基于时变加权马尔科夫链的网络异常检测模型

王 笑 戚 湧 李千目

(南京理工大学计算机科学与工程学院 南京 210094)

摘 要 随着互联网技术的迅猛发展,网络入侵事件日益频发,入侵检测对于保障网络安全具有重要意义。针对网络入侵检测的迫切需求,提出一种基于时变加权马尔科夫链的网络异常检测模型,使用组合状态转移概率矩阵来描述状态转移。利用DARPA 2000数据集在NT系统上重放时产生的事件log作为实验数据以验证该模型的效果,并与普通时变加权马尔科夫链模型进行比较,仿真实验结果表明该模型能够对网络进行实时入侵检测,具有较高的准确性和较强的鲁棒性,并且能够有效降低误测率和漏测率。

关键词 网络安全,加权马尔科夫,时变模型,入侵检测

中图法分类号 TP393.08 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.09.027

Network Anomaly Detection Model Based on Time-varying Weighted Markov Chain

WANG Xiao QI Yong LI Qian-mu

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract With the rapid development of the Internet, the network intrusion events are becoming more and more frequent, and the instruction detection is of great significance to the protection of network security. In view of the urgent demand of real-time instruction detection, a model of network instruction detection based on time-varying weighted Markov chain model was proposed in this paper. This model uses the combined state sequence to describe state transition. The log event generated by the DARPA2000 data set on the NT system was used as the experimental data to carry out simulation experiments, and the time-varying weighted Markov chain model were compared. The simulation results show that the model mentioned in this paper can be used for real-time instruction detection, which has high accuracy, strong robustness, and can effectively reduce the false detection rate.

Keywords Network security, Weighted Markov, Time varying model, Instruction detection

1 引言

互联网的多样性、开放性特点使得网络安全容易受到各种攻击的威胁[1-4]。目前,完全避免网络入侵事件的发生并不现实,网络安全人员须尽力发现和察觉入侵行为及攻击企图,以便采取有效的安全策略。入侵检测是及时发现入侵行为的技术,通过收集计算机网络或计算机系统中的若干关键点信息,并对收集的信息进行分析,从而判断计算机网络或系统中是否有异常行为和攻击迹象,是安全防御体系的一个重要组成部分[5-6]。

入侵检测技术分为两种类型:异常检测(anomaly detection)和误用检测(misuse detection)[7-9]。异常检测假定所有的入侵行为都与正常行为不同,若实时获得的当前用户行为与正常值的差异超出指定的阈值,则认为受到了攻击,进行入侵报警。误用检测假定所有的攻击形式都可以表达为一种特定的模式特征,可以通过模式匹配的方法直接检测出已知种类的入侵行为。基于上述两种检测类型的特点,Ambusaidi M A等[10]曾结合误用检测与异常检测两种方法提出混合入侵检测模型。本文提出的入侵检测方法属于异常检测,目前已研究的异常检测算法和模型主要包括:统计异常检测[11]、基于神经网络的异常检测[12]和基于数据挖掘的异常检测[13]。

近年来,在网络入侵检测领域中具有代表性的研究工作包括:杨雅辉等[14]针对传统的网络入侵检测方法不能识别网络上新出现的攻击类型的缺陷,对GHSOM神经网络模型进行了扩展,提出了一种基于增量式GHSOM神经网络模型的网络入侵检测方法,实现了对入侵检测模型的动态扩展。陈行、王辉等[15-16]设计了基于贝叶斯博弈理论的入侵检测模型。段雪涛[17]等提出了一种变长短序列的语义模式切分方法,并基于此语义模式切分方法提出了基于层次隐马尔科夫模型的入侵检测方法。以上研究工作及方法虽然能够对网络进行入侵检测,但是都具有一定局限性。基于神经网络的入

到稿日期:2016-08-08 返修日期:2016-12-19 本文受国家自然科学基金项目(61272419)、赛尔下一代互联网创新项目(NGII20160122)、中兴通讯产学研合作论坛合作项目(2016ZTE04-11)资助。

王 笑(1992-),女,硕士生,主要研究方向为信息安全;戚 湧(1970-),男,博士,教授,博士生导师,CCF高级会员,主要研究方向为信息安全,E-mail:790815561@qq.com(通信作者);李千目(1979-),男,博士,教授,博士生导师,主要研究方向为信息安全。

侵检测的缺点包括系统建立速度慢、模型更新代价高、实时性不足等。基于贝叶斯博弈理论的入侵检测存在参数选取难、收敛速度慢等缺陷。基于层次隐马尔科夫模型的入侵检测方法存在建模复杂度较高、建模耗时长的缺点，影响了网络入侵检测的实时性。张勇、席荣荣等人[18-19]通过对威胁、管理员和普通用户的行为进行博弈分析，建立了三方参与的 Markov 博弈模型，以动态评估系统安全态势。冯学伟等人[20]提出了一种基于马尔科夫性质的因果知识挖掘方法，为因果知识挖掘提供了一种新的方法。邓鑫洋等人[21]针对边界数据状态难以确定的问题，通过将信度理论引入马尔科夫形成了信度马尔科夫模型。李方伟等人[22]使用时变加权马尔科夫链模型进行网络安全的实时检测。以上马尔科夫相关研究和进展都具有重要意义，但在网络入侵检测领域仍具有一定局限性。马尔科夫博弈模型的建模过程过于复杂；信度马尔科夫还未在入侵检测领域有所验证；时变加权马尔科夫链模型考虑了不同时间段的状态间客观存在的相依关系，可以提高检测的精度，且建模复杂度可控。网络攻击的发生往往是一系列行为的集合，连续的几个安全攻击行为才能够导致系统在未来的安全状态发生变化。以上几个基于马尔科夫的模型均认为几个单点时刻的状态能够在一定程度上确定未来某一时刻的状态，未能充分发挥历史数据的作用，检测结果依然缺乏可信度。

