基于隐马尔可夫模型的用户行为异常检测新方法

邬书跃[1]，田新广[2,3]

（1. 湖南科技大学 信息与电气工程学院,湖南 湘潭 411201; 2. 北京交通大学 计算技术研究所, 北京 100029;
3. 海军装备研究院 博士后工作站, 北京 100073）

摘 要：提出一种基于隐马尔可夫模型的用户行为异常检测方法，主要用于以 shell 命令为审计数据的主机型入侵检测系统。与 Lane T 提出的检测方法相比，所提出的方法改进了对用户行为模式和行为轮廓的表示方式，在 HMM 的训练中采用了运算量较小的序列匹配方法，并基于状态序列出现概率对被监测用户的行为进行判决。实验表明，此方法具有很高的检测准确度和较强的可操作性。

关键词：入侵检测；异常检测；行为模式；隐马尔可夫模型

中图分类号：TP18；TP393 **文献标识码**：A **文章编号**：1000-436X(2007)04-0038-06

Method for anomaly detection of user behaviors based on hidden Markov models

WU Shu-yue[1], TIAN Xin-guang[2,3]

(1. School of Information and Electronic Engineering, Hunan University of Science and Technology, Xiangtan 411201, China;
2. Research Institute of Computing Technology，Beijing Jiaotong University, Beijing 100029, China;
3. Postdoctoral Working Station of Navy Equipment Academe, Beijing 100073, China)

Abstract: A method for anomaly detection of user behaviors was presented for host-based intrusion detection systems with shell commands as audit data. The method constructs specific hidden Markov models(HMMs) to represent the behavior profiles of users. The HMMs were trained by a sequence matching algorithm which was much simpler than the classical Baum-Welch algorithm. A decision rule based on the probabilities of short state sequences was adopted while the particularity of the states was taken into account. The results of computer simulation show the method presented can achieve high detection accuracy and practicability.

Key words: intrusion detection; anomaly detection; behavior pattern; hidden Markov model

1 引言

异常检测是目前入侵检测系统（IDS, intrusion detection system）研究的主要方向。这种检测技术建立系统或用户的正常行为模式，通过被监测系统或用户的实际行为模式和正常模式之间的比较和匹配来检测入侵，其特点是不需要过多有关系统缺陷的知识，具有较强的适应性，并且能够检测出未知的入侵模式。近年来，以 shell 命令或系统调用为审计数据的异常检测得到了较多的研究和应用[1~10]。文献[1]提出一种基于隐马尔可夫模型（HMM）的异常检测方法，该方法利用 HMM 在用户界面层（user interface level）建立合法用户的正常行为轮廓，并采用 Baum-Welch 算法对 HMM 进行训练，在检测

收稿日期：2006-04-04；修回日期：2006-12-20
基金项目：国家高技术研究发展计划（“863”计划）基金资助项目（863-307-7-5）
Foundation Item: The National High Technology Research and Development Program of China(863Program) (863-307-7-5)

阶段利用近似的前向后向算法并根据贝叶斯准则对用户当前行为进行判决；该方法的主要优点是检测准确率高，但 HMM 训练和工作中所需要的计算量很大，检测效率和实时性较差，这在一定程度上限制了它在实际系统中的应用[1]。文献[3,5]提出了基于实例学习的用户行为异常检测方法，该方法的优点是原理较为简单，可操作性强，对用户行为的变化有较强的适应能力，但是，它在检测阶段仅利用特定的相似度函数来计算行为模式之间的相似度，而没有考虑行为模式在训练数据中的出现频率和不同行为模式之间的相关性，因而检测的准确率相对较低[1, 3]。

在以上检测方法的基础上，本文提出一种新的基于 HMM 的用户行为异常检测方法，该方法在用户行为轮廓的表示方式、HMM 的训练以及判决准则的选取等方面与现有方法均有较大不同，其计算复杂度低于文献[1]中 Lane T 提出的基于 HMM 的检测方法，因而检测的效率和实时性相对较高；而与文献[3,5]中基于实例学习的检测方法相比，该方法则在检测准确率方面具有较大优势。

