基于自修正系数修匀法的网络安全态势预测

杨宏宇，张旭高
（中国民航大学计算机科学与技术学院，天津 300300）

摘 要：针对目前网络安全态势预测方法的精确度不足问题，以自修正系数修匀法为基础提出一种新的网络安全态势预测模型。首先，设计一种网络安全态势评估量化方法，基于熵关联度将警报信息转化为态势实际值时间样本序列。然后，计算静态修匀系数自适应解并利用可变域空间获取预测初始值。最后，为了进一步提高预测精度，基于偏差类别并采用时变加权马尔可夫链对网络安全态势初始预测结果进行修正。采用 LL_DOS_1.0 数据集检验预测效果，实验结果表明，所提模型面向网络态势时间序列具有较高的自适应性和预测精度。

关键词：安全态势；量化方法；可变域空间；修正；多重系数修匀
中图分类号：TP309
文献标识码：A
doi:10.11959/j.issn.1000−436x.2020092

Self-corrected coefficient smoothing method based network security situation prediction

YANG Hongyu, ZHANG Xugao
School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China

Abstract: In order to solve the problem of insufficient accuracy of current network security situation prediction methods, a new network security situation prediction model was proposed based on self-correcting coefficient smoothing. Firstly, a network security assessment quantification method was designed to transform the alarm information into situation real value time series based on the entropy correlation degree. Then, the adaptive solution of the static smoothing coefficient was calculated and the predicted initial value was obtained by using the variable domain space. Finally, based on the error category, the time-changing weighted Markov chain was built to modify the initial network situation prediction result and the prediction accuracy was further raised. The prediction model was tested with LL_DOS_1.0 dataset and the experimental results show that the proposed model has higher adaptability and prediction accuracy for network situation time series.

Key words: security situation, quantification method, variable domain space, modify, multiple coefficient smoothing

1 引言

网络安全态势预测方法通过对网络中各种安全预警（报警）信息和关联信息的处理生成时间样本序列，通过对相关信息的进一步处理和分析获取一定时间段内的网络安全总体情况和可能变化，对及时发现网络中存在的高危态势具有积极作用。目前，灰色预测法、机器学习预测法和时间序列预测法为常见的网络安全态势预测方法[1]。

Cipriano 等[2]基于以往警报提出一种网络攻击行为预测模型。该模型将以往警报作为训练集，通过机器学习方法获得警报知识库，再根据现有警报序列预测攻击者下一步攻击行为，为实时评估网络安全态势提供参考。Xiao 等[3]提出了基于 MEA-BP（mind evolution algorithm-back propagation）的网络安全态势预测方法。该方法通过对网络权重和阈值

收稿日期：2019−11−14；**修回日期**：2020−04−11
基金项目：国家自然科学基金资助项目（No.U1833107）
Foundation Item: The National Natural Science Foundation of China (No.U1833107)

进行改进提高了安全态势的预测准确率和效率，但对以往数据的标准化不够完善。Sun[4]提出了基于复杂网络的 Markov 预测模型。该模型将网络安全状况的转换关系构造成复杂网络，并利用加权马尔可夫链预测安全态势，可在一定程度上反映网络的安全状态，但面对多状态的网络，所构造出的状态转移概率矩阵规模过大。Leau 等[5]提出一种经卡尔曼滤波方程修正的网络安全态势预测模型。该模型基于层次分析法生成网络安全态势值序列，并通过灰色 Verhulst-Kalman 方法动态预测网络安全态势，但局限于安全态势为单峰变化的情况。Schatz 等[6]提出一种减少不确定性的安全预测方法。该模型基于信息安全领域内具有不同程度专业知识受访者对网络安全威胁的认知语料，利用概率主题建模方法预测网络安全威胁，但受访人群的层次、经验的离散性会影响预测精度。孙卫喜等[7]提出一种网络安全态势预测方法，提高了网络安全态势预测的准确率和有效性，但所需源数据维度较多。周新卫等[8]通过灰熵关联法提取影响网络安全的主要因素，并在此基础上建立卡尔曼滤波方程，提高了安全态势预测的精度。韩晓露等[9]提出基于直觉模糊集的非线性自回归神经网络预测模型（IFS-NARX, nonlinear autoregressive neural network with exogenous inputs based on intuitionistic fuzzy set），对网络安全态势预测可靠性的提升途径做了有益的探索。

针对上述网络安全态势预测方法中存在的数据质量参差不齐以及对多峰变化的网络安全态势预测准确度降低的不足，为解决目前网络安全态势预测方法的准确性和有效性不足的问题，本文提出一种基于自修正系数修匀法的网络安全态势预测模型。

2　网络安全态势预测模型

本文提出的基于自修正系数修匀法的网络安全态势预测模型如图 1 所示。其中，初始预测部分由可变域空间内的安全时间样本序列建立多重系数修匀模型以得到初始预测值；预测修正部分通过初始预测值和真实结果的偏离建立时变加权马尔可夫链，通过该模型对偏差值进行预测并修正初始预测值，最终得到网络安全态势预测结果。

