基于 Shell 命令和多阶 Markov 链模型的用户伪装攻击检测

肖 喜[1],翟起滨[1],田新广[2],陈小娟[3],叶润国[4]
(1.中国科学院研究生院信息安全国家重点实验室,北京 100049;
2.中国科学院计算技术研究所网络科学与技术重点实验室,北京 100190;
3.北京工商大学计算机与信息工程学院,北京 100037;4.北京启明星辰信息安全技术有限公司,北京 100193)

摘 要: 伪装攻击是指非授权用户通过伪装成合法用户来获得访问关键数据或更高层访问权限的行为.提出一种新的用户伪装攻击检测方法.该方法针对伪装攻击用户行为的多变性和审计数据 shell 命令的相关性,利用特殊的多阶齐次 Markov 链模型对合法用户的正常行为进行建模,并通过双重阶梯式归并 shell 命令来确定状态,提高了用户行为轮廓描述的准确性和检测系统的泛化能力,并大幅度减少了存储成本.检测阶段根据实时性需求,采用运算量小的、仅依赖于状态转移概率的分类值计算方法,并通过加窗平滑处理分类值序列得到判决值,进而对被监测用户的行为进行判决.实验表明,同现有的典型检测方法相比,该方法在虚警概率相同的情况下大幅度提高了检测概率,并有效减少了系统计算开销,特别适用于在线检测.

关键词: 网络安全; 伪装攻击; 入侵检测; shell 命令; 异常检测; 多阶 Markov 链
中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2011) 05-1199-06

Masquerade Detection Based on Shell Commands and High-Order Markov Chain Models

XIAO Xi[1], ZHAI Qi-bin[1], TIAN Xin-guang[2], CHEN Xiao-juan[3], YE Run-guo[4]
(1. *State Key Laboratory of Information Security, Graduate University of Chinese Academy of Sciences, Beijing* 100049, *China*;
2. *Key Laboratory of Network Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190, *China*;
3. *College of Computer and Information Engineering, Beijing Technology and Business University, Beijing* 100037, *China*;
4. *Beijing Venustech Company Ltd, Beijing* 100193, *China*)

Abstract: Masquerade attacks are attempts by unauthorized users to gain access to confidential data or greater access privileges, while pretending to be legitimate users. This paper proposes a novel method to distinguish legitimate users from masqueraders. The uncertainty of the user's behavior and the relevance of the operation of shell commands are thoroughly considered. The method constructs specific high-order homogeneous Markov chain models to represent the normal behavior profiles of valid users. It defines the states by twofold hierarchical merging shell commands. Therefore this method increases the accuracy of describing the normal behavior profiles, improves the generalization of the detection system and sharply reduces the storage space. In the detection period, taking the real-time performance into account, it computes the categorical boolean variables only using the transition probabilities, which has little computation workload, and then smoothes them to get the decision values used to determine whether the monitored user's behavior is normal or anomalous. Its performance is tested in computer simulation, showing higher detection accuracy and fewer computation costs than related methods'. The proposed method is especially suitable for on-line detection.

Key words: network security; masquerade attack; intrusion detection; shell command; anomaly detection; high-order Markov chain

1 引言

伪装攻击(masquerade attack)是指非授权用户通过伪装成合法用户来获得访问关键数据或更高层访问权限的行为[1,2],是信息系统面临的最为严重的安全威胁之一.伪装攻击检测由 Smaha[1] 于 1988 年首次提出,目前已经成为入侵检测领域研究的热点内容,并在网络信息安全工程中发挥着越来越大的作用[2~9].伪装攻击检测系统当前大多采用异常检测技术.