2　HMM 简述

HMM 是双重随机过程，其中一个是隐含的有限状态马尔可夫链，它描述状态的转移；另一个随机过程描述状态与观测值之间的统计对应关系[9]。HMM 一般有 3 个假设：当前状态只同上一状态相关；状态之间的转移概率同状态所处的具体时间无关；观测值只与当前状态有关。这 3 个假设大大降低了模型的复杂度。设观测值序列为 $O=(O_1, O_2, \cdots, O_T)$，相应的状态序列为 $q=(q_1, q_2, \cdots, q_T)$，其中 $O_i \in \Omega_O=\{v_1, v_2, \cdots, v_M\}$，$q_i \in \Omega_q=\{\theta_1, \theta_2, \cdots, \theta_N\}$，$\Omega_O$ 和 Ω_q 分别表示观测值集合和状态集合。HMM 通常用五元组 $\mu=(\Omega_q, \Omega_O, \boldsymbol{A}, \boldsymbol{B}, \pi)$ 来表示，$\boldsymbol{A}$ 为状态转移概率矩阵，$\boldsymbol{A}=(a_{ij})_{N\times N}$，其中

$$a_{ij}=P(q_{t+1}=\theta_j \mid q_t=\theta_i) \quad , \quad 1\leqslant i, j\leqslant N \tag{1}$$

$\boldsymbol{B}$ 为观测值概率矩阵，$\boldsymbol{B}=(b_{jk})_{N\times M}$，其中

$$b_{jk}=P(O_t=v_k \mid q_t=\theta_j), 1\leqslant j\leqslant N, \quad 1\leqslant k\leqslant M \tag{2}$$

π 为初始状态概率矢量，$\pi=(\pi_1, \pi_2, \cdots, \pi_N)$，其中

$$\pi_i=P(q_1=\theta_i) \quad , \quad 1\leqslant i\leqslant N \tag{3}$$

训练、解码和评估是 HMM 的 3 个基本问题。训练是指给定观测值序列 O，确定模型参数 $\lambda=(\boldsymbol{A}, \boldsymbol{B}, \pi)$，使得 $P(O/\lambda)$ 最大；解码是对于给定的 λ 和 O，求使 $P(q/O, \lambda)$ 最大的状态序列 q；评估则是指给定模型参数 λ，求观测值序列 O 的出现概率 $P(O/\lambda)$。HMM 训练、解码和评估的经典算法分别是 Baum-Welch 算法、Viterbi 算法和前向后向算法。利用 Baum-Welch 算法训练 HMM 时，不同的初始参数值会产生不同的训练结果；对于初始参数的选择至今没有广泛适用的方法，一般认为 π 和 A 的初值对训练结果的影响较小，在满足一定约束条件的前提下可以随机选择或均匀取值，而 $\boldsymbol{B}$ 的初值对训练结果影响较大，一般倾向于采用较为复杂的方法来选择该值[9]。

3　基于 HMM 的用户行为异常检测新方法

3.1　审计数据的描述及预处理

本文提出的基于 HMM 的用户行为异常检测方法主要用于 UNIX 平台上的主机型入侵检测系统，它采用系统用户执行的 shell 命令作为审计数据，主要原因在于[1, 9]：1）同 UNIX 平台上其他审计数据（如 CPU 使用量、内存占用率）相比，shell 命令能够更直接地反映用户的行为；2）shell 命令比较容易收集，也便于分析；3）在 UNIX 平台上，shell 是终端用户与操作系统之间最主要的界面，很大比例的用户活动都是利用 shell 完成的。

与文献[1]和文献[10]的检测方法相同，本文的检测方法对用户在 shell 会话中所执行的原始 shell 命令行进行了预处理，即提取出 shell 命令的名称及参数，将 shell 命令行中的主机名、网址等信息用统一格式的标识符号来代替；各命令符号按照在 shell 会话中的出现次序进行排列，不同的 shell 会话按照时间顺序进行连接，每个会话开始和结束的时间点上插入了标识符号[1, 10]。经预处理后，原始的 shell 命令行数据在形式上成为 shell 命令流。

