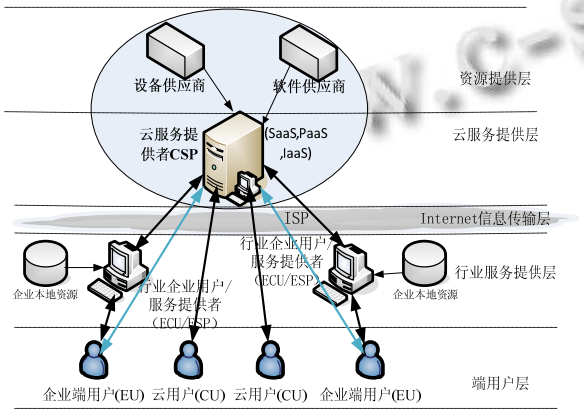

图 1　云计算基本架构中的用户和服务提供者

在云计算环境下, 传统的身份认证技术表现出一定的缺陷, 由于传统的身份认证技术只是对人的认证, 只验证了用户的身份, 设计一个基于硬件的安全性高的身份认证方案来保护云计算环境中的数据免受非法用户的恶意入侵与窃取. 然而, 一旦入侵者以合法用户的身份进入系统, 传统的身份认证机制就缺乏相应的针对手段, 无法保证系统的安全. 因此, 为有效遏制以合法用户身份进行的入侵破坏行为, 本文提出了需要结合身份认证和行为认证的新的认证机制, 以保证用户身份和行为的双重可信[2-3]. 用户行为认证是指用户在使用网络资源时通过用户的行为对行为的可信性和行为的身份再确认过程. 这一过程是服务提供者通过与用户的交互获得行为证据, 然后提交给认证服务器, 后者将行为证据与存储在数据库里的用户行为认证集进行核对处理, 根据比较结果确认用户行为的是否可信, 用户身份的是否真实[4-6].

由于用户行为认证机制是利用历史和实时用户信息实现的, 而想要实现对用户行为的有效控制, 还需要能有效预测用户未来的信任等级, 因此, 本文在用户行为认证机制的基础上建立了贝叶斯网络模型, 利用贝叶斯网络对用户行为认证等级进行预测. 本文提供的机制不仅可以预测单属性不同条件下的行为认证等级, 而且可以预测多属性不同条件下的行为信任等级, 因此可以满足不同条件的要求, 并可以根据需要灵活设置[7].

2　用户行为认证机制

2.1 用户行为认证集

用户行为认证过程中需要将获得的行为证据与用户行为认证集进行核对处理, 用户行为认证成功的概率大小取决于行为认证集的划分、定义和行为相关的集合的覆盖率, 因此确定行为认证集是行为认证过程中的重要内容.

用户行为认证集包含以下几个方面: 行为状态认证集、行为内容认证集、行为习惯认证集、行为安全认证集、行为契约认证集[8]. 行为状态认证集是指用户访问系统的状态行为, 如使用的操作系统的版本、上网的时间、地点、IP 地址、MAC 地址等. 行为内容认证集是指用户访问系统的内容行为, 如在电子资源中访问的专业方向, 或者在电子购物中的购物价格区间等. 行为习惯认证集是指用户访问系统的习惯性行为, 如使用资源时的操作习惯、操作序列、操作流程、释放资源的操作、用户键入特征、用户浏览时间等. 行为安全认证集是指用户访问系统的安全行为, 行为安全认证集是检测行为安全异常的对照集, 可根据目前的异常入侵检测规则作为我们行为安全检测的认证集. 行为契约认证集是指服务提供方和用户在提供服务之前签订的契约, 规定服务的内容、时间、禁止的行为和收费的标准等.

2.2 用户行为认证机制的实现过程

用户行为认证基本机制如图 2 所示, 具体步骤为:

①通过用户终端监测获取用户行为证据, 并通过网络传输用户行为证据到用户行为认证服务器端;

②服务器对用户行为进行认证;

③返回认证结果.