伪装攻击检测面临的主要困难在于用户行为的复

收稿日期:2010-07-08;修回日期:2010-09-25
基金项目:国家 863 高技术研究发展计划(No.2006AA01Z452);国家 242 信息安全计划(No.2005C39)

杂性和多变性,即用户行为会随着工作内容、个人兴趣、工作时间和其它不确定性因素的变化而改变[8,10,11].Lane 等人[11,12]研究了基于实例学习的检测方法,用特定的相似度函数刻画当前行为与正常行为模式之间的相似性,有较强的适应能力,但检测准确率较低.孙宏伟等人[13]在 Lane 方法的基础上改进了对用户行为模式的表示方式,以 shell 命令为单位进行相似度赋值,改善了检测性能.Tian 等人[7]以 shell 命令序列为单位进行相似度赋值,克服了文献[13]相似度赋值的不足,提高了检测准确率.Schonlau 等人[14]研究了基于统计理论的伪装攻击检测方法,比较分析了 6 种方法的优势和局限性.Maxion 等人[10]引入贝叶斯分类算法,对 Schonlau 的方法进行了改进.最近,Dash 等人[8]提出延迟检测概念,利用适应性朴素贝叶斯方法进行检测,进一步提高了性能.曾剑平等人[3]研究了基于区间值 2 型模糊集的伪装入侵检测算法.Wu 等人[4]把主成份分析法运用于伪装攻击检测.Tian 等人[5]提出了基于 Markov 链模型的方法,考虑了 shell 命令(符号)的相互转移关系,具有良好的检测性能.此外,支持向量机[6,15]、数据挖掘[16]、生物信息技术[9]等技术也已应用于伪装攻击检测.

文献[5]假设每个 shell 命令仅与它前面的一个命令有关,采用 1 阶齐次 Markov 链模型.实际上,每个命令还会它与之前的多个命令相关.由此本文引入多阶齐次模型.但如果只在不同 shell 命令对应不同状态[5]的基础上采用一般的多阶模型,状态转移概率矩阵的存储空间和相应的计算复杂度会随模型的阶数呈指数级增加.本文建立特殊的多阶齐次 Markov 链模型,以克服上述不足.本文方法充分考虑了用户行为的多变性和 shell 命令的相关性,分别根据 shell 命令符号和 shell 命令序列的出现频率进行阶梯式数据归并来确定多阶 Markov 链的状态和序列状态,同现有的 Markov 链方法[5]相比,提高了用户行为轮廓描述的准确性和检测系统的泛化能力,并大幅度减少了存储空间.在检测阶段考虑到实时性需求,只利用状态转移概率来计算分类值,然后加窗平滑处理分类值序列得到判决值,来分析用户行为的异常程度,有效减少了系统计算开销.试验表明,同现有的 4 种典型检测方法相比,本文方法降低了存储成本和计算成本,在虚警概率相同的情况下大幅度提高了检测概率,改善了系统的整体性能,具有较强的实用性和可操作性,特别适用于在线检测.

2 相关知识

2.1 多阶 Markov 链[17,18]

定义 1 假设 Markov 链$\{X_t, t=1,2,\cdots\}$有 N 个状态,其状态空间为 $\Omega=\{1,2,\cdots,N\}$,如果其在 $m(m>n)$时刻所处的状态为 q_m 的概率只与前面 n 个状态有关,即

$$P(X_m=q_m \mid X_{m-1}=q_{m-1}, X_{m-2}=q_{m-2},\cdots,X_1=q_1)$$
$$=P(X_m=q_m \mid X_{m-1}=q_{m-1}, X_{m-2}=q_{m-2},\cdots,X_{m-n}=q_{m-n}) \quad (1)$$

则称该 Markov 链为 n 阶 Markov 链.式中 $q_1,q_2,\ldots,q_m\in\Omega=\{1,2,\cdots,N\}$,$(q_{m-n},q_{m-n+1},\cdots,q_{m-1})$称作该 Markov 链的(长度为 n 的)序列状态.所有序列状态记为 $\zeta_1,\zeta_2,\cdots,\zeta_M$,共 M 个,则

$$M=N^n \quad (2)$$

定理 1 如果 Markov 链$\{X_t, t=1,2,\cdots\}$是 n 阶 Markov 链,则该链也是 $n+1$ 阶 Markov 链.