3.2　建模

首先建立 2 个 HMM，其中一个 HMM 用于描述一个或一组合法用户的正常行为轮廓，另一个 HMM 用于描述（入侵者或合法用户的）异常行为轮廓。2 个 HMM 的状态集合以及各状态对应的观测值集合相同，其状态对应于合法用户的行为模式类型。按照行为模式所对应的 shell 命令序列的长度对其进行分类，并根据合法用户的正常训练数据（历史上的正常行为）确定每个状态对应的观测值集合。将

shell 命令序列的长度作为行为模式分类的依据，把长度相同的 shell 命令序列所表示的行为模式划为同一种类。建模的首要问题是确定合法用户正常行为模式的种类个数W，以及相应的 shell 命令序列长度集合$C=\{l(1), l(2), \cdots, l(W)\}$，其中$l(i)$表示第$i$类正常行为模式对应的 shell 命令序列的长度，且$l(1)>l(2)>\cdots>l(W)$。W和C对检测性能有直接影响，在选择它们时，需充分考虑合法用户的行为特点，同时还要考虑模型的复杂度及检测效率（W和$l(i)$越大，检测系统的存储量和工作中的运算量也会越大）。将 HMM 的状态个数设为$N=W+1$，状态集合设为$\Omega_q=\{1, 2, \cdots, W, W+1\}$，其中前 W 个状态同合法用户的W类正常行为模式一一对应，第$W+1$个状态为附加状态，它对应于合法用户的正常历史行为（正常训练数据）中未出现过的行为模式（类型），并规定这类行为模式对应的命令序列长度$l(W+1)=1$。

此外，需根据合法用户的正常训练数据确定 HMM 各状态对应的观测值集合$\Omega_O=\{L(1), L(2), \cdots, L(W+1)\}$，其中$L(i)$为状态$i$对应的观测值集合，即第$i$类行为模式对应的命令序列集合，它包含若干个长度为$l(i)$的命令序列；这里，HMM 状态所对应的观测值（或称观测事件）是命令序列。设一个合法用户的正常训练数据为$R=(s_1, s_2, \cdots, s_r)$，它是该用户在正常操作时所执行的长度为$r$的 shell 命令流，其中$s_j$表示按时间顺序排列的第$j$个 shell 命令；$R$对应的长度为$l(i)$（$1\leqslant i\leqslant W+1$）的命令序列流可表示为$S^i=(Seq_1^i, Seq_2^i, \cdots, Seq_{r-l(i)+1}^i)$，其中命令序列$Seq_j^i=(s_j, s_{j+1}, \cdots, s_{j+l(i)-1})$。设定一个概率门限$\eta$，将$S^i$（$1\leqslant i\leqslant W$）中出现概率大于$\eta$的命令序列视为合法用户的（正常）行为模式，$L(i)$即由这些命令序列组成（一个序列的出现概率是指此序列在相应序列流中的出现次数与该序列流中的序列总数之比）。附加状态对应的观测值集合$L(W+1)$包括两部分，一部分是由正常训练数据R中未出现过的命令组成的长度为 1 的序列，另一部分则有所区别，当$l(W)=1$时(此时$S^W=S^{W+1}$)，它是S^{W+1}中出现概率小于或等于η的序列，当$l(W)\neq 1$时，它是S^{W+1}中的所有序列。当$i\neq j$时（$1\leqslant i, j\leqslant W+1$），$L(i)\bigcap L(j)=\phi$，即不同状态对应的观测值集合是不相交的，这和一般的 HMM 不同，也是此方法的一个主要特点。需要指出，合法用户可以只有一个，也可以有多个；当有多个合法用户时，可将这些用户的正常训练数据组合在一起构成总的训练数据。

3.3 HMM 训练

HMM 的训练，或称参数估计问题，是 HMM 在用户行为异常检测中应用的关键问题；Baum-Welch 算法只是解决这一问题的经典方法，但并不是惟一的，也不是最完善的方法[3, 9]。设描述合法用户正常行为的 HMM 参数为$\lambda_1=(\boldsymbol{A}, \boldsymbol{B}, \pi)$，其中$\boldsymbol{A}$和$\pi$的计算方法如下：