模型的具体处理和分析过程设计如下。

步骤 1　基于熵关联度将网络警报信息转化为安全态势值非线性时间序列。

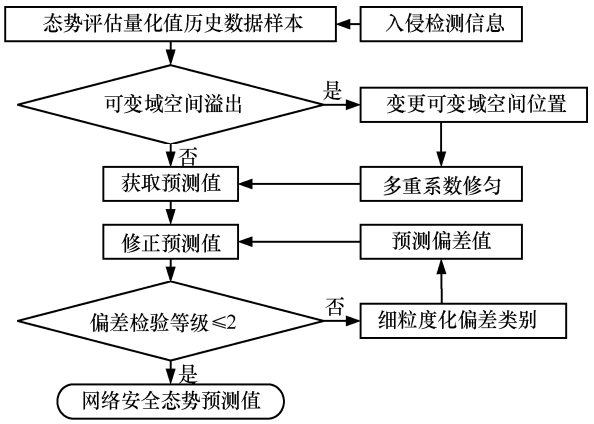


图 1　网络安全态势预测模型

步骤 2　利用可变域空间划分网络安全态势值序列片段，每更新一个安全态势值，可变域空间即向后移动一个单位。

步骤 3　基于可变域空间内的安全态势序列建立多重系数修匀预测模型，并通过自适应调整静态修匀系数α以初步提高预测精度。

步骤 4　计算可变域空间内的安全态势预测值与实际值的偏差，将偏差划分为 k 个偏差区间或分区。采用时变加权马尔可夫链模型对预测值进行处理，对偏差值进行预测并对原始预测值进行二次修正。

步骤 5　检验偏差，若未满足阈值条件，则返回步骤 4，并将偏差类别划分为 k+1 个；若满足阈值条件，则按步骤 1~步骤 4 得到下一周期的安全态势值。

本文模型通过动态调整静态修匀系数α初步提高态势值预测精度，再通过调整偏差类别数量提高时变加权马尔可夫模型对偏差的预测精度，最终完成对安全态势预测值的自适应修正目标。

3　网络安全态势评估量化

首先，基于开源入侵检测系统获取警报信息。然后，基于熵关联度计算各量化周期内的网络安全态势值。具体方法设计如下。

各周期网络安全态势量化值依据具有最高质量值的警报确定[10]。在 C 个量化周期内，Z_i（i=1,2,…,C）为周期 i 的量化值，Q_i为周期 i 内质量值最高的警报，则 Z_i=Z_i（AO_{Q_i},AM_{Q_i},AN_{Q_i}）（i=1,2,…, C），其中，警报发生率（AO, alarm occurance）为

$$\mathrm{AO}_{Q_i}=\frac{Z_i\text{内}Q_i\text{警报数}}{Z_i\text{内所有警报数}} \tag{1}$$

Q_i 的警报致变程度（AM, alarm mutagenicity）为 AM_{Q_i}，表示 Q_i 引发网络安全状态变更的难易程度。AM_{Q_i} 越低，则变更难度越大。AM_{Q_i} 优先级设为 1、2、3，分别对应警报 Q_i 为周期 i 内发生、周期 i–M 至周期 i–1 内发生和周期 i–M 至周期 i–1 内未发生，本文取 M=2[11]。

Q_i 的警报负面程度（AN, alarm negativity）为 AN_{Q_i}，该值越小，则网络安全状态受 Q_i 影响程度越小。AN_{Q_i} 优先级设为 3、2、1，分别对应警报负面程度为高危、中危、低危。

网络安全态势依据评价关联度矩阵 $\boldsymbol{R}$（如表 1 所示）量化。令 $Y_1=\mathrm{AO}_{Q_i}$，$Y_2=\mathrm{AM}_{Q_i}$，$Y_3=\mathrm{AN}_{Q_i}$，则 Y_1、Y_2、Y_3 分别对应周期 i 内警报质量最高的警报 Q_i 的 3 个量化指标，即警报发生率、警报致变程度及警报负面程度。表 1 中 r_{ij} 为第 i 个指标关联第 j 个评价（$i,j\in\{3,2,1\}$）的密切程度。

表 1　　评价关联度矩阵 $\boldsymbol{R}$

指标	高危	中危	低危
Y_3	r_{33}	r_{32}	r_{31}
Y_2	r_{23}	r_{22}	r_{21}
Y_1	r_{13}	r_{12}	r_{11}

为区分指标 Y_1 对网络安全威胁严重程度，设定警报发生率区间 o_j 如表 2 所示，基于 Y_1 值和警报发生率区间端点偏离距离计算当前时刻 Y_1 和区间 o_j（j=3, 2, 1）的相关度，该相关度即为 Y_1 对表 1 内各评价的关联度。

表 2　　警报发生率区间 o_j

o_j	警报负面程度
[0.7,1)	高危
[0.3,0.7)	中危
[0,0.3)	低危

设 $Y_1=y$，则特定指标与每个评价之间的关联度为

$$r_{1j}=\frac{1-\left|y-\dfrac{L_j+U_j}{2}\right|+\dfrac{U_j-L_j}{2}}{\displaystyle\sum_{j=1}^{3}\left[1-\left|y-\frac{L_j+U_j}{2}\right|+\frac{U_j-L_j}{2}\right]} \tag{2}$$