基于以上不足，本文提出一种基于组合状态的时变加权马尔科夫链网络异常检测模型，使用 DARPA2000 数据集在 NT 系统上重放时产生的事件 log 作为实验数据，并进行仿真实验验证该模型的效果，同时与普通时变加权马尔科夫链模型进行比较。

2　时变加权马尔科夫链模型

马尔科夫过程（Markov Process）是一个典型的随机过程，当过程在 t_0 时刻所处的状态已知时，$t(t>t_0)$ 时刻所处的状态与过程在 t_0 时刻之前的状态无关，上述特性称为无后效性。马尔科夫过程中的时间和状态可以是离散的，也可以是连续的。时间离散、状态离散的马尔科夫过程被称为马尔科夫链。马尔科夫链中各个时刻的状态的转变由状态转移概率矩阵控制，传统的马尔科夫链模型往往假设系统状态转移概率矩阵是不随时间变化的，然而在许多实际问题中，状态的转移概率矩阵 P 是随时间不断变化的，符合该特性的马尔科夫链模型是时变马尔科夫链模型。

在实际应用中，各事物在各个时段的数值、状态具有一定的关联关系。传统的马尔科夫链模型忽略了事物各状态间的关联关系，加权马尔科夫链模型考虑了状态间客观存在的关联关系，并将这种关联关系的强弱进行量化，作为马尔科夫链中的权重，本文的研究工作基于时变加权马尔科夫链模型进行。

时变加权马尔科夫链模型具体如式(1)所示：

$$\lambda=\{S,P,w,\pi,m\} \tag{1}$$

其中：

1）S 为系统的状态空间，是由系统所有可能的状态所组成的非空的状态集。

2）$P_k=[p_{ij}(t,t+k)]$ 为系统的 K 阶状态转移概率矩阵，$p_{ij}(t,t+k)=P\{X_{t+k}=j \mid X_t=i\}$ 表示系统在时刻 t 处于状态 i，但经过 k 步之后（即处于时刻 $t+k$ 时）状态转移至状态 j 的概率，$i,j\in S$，且存在对于任意 $i\in S$，满足 $\sum_{j=1}^{n}p_{ij}(t,t+k)=1$，$0\leqslant p_{ij}(t,t+k)\leqslant 1,i,j\in S$，$P_k$ 随时间变化实时更新。

3）$w=[w_1,w_2,\cdots,w_n]$ 为系统各阶状态转移概率矩阵所占权重。

4）$\pi=[\pi_1\ \pi_2,\cdots,\pi_n]$ 为系统的初始概率分布矩阵，π_i 表示系统在初始时刻处于状态 i 的概率，且满足 $\sum_{i=1}^{n}\pi_i=1$。

5）m 代表一共有 m 阶状态转移矩阵。

使用时变加权马尔科夫链进行入侵检测时需要确定 k 步转移概率矩阵 P_k 以及各阶矩阵所占权重，$k=1,2,\cdots,m$，m 的取值根据实际情况确定；然后，结合 k 步之前的状态、各阶状态转移概率矩阵及其所占权重，计算出下一步转移到各个状态的概率，概率越大，处于该状态的可能性越大。

3　基于组合状态的时变加权马尔科夫链异常检测模型

3.1　模型定义

针对普通时变加权马尔科夫链模型的不足，本文使用组合状态序列来描述状态转移，提出一种基于组合状态的时变加权马尔科夫链模型，如式(2)所示：

$$\lambda=\{S,P,w,\pi,m\} \tag{2}$$

其中：

1）S 为系统的状态空间，是由系统所有可能的状态所组成的非空的状态集。

2）$P_k=[p_{ijs}(t,t+k)]_{n*n*n}$ 为系统的 k 阶状态转移概率矩阵，矩阵行数为 $n*n$，列数为 n，$p_{ijs}(t,t+k)=p\{X_{t+k}=s \mid X_{t-1}=i,X_t=j\}$，$i,j\in S$ 表示系统在时刻 $t-1$ 处于状态 i，t 时刻处于状态 j，经过 k 步之后（即处于时刻 $t+k$ 时）状态转移至状态 s 的概率，且存在对于任意 $i,j\in S$，满足 $\sum_{s=1}^{n}p_{ijs}(t,t+k)=1$，$0\leqslant p_{ijs}(t,t+k)\leqslant 1,i,j,s\in S$，$P_k$ 随时间变化实时更新。

3）$w=[w_1,w_2,\cdots,w_n]$ 为系统各阶状态转移概率矩阵所占权重。

4）$\pi=[\pi_1,\pi_2,\cdots,\pi_n]$ 为系统的初始概率分布矩阵，π_i 表示系统在初始时刻处于状态 i 的概率，且满足 $\sum_{i=1}^{n}\pi_i=1$。

5）m 代表一共有 m 阶状态转移矩阵。

使用该模型进行入侵检测时同样需要确定 k 步转移概率矩阵 P_k 以及各阶矩阵所占权重，不同的是这个转移概率矩阵 P_k 中的数记录的是从组合状态 $X_{t-1}X_t$ 转移至各状态 X_{t+k} 的概率。结合 k 步之前的组合状态 $X_{t-k}X_{t-k+1}$ 和各阶状态转移概率矩阵及其所占权重，可以得到系统下一步转移到各个状态的概率。