定义 2 称 n 阶 Markov 链$\{X_t, t=1,2,\cdots\}$是齐次的,如果式(1)与 m 无关.

记 n 阶齐次 Markov 链$\{X_t, t=1,2,\cdots\}$状态转移概率矩阵 $\boldsymbol{A}=(a_{ij})_{M\times N}$,其中 a_{ij}表示当前序列状态为 ζ_i 下一个状态为 j 的概率($1\leqslant i\leqslant M, 1\leqslant j\leqslant N$).

注:(1)由定理 1 知,1 阶 Markov 链条件最强,它是任意阶的 Markov 链,但现实世界中很多事物之间的联系很复杂,不满足 1 阶的条件,所以人们要使用多阶($n>1$)Markov 链.(2)多阶 Markov 链的状态转移概率矩阵的元素有 $M\times N=N_{n+1}$个,其元素个数太多妨碍了它在实际生活中的应用[17].

2.2 审计数据的分析及预处理

与文献[2~16]中的检测方法相同,本文方法采用 Unix 平台上的 shell 命令作为审计数据.主要用于工作站模式的网络(如某些军队网络)以及特定研究或开发环境(如嵌入式系统)下 Unix 或 Linux 系统的伪装攻击检测;在此类平台上,shell 是终端用户与操作系统之间最主要的界面,能反映用户的行为,且 shell 命令容易收集,便于分析.本文方法在训练和检测阶段,需要对原始 shell 命令数据进行预处理,主要有两种方式:第一种是 Purdue 大学试验数据采用的方式,详见文献[5,7,11,12];第二种方式较为简洁,被 AT&T Shannon 试验室试验数据所采用,详见文献[10,14].

3 基于多阶 Markov 链模型的用户伪装攻击检测方法

3.1 训练

本文建立特殊的 n 阶齐次 Markov 链:预先设定 Markov 链的状态个数 N 和序列状态个数 M,然后根据单个 shell 命令确定 Markov 链的状态,根据长度为 n 的 shell 命令序列确定 Markov 链的长度为 n 的序列状态(为方便描述,以下用自然数 $1,2,\cdots,M$ 来表示序列状态).新确定的 $\boldsymbol{A}=(a_{ij})_{M\times N}$描述了长度为 n 的 shell 命

令序列与单个 shell 命令之间的转移关系,M 和 N 不满足关系式(2).

定义 3　设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_m)$为一个长度为 m 的有序符号串,称$\overline{x_i}=(x_i,x_{i+1},\cdots,x_{i+l-1})$为在 $\boldsymbol{x}$ 上以 l 为窗长截取出来的第 i 个序列$(1\leqslant i\leqslant m-l+1)$,其长度为 l.

定义 4　称由序列$\overline{x_i}$构成的序列$\overline{\overline{x}}=(\overline{x_1},\overline{x_2},\cdots,\overline{x_{m+l-1}})$为由 $\boldsymbol{x}$ 以 l 为窗长生成的序列流,简称为 x 的序列流.序列里的符号和序列流里的序列都是按时间先后次序排列的.

定义 5　以有序符号串 $\boldsymbol{x}=(x_1,x_2,\cdots,x_m)$为模版以$\{1,2,\cdots,K\}$为状态集合,符号 $x_{\#}$ 的状态 q 定义如下:

(1)提取出有序符号串 $\boldsymbol{x}$ 中互不相同的符号,并根据频率降序排列.设 $\boldsymbol{x}$ 中互不相同的符号共有 W 个$(W\leqslant m)$,符号 $x_{\&}$ 在 $\boldsymbol{x}$ 中出现的次数为 e,则 $x_{\&}$ 在 $\boldsymbol{x}$ 中出现的频率 f 定义为

$$f=e/m. \tag{3}$$

(2)把排序后的符号按频率从大到小阶梯式归并成 $K-1$ 个集合.设 b 是不大于 $W/(K-1)$的最大整数,$v=b+1$,$h=W-(K-1)b$,则 $W=hc+(K-h-1)b$,前 hc 个符号按 c 个一组归并成 1 个集合,剩下的符号按 b 个一组归并成 1 个集合.