第 1 步：根据R得到S^i（$1\leqslant i\leqslant W$）。设定$\pi=(\pi_1, \pi_2, \cdots, \pi_{W+1}):=0$，$\boldsymbol{A}=(a_{ij})_{(W+1)\times(W+1)}:=0$，$m:=1$，$j:=1$，$n:=0$。

第 2 步：如果$m\leqslant r-l(1)+1$,将Seq_m^j与$L(j)$进行比较；否则，$X:=n$,跳至第 5 步。

第 3 步：如果$Seq_m^j\in L(j)$，且$m=1$，则$m:=m+l(j)$，$i:=j$，$j:=1$，返回执行第 2 步；如果$Seq_m^j\in L(j)$，且$m>1$，则$\pi_i:=\pi_i+1$，$a_{ij}:=a_{ij}+1$，$n:=n+1$，$q_n:=i$，$m:=m+l(j)$，$i:=j$，$j:=1$，返回执行第 2 步；如果$Seq_m^j\notin L(j)$，则$j:=j+1$。

第 4 步：如果$j\neq W+1$，返回执行第 2 步；如果$j=W+1$(此时$Seq_m^j\in L(W+1)$)，且$m=1$，则$m:=m+1$，$i:=j$，$j:=1$，返回执行第 2 步；如果$j=W+1$，且$m>1$，则$\pi_i:=\pi_i+1$，$a_{ij}:=a_{ij}+1$，$n:=n+1$，$q_n:=i$，$m:=m+1$，$i:=j$，$j:=1$，返回执行第 2 步。

第 5 步：对于$1\leqslant i, j\leqslant W+1$，$a_{ij}:=a_{ij}/\pi_i$。对于$1\leqslant i\leqslant W+1$，$\pi_i:=\pi_i/X$。

设描述异常行为的 HMM 参数为$\lambda_2=(\boldsymbol{A}^*, \boldsymbol{B}^*, \pi^*)$，异常训练数据为$R^*=(s_1^*, s_2^*, \cdots, s_r^*)$，它是入侵者（非法用户）或合法用户在非法操作或误操作时所执行的 shell 命令流，$\boldsymbol{A}^*$和π^*可根据R^*同样采用以上的序列匹配方法进行计算。

3.4 检测

在检测阶段，首先要得到被监测用户在被监测时间内所执行的 shell 命令流。设被监测用户在被监测时间内所执行的 shell 命令流为$\overline{R}=(\overline{s}_1, \overline{s}_2, \cdots, \overline{s}_r)$。检测时要利用前面参数计算中的序列匹配方法，由

$\overline{R}$ 得到状态序列 $q=(q_1, q_2, \cdots, q_X)$ 及其对应的观测值序列 $O=(O_1, O_2, \cdots, O_X)$，其中 X 为 $\overline{R}$ 中的状态总数，$r/l(1) \leqslant X \leqslant r$，$q_j$ 表示按时间顺序排列的第 j 个状态，O_j 表示与 q_j 对应的观测值(命令序列)，O_j 的长度为 $l(q_j)$（$1 \leqslant q_j \leqslant W+1$）。

为了实时监测用户的行为，我们用滑动窗在 O 中截取短序列，以短序列为数据单元进行分析。设短序列为 $So_j=(O_j, O_{j+1}, \cdots, O_{j+u-1})$，其中 u 表示短序列的长度（$u<X$），$1 \leqslant j \leqslant X-u+1$。相应的状态短序列为 $Sq_j=(q_j, q_{j+1}, \cdots, q_{j+u-1})$。$q$ 和 O 对应的短序列流可分别表示为 $So=(So_1, So_2, \cdots, So_{X-u+1})$ 和 $Sq=(Sq_1, Sq_2, \cdots, Sq_{X-u+1})$。