3.2　基于组合状态的时变加权马尔科夫链的异常检测

基于组合状态的时变加权马尔科夫链模型的异常检测步骤如下：

Step1　量化网络风险值，得到一个具有时序性的风险值序列，对风险值进行聚类，参照聚类结果确定安全状态分类标准。

Step2 根据分类标准对网络安全状态进行判定，并生成状态序列。

Step3 选取状态序列中相邻的两个状态为一组，本文称它为组合状态。统计每一个状态组合以及每个组合经过 k 步之后转移到各状态的数量，$k=1,2,\cdots,m$，m 的取值根据实际情况确定，在任何情况下都采取统一的阶数是存在问题的。过小的阶数能发挥的作用十分有限，而过大的阶数则会造成很大的稀疏，导致算法效率大幅下降。因此需要根据具体环境确定阶数。当日志中异常日志出现的频率较小且稳定时，选取小阶数；当日志中异常日志出现的频率较大时，选取大阶数。可以参照权重进行调整，Step6 中给出了进一步调整 m 的方法。

Step4 根据 Step3 获得的组合状态和转移数据，根据频率近似概率的原理计算转移概率，生成 $1-m$ 阶组合状态转移概率矩阵，训练马尔科夫链模型。以 3 个状态为例，生成的组合状态转移概率矩阵形式如式(3)所示，表示从组合状态经 k 步后转移到各状态的概率：

$$P_k=\begin{bmatrix} P_{000} & P_{001} & P_{002} \\ P_{010} & P_{011} & P_{012} \\ P_{020} & P_{021} & P_{022} \\ P_{100} & P_{101} & P_{102} \\ P_{110} & P_{111} & P_{112} \\ P_{120} & P_{121} & P_{122} \\ P_{200} & P_{201} & P_{202} \\ P_{210} & P_{211} & P_{212} \\ P_{220} & P_{221} & P_{222} \end{bmatrix} \tag{3}$$

Step5 确定各阶转移矩阵的权重，步长为 k 时，组合状态 $x_{t-1}x_t$ 与 x_{t+k} 之间的相关系数记为 q_k，权重记为 w_k，具体如下：

$$q_k=\frac{\sum_{t=1}^{n-k}(x_{t-1}+x_t-2\bar{x})(x_{t+k}-\bar{x}))}{\sqrt{\sum_{t=1}^{n-k}(x_{t-1}+x_t-2\bar{x})^2}*\sqrt{\sum_{t=1}^{n-k}(x_{t+k}-\bar{x})^2}} \tag{4}$$

$$w_k=\frac{|q_k|}{\sum_{k=1}^{m}|q_k|} \tag{5}$$

式(4)中 $\bar{x}$ 表示马尔科夫链序列值的平均值，n 表示序列值的总数。

Step6 根据权重调整有效转移概率矩阵的阶数范围 $1\sim m$，设定权重阈值为 0.05，如果 k 阶转移概率矩阵的权重小于阈值，则可以将它舍弃，继续对比 $k-1$ 阶直至找到不小于阈值的矩阵为止。根据调整后的阶数，重新计算各阶矩阵权重。

Step7 利用 Step4 生成的各阶状态转移概率矩阵以及当前组合状态 X_{t-1} X_t 和 Step6 调整后的权重，基于加权的方式，得出下一时刻处于各安全状态的概率，即 $p_{t+1}=\sum_{i=1}^{m}p^i_{(X_{t-i}X_{t-i+1})}*w_i$。其中 p_{t+1} 表示系统在 $t+1$ 时刻处于各状态的概率；$p^i_{(X_{t-i}X_{t-i+1})}$ 表示 i 阶状态转移概率矩阵中 $t-i$ 时刻的状态为 X_{t-i}，$t-i+1$ 时刻的状态为 X_{t-i+1} 对应的行(转移到各状态的概率)；w_i 是 i 阶状态转移概率矩阵所占权重；m 表示一共有 m 阶状态转移矩阵，$1\leqslant i\leqslant m$。

通过以上步骤可以得到系统下一时刻处于各状态的概率 p_i，$i\in S$，概率最大的状态即被认为是系统下一时刻所处状态。

根据攻击的阶段及网络所处的风险状态的不同，将网络安全状态划分为：正常状态 L_0(即安全状态，认为此时无风险)、轻微风险状态 L_1(即网络被扫描探测)、低风险状态 L_2(此时网络中的漏洞可能被发现利用)、较严重风险状态 L_3(网络已经受到了攻击)以及严重风险状态 L_4(受到的攻击很大，网络已被攻陷)。这些状态之间以一定的概率相互转移，从而构成基于组合状态的时变加权马尔科夫链模型的状态空间，即 $S=\{L_0,L_1,L_2,L_3,L_4\}$，由此得到网络安全状态转移如图 1 所示。

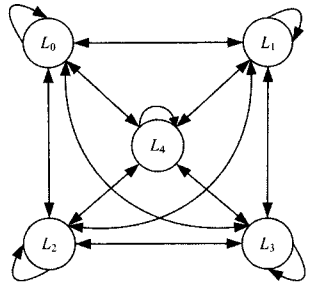

图 1 网络安全状态转移

在改进模型中，考虑从组合状态转移到某一状态的情况，可能存在的组合状态有 25 种，组成的集合为 $Z=\{L_0L_0,L_0L_1,L_0L_2,\cdots,L_4L_2,L_4L_3,L_4L_4\}$，构成的网络安全组合状态转移如图 2 所示。