其中，L_j 和 U_j 分别为 o_j 的下端点和上端点，j=3, 2, 1。

由于指标 Y_2、Y_3 均根据优先级判定指标对网络安全威胁程度，故指标 Y_2、Y_3 对各评价的关联度设定如表 3 所示。

表 3　　Y_2、Y_3 对各评价的关联度设定

优先级	高危	中危	低危
3	0.5	0.333	0.167
2	0.25	0.5	0.25
1	0.167	0.333	0.5

表 3 中，优先级越高，则指标威胁程度越大，故低优先级对评价高危、中危、低危的关联度递增，反之则递减。当指标 Y_i（$i\in\{2,3\}$）优先级为 j 时，取表 3 与 j 同一行内的关联度作为表 1 内 Y_i（$i\in\{2,3\}$）对应的关联度。警报各指标的绝对熵值为

$$E_i=-\sum_{j=1}^{n}r_{ij}\ln r_{ij} \tag{3}$$

当 $r_{i1}=r_{i2}=\cdots=r_{in}$ 时，$E_{\max}=\ln\ n$，则警报各指标的相对熵值为

$$\psi_i=-\frac{1}{\ln n}\sum_{j=1}^{n}r_{ij}\ln r_{ij} \tag{4}$$

某指标相对熵值越大，则表示该指标对警报的量化值的影响越小，则以 $1-\psi_i$ 表示对应指标的权值，即

$$\eta_i=\frac{1}{n-\displaystyle\sum_{i=1}^{n}\psi_i}(1-\psi_i) \tag{5}$$

其中，$\eta_i\in[0,1]$为指标 Y_i 的熵权系数，且 $\eta_1+\cdots+\eta_n=1$。

各评价权值[12]为 $\boldsymbol{S}=(s_{\text{高危}},s_{\text{中危}},s_{\text{低危}})=\left(\dfrac{7}{15},\dfrac{1}{3},\dfrac{1}{5}\right)$。

在第 i 个周期，计算得到其网络安全态势的量化结果[13]为

$$Z_i=\rho\eta\boldsymbol{R}\boldsymbol{S}^{\mathrm{T}} \tag{6}$$

其中，态势放大系数 ρ=10 000。态势量化值越高，则网络安全状况越差。

4　网络安全态势预测方法

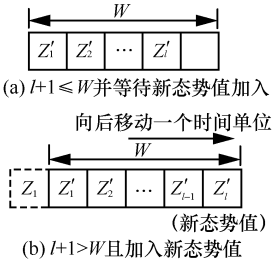
通过自适应调整静态修匀系数 α，使基于多重系数修匀法获取的初始预测结果精度较高。初始预测值计算步骤如下。

步骤 1　利用可变域空间划分以往数据时间序列片段。设 W 为域空间宽度，$Z_1,Z_2,\cdots,Z_p$（p 为正整数）为当前网络安全态势评估量化值序列，可变域空间工作过程如下。

1) 定义当前域空间宽度内的态势值个数为 l（1≤l≤p），则在该域空间内的该值的时间序列为 Z_1',Z_2',…,Z_l'。若 l+1≤W，域空间位置固定，则计算第 l+1 周期态势值，然后在域空间内输入下一个态势值。

2)若 l+1>W，输入下一个态势值到以往序列中，并将域空间向后移动一个时间单位，以域空间内新态势值序列片段为对象，计算第 p+1 个周期的态势值。

可变域空间的移动与取值变化如图 2 所示。该机制保证多重系数修匀法所基于的时间序列长度不超过 W，从而保证新的安全态势值加入以往序列后多重系数修匀法仍能正常预测，且可提高安全态势值预测的准确性与动态性。



(a) l+1≤W并等待新态势值加入

(b) l+1>W且加入新态势值

图 2　可变域空间的移动与取值变化

步骤 2　计算静态修匀系数。设当前网络安全态势值序列为 Z_1,Z_2,…,Z_p，域空间内态势值个数为 l。若 p≤W，则 Z_1'=Z_1，Z_l'=Z_p；若 p>W，则 Z_1'=Z_{p-W+1}，Z_l'=Z_p。多重系数修匀法为

$$Z_{t+X}^1 = d_t + e_t X + f_t X^2 \tag{7}$$

其中，周期 t+X 的安全态势量化预测结果为 Z_{t+X}^1，预测周期提前量为 X，d_t、e_t、f_t 为周期 t 的预测系数。

$$d_t = 3v_t^{(1)} - 3v_t^{(2)} + v_t^{(3)} \tag{8}$$

其中，$v_t^{(1)}$、$v_t^{(2)}$、$v_t^{(3)}$分别为周期 t 的 1、2、3 次修匀系数。

$$e_t = \frac{\alpha}{2(1-\alpha)^2}[(6-5\alpha)v_t^{(1)} - 2(5-4\alpha)v_t^{(2)} + (4-3\alpha)v_t^{(3)}] \tag{9}$$

$$f_t = \frac{\alpha^2}{2(1-\alpha)^2}(v_t^{(1)} - 2v_t^{(2)} + v_t^{(3)}) \tag{10}$$

其中，α∈[0, 1]为静态修匀系数。

设 $v_{t-1}^{(1)}$、$v_{t-1}^{(2)}$、$v_{t-1}^{(3)}$为周期 t 的 1、2、3 次系数修匀初始值，则

$$v_t^{(1)} = \alpha Y_t + (1-\alpha)v_{t-1}^{(1)} \tag{11}$$

$$v_t^{(2)} = \alpha v_t^{(1)} + (1-\alpha)v_{t-1}^{(2)} \tag{12}$$

$$v_t^{(3)} = \alpha v_t^{(2)} + (1-\alpha)v_{t-1}^{(3)} \tag{13}$$

其中，Y_t 为周期 t 的真实态势值

设 3 次修匀的周期预测均值为 $\boldsymbol{Z}'=\frac{Z_1'+Z_2'+Z_3'}{3}$，则修匀系数初始值 $v_0^{(1)}$、$v_0^{(2)}$、$v_0^{(3)}$均为 Z'。