(3)定义符号 $x_{\#}$ 的状态 q:如果存在 $1\leqslant i\leqslant K-1$,使 $x_{\#}\in\Lambda_i$,则 $q=i$;否则,$q=K$.在实际操作中为节约时间,可利用频率优先匹配方法确定 $x_{\#}$ 的状态 q:依次在集合 $\Lambda_1,\Lambda_2,\cdots,\Lambda_{k-1}$中查找 $x_{\#}$,如果在第 $i(1\leqslant i\leqslant K-1)$个集合 Λ_i 中查找到 $x_{\#}$,则 $q=i$;如果在所有 $K-1$ 个集合中都查找不到 $x_{\#}$,则 $q=K$.

训练阶段的主要工作是计算特殊的 n 阶 Markov 链的状态转移概率矩阵,具体步骤如下:

(1)获得合法用户的正常行为训练数据,并预处理为 shell 命令有序符号串 $\boldsymbol{s}=(s_1,s_2,\cdots,s_r)$.

(2)确定 n 阶 Markov 链的状态.以 shell 命令有序符号串 $\boldsymbol{s}$ 为模版,以$\{1,2,\cdots,N\}$为状态集合,利用定义 5 的方法确定 $\boldsymbol{s}$ 中每个 shell 命令符号的状态(符号状态)作为 Markov 链的状态.

(3)确定 n 阶 Markov 链的序列状态.首先由 shell 命令有序符号串 $\boldsymbol{s}$ 以 n 为窗长生成 shell 命令序列流 $\overline{\overline{s}}=(\overline{s_1},\overline{s_2},\cdots,\overline{s_{r-n+1}})$.把$\overline{s_i}$和$\overline{\overline{s}}$ 分别当作“符号”和“有序符号串”,以 $\overline{\overline{s}}$ 为模版,以$\{1,2,\cdots,M\}$为状态集合,确定 $\overline{\overline{s}}$ 中每个 shell 命令序列的“状态”(序列状态)作为 Markov 链的序列状态.

(4)计算 n 阶 Markov 链的状态转移概率矩阵 $\boldsymbol{A}=(a_{ij})_{M\times N}$.设在训练数据中,序列状态为 i 的(长度为 n 的)shell 命令序列之后的下一个 shell 命令符号的状态为 j 的次数(即序列状态 i 向符号状态 j 转移的次数)为 Z_{ij},序列状态 i 向各个符号状态转移的总次数为 Y_i,则

$$a_{ij}=\begin{cases}Z_{ij}/Y_i, & 1\leqslant i\leqslant M-1,1\leqslant j\leqslant N-1\\ 0, & i=M \quad \text{or} \quad j=N\end{cases} \tag{4}$$

3.2　检测

检测阶段的工作是利用特定的检测模型来识别被监测用户当前行为中的异常,步骤如下:

(1)得到被监测用户执行的 shell 命令行,并预处理成 shell 命令有序符号串 $\boldsymbol{c}=(c_1,c_2,\cdots,c_u)$,然后由 shell 命令有序符号串 $\boldsymbol{c}$ 以 $n+1$ 为窗长生成 shell 命令序列流 $\overline{\overline{c}}=(\overline{c_1},\overline{c_2},\cdots,\overline{c_{u-n}})$.

(2)根据状态转移概率矩阵,计算 shell 命令序列流 $\overline{\overline{c}}$ 中的每个序列的“状态转移概率”.