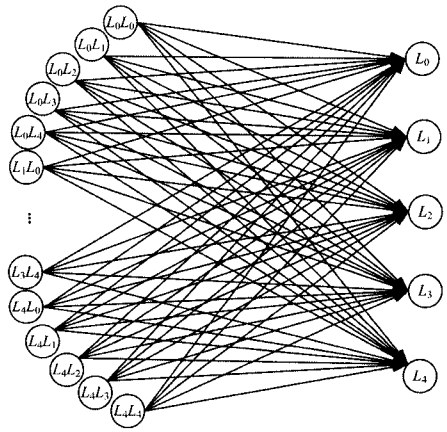

图 2 网络安全组合状态转移

计算 k 阶组合状态转移矩阵 P_k，$k=1,2,\cdots,m$，转移矩阵 P_k 中的每一个元素对应每个组合状态经过 k 步后转移到其他任何一个状态的转移概率，采用频率近似概率的原理可以确定转移概率矩阵，如式(6)所示：

$$P_k=\begin{bmatrix} P_{L_0L_0L_0} & P_{L_0L_0L_1} & P_{L_0L_0L_2} & P_{L_0L_0L_3} & P_{L_0L_0L_4} \\ P_{L_0L_1L_0} & P_{L_0L_1L_1} & P_{L_0L_1L_2} & P_{L_0L_1L_3} & P_{L_0L_1L_4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{L_4L_4L_0} & P_{L_4L_4L_1} & P_{L_4L_4L_2} & P_{L_4L_4L_3} & P_{L_4L_4L_4} \end{bmatrix} \tag{6}$$

当有新的数据加入时，结合历史数据和当前数据进行统计，实时更新转移概率矩阵 P。同时考虑到在进行入侵检测

时采用过小或者过大的阶数都会影响算法效率，在入侵检测过程中需要根据具体环境确定并调整阶数。

使用加权方式得出下一时刻处于各安全状态的概率 p_i，$i \in S$。引入代价向量 C 表示网络在每个状态的风险值，对代价向量 $C=\{c_1, c_2, \cdots, c_n\}$ 及所处状态的概率加权的方式进行定量分析，利用公式 $R=\sum_{i=1}^{n} p_i c_i$ 量化得到网络的综合风险值。

4　仿真实验

4.1　数据集介绍

为了验证提出的基于组合状态的时变加权马尔科夫链模型对入侵检测的有效性，将其与普通的时变加权马尔科夫链模型进行比较。本文使用 DARPA 2000 数据集在 NT 系统上重放时产生的事件 log 作为实验数据进行仿真实验，DARPA 2000 数据集的攻击场景如图 3 所示。

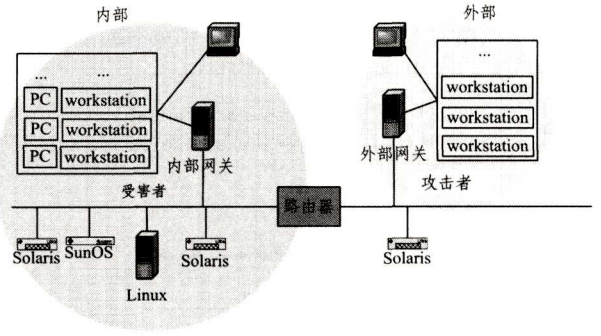

图 3　DARPA 2000 数据集的攻击场景

DARPA 2000 攻击场景测试数据集包含一系列的攻击，整个攻击过程通过 DDoS 攻击实现。攻击者先通过 IP Sweep 进行活动主机的探测；然后进行端口扫描，查找到具有 Sadmind 漏洞的主机之后，攻击具有该漏洞的 3 台主机：Pascal (172.16.112.50)，Mill(172.16.115.20)和 Locke(172.16.112.10)，使之成为傀儡机；再在傀儡机上安装实施 DDoS 攻击的木马软件，利用被控主机对攻击目标发起 DDoS 攻击。

DARPA 2000 数据集的攻击步骤如图 4 所示。

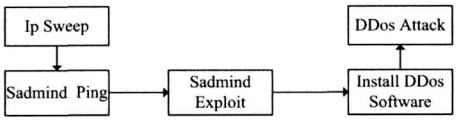

图 4　DARPA 2000 数据集的攻击步骤

攻击步骤具体如下。

Step1　IP Sweep。攻击者对目标网络进行扫描以搜寻活跃的主机。

Step2　Sadmind Ping。对 Step1 中发现的活跃主机进行探测，了解有哪些主机在执行 Sadmind 远端管理者工具，进而锁定攻击目标。

Step3　Sadmind Exploit。对于 Step2 中锁定的 3 台主机 Mill，Pascal 和 Locke，攻击者不停地尝试利用 Sadmind 的漏洞进行攻击，直至成功入侵。

Step4　Install DDos Software。在 3 台傀儡机上安装 DDoS 攻击工具，攻击者通过 RSH 服务远程登录傀儡机，并在所有傀儡机上安装会产生真正 DDoS 攻击包的攻击工具，同时在其中一台受害主机上安装一个攻击代理，该代理提供一个使用者界面并能控制安装在其他受害主机上的攻击工具。

Step5　DDos Attack，即 DDoS 攻击。攻击者远程登录到安装了攻击代理的主机，控制所有安装了攻击工具的受控机器伪造 IP 地址，并一起对远程服务器进行 DDoS 攻击。