在上述处理过程中，α的取值间接影响最终预测结果的准确性和精度。通常，当实际值序列呈水平趋势时，α∈[0.05, 0.2]；当实际值序列存在波动，但长期波动较小时，α∈[0.3, 0.5]；当实际值序列波动很大，呈明显的上升或下降趋势时，α∈[0.6, 0.8]。α值越大，表明远期数据对预测值的影响越大。因态势实际值序列片段随可变域空间位置发生变化，本文通过最小化实际值和预测值的偏差绝对值之和求得α自适应解。α自适应解求解步骤如下。

1) 设当前可变域空间内的 l 个网络安全态势实际值组成向量 $\boldsymbol{Z}'$=（Z_1',Z_2',…,Z_k'），静态修匀系数α初值为 0。

2) 已知修匀系数初始值 $v_0^{(1)}$、$v_0^{(2)}$、$v_0^{(3)}$均为 Z'，Y_1=Z_1'。首先，由式(11)~式(13)计算得到 $v_t^{(1)}$、$v_t^{(2)}$、$v_t^{(3)}$；然后，由式(8)~式(10)计算得到 d_t、e_t、f_t，其中，t=0,1,…,l。

3) 设 t=0,1,…,l–1，预测周期提前量 X=1，由式(7)计算得到经α修匀的预测值序列 Z^1。

4)设预测值序列与实际值序列的偏差绝对值之和为 V，则有

$$V = \sum_{i=1}^{l} \left| Z_i^1 - Z_i' \right| \tag{14}$$

5) 循环 1)~4)，若α=1，则转到步骤 3；否则继续循环 1)~4)。

设第 j 次循环后得到的偏差的绝对值为 V_j，计算得到 V_j最小值条件下的α值静态修匀系数自适应解为α_a。

步骤 3　计算网络安全态势初始预测值。令 t=l=p，α=α_a，X=1，由式(7)~式(13)求得第 p+1 个周期的安全态势值。

5　预测值的修正

通过网络安全态势预测子模块，得到可变域空间内各周期网络安全态势初始预测值。根据常识可知，该值与同域空间内的已知安全态势实际值存在偏差，且偏差大小与可变域空间内安全态势波动大

小有关。本文将预测值与实际值偏差划分为若干偏差类别，并通过时变加权马尔可夫链预测偏差值。

5.1 偏差类别划分

处于不同时刻的网络所面临的漏洞、威胁将发生变化，可能出现如下的情况。

1)在短时间内网络遭受集中攻击，导致其安全态势出现较大波动，安全态势预测值与实际值偏差上、下限值距离较大。

2)网络面临常规漏洞，故其安全态势在一定时间内会较为平缓或出现较小波动，安全态势预测值与实际值偏差上、下限值距离较小。

新态势值加入以往序列引发可变域空间移动，改变可变域空间内态势值序列片段波动离散程度和最大、最小偏差距离。

设 i=1,2,3,⋯,l，则当前域空间中的态势实际值片段为 $Z=\{Z_i'\}$，态势预测值片段为 $Z^1=\{Z_i^1\}$，最小偏差值为 $D^L=\min\{Z_i^1-Z_i'\}$，最大偏差值为 $D^U=\max\{Z_i^1-Z_i'\}$，偏差距离为 DL=D^U–D^L。划分偏差类别步骤如下。

步骤 1 划分偏差距离为 k 个区间，区间宽度为 $\frac{\mathrm{DL}}{k}$，区间元素为 $\left[D^L,D^L+\frac{\mathrm{DL}}{k}\right)$，$\left[D^L+\frac{\mathrm{DL}}{k},D^L+2\frac{\mathrm{DL}}{k}\right)$，⋯，$\left[D^L+(k-1)\frac{\mathrm{DL}}{k},D^U\right]$。

步骤 2 设 i=1,2,3,⋯,l，当前域空间内偏差时间序列为 $D=\{D_i=Z_i^1-Z_i'\}$，若 $D_i\in\left[D^L+(j-1)\frac{\mathrm{DL}}{k},D^L+j\frac{\mathrm{DL}}{k}\right)$，则偏差 D_i 属于偏差类别 j，$j\in$1,2,⋯,k。当 D_i=D^U 时，则可将 D_i 划归为类别 k。

步骤 3 若修正后的预测值不满足偏差检验等级要求，则偏差类别数 k=k+1，使偏差修正细粒度化。

5.2 偏差预测方法

针对域空间内的偏差类别的样本，采用时变加权马尔可夫链模型对安全态势的偏差进行预测，具体步骤设计如下。

步骤 1 安全态势的偏差类别转移概率矩阵获取。假设当前安全态势的偏差类别为 k 个，当前时刻为 x，相邻时刻偏差类别为 $d_{x-1}d_x$，m 个时刻后的偏差类别为 d_{x+m}，则有

$$p_a=p\{d_{x+m}=c\mid d_{x-1}=a,\ d_x=b\},\ a,b,c\in1,2,\cdots,k$$

其中，p_a 为偏差类别转移概率，a 为时刻 x–1 的偏差类别，b 为时刻 x 的偏差类别，c 为时刻 x+m 的偏差类别。

设 k 为偏差类别数，当 k=3 时，转移概率矩阵为

$$\boldsymbol{P}_{(k\times k)\times k}^{m}=\begin{pmatrix}p_{111}&p_{112}&\cdots&p_{11k}\\p_{121}&p_{122}&\cdots&p_{12k}\\\vdots&\vdots&\ddots&\vdots\\p_{kk1}&p_{kk2}&\cdots&p_{kkk}\end{pmatrix}\tag{15}$$