其中服务器对用户行为进行认证的实现过程详细步骤如下:

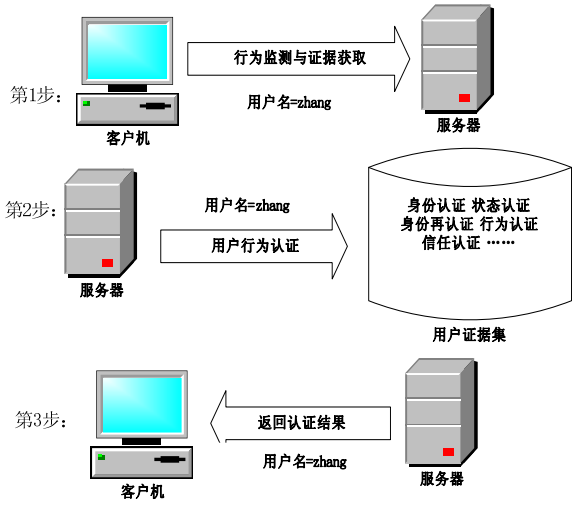

图 2 用户行为认证机制

①当终端用户请求到达时，行为认证机制首先进行身份认证，对于身份认证不通过的用户，拒绝访问；对于通过身份认证且并非是首次访问的用户，执行②；对于通过身份认证并且是首次访问的用户，容许授权访问，进行重点实时行为监控，执行④.

②获取用户状态信息，进行基于状态的认证，对于状态认证不成功的用户则拒绝访问；对于通过状态认证的用户，执行③.

③根据用户历史行为状态信息，用户历史行为状态若为可信，则容许授权访问，并进行正常实时行为监控，执行④；用户历史行为状态若为陌生，则进入预警防范，容许授权访问，但进行重点实时行为监控，执行④；用户历史行为状态若为不可信，则拒绝访问.

④获取用户实时行为证据，并进行基于安全认证集的认证，认证成功则执行⑤；认证失败将终止用户访问.

⑤基于其他认证集的认证，认证成功，则允许用户继续访问，执行⑦;认证失败，则执行⑥;

⑥博弈分析：结合历史行为可信认证状态进行博弈分析，利用贝叶斯网络预测模型对用户行为认证等级进行预测，将用户行为信任及各认证集划分为 L 个信任等级，根据本次用户行为及用户的历史行为对本次用户行为的可信概率进行预测，得到各信任等级的概率. 然后结合预测结果和博弈理论分析用户的决策概率，当服务提供者的收益值大于零时就接受访问，执行⑦；否则就拒绝用户的继续访问.

⑦如果用户访问结束，监测系统实时获取整体行为证据，进行本次用户访问的整体行为可信认证，并更新用户的行为认证子集及用户的长期行为可信认证状态；若用户继续访问，执行④.

用户行为认证流程图如图 3 所示.

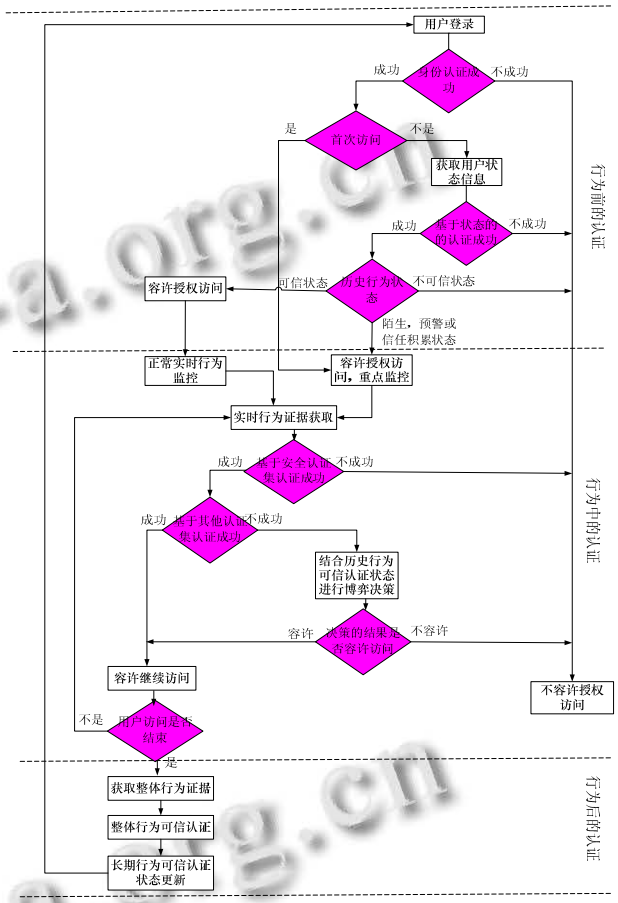

图 3 用户行为认证流程图

3 用户行为认证等级预测

3.1 用户行为认证等级预测的贝叶斯网络模型

用户行为认证等级预测是基于过去交往的行为证据基础上对未来用户的用户行为认证等级进行预测，从可信网络[9]的定义(网络，服务提供者和用户的行为及其结果总是可以预期与可管理的)可知，对用户行为认证等级进行预测是实现可信网络的重要内容之一，是实现对用户行为进行控制的前提.