通过对 DARPA 2000 的攻击步骤进行研究，参考 MIT 实验室发布的 High-Level Attack Truth File，可以得出：1～70min 处于安装攻击软件、收集信息等准备阶段；IP Sweep 和端口扫描发生在 70～125min 之间；Sadmind Exploit 发生在 126～240min 之间；Install DDos Software 发生在 241～319min；最后发起攻击的时间在 320min。

4.2　实验数据处理

采用 MIT 实验室公布的 DARPA 2000 数据集在 NT 系统上重放时产生的事件 log 作为实验数据，该 log 持续时间约为 386min，其中有信息、警告、错误 3 种级别的事件，出现的事件 ID 有 29 种，对应的事件信息如表 1 所列。

表 1　DARPA 2000 数据集产生的事件 log 类型

事件 ID	级别	任务类别	关键字	来源
592	信息	详细追踪	经典，审核成功	Security
593	信息	详细追踪	经典，审核成功	Security
594	信息	详细追踪	经典，审核成功	Security
595	信息	详细追踪	经典，审核成功	Security
0	信息	无	经典	TimeServ
512	信息	系统事件	经典，审核成功	Security
514	信息	系统事件	经典，审核成功	Security
515	信息	系统事件	经典，审核成功	Security
517	信息	系统事件	经典，审核成功	Security
576	信息	特权使用	经典，审核成功	Security
577	信息	特权使用	经典，审核成功	Security
578	信息	特权使用	经典，审核失败	Security
528	信息	登录/注销	经典，审核成功	Security
538	信息	登录/注销	经典，审核成功	Security
2001	信息	—3	经典	Mail Service
2003	信息	—3	经典	Mail Service
1000	信息	—4	经典	Mail Service
5722	错误	无	经典	NETLOGON
7001	错误	无	经典	Service Control Manager
3216	错误	无	经典	REPLICATOR
560	信息	对象访问	经典，审核成功	Security
562	信息	对象访问	经典，审核成功	Security
564	信息	对象访问	经典，审核成功	Security
3007	警告	—4	经典	Mail Service
1008	错误	无	经典	Perflib
632	信息	帐户管理	经典，审核成功	Security
633	信息	帐户管理	经典，审核成功	Security
636	信息	帐户管理	经典，审核成功	Security
637	信息	帐户管理	经典，审核成功	Security

参见文献[23-24]，可以通过给每个事件 log 确定风险代价的方式量化系统当前的安全状态。由于涉及的事件 log 的类型并不多，且很多事件 log 属于同类操作，如 560，562，564 均为对象访问，因此对安全风险产生的效果极为类似。为了将评估结果数值化，本文将同类事件合并，并根据每条告警的威胁程度设置事件的风险值，风险值的设置参考文献[23-24]，设置级别为："信息"的事件的风险值范围为 1～3，"警告"级别的风险值范围为 4～5，"错误"级别的风险值范围为6～8。在同一级中又可以再细分，如"信息"级重点检查的事

件 ID 为 528/538:用户成功登录/注销计算机;578:特权获取失败等,具体设置如表 2 所列。

表 2 log 事件对应的风险值

事件 ID	级别	风险值
592,593,594,595	信息	1
0	信息	1
512,514,515,517	信息	1
576,577	信息	2
578	信息	3
528,538	信息	2
1000,2001,2003	信息	1
632,633,636,637	信息	3
560,562,564	信息	1
3007	警告	5
5722	错误	6
7001	错误	8
3216	错误	6
1008	错误	6

通过计算实验数据中每分钟的日志产生的积累风险值形成具有时序性的风险值序列,序列跨度为 386min。使用k-means 方法对风险值进行聚类,基于聚类结果确定模型状态空间的分类标准。

根据分类标准,如果网络的风险值在 1～300 之间,则表示该网络处于安全的状态(L_0);如果风险值在 300～1350 之间,则表示网络很可能被探测到(L_1);如果风险值在 1350～3350 之间,则表示网络可能遭受到攻击(L_2);如果风险值在 3350～5500 之间,则表示网络已经遭受到攻击(L_3);如果风险值超过 5500,则表示网络受到的攻击已经非常严重(L_4)。

基于组合状态的时变加权马尔科夫链模型的参数具体设置如下:初始概率 $\pi=\{1,0,0,0\}$,代价向量 $C=\{39,600,2120,4365,12275\}$。其中代价向量 C 中的具体数值参照每个分类状态中的风险值数据的平均值。

为方便以图表展示,本文以分钟为单位,对一分钟内产生的风险值进行统计,得出网络安全风险值序列。整个攻击从 8:00 开始,至 14:26 结束,持续时间为 386min,风险值在 0～16000 之间。

4.3 实验结果分析

图 5—图 9 示出了使用普通时变加权马尔科夫链和基于组合状态的时变加权马尔科夫链对网络进行入侵检测的结果。

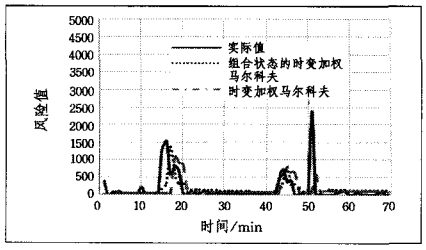

图 5 网络入侵检测结果(0～69min)

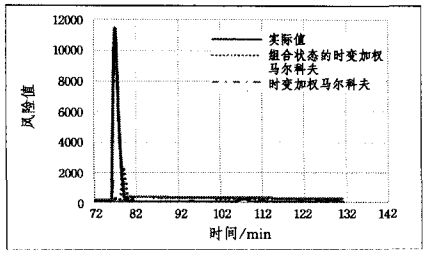

图 6 网络入侵检测结果(70～130min)