其中，k=1,2,⋯,φ。φ值通过步骤 3 调整，其初值由可变域空间宽度 W 确定，本文取$\varphi_0=\left\lceil\frac{W}{3}\right\rceil$。

步骤 2 确定偏差类别转移概率矩阵权值。首先计算 $d_{x-1}d_x$ 和 d_{x+m} 间的相关系数χ_q 为

$$\chi_q=\frac{\sum_{x=1}^{k-m}\left(q_{x-1}+q_x-2\overline{q}\right)\left(q_{x+m}-\overline{q}\right)}{\sqrt{\sum_{x=1}^{k-m}\left(q_{x-1}+q_x-2\overline{q}\right)^2\sum_{x=1}^{k-m}\left(q_{x+m}-\overline{q}\right)^2}}\tag{16}$$

其中，q_{x-1}、q_x 和 q_{x+m} 分别为域空间内时刻 x–1、时刻 x 和时刻 x+m 的偏差值，$\overline{q}$ 为域空间内偏差序列片段均值，m=1,2,⋯,φ。则 m 阶偏差类别转移概率矩阵权值μ_m 为

$$\mu_m=\frac{|\chi_m|}{\sum_{m=1}^{\varphi}|\chi_m|},\ m=1,2,\cdots,\varphi\tag{17}$$

步骤 3 根据μ_φ值调整φ值。当μ_φ<0.05[14]时，去除对预测偏差作用较小的φ阶偏差类别转移概率矩阵，令φ=φ–1 并更新μ_φ值，当$\mu_\varphi\geqslant$0.05 时，取 $m_{\max}$=φ。

步骤 4 计算偏差预测值。x+1 时刻偏差值属于偏差类别 c，c=1,2,⋯,k 的概率 $p_{c(x+1)}$为

$$p_{c(x+1)}=\sum_{m=1}^{\varphi}p_{abc}{}^{(m)}\mu_m\tag{18}$$

其中，m=1,2,⋯,φ；$a,b\in\{1,2,\cdots,k\}$，$p_a{}^{(m)}$根据 m 阶偏差类别转移概率矩阵 $\boldsymbol{P}^m$ 确定，表示由相邻偏差类别组 d_{x-m}=a、d_{x-m+1}=b 转移至偏差类别 d_{x+1}=c 的概率。μ_m 为 m 阶偏差类别转移概率矩阵权值，x+1 时刻的安全态势预测偏差的类别概率分布向量为 $\boldsymbol{P}_{c(x+1)}=\{p_{1(x+1)},p_{2(x+1)},\cdots,p_{k(x+1)}\}$。

设由各偏差区间中值组成的偏差中值向量为

$$\boldsymbol{D}_{\mathrm{mid}}=\left\{\frac{D^L+\left(D^L+\frac{\mathrm{DL}}{k}\right)}{2},\frac{D^L+\frac{\mathrm{DL}}{n}+\left(D^L+\frac{2\mathrm{DL}}{k}\right)}{2},\cdots,\frac{D^L+(k-1)\frac{\mathrm{DL}}{k}+D^U}{2}\right\}$$

则 x+1 时刻偏差预测值算子为

$$D'_{(x+1)} = \boldsymbol{P}_{c(x+1)}\boldsymbol{D}_{\mathrm{mid}} \tag{19}$$

x+1 时刻预测值修正结果为

$$Z_{f(x+1)} = Z^1_{(x+1)} - D'_{(x+1)} \tag{20}$$

其中，$Z^1_{(x+1)}$为基于第 4 节网络安全态势预测方法获取的初始预测值。

5.3　偏差检验

分析修正后的安全态势预测值与实际值的接近程度，从而判断偏差类别划分数量 k 是否足够。已知某域空间内的修正后的安全态势预测值序列与实际值序列如表 4 所示。

表 4　　预测值序列与实际值序列

预测值	实际值
$Z_{f(2)}$	Z_2
$Z_{f(3)}$	Z_3
⋮	⋮
$Z_{f(l)}$	Z_l

本文偏差检验方法介绍如下。

1) 后验差检验

残差 R_i=Z_i–$Z_{f(i)}$，i=2,3,⋯,l 为实际值和经修正的预测值之差。当前安全态势序列片段内安全态势值方差 S_1^2 为

$$S_1^2 = \frac{1}{l}\sum_{i=2}^{l}\left(Z_i - \frac{1}{l}\sum_{i=2}^{l}Z_i\right)^2 \tag{21}$$

残差序列方差 S_2^2 为

$$S_2^2 = \frac{1}{l}\sum_{i=2}^{l}\left(R_i - \frac{1}{l}\sum_{i=2}^{l}R_i\right)^2 \tag{22}$$

则后验差比值 θ=$\dfrac{S_2}{S_1}$，该值越大，表明本文预测模型偏差越大，预测精度越不理想。

2) 小概率检验

小概率检验结果 τ为

$$\tau = P\left(\left|R_i - \frac{1}{l}\sum_{i=2}^{l}R_i\right| < 0.674\,5S_1\right) \tag{23}$$