对 $\overline{\overline{c}}$ 中第 $i(1\leqslant i\leqslant u-n)$个长度为 $n+1$ 的序列$\overline{c_i}=(c_i,c_{i+1},\cdots,c_{i+n})$,利用训练阶段(2)(3)里的结果和频率优先匹配法,可确定符号 c_{i+n}的符号状态为 $h(1\leqslant h\leqslant N)$,及长度为 n 的子序列$(c_i,c_{i+1},\cdots,c_{i+n-1})$的序列状态为 $g(1\leqslant g\leqslant M)$,则$\overline{c_i}$的“状态转移概率”为 $P(c_{i+n}\mid c_i,c_{i+1},\cdots,c_{i+n-1})=a_{gh}$.

(3)根据 shell 命令序列流 $\overline{\overline{c}}$ 中每个序列对应的“状态转移概率”,计算每个序列的分类值.

分类值的计算公式为:

$$class(\overline{c_i})=\begin{cases}1, & P(c_{i+n}\mid c_i,c_{i+1},\cdots,c_{i+n-1})>\lambda\\ 0, & \text{others}\end{cases} \tag{5}$$

式中,$class(\overline{c_i})$表示 shell 命令序列$\overline{c_i}$对应的分类值,取值“1”表示正常转移,“0”表示异常转移,λ 为概率门限需预先设定.经过以上计算,可得到分类值序列$(class(\overline{c_1}),class(\overline{c_2}),\cdots,class(\overline{c_{u-n}}))$.

(4)对分类值序列$(class(\overline{c_1}),class(\overline{c_2}),\cdots,class(\overline{c_{u-n}}))$进行加窗平滑处理,获得判决值.

判决值的计算公式为:

$$D(k)=\frac{1}{w}\sum_{i=k-w+1}^{k}class(\overline{c_i}) \tag{6}$$

式中,$D(k)$表示 shell 命令序列$\overline{c_k}$对应的判决值,w 为窗长度,且 $w\leqslant k\leqslant u-n$,$k$ 的增长步长为 1.我们不直接利用分类值对用户行为进行判决,是考虑到用户在短时间内的行为可能会偏离其历史行为.

(5)根据判决值和预先设定的判决门限对用户行为进行判决.

设判决门限为 d,判决方法为:如果 $D(k)\geqslant d$,将被监测用户的“当前行为”判为正常行为,否则,将其判为异常行为.这里,“当前行为”是相对于 shell 命令序列$\overline{c_k}$而言的,它是指被监测用户执行的以 shell 命令序列$\overline{c_k}$

为终点的 w 个 shell 命令序列 $\overline{c_{k-w+1}}, \overline{c_{k-w+2}}, \cdots, \overline{c_k}$.

在在线检测的情况下,被监测用户所执行的 shell 命令行的获取和预处理,"状态转移概率"的计算,分类值序列的获得,判决值的计算以及对用户行为的判决都是同步进行的.

4 实验设计与结果分析

4.1 在 Purdue 大学数据上的试验

本组实验采用 Purdue 大学数据[11,12]中的 4 个用户 user1、user2、user3、user4 的数据.将 user1、user2、user4 设为伪装用户,将 user3 设为合法用户.每个用户的 shell 命令流中各有 15000 个命令,user3 的前 10000 个命令作为训练数据用于正常行为建模,而每个用户的后 5000 个命令作为测试数据用于性能测试.参数设置为 $M=3, N=3, n=2, w=91, \lambda=10^{-4}$.实验时,正常行为训练数据中互不相同的 shell 命令符号共有 200 个,Markov 链的状态个数为 3,互不相同长度为 2 的 shell 命令序列共有 843 个,Markov 链的长度为 2 的序列状态个数为 3,所需存储单元的个数为 $200+843\times2+3\times3=1895$.Markov 链的状态转移概率矩阵:

$$\boldsymbol{A}=\begin{pmatrix}0.989 & 0.011 & 0\\ 0.955 & 0.045 & 0\\ 0 & 0 & 0\end{pmatrix}$$

4.1.1 检测准确率分析

图 1 示出了由式(6)计算出的判决值曲线.可见,我们的判决值曲线具有很好的可分性.