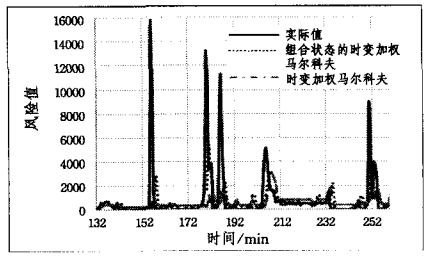

图 7 网络入侵检测结果(131～260min)

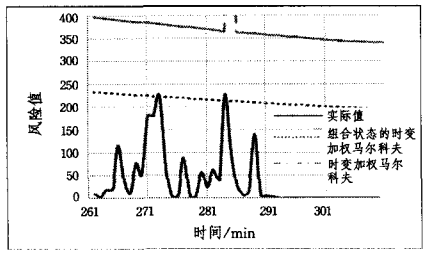

图 8 网络入侵检测结果(261～310min)

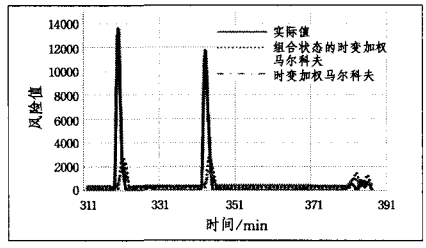

图 9 网络入侵检测结果(310min 以后)

图 5—图 9 中实线为安全状态的原始曲线,虚线为使用基于组合状态的时变加权马尔科夫链的入侵检测曲线,点划线为使用普通时变加权马尔科夫链的入侵检测曲线。通过分析图 5—图 9 可以看出,入侵检测曲线的走势与原始曲线(实际值)相符,即符合网络安全状态的实际变化,可以得到两种入侵检测方法的效果对比,具体分析如下:

(1)在 0～69min 阶段,网络风险值不高,攻击者进行攻击软件的安装以及信息探测等准备工作。两条曲线在该阶段的检测值都符合实情。

(2)在 70～130min 阶段,网络受到攻击者的端口 sweep 扫描和 Sadmindping 试探,网络风险值出现了突然增高的现象。基于组合状态的时变加权马尔科夫链的入侵检测比普通时变加权马尔科夫链更好地检测到了该次风险突增。

(3)在 131～260min 阶段。攻击者锁定 3 台主机 Mill,Pascal 和 Locke,并尝试利用 Sadmind 漏洞攻击,直至入侵成功并获得完全控制权限。这一阶段出现了比较高的网络风险值,且风险值的波动性较大。可以看出,在出现高风险值的情况下,点线的拟合度比点划线更高,即基于组合状态的时变加权马尔科夫链的入侵检测比普通时变加权马尔科夫链在高风险检测方面有更好的表现,漏测率更低。

(4)在 261～310min 阶段,网络风险值明显降低。经过上一阶段的攻击,攻击者已经获得了 3 台主机的管理员权限,使之成为傀儡机,这一阶段攻击者将登录到 3 台主机上,分别在 3 台主机上安装攻击程序。事实上,这一步的风险已经很大,但是由于我们采用了 log 日志作为实验数据,而上一步攻击者已经获得了权限,因此主机并没有意识到这是一个攻击行

为，也就没有对这一步的攻击行为产生告警或者错误类型的日志，属于不可控因素，这间接影响了入侵检测的准确性。在此背景下我们分析两条曲线的拟合情况，可以看出点线的检测误差显然要比点划线小。

(5)在 310min 以后，攻击者发起了 DDoS 攻击，出现了非常高的网络风险值。同样可以看出，基于组合状态的时变加权马尔科夫链的入侵检测方法比普通时变加权马尔科夫链在高风险检测方面有更好的表现，漏测率更低。

本文将实际网络风险值处于安全状态而检测值已经达到 400 的情况视为误测，将实际网络风险值已经达到 L_3 标准而检测值还处于安全状态的情况视为漏测，对使用时变加权马尔科夫链模型以及基于组合状态的时变加权马尔科夫链模型进行入侵检测的仿真实验效果进行对比，结果如表 3 所列。

表 3　仿真实验效果对比/%

	时变加权马尔科夫链	基于组合状态的时变加权马尔科夫链
漏测	3.1	1.9
误测	9.3	4.3

仿真实验结果表明：

(1)基于组合状态的时变加权马尔科夫链模型能够以较高的准确度检测网络受到的攻击及面临的风险，在所监控的网络受到攻击时，风险值显著增加，当网络受到的攻击减少时，风险值也降低。风险值的变化在时间上与原始信息非常吻合。

(2)网络风险值具有波动性，且在被攻击环境下的变化具有一定的规律性，组合状态的引入能够使得历史数据得到更加充分的利用，进而使得该模型的入侵检测更准确，相较于普通的时变加权马尔科夫链模型，能够更好地平滑离值波动，与原数据的误差更小。

(3)相较于普通的时变加权马尔科夫链模型，基于组合状态的时变加权马尔科夫链模型出现误测的比例为 4.3%，低于原时变加权马尔科夫链模型的 9.3%；与时变加权马尔科夫链模型相比，基于组合状态的时变加权马尔科夫链模型出现漏测的比例也从 3.1%降低至 1.9%。