偏差检验等级如表 5 所示。通过表 5 判断是否需增加偏差类别划分数量。若偏差等级为 1 级或 2 级，满足偏差等级检验要求，不需增加偏差类别数量，否则偏差类别数量为 k+1。偏差等级小，表明预测的态势值结果偏差小。

6　实验与结果

6.1　实验场景

采用林肯实验室的标准数据集 LL_DOS_1.0 验证本文模型的预测有效性。LL_DOS_1.0 攻击过程如下。

表 5　　偏差检验等级

偏差等级	τ	θ
4	(0.65, 0.70]	[0.65, 0.70)
3	(0.70, 0.80]	[0.50, 0.65)
2	(0.80, 0.95]	[0.35, 0.50)
1	(0.95, 1]	[0, 0.35)

1~70 min：攻击者安装相关攻击软件，并通过 IP Sweep 扫描实验网络拓扑以寻找当前活跃主机。

70~125 min：利用 Sadmind Ping 查找存在 Sadmind 漏洞的主机。

126~240 min：攻击者利用 Sadmind Exploit 攻击经 70~125 min 锁定的 3 台主机 Pascal、Mill 和 Locke 直至入侵各主机系统。

241~319 min：攻击者在受到入侵的 3 台主机上安装 DDoS 木马程序。

320 min 以后：攻击者对远程服务器发动 DDoS 攻击。

6.2　数据处理

在 Ubuntu16.04 操作系统下，采用 Tcpreplay 技术重放 LL_DOS_1.0 数据分组，并在 Windows10 操作系统下通过 Snort 入侵检测系统针对重放流量生成告警日志。

基于第 3 节网络安全态势评估量化方法生成态势实际值序列。将 1~360 min 按时间间隔 T=4 min 划分为 90 个量化周期，各周期内的态势量化值区间为[2 800, 4 000]。初始态势值序列由 1~40 min 的 10 个量化值组成，通过比较剩余的 80 个态势实际值与对应预测值拟合程度验证本文模型有效性。

以 41~360 min 某 40 min 时间段内的 10 个安全态势值为例，说明本文模型预测过程。

某量化周期 T=4 min 内的优选警报 Q、警报负面程度 AN_Q 级别、警报发生率 AO_Q 级别和警报致变程度 AM_Q 级别的属性如表 6 所示。

表 6　　警报属性样例

Q	AO_Q	AM_Q	AN_Q
敏感数据电子邮件地址	0.25	3	2

依据式(2)及表 1~表 3，得到评价关联度，如表 7 所示。

表 7 评价关联度

指标	高危	中危	低危
AN_Q	0.25	0.5	0.25
AM_Q	0.5	0.333	0.167
AO_Q	0.22	0.37	0.41

由式(2)~式(6)计算可得，该周期的网络安全态势量化值为 Z=3 504，其他量化周期的量化过程在此不再赘述。则得到该 40 min 时段内的 10 个安全态势序列如表 8 所示。

表 8 安全态势序列

T_i	Z_i
T_1	3 504
T_2	3 485
T_3	3 285
T_4	2 919
T_5	3 582
T_6	3 306
T_7	2 921
T_8	3 070
T_9	3 321
T_{10}	2 926

6.3 安全态势值的预测与修正

由第 4 节网络安全态势预测方法获取 T_2~T_{10} 的安全态势预测值。设可变域空间宽度为 W=10，得当前态势值序列下静态修匀系数自适应解 α_a=0.126，初始预测值与实际值对比如表 9 所示，经计算得 Z_{11}^1=3 115。

表 9 初始预测值与实际值对比

T_i	Z_i	Z_{i1}^1
T_1	3 504	—
T_2	3 485	3 454.7
T_3	3 285	3 469.9
T_4	2 919	3 405.4
T_5	3 582	3 218.3
T_6	3 306	3 329.3
T_7	2 921	3 310.1
T_8	3 070	3 151.2
T_9	3 321	3 089.7
T_{10}	2 926	3 141.2

T_2~T_{10} 周期偏差序列如表 10 所示，偏差区间 $[D^L,D^U]$=[−363.7, 486.4]。根据第 5 节的预测值修正方法，当划分偏差类别数量 k=8、φ=4 时，初始预测值修正后，可满足后验差检验与小概率检验条件。偏差类别区间如表 11 所示。

表 10 偏差序列

d_i	Z_i
d_2	−30.3
d_3	184.9
d_4	486.4
d_5	−363.7
d_6	23.3
d_7	389.1
d_8	81.2
d_9	−231.3
d_{10}	215.2

表 11 偏差类别区间划分

偏差类别 i	偏差值区间
1	[−363.7, −257.4)
2	[−257.4, −151.1)
3	[−151.1, −44.8)
4	[−44.8, 61.5)
5	[61.5, 167.8)
6	[167.8, 274.1)
7	[274.1, 380.4)
8	[380.4, 486.4)

由式(14)~式(19)计算得 T_2~T_{10} 修正后的安全态势预测值，其中偏差类别初始概率分布向量由在 T_1 周期之前的 10 个量化周期的安全态势值确定。安全态势预测修正值与实际值对比如表 12 所示。

表 12 预测修正值与实际值对比

T_i	Z_i	$Z_{f(i)}$
T_2	3 485	3 467
T_3	3 285	3 302.1
T_4	2 919	3 142
T_5	3 582	3 443.9
T_6	3 306	3 320.6
T_7	2 921	3 036
T_8	3 070	3 099.7
T_9	3 321	3 239.6
T_{10}	2 926	2 993.3