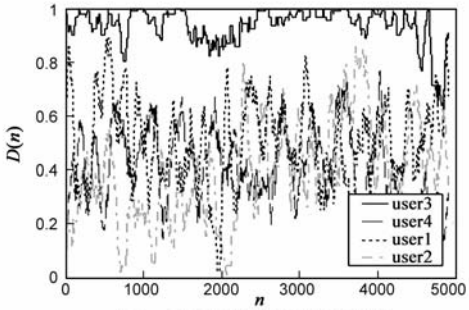

图1 式(6)对应的判决值曲线

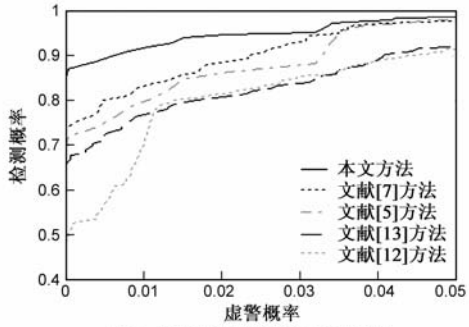

图2 5种不同方法的ROC曲线

图 2 示出了本文和文献[5,7,12,13]方法的 ROC 曲线.图中各种方法的参数是在保证平均检测时间基本相同的前提下设置的[19].由图可见,本文方法比其它 4 种方法的检测准确率均有大幅度的提高.

4.1.2 状态个数对检测性能的影响

图 3 示出了序列状态个数 M 和状态个数 N 同时变化的 ROC 曲线.$(M,N)=(2,2),(3,3),(5,5),(9,9)$时的检测准确率很接近(ROC 曲线重合了),$(17,17)$时的检测准确率稍好一点,$(17,17),(33,33),(65,65)$时的检测准确率越来越差.对某些具体用户而言,状态个数越多本文方法往往能够对用户行为描述得更加精细,而对用户复杂多变的行为适应性往往较差,训练数据不充分会引起检测准确率的降低.由图 2 和图 3 知本文方法(2,2)也比其它 4 种方法的检测准确率均有大幅度的提高.考虑到方法的普遍性和计算及存储成本,试验中取$(M,N)=(3,3)$.

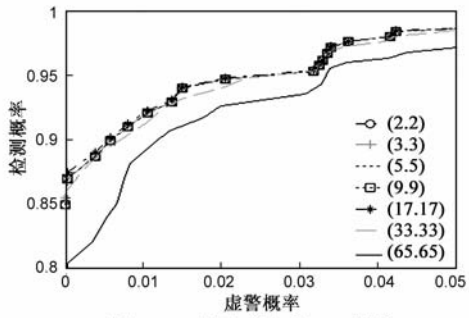

图3 *M,N*同时变化的ROC曲线

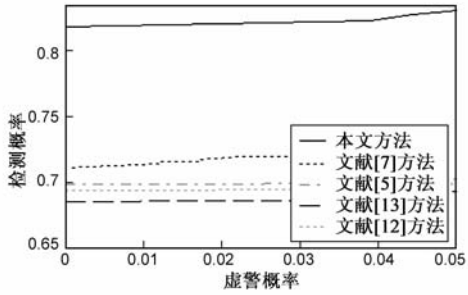

图4 在AT&T Shannon实验室数据上5种不同方法的ROC曲线

4.1.3 存储空间和试验时间分析

表 1 列出了本文方法和其它 4 种不同方法的存储空间和试验时间(相同试验条件下测量).

从表 1 可看出本文方法存储空间是所有方法中最少的,仅是文献[5]方法的 4.67%,减少了 2 个数量级.文献[5]方法状态个数是 201,而本文试验方法多阶 Markov 链的状态个数和序列状态个数都是 3,减少了 2 个数量级,状态转移概率矩阵的存储量仅为文献[5]方法的 0.02%,减少了 4 个数量级.实验时间是指实验中进行训练和检测所需要的时间,它与检测方法的计算成本成正比,并在一定程度上反映了检测的实时性.从表 1 可看出本文方法实验时间也是所有方法中最少的.