该实验证明，基于组合状态的时变加权马尔科夫链模型具有很高的灵敏度及良好的网络入侵检测精度，相较于普通的时变加权马尔科夫链模型能够降低误测率和漏测率。

结束语　针对网络安全事件频发对网络安全技术提出的新要求，本文分析了时变加权马尔科夫链模型在异常检测中的不足，对时变加权马尔科夫链模型进行改进，以相邻的状态序列组成的组合为基本单位计算状态转移概率，提出一种基于组合状态的时变加权马尔科夫链网络入侵检测模型，利用 DARPA 2000 在 NT 系统上重放时产生的事件 log 作为实验数据对该模型的入侵检测效果进行验证。实验结果表明，该模型具有很好的实时性、较高的准确度及较强的鲁棒性，能够更好地平滑离值波动，有效降低误测率和漏测率。

但是该模型存在状态空间分类的稳定性、实时性问题。本文提出的模型的状态空间分类依赖于聚类，是对大量数据进行一次性聚类，以此确定状态空间的分类标准。该模型没有涉及聚类的更新，从而影响了入侵检测效果的实时性，而聚类算法的稳定性也需纳入考虑。下一步工作将开展深入研究，优化模型的建立过程，以期能够更加高效、准确地对网络进行实时的入侵检测。

参 考 文 献

[1] LIANG Y J, XU L L, TANG W. CNCERT released the 2013 Internet network security Posture Review[J]. China Information Security, 2014, 26(4): 20. (in Chinese)
梁玉坚，徐玲玲，唐雯. CNCERT 发布《2013 年互联网网络安全态势综述》[J]. 中国信息安全，2014，26(4)：20.

[2] National computer network emergency technology coordination center. Review of China's Internet security situation in 2015 [J]. Secrecy Science and Technology, 2016(4): 12-16. (in Chinese)
国家计算机网络应急技术处理协调中心. 2015 年我国互联网网络安全态势综述[J]. 保密科学技术，2016(4)：12-16.

[3] SAHA D, MUKHERJEE A. Pervasive Computing: A Paradigm for the 21st Century[J]. Computer, 2003, 36(3): 25-31.

[4] MARY M. Internet Trends 2016[EB/OL]. [2016-09-28]. http://www.kpcb.com/internet-trends.

[5] HUANG J Z, ZHU M L. Review of anomaly detection based on program [J]. Computer Science, 2011, 38(6): 7-13. (in Chinese)
黄金钟，朱淼良. 基于程序的异常检测研究综述[J]. 计算机科学，2011，38(6)：7-13.

[6] QING S H, JIANG J C, MA H T, et al. Survey of intrusion detection technology [J]. Journal of Communication, 2004, 25(7): 19-29. (in Chinese)
卿斯汉，蒋建春，马恒太，等. 入侵检测技术研究综述[J]. 通信学报，2004，25(7)：19-29.

[7] GOVINDARAJAN M, CHANDRASEKARAN R. Intrusion detection using neural based hybrid classification methods [J]. Computer Networks, 2011, 55(8): 1662-1671.

[8] GARC, A-TEODORO P, AZ-VERDEJO J, et al. Anomaly-based network intrusion detection: Techniques, systems and challenges [J]. Computers & Security, 2009, 28(1/2): 18-28.

[9] MOHAMMAD M N, SULAIMAN N, MUHSIN O A. A novel intrusion detection system by using intelligent data mining in weka environment[J]. Procedia Computer Science, 2011, 3(1): 1237-1242.

[10] AMBUSAIDI M A, HE X, NANDA P, et al. Building an intrusion detection system using a filter-based feature selection algorithm[J]. IEEE Transactions on Computers, 2016, 65(10): 2986-2998.

[11] BIERMANN E, CLOETE E, VENTER L M. A comparison of Intrusion Detection systems[J]. Computers & Security, 2001, 20(8): 676-683.

[12] PALOMO E J, DOMÍNGUEZ E, LUQUE R M, et al. An Intrusion Detection System Based on Hierarchical Self-Organization [C]// International Workshop on Computational Intelligence in Security for Information Systems (Cisis'08). Genova, Italy, October. DBLP, 2008: 139-146.

[13] HAN S J, CHO S B. Detecting intrusion with rule-based integration of multiple models[J]. Computers & Security, 2003, 22(7): 613-623.

(下转第 161 页)

[2] JIANG W, CLIFTON C. A secure distributed framework for achieving k-anonymity [J]. Journal of Very Large Data Bases J, 2006, 15(4): 316-333.

[3] MOHAMMED N, FUNG B C M, DEBBABI M, et al. Anonymity meets game theory: secure data integration with malicious participants [J]. Journal of Very Large Data Bases, 2011, 20(4): 567-588.

[4] SONG J L, HUANG L M, LIU G H. Algorithm for Finding Quasi-identifiers in the k-anonymity Method [J]. Journal of Chinese Mini-Micro Computer Systems, 2008, 29(9): 1688-1693. (in Chinese)
宋金玲，黄立明，刘国华. k-匿名方法中准标示符的求解算法[J]. 小型微型计算机系统，2008，29(9)：1688-1693.

[5] FUNG B C M, WANG K, YU P S, et al. Anonymizing Classification Data for Privacy Preservation [J]. IEEE Transaction on Data Engineering, 2007, 19(5): 711-725.

[6] WANG P S, MA Q J. Research on k-anonymity algorithm for privacy preservation [J]. Computer Engineering and Applications, 2011, 47(28): 117-119. (in Chinese)
王平水，马钦娟. 隐私保护 k-匿名算法研究[J]. 计算机工程与应用，2011，47(28)：117-119.

[7] YANG X C, WANG Y Z, WANG B, et al. Privacy Preserving Approaches for Multiple Sensitive Attributes in Data Publishing [J]. Chinese Journal of Computers, 2008, 31(4): 574-587. (in Chinese)
杨晓春，王雅哲，王斌，等. 数据发布中面向多敏感属性的隐私保护方法[J]. 计算机学报，2008，31(4)：574-587.