预测修正值 $Z_{f(i)}$序列的后验差比值 θ=0.42，小概率检验结果 τ=0.89。由表 5 可知，该模型偏差等

级为 2 级，满足偏差检验条件，则 T_{11} 的安全态势预测值为 $Z_{f(11)}=Z^{1}_{11}-D'_{(11)}$=3 115−141.6=2 973.4，与原态势序列中同周期的安全态势值 Z_{11}=2 920 的相对偏差为 1.8%，表明该预测精度较高。

对于其他周期安全态势值的预测和偏差检验，重复 6.2 节和 6.3 节过程，共计生成 80 个安全态势预测值。

6.4　最佳可变域空间宽度选取

静态修匀系数α自适应解受可变域空间内以往数据序列片段长度影响，改变初始预测精度。本文选取最佳可变域空间宽度以提高初始预测值和实际值的拟合程度。由于多重系数修匀法单次预测精度在 15 个以往数据以内较高，故本文试用可变域空间宽度集合为{W|W=5, 10, 15}，不同域空间宽度下预测值对比如图 3 所示。

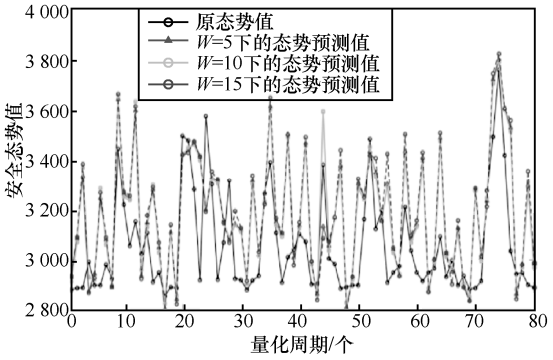


图 3　不同域空间宽度下预测值对比

由图 3 可得，最佳可变域空间宽度为 W=10，此时初始预测值精度更高。原因分析如下。

1) 当 W=5 时，域空间宽度较小，以往样本数据片较短，最近样本数据的影响更加显著。

2) 当 W=10 时，可变域空间宽度居中，以往样本数据段内异常波动数据和平缓波动数据的数量差距减小，远期、近期数据均衡影响安全态势预测，从而提高了初始预测值精度。

3) 当 W=15 时，域空间宽度大，以往样本数据片较长，域空间内少量异常波动数据和其他平缓波动数据相比，对静态修匀系数自适应解影响作用更小，降低了态势突变处的初始预测值精度。

6.5　态势预测对比实验与分析

实验数据集为 LL_DOS_1.0 数据集，分别采用本文模型、IFS-NARX 模型[9]和传统马尔可夫模型生成安全态势预测值序列，如图 4 所示，安全态势预测值绝对偏差序列如图 5 所示。

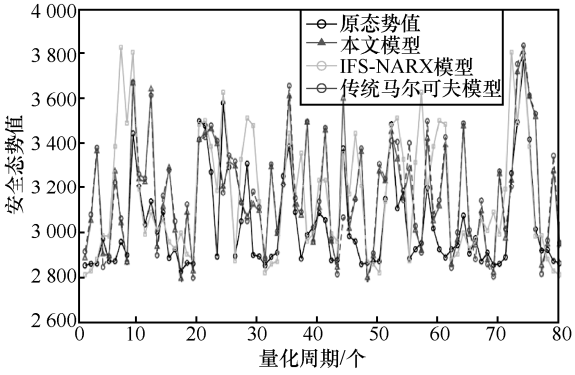


图 4　安全态势预测值序列

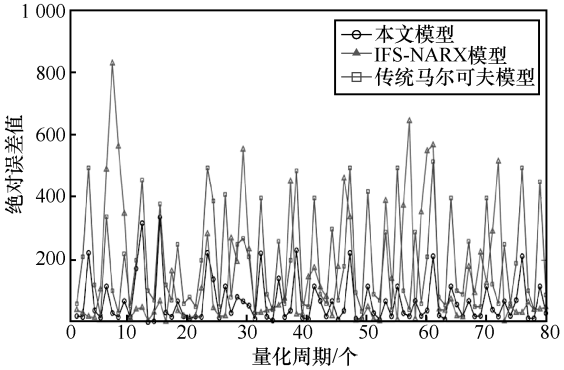


图 5　安全态势预测值绝对偏差序列

从图 4 和图 5 可知，由本文模型获取的态势预测结果更加符合原始的网络安全态势情况，绝对偏差更小。原因分析如下。

1) 传统马尔可夫状态转移概率矩阵随以往数据增加而收敛，故传统马尔可夫模型面向较短时间序列预测效果理想，当时间序列较长时，绝对偏差增大且偏差峰值周期性出现。

2) 将警报发生率、警报致变程度和警报负面程度作为 IFS-NARX 模型输入特征，非线性自回归神经网络参数由经验公式确定，该模型面向短序列预测因样本数量较少而预测精度不佳，当序列长度增加时，样本数量提高，模型预测精度提升。

3) 本文模型中，可变域空间位置随新态势值加入以往序列而发生移动，更新域空间内态势序列片段，调整静态修匀系数自适应解取值、偏差类别划分数量和偏差类别转移概率矩阵，使预测精度在不同长度时间序列下保持较高水平。