本文方法实验时间仅是文献[12]Lane 方法的 0.89%，　　减少了 2 个数量级;是文献[5]方法的 85.32%.

表 1　5 种不同方法的存储空间和试验时间

	文献[12]方法	文献[13]方法	文献[7]方法	文献[5]方法	本文方法
存储单元(个)	2544	3512	3512	40601	1895
试验时间(秒)	1743.1	30	28.609	18.094	15.437

综合考虑检测准确率、存储空间和试验时间,本文方法的整体性能优于已有的 4 种方法.

4.2　在 AT&T Shannon 实验室数据上的试验

本组实验采用 AT&T Shannon 实验室数据[14]中前 4 个用户 user1、user2、user3、user4 的数据,每个用户有 5000 个 shell 命令.实验时将 user4 设为合法用户,他的前 4000 个命令作为训练数据用于正常行为建模,其他 3 个用户设为伪装用户,他们的 5000 个 shell 命令和 user4 的后 1000 个命令均作为测试数据用于性能测试.参数设置为 $M=3, N=3, n=2, w=100, \lambda=0.74$.图 4 示出了本文和文献[5,7,12,13]方法的 ROC 曲线.可见,本文方法比其它 4 种方法的检测准确率均有大幅度的提高.

5　结束语

本文提出一种基于多阶 Markov 链模型的高效的伪装攻击检测方法,主要用于以 shell 命令为审计数据的主机入侵检测系统.该方法充分考虑了用户行为的多变性和审计数据的相关性,改进了对用户正常行为模式的表示方式,仅根据状态转移概率对用户行为进行判决.实验表明,同已有的 4 种典型检测方法相比,本文方法在降低存储成本和计算成本的同时,提高了检测准确率,具有很强的可操作性,特别适用于在线检测.而且,在实际应用中还可以通过优化参数设置进一步提高性能.

参考文献

[1] Smaha S E. Haystack: An intrusion detection system[A]. Proceedings of the IEEE forth Aerospace Computer Security Applications Conference[C]. Austin, Texas: Tracor Applied Science Inc, 1988. 37 – 44.

[2] 田新广,段▮毅,程学旗.基于 shell 命令和多重行为模式挖掘的用户伪装攻击检测[J].计算机学报,2010,33(4):697 – 705.
Tian Xin-guang, Duan Mi-yi, Cheng Xue-qi. Masquerade detection based on shell commands and multiple behavior pattern mining [J]. Chinese Journal of Computers, 2010, 33(4): 697 – 705. (in Chinese)

[3] 曾剑平,郭东辉.基于区间值 2 型模糊集的伪装入侵检测算法[J].电子学报,2008,36(04):777 – 780.
Zeng Jiang-ping, Guo Dong-hui. Masquerade intrusion detection algorithm based on interval type-2 fuzzy set[J]. Acta Electronica Sinica, 2008, 36(4): 777 – 780. (in Chinese)

[4] Wu H C, Huang S H S. User behavior analysis in masquerade detection using principal component analysis[A]. Proceedings of the 2008 Eighth International Conference on Intelligent Systems Design and Applications[C]. Washington DC, USA: IEEE Computer Society, 2008. 201 – 206.

[5] Tian X G, Duan M Y, Li W F, et al. Anomaly detection of user behavior based on shell commands and homogeneous Markov chains[J]. Chinese Journal of Electronics, 2008, 17(2): 231 – 236.

[6] Shim C Y, Kim J Y, Gantenbein R E. Practical user identification for masquerade detection[A]. Advances in Electrical and Electronics Engineering-IAENG Special Edition of the World Congress on Engineering and Computer Science 2008[C]. San Francisco, California, USA: IEEE Press, 2008. 47 – 51.

[7] Tian X G, Gao L Z, Sun C L, et al. A method for anomaly detection of user behaviors based on machine learning[J]. The Journal of China Universities of Post and Telecommunications, 2006, 13(2): 61 – 65, 78.