[8] MACHANAVAJJHALA A, GEHRKE J, KIFER D. l-diversity: privacy beyond anonymity [C]// The 22nd International Conference on Data Engineering. New York, ACM Press, 2006: 24-35.

[9] TRAIAN T M, BINDU V. Privacy protection: p-sensitive k-anonymity property [C]// The 22nd International Conference on Data Engineering. New York: ACM Press, 2006: 94.

[10] XIAO X K, TAO Y E. Personalized privacy preservation [C]// Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. Chicago, Illinois, USA: ACM Press, 2006: 229-240.

[11] SWEENEY L. Achieving k-anonymity privacy protection using generalization and suppression[J]. International Journal of Uncertainty Fuzziness, 2012, 10(5): 571-588.

[12] JURCZYK P, XIONG L. Distributed Anonymization: Achieving Privacy for Both Data Subjects and Data Providers [C]// LFIP Wag 11. 3 Working Conference on Data& Applications Security XXiii. 2009: 191-207.

[13] TAKENOUCHI T, KAWAMURA T, OSUNA A. Distributed Anonymization Method with Hiding the Presence of Individuals [J]. Ibices Transactions on Information & Systems, 2013, 96(3): 596-610.

[14] HAN J M, YU J, YU H Q, et al. Individuation Privacy Preservation Oriented to Sensitive Values[J]. Acta Electronica Sinica, 2010, 38(7): 1723-1728. (in Chinese)
韩建民，于娟，虞慧群，等. 面向敏感值的个性化隐私保护[J]. 电子学报，2010，38(7)：1723-1728.

(上接第 141 页)

[14] YANG Y H, HUANG H Z, SHEN Q N, et al. Intrusion detection based on incremental GHSOM neural network model[J]. Journal of Computer Science, 2014(5): 1216-1224. (in Chinese)
杨雅辉，黄海珍，沈晴霓，等. 基于增量式 GHSOM 神经网络模型的入侵检测研究[J]. 计算机学报，2014(5)：1216-1224.

[15] CHEN X, TAO J, et al. Intrusion detection algorithm based on Bias game model in wireless networks [J]. Journal of Communication, 2010, 31(2): 107-112(in Chinese)
陈行，陶军，等. 无线网络中基于贝叶斯博弈模型的入侵检测算法研究[J]. 通信学报，2010，31(2)：107-112.

[16] WANG H, CHEN H Y, LIU S F, et al. Intrusion detection system based on improved naive Bayes algorithm [J]. Computer Science, 2014, 41(4): 111-115. (in Chinese)
王辉，陈泓予，刘淑芬，等. 基于改进朴素贝叶斯算法的入侵检测系统[J]. 计算机科学，2014，41(4)：111-115.

[17] DUAN X T, JIA C F, LIU C B. Detection method of hierarchical hidden Markov model and variable length semantic model based on Intrusion [J]. Journal of Communication, 2010, 31(3): 109-114. (in Chinese)
段雪涛，贾春福，刘春波. 基于层次隐马尔科夫模型和变长语义模式的入侵检测方法[J]. 通信学报，2010，31(3)：109-114.

[18] ZHANG Y, TAN X B, CUI X L, et al. Network security situation awareness method based on Markov game model [J]. Chinese Journal of Software, 2011, 22(3): 495-508. (in Chinese)
张勇，谭小彬，崔孝林，等. 基于 Markov 博弈模型的网络安全态势感知方法[J]. 软件学报，2011，22(3)：495-508.

[19] XI R R, YUN X C, ZHANG Y Z, et al. An improved quantitative evaluation method of network security situation [J]. Chinese Journal of Computers, 2015, 38(4): 749-758. (in Chinese)
席荣荣，云晓春，张永铮，等. 一种改进的网络安全态势量化评估方法[J]. 计算机学报，2015，38(4)：749-758.

[20] FENG X W, WANG D X, HUANG M H, et al. A method of causal knowledge mining based on Markov [J]. Computer Research and Development, 2014, 51(11): 2493-2504. (in Chinese)
冯学伟，王东霞，黄敏桓，等. 一种基于马尔科夫性质的因果知识挖掘方法[J]. 计算机研究与发展，2014，51(11)：2493-2504.

[21] DENG X Y, DENG Y, ZHANG Y J, et al. A Markov reliability model and application [J]. Journal of Automation, 2012, 38(4): 666-672. (in Chinese)
邓鑫洋，邓勇，章雅娟，等. 一种信度马尔科夫模型及应用[J]. 自动化学报，2012，38(4)：666-672.

[22] LI F W, DENG W, ZHU J. A network security situation prediction mechanism based on complex network [J]. Computer Application Research, 2015, 32(4): 1141-1144. (in Chinese)
李方伟，邓武，朱江. 一种基于复杂网络的网络安全态势预测机制[J]. 计算机应用研究，2015，32(4)：1141-1144.

[23] DONG J. Research on improved HMM network security risk assessment method [D]. Wuhan: Huazhong University of Science and Technology, 2008. (in Chinese)
董静. 改进的 HMM 网络安全风险评估方法研究[D]. 武汉：华中科技大学，2008.

[24] LEI J. Research on network security threat and situation assessment [D]. Wuhan: Huazhong University of Science and Technology, 2008. (in Chinese)
雷杰. 网络安全威胁与态势评估方法研究[D]. 武汉：华中科技大学，2008.