7　结束语

本文提出一种基于自修正系数修匀法的网络安全态势预测模型。通过熵关联度量化若干周期的网络安全态势值，采用可变域空间机制对按时序排

列的安全态势值进行片段化处理，运用自适应多重系数修匀法初步生成安全态势预测结果，运用时变加权马尔可夫链对偏差进行预测并修正安全态势预测值。实验结果表明，本文模型预测自适应性较强，预测精度较高。下一步主要分析态势序列线性性质并将长短期记忆网络模型和动态信誉机制[15]、安全规则集合[16]相结合，以提高本文模型对态势突变处的适应性。

参考文献：

[1] LEAU Y B, MANICKAM S. Network security situation prediction: a review and discussion[J]. Communications in Computer & Information Science, 2015, 516: 424-435.

[2] CIPRIANO C, ZAND A, HOUMANSADR A, et al. Nexat: a history-based approach to predict attacker actions[C]//Proceedings of the 27th Annual Computer Security Applications Conference. New York: ACM Press, 2011: 383-392.

[3] XIAO P, XIAN M, WANG H M. Network security situation prediction method based on MEA-BP[C]//3rd International Conference on Computational Intelligence & Communication Technology. Piscataway: IEEE Press, 2017: 1-5.

[4] SUN S X. The research of the network security situation prediction mechanism based on the complex network[C]//International Conference on Computational Intelligence and Communication Networks. Piscataway: IEEE Press, 2015: 1183-1187.

[5] LEAU Y B, KHUDHER A A, MANICKAM S, et al. An adaptive assessment and prediction mechanism in network security situation awareness[J]. Journal of Computer Science, 2017, 13(5): 114-129.

[6] SCHATZ D, BASHROUSH R. Security predictions-a way to reduce uncertainty[J]. Journal of Information Security and Applications, 2019, 45: 107-116.

[7] 孙卫喜, 孙欢. 网络安全态势预测技术研究[J].计算机技术与发展, 2019, 29(4): 100-104.
SUN W X, SUN H. Research on network security situation prediction technology[J]. Computer Technology and Development, 2019, 29(4): 100-104.

[8] 周新卫, 李小玲. 基于改进 G-K 算法的多节点网络安全态势预测模型[J]. 科学技术与工程, 2018, 18(25): 72-77.
ZHOU X W, LI X L. Multi node network security situation prediction model based on improved G-K algorithm[J]. Science Technology and Engineering, 2018, 18(25): 72-77.

[9] 韩晓露, 刘云, 张振江, 等. 基于 IFS-NARX 模型的网络安全态势预测[J]. 吉林大学学报(工学版), 2019, 49(2): 592-598.
HAN X L, LIU Y, ZHANG Z J, et al. Network security situation prediction method based on IFS-NARX model[J]. Journal of Jilin University (Engineering and Technology Edition), 2019, 49(2): 592-598.

[10] 席荣荣, 云晓春, 张永铮. 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.
XI R R, YUN X C, ZHANG Y Z, et al. An improved quantitative evaluation method for network security[J]. Chinese Journal of Computers, 2015, 38(4): 749-758.

[11] DEBAR H, WESPI A. Aggregation and correlation of intrusion-detection alerts[C]// International Symposium on Recent Advances in Intrusion Detection. Berlin: Springer, 2001: 85-103.

[12] 赵冬梅, 张玉清, 马建峰. 熵权系数法应用于网络安全的模糊风险评估[J]. 计算机工程, 2004, 30(18): 21-23.
ZHAO D M, ZHANG Y Q, MA J F. Fuzzy risk assessment of entropy-weight coefficient method applied in network security[J]. Computer Engineering, 2004, 30(18): 21-23.

[13] 付钰, 吴晓平, 叶清. 基于模糊集与熵权理论的信息系统安全风险评估研究[J]. 电子学报, 2010, 38(7): 1489-1494.
FU Y, WU X P, YE Q. An approach for information systems security risk assessment on fuzzy set and entropy-weight[J]. Chinese Journal of Electronics, 2010, 38(7): 1489-1494.

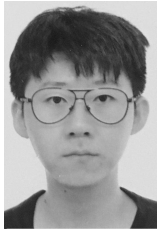
[14] 王笑, 戚湧, 李千目. 基于时变加权马尔可夫链的网络异常检测模型[J]. 计算机科学, 2017,44(9): 136-141, 161.
WANG X, QI Y, LI Q M. Network anomaly detection model based on time-varying weighted Markov chain[J]. Computer Science, 2017, 44(9): 136-141, 161.

[15] 杨宏宇, 韩越. 基于动态信誉的无线 Mesh 网络安全路由机制[J]. 通信学报, 2019, 40(4): 195-201.
YANG H Y, HAN Y. Wireless Mesh network secure routing mechanism based on dynamic reputation[J]. Journal on Communications, 2019, 40(4): 195-201.

[16] 杨宏宇, 王在明. Android 共谋攻击检测模型[J]. 通信学报, 2018, 39(6): 27-36.
YANG H Y, WANG Z M. Android collusion attack detection model[J]. Journal on Communications, 2018, 39(6): 27-36.

[作者简介]

杨宏宇（1969– ），男，吉林长春人，博士，中国民航大学教授，主要研究方向为网络信息安全。



张旭高（1993– ），男，山东威海人，中国民航大学硕士生，主要研究方向为网络信息安全。