[8] Dash S K, Reddy K S, Pujari A K. Adaptive Naive Bayes method for masquerade detection[J]. Security and Communication Networks, 2010, DOI: 10.1002/sec.168.

[9] Coull S E, Branch J W, Szymanski B K, et al. Sequence alignment for masquerade detection[J]. Computational Statistics & Data Analysis, 2008, 52(8): 4116 – 4131.

[10] Maxion R A, Townsend T N. Masquerade detection using truncated command lines[A]. Proceedings of the International Conference on Dependable Systems and Networks[C]. Los Alamitos, California: IEEE Computer Society, 2002. 219 – 228.

[11] Lane T. Machine Learning Techniques for The Computer Security Domain of Anomaly Detection[D]. West Lafayette: Purdue University, 2000.

[12] Lane T, Brodley C E. An empirical study of two approaches to sequence learning for anomaly detection[J]. Machine Learning, 2003, 51(1): 73 – 107.

[13] 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 Hong-wei, Tian Xin-guang, Li Xue-chun, et al. An improved anomaly detection model for IDS[J]. Chinese Journal

of Computers, 2003, 26(11): 1450 – 1455. (in Chinese)
[14] Schonlau M, DuMouchel W, Ju W H, et al. Computer intrusion: detecting masquerades[J]. Statistical Science, 2001, 16(1): 58 – 74.
[15] Kim H S, Cha S D. Empirical evaluation of SVM-based masquerade detection using UNIX commands[J]. Computers & Security, 2005, 24(2): 160 – 168.
[16] Szymanski B K, Zhang Y Q. Recursive data mining for masquerade detection and author identification[A]. Proceedings of the 5th IEEE System, Man and Cybernetics Information Assurance Workshop[C]. Los Alamitos: IEEE CS Press, 2004. 424 – 431.
[17] Berchtold A, Raftery A. The mixture transition distribution model for high-order Markov chains and non-Gaussian time series[J]. Statistical Science, 2002, 17(3): 328 – 356.
[18] Apostolov S S, Mayzelis Z A, Usatenko O V, et al. Isotropy properties of the multi-step Markov symbolic sequences[J]. Physica A: Statistical Mechanics and its Applications, 2007, 376: 165 – 172.
[19] 田新广. 基于主机的入侵检测方法研究[D]. 长沙: 国防科学技术大学, 2005.
Tian Xin-guang. Anomaly Detection Methods for Host-Based Intrusion Detection Systems[D]. Changsha, China: National University of Defense Technology, 2005. (in Chinese)

作者简介

肖 喜 男, 1979 年生于湖南宜章, 中国科学院研究生院信息安全国家重点实验室博士生, 主要研究方向为入侵检测、信息安全和密码应用技术.
E-mail: xiaoxi_ac@163.com

翟起滨 男, 1947 年生于黑龙江哈尔滨, 中国科学院研究生院信息安全国家重点实验室教授, 博士生导师. 主要研究方向为密码学、信息安全和入侵检测. 从事密码技术领域工作近三十年, 获政府特殊津贴. 自 1987 年以来一直担任《中国科学》,《数学年刊》,《通信学报》等杂志的审稿专家. 承担国家密码与信息安全研究项目, 担任国家密码专项课题研究主持人.

田新广 男, 1976 年生于河北吴桥, 中国科学院计算技术研究所博士后, 中国计算机学会高级会员(计算机安全专业委员会委员), 主要研究方向为网络安全、入侵检测、智能信息处理. 在国内外重要学术期刊发表论文 80 余篇, 拥有 8 项发明专利, 先后主持了多项国家和部委级重大科研项目.

陈小娟 女, 1977 年生于江苏南通, 北京工商大学计算机与信息工程学院实验师, 主要研究方向为通信工程、数字信号处理.

叶润国 男, 1976 年生于江西萍乡, 博士后, 北京启明星辰信息安全技术有限公司资深安全工程师, 主要研究方向为数据挖掘和网络安全, 拥有 10 项发明专利.