然后，计算 $P(\lambda_1/Sq_j)$

$$P(\lambda_1/Sq_j)=P(Sq_j/\lambda_1)P(\lambda_1)/P(Sq_j)=\frac{P(\lambda_1)P(q_j/\lambda_1)}{P(Sq_j)}\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_1) \quad (4)$$

其中，$P(\lambda_i)$ 表示参数为 λ_i（i 的取值为 1 或 2）的 HMM 所描述的行为的出现概率。

考虑到用户在短时间内的行为可能会偏离其历史行为，检测中我们并不直接利用 $P(\lambda_1/Sq_j)$ 对被监测用户的行为进行判决，而是对其做了如下的加窗平滑处理

$$D(n)=\frac{1}{w}\sum_{j=n-w+1}^{n}\mathrm{sgn}[\ P(\lambda_1/Sq_j)-P(\lambda_2/Sq_j)]$$

$$=\frac{1}{w}\sum_{j=n-w+1}^{n}\mathrm{sgn}\left[\frac{P(\lambda_1)P(q_j/\lambda_1)}{P(Sq_j)}\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_1)-\frac{P(\lambda_2)P(q_j/\lambda_2)}{P(Sq_j)}\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_2)\right] \quad (5)$$

此外，还可以对 $P(\lambda_1/Sq_j)$ 做以下形式的加窗和处理来得到判决值

$$D(n)=\frac{1}{w}\sum_{j=n-w+1}^{n}\lg\ [P(\lambda_1/Sq_j)/P(\lambda_2/Sq_j)]$$

$$=\frac{1}{w}\sum_{j=n-w+1}^{n}\lg\{\left[\frac{P(\lambda_1)P(q_j/\lambda_1)}{P(Sq_j)}\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_1)\right]/\left[\frac{P(\lambda_2)P(q_j/\lambda_2)}{P(Sq_j)}\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_2)\right]\} \quad (6)$$

一般情况下，我们假设 $P(\lambda_1)=P(\lambda_2)$，此时式（5）可简化为

$$D(n)=\frac{1}{w}\sum_{j=n-w+1}^{n}\mathrm{sgn}[\left[P(q_j/\lambda_1)\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_1)-P(q_j/\lambda_2)\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_2)\right] \quad (7)$$

式（6）可简化为

$$D(n)=\frac{1}{w}\sum_{j=n-w+1}^{n}\left\{\lg\left[\ [P(q_j/\lambda_1)\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_1)\right]-\lg\left[P(q_j/\lambda_2)\prod_{k=j}^{j+u-2}P(q_{k+1}/q_k,\lambda_2)\right]\right\} \quad (8)$$

式（7）、式（8）中，$D(n)$表示状态序列 Sq_n 对应时刻的判决值，$w \leqslant n \leqslant X-u+1$，$w$ 为窗长度（Sq 中第 w 个状态短序列及其后面的每个短序列所对应的时间点上都有一个判决值输出）。对 $D(n)$ 设定一个门限，若它大于这个门限，将被监测用户的当前行为判为正常行为（或将此用户判为合法用户），否则，将其判为异常行为（或将此用户判为非法用户）。需要指出，在实时检测（在线检测）的情况下，被监测用户所执行的 shell 命令流的获取，shell 命令序列的匹配与相应状态的确定，状态短序列出现概率的计算，判决值的计算，以及对用户行为的判决都是同步进行的。

3.5　特点分析

以上基于 HMM 的用户行为异常检测方法主要有以下几个特点：

1) 在该方法中，描述正常行为和异常行为的 HMM 状态以及各状态对应的观测值集合都是根据合法用户的正常行为训练数据确定的，描述正常行为的 HMM 参数也是根据正常行为训练数据计算得到的(在计算描述异常行为的 HMM 参数时，需要用到异常行为训练数据)。因此，该方法是一种异常检测方法。

2) HMM 的状态具有明确的含义，状态个数是根据行为模式种类的个数来确定的；状态对应的观测值(或称观测事件)是多种长度的 shell 命令序列。该方法所用的 HMM 中，状态的“隐含”是指观测数据（被监测用户执行的 shell 命令流）中的状态并非直接可见，而是需要通过序列匹配来确定。