用户行为认证等级预测基于贝叶斯网络[10-11]，贝叶斯网络是基于概率论的专家系统，网络拓扑结构用于定性分析，条件概率表用于定量分析. 它是图论与概率论的有机结合，直观的知识表示形式使它具有良

好的可理解性和逻辑性，并具有因果和概率语义，可以有机结合先验知识和样本数据，将主观与客观有机地结合起来，更加全面客观地反映数据对象内在的联系与本质. 贝叶斯网络中节点之间是相互影响的，任何节点观测值的获得或者节点信息的改变，都会影响其他的节点，因而具有良好的推理和预测能力.

根据贝叶斯网络模型的特色，贝叶斯网络中的任何一个节点的状态确定，网络就可以利用贝叶斯公式进行正向或逆向计算，从而得出网络中任意一个节点的概率. 用户行为认证等级预测的贝叶斯基本网络模型是一个有向无环图(如图4所示)，根据行为认证集对用户 U 分别从行为状态 S、行为内容 C、行为习惯 H、行为安全 D 和行为契约 R 五个认证集方面进行评估预测.

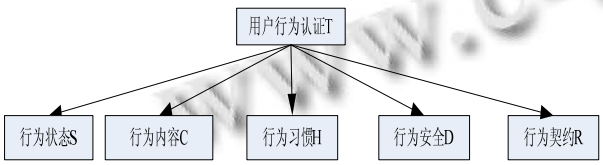

图 4　用户行为认证预测的贝叶斯网络基本模型

3.2 用户行为认证等级的先验概率

在利用贝叶斯网络对用户行为认证等级进行预测之前，必须先计算出用户行为认证等级的先验概率，用户行为认证集等级的先验概率. 我们将用户行为认证等级 T、行为状态认证集 S 和行为内容认证集 C 等节点划分为 L 个信任等级，并对这些信任等级从高到低进行顺序编号为整型变量 i ($i \in [1, L]$)，它们所代表的信任区间范围从高到低的顺序分别是

$$\left[1-\frac{TH_1}{L-1},1\right],\left[1-2\times\frac{TH_1}{L-1},1-\frac{TH_1}{L-1}\right],\ldots\left[TH_0,1-(L-2)\times\frac{TH_1}{L-1}\right],[0,TH_0]$$

其中 TH_0 是信任阈值，即当结点的行为信任值小于 TH_0 时，服务提供者就不信任用户了，且 $TH_0+TH_1=1$.

用户行为认证等级的先验概率，其计算公式见公式 1:

$$p(T_i)=\frac{|T_i|}{n}(1\le i\le L)，并且\sum_{i=1}^{L}p(T_i)=1 \quad (1)$$

其中 n 表示与所预测用户交往的总次数，$|T_i|$ 表示与所预测用户的交往历史中整体认证的值落在 T_i 范围内的次数，$p(T_i)$ 表示它的概率.

计算用户行为认证集等级的先验概率，其计算方法与计算用户行为认证等级的先验概率的方法相同.

行为状态认证集等级的先验概率 $p(S_i)$ 为:

$$p(S_i)=\frac{|S_i|}{n}(1\le i\le L)，并且\sum_{i=1}^{L}p(S_i)=1 \quad (2)$$

行为内容认证集等级的先验概率 $p(C_i)$ 为:

$$p(C_i)=\frac{|C_i|}{n}(1\le i\le L)，并且\sum_{i=1}^{L}p(C_i)=1 \quad (3)$$

行为习惯认证集等级的先验概率 $p(H_i)$ 为:

$$p(H_i)=\frac{|H_i|}{n}(1\le i\le L)，并且\sum_{i=1}^{L}p(H_i)=1 \quad (4)$$

行为安全认证集等级的先验概率 $p(D_i)$ 为:

$$p(D_i)=\frac{|D_i|}{n}(1\le i\le L)，并且\sum_{i=1}^{L}p(D_i)=1 \quad (5)$$

行为契约认证集等级的先验概率 $p(R_i)$ 为:

$$p(R_i)=\frac{|R_i|}{n}(1\le i\le L)，并且\sum_{i=1}^{L}p(R_i)=1 \quad (6)$$

3.3 各行为认证集等级的条件概率

除了计算先验概率外，还必须计算各认证集等级的条件概率. 认证集等级的条件概率可由公式(7)来计算:

$$p(e/h)=\frac{p(h,e)}{p(h)} \quad (7)$$

它表示在满足 h 的条件下满足 e 条件的概率.
以计算 $p(C_i/T_j)$ 条件概率为例，由公式(7)得:

$$p(C_i/T_j)=\frac{p(C_i,T_j)}{p(T_j)}=\frac{|C_i\cap T_j|/n}{|T_j|/n}=\frac{|C_i\cap T_j|}{|T_j|} \quad (8)$$

3.4 用户行为认证等级预测

计算出用户行为认证等级的先验概率，用户行为认证集等级的先验概率及其条件概率，就可以预测在某个特定条件下的用户行为认证等级的概率，用户行为认证等级预测的流程图如图 5 所示.

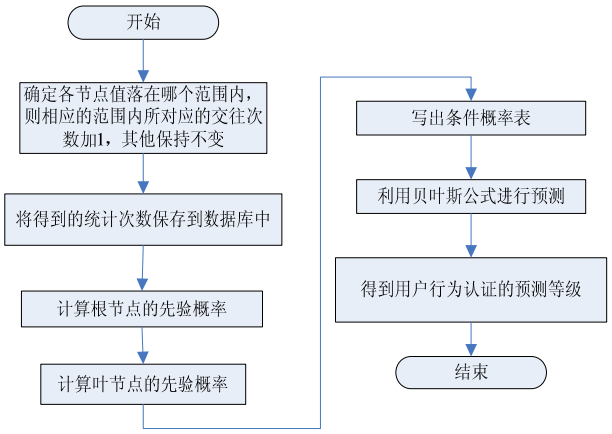

图 5　基于贝叶斯网络的用户行为认证预测流程

4 用户行为认证等级的预测实例与分析

4.1 用户交互行为数据统计

上面给出了基于贝叶斯网络模型的用户行为等级认证的分析预测步骤, 下面将通过一个实例来演示模型预测过程, 论证模型的有效性. 现有服务器 S 与预测的用户 U 的交互统计数据为 200 组, 行为认证分为三个信任等级, 分别是: “信任”、“基本信任”和“不信任”. 某时刻, 用户 U 请求访问服务器 S, 现通过模型预测在某些特定行为认证集的条件下用户 U 的行为认证等级. 各节点的值落在不同区间的次数见表 1.

表 1 各节点值落在不同区间的次数

范围	T 值频数	范围	S 值频数	范围	C 值频数
T_1	98	S_1	149	C_1	90
T_2	83	S_2	37	C_2	83
T_3	19	S_3	14	C_3	27
范围	H 值频数	范围	D 值频数	范围	R 值频数
H_1	127	D_1	106	R_1	153
H_2	56	D_2	83	R_2	39
H_3	17	D_3	11	R_3	8

对于节点值同时落在两个不同认证集范围内的次数用二维数组表示, 如 T_iS_j (其中 $i,j \in \{1,2,3\}$), 同时落在三个不同认证集的次数用三维数组表示, 如 $T_iS_jC_k$(其中 $i,j,k \in \{1,2,3\}$), 以此类推, 则由 200 次交互过程的统计数据可以得出相应的二维到六维数据的数据, 数据值这里就不再一一写出.

4.2 基于不同条件的用户行为认证等级的预测

根据以上得到的用户交互行为数据统计信息, 利用公式(1)~ (6)就可以得到网络中个节点先验概率, 见表 2, 再由公式(7)就可计算各节点的条件概率, 有了所有节点的先验概率及叶节点的条件概率表, 在进行用户行为认证等级预测的时候, 对于任意一组观测值的状态, 我们都有对应的先验概率及条件概率, 分别将其代入式(8)就可求得所需的后验概率, 以预测用户行为的认证等级.

通过反复应用贝叶斯公式和乘积与求和公式, 我们可以得到网络中任意想知道的概率. 根据历史数据统计信息和贝叶斯网络预测模型所得到的后验概率, 我们可以分析得知不同行为认证集对单个用户行为认证等级的影响的概率分布, 实现对各个行为认证集进行不同层级的控制和预警, 体现用户行为认证等级预测的个性化; 对于多个用户而言, 对比同一行为认证集下的预测值, 从而获得用户在相同行为认证集下认证等级排序, 设置不同的优先级, 实现不同等级用户的访问权限控制.

表 2 各节点的先验概率

节点	先验概率	节点	先验概率	节点	先验概率
T_1	49/100	T_2	83/200	T_3	19/200
S_1	149/200	S_2	37/200	S_3	7/100
C_1	9/20	C_2	83/200	C_3	27/200
H_1	127/200	H_2	7/25	H_3	17/200
D_1	53/100	D_2	83/200	D_3	11/200
R_1	153/200	R_2	39/200	R_3	1/25

4.3 结果分析

在计算机和互联网络世界里, 身份认证是一个最基本的安全特性, 也是整个信息安全的基础. 用户身份认证是保证用户真实身份的网络安全机制, 而在云计算环境下, 传统的身份认证技术并不能有效阻止身份认证失效或具有合法身份的恶意用户行为对系统的破坏.

本文论述了云计算环境下, 结合用户身份认证结果和用户行为信息的用户行为认证机制, 并利用贝叶斯网络模型对未来用户行为认证等级进行了预测. 贝叶斯网络结构中的每个节点, 在已知其父节点的条件下独立于所有其余的非子代节点. 根据条件独立性, 贝叶斯网络将联合概率分解成若干个条件概率的乘积, 使得概率推理更加直观和便捷.

贝叶斯网络模型将用户行为认证等级预测的因果知识用有向图自然直观地表示出来, 同时将以往用户行为各认证集的统计数据以条件概率的形式融入模型, 实现了用户行为的先验知识和样本数据的完美结合, 可以有效的避免对数据的过分拟合. 我们通过实例分析, 论证了所提出的模型的有效性, 从而实现了对未来用户行为认证等级的预测.

参考文献

1 林闯,苏文博,孟坤,刘渠,刘卫东.云计算安全:架构、机制与模型评价.计算机学报,2013,36(9):1765−1784.

2 田立勤,林闯.可信网络中一种基于行为信任预测的博弈控制机制.计算机学报,2007,30(11):1930−1938.

3 S Hong. User behavior based authentication on mobile network. International Journal of Advancements in Computing Technology, 2013, 5(11): 233−237.

4 陈亚睿,田立勤,杨扬.云计算环境下动态用户行为认证的机制、模型与分析.系统仿真学报,2011,23(11):2302−2307.

5 田立勤.网络安全的特性、机制与评价.第一版.北京:清华大学出版社/北京交通大学出版社,2013:261-287.
6 Kent AD, Liebrock LM, Neil JC. Authentication graphs: Analyzing user behavior within an enterprise network. Computers & Security, 2015, 48: 150-166.
7 林闯,田立勤,王元卓.可信网络中用户行为可信的研究.计算机研究与发展,2008,45(12):2033-2043..
8 王晓菊,田立勤,赵竞雄.基于物联网的用户行为认证机制与分析.南京理工大学学报,2015,39(1):70-77.
9 林闯,彭雪海.可信网络研究.计算机学报,2005,28(5):751-758.
10 叶跃祥,糜仲春,王宏宇等.基于贝叶斯网络的不确定环境下多属性决策方法.系统工程理论与实践,2007,27(4):107-113,125.
11 赵洁,肖南峰,钟军锐.基于贝叶斯网络和行为日志挖掘的行为信任控制.华南理工大学学报:自然科学版,2009,37(5):94-100.
12 冯登国,张敏,张妍,徐震.云计算安全研究.软件学报,2011,22(1):71-83.
13 Tian LQ, Lin C, Ji TG. Quantitative Analysis of Trust Evidence in Internet. 2006 IEEE International Conference on Communication Technology. USA. IEEE. 2006. 1-5.
14 Deutschmann I, Nordstrom P, Nilsson L. Continuous authentication using behavioral biometrics. IT Professional Magazine, 2013, 15 (4): 12-15.