3) 由于不同状态对应的观测值集合互不相交，HMM 的训练和解码均采用了序列匹配方法，同 Lane T 的传统方法相比，较大程度地减小了计算量，缩短了训练和解码的时间。

4) 根据 HMM 状态的特点及实际含义，采用了

基于状态序列出现概率的判决准则，减小了判决中的计算量，提高了检测的实时性。

4 实验结果及分析

通过实验对本文提出的用户行为异常检测方法的性能进行了测试，实验中采用了普渡大学公开发布的shell命令实验数据[1]。其数据包含8个UNIX用户在2年时间内的活动记录。每个用户的数据文件中的主机名、网址等信息用统一格式的标识符号来代替，仅保留了shell命令的名称及参数；用户命令流中的命令按照在 shell 会话中的出现次序进行排列，不同的shell会话按照时间顺序进行连接。我们采用其中4个用户（分别为user1、user2、user3、user4）的数据进行实验，将user3设为合法用户（其行为设为正常行为），将user1、user2、user4设为非法用户(其行为设为异常行为)。每个用户各有15 000个命令，user3的前10 000个命令作为正常训练数据 R 用于两个 HMM 状态和观测值集合的确定以及描述正常行为的HMM参数的计算，user1、user2、user4的前 10 000 个命令组合在一起作为异常训练数据 R^* 用于计算描述异常行为的HMM参数，每个用户的后5 000个命令用作测试数据。

实验的参数设置为 $W=3$， $C=\{3, 2, 1\}$，$\eta=0.02\%$， $u=6$， $w=44$，并假设 $P(\lambda_1)=P(\lambda_2)=0.5$（即正常行为和异常行为的出现概率相等）。在检测阶段，user3的后5 000个命令中共出现1 939个状态，状态对应的命令序列（观测值）的平均长度为2.6；在user1、user2、user4的后5 000个命令中，分别出现2 746、3 291、3 063个状态（其中相当一部分为附加状态），状态对应的命令序列的平均长度为1.6，这表明长度为3和2的命令序列所表示的合法用户的正常行为模式在 3 个非法用户的测试数据中较少出现。图1给出了由式(7)计算出的判决值曲线，图2给出了根据式（8）计算出的2条判决值曲线（图中上方的实线为合法用户user3的正常行为测试数据对应的判决值曲线，下方的3条虚线是3个非法用户的异常行为测试数据对应的判决值曲线）。由2图可见，正常行为和异常行为测试数据对应的判决值曲线具有良好的可分性。

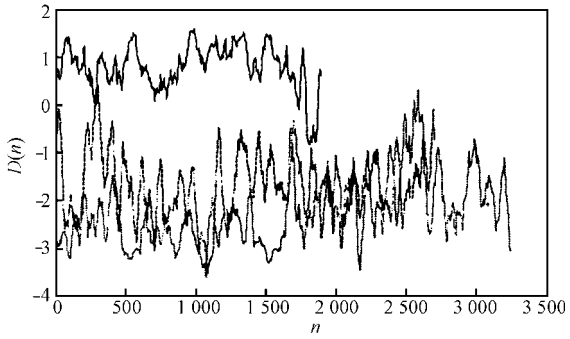

图1 式（7）对应的判决值曲线

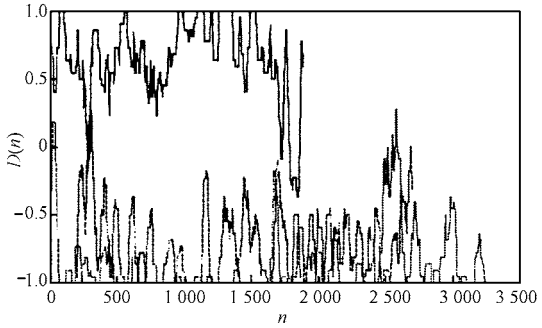

图2 式（8）对应的判决值曲线

在实验的检测阶段，通过调整判决门限可以得到不同虚警概率条件下对3个非法用户的异常行为的平均检测概率。表 1 给出了式（7）和式（8）2种判决值计算方法对应的实验结果。

表1 2种判决值计算方法对应的实验结果

虚警概率	0	0.001	0.005	0.010	0.050
式（7）对应的平均检测概率	0.928	0.930	0.943	0.951	0.985
式（8）对应的平均检测概率	0.932	0.933	0.938	0.944	0.995

根据表1中的实验数据，在虚警概率为0的条件下两种判决值计算方法对应的平均检测概率均可达到90%以上。而且，在虚警概率较低的区间，式（7）和式（8）对应的平均检测概率非常接近，这说明该方法所采用的判决准则具有较好的稳健性，并且可获得很高的检测准确度。

5 结束语

本文提出一种基于隐马尔可夫模型的用户行为异常检测新方法，并通过实验对该方法的性能进行了测试。实验表明，该方法具有很高的检测准确率和较强的可操作性。需要指出，该方法中的一些检测思想还适用于以系统调用为审计数据的程序行为异常检测，但是具体的操作方式及检测性能还有待分析和验证。

参考文献：

[1] LANE T. Machine Learning Techniques for the Computer Security Domain of Anomaly Detection [D]. Purdue University,2000.

[2] LEE W, DONG X. Information-theoretic measures for anomaly detection [A]. Proceedings of the 2001 IEEE Symposium on Security and Privacy[C]. Oakland, USA, 2001. 130-134.

[3] LANE T, BRODLEY C E. Temporal sequence learning and data reduction for anomaly detection［J］. ACM Transactions on Information and System Security, 1999, 2(3): 295-331.

[4] WARRENDER C, FORREST S, PEARLMUTTER B. Detecting intrusions using system calls: alternative data models [A]. Proceedings the 1999 IEEE Symposium on Security and Privacy[C]. Berkely, USA: IEEE Computer Society, 1999.133-145.

[5] LANE T，BRODLEY C E. An application of machine learning to anomaly detection [A]. Proceedings of the 20th National Information Systems Security Conference[C]. Baltimore, USA, 1997.366-377.

[6] 孙宏伟, 田新广, 李学春等. 一种改进的IDS异常检测模型[J].计算机学报, 2003, 26(11): 1450-1455.
SUN H W, TIAN X G, LI X C, *et al*. An improved anomaly detection model for IDS[J]. Chinese Journal of Computer, 2003, 26(11): 1450-1455.

[7] 连一峰, 戴英侠, 王航. 基于模式挖掘的用户行为异常检测[J]. 计算机学报, 2002, 25(3) : 325-330.
LIAN Y F, DAI Y X , WANG H. Anomaly detection of user behaviors based on profile mining[J]. Chinese Journal of Computer, 2002, 25(3): 325-330.

[8] 田新广,高立志,李学春等.一种基于隐马尔可夫模型的 IDS 异常检测新方法[J].信号处理,2003,19(5):420-424.
TIAN X G, GAO L Z, LI X C, *et al*. A new anomaly detection method based on hidden Markov models for IDS[J]. Signal Processing, 2003, 19(5)：420-424.

[9] 田新广. 基于主机的入侵检测方法研究[D]. 长沙：国防科技大学研究生院, 2005.
TIAN X G. Anomaly Detection Methods for Host-based Intrusion Detection Systems [D]. Changsha: Graduate School of National University of Defense Technology, 2005.

[10] 田新广, 高立志, 张尔扬. 新的基于机器学习的入侵检测方法[J]. 通信学报, 2006, 27(6):108-114.
TIAN X G, GAO L Z, ZHANG E Y. Intrusion detection method based on machine learning[J]. Journal on Communications, 2006, 27(6): 108-114.

作者简介：

邬书跃（1963-），男，湖南常德人，硕士，湖南科技大学教授，主要研究方向为网络安全、数字信号处理、移动通信。

田新广（1976-），男，河北吴桥人，博士，北京交通大学与海军装备研究院联合培养博士后，主要研究方向为信号处理、网络安全、入侵检测。