引用格式:王劲松,吴天昊,朱兴奎,等. 基于NH-DBNs的网络空间态势预测[J]. 电光与控制,2019,26(2):44-48. WANG J S, WU T H, ZHU X K, et al. NH-DBNs based cyberspace state prediction[J]. Electronics Optics & Control, 2019, 26(2):44-48.

基于NH-DBNs的网络空间态势预测

王劲松[1], 吴天昊[1,2], 朱兴奎[1], 颜文琦[1]
(1. 战略支援部队信息工程大学,郑州 450001; 2. 中国人民解放军32553部队,海口 570100)

摘 要: 结合网络空间态势研判的实际问题,对网络态势变量预测进行研究,提出基于NH-DBNs模型的网络态势预测方法,以解决现有模型单时序回归和忽略多变量之间相互影响的问题。给出了网络态势的呈现方式和态势预测的方法与流程。在同一个环境下进行对比实验,证明该算法符合实际应用环境,且能够提升网络空间态势指标预测的精确性,对辅助指挥者科学、合理、精确决策具有一定参考价值。

关键词: 网络空间; 网络态势; 态势预测; NH-DBNs; 态势研判; 辅助决策
中图分类号: TN956; O221.6 **文献标志码**: A **doi**:10.3969/j.issn.1671-637X.2019.02.009

NH-DBNs Based Cyberspace State Prediction

WANG Jin-song[1], WU Tian-hao[1,2], ZHU Xing-kui[1], YAN Wen-qi[1]
(1. Information Engineering University of Strategy Support Army, Zhengzhou 450001, China;
2. No. 32553 Unit of PLA, Haikou 570100, China)

Abstract: To address the practical problems in the prediction and assessment of cyberspace state, this paper studies the prediction of variables in cyberspace, and presents a state prediction method based on Non-Homogeneous Dynamic Bayesian Networks (NH-DBNs), so as to solve the problems in the existing model of single time-sequence regression and the ignorance of the inter-influence between multiple variables. The presenting form of the network state is given, and the method and process of state prediction are presented. The comparison experiment conducted in the same environment has proved that the algorithm accords with the practical application environment, and can improve the prediction accuracy of the indexes of the cyberspace state. The algorithm has certain value for the conductor to make scientific, reasonable and accurate decisions.

Key words: cyberspace; cyberspace state; state prediction; NH-DBNs; state assessment; auxiliary decision-making

0 引言

在网络空间对抗中,对网络战场态势的精确预测与研判是网络空间下指挥决策的基石。网络空间是信息环境中的一个全球域,具有高度抽象性、层次性、拓扑性、瞬时性等特性,其网络结构多样复杂,信息数据动态多变,对未知不确定网络态势的完全掌控显然难以实现,但现阶段对非稳态时序数据的研究已经较为成熟,可以通过对先验信息的训练学习,实现对网络态势走向趋势进行概率上的分析与预测[1]。

目前,经典的稳态时间序列预测模型有AR(p),MA(q)和ARMA(p,q)等模型,非稳态时间序列有GLM广义回归模型、*ARIMA*(p,d,q)等模型。以上模型的基本思想是通过建立数个随机序列,以达到序列方差最小化的目的来构造模型,并结合先验信息来预测未来数据走向。然而,在上述预测模型中,$AR(p)$和MA(q)无法对动态网络态势进行预测,ARIMA是针对动态变量的预测模型,且已广泛应用于股市趋势的预测,但ARIMA属于单变量模型,仅考虑目标变量的变化,忽略其他变量对目标变量的影响[2],在网络空间态势研判中显然不能满足精确预测的要求。本文针对网络空间对抗特点建立网络模型,提出一种非同构动态贝叶斯网络模型(Non-homogeneous Dynamic Bayesian Networks,NH-DBNs)对网络空间态势中动态变量进行预测判断,并通过仿真实例验证该方法提升网络态势预测精度的可行性。

收稿日期:2018-03-27 **修回日期**:2018-04-14
基金项目:军事科学研究基金项目(2016605907)
作者简介:王劲松(1965—),男,河南驻马店人,硕士,教授,研究方向为信息作战指挥理论与应用。

1 经典非平稳时间序列模型

为验证NH-DBNs算法的优势,本文采用ARIMA

模型、传统动态贝叶斯网络模型(DBN)和 NH-DBNs 模型在同一个网络拓扑网络及环境下进行对比仿真实验,由于前两者在指标预测及目标选择上应用较为广泛,且与本文提出的模型有相似的算法基础,本文对 ARIMA 和 DBN 两种算法原理进行简要说明。

1.1 ARIMA 模型

ARIMA 模型的全称为自回归积分滑动平均模型(Autoregressive Integrated Moving Average Model),是一类以差分操作为依据的非平稳时间序列模型,其本质是 1 阶差分操作的自回归过程,用 $t-1$ 时刻的数据作为自变量来拟合当前序列值[2]。对于非平稳时间序列在差分操作之后显示出平稳状态的现实问题,可以选择采用 ARIMA 模型进行建模仿真。

差分操作即 t 时刻与 $t-1$ 时刻指标参数差,用 y 表示该时刻的指标值,差分操作的 1 阶和 d 阶表达式分别为

$$\nabla y_t = y_t - y_{t-1} \tag{1}$$

$$\nabla y_t = y_t - y_{t-d}\circ \tag{2}$$

以 p,d,q 为参数的 ARIMA 模型表达式为

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{3}$$

式中:p 为自回归的阶数;d 为序列差分的阶数;q 为移动平均阶数;y_t 为时间 t 时的态势值;e_t 为白噪声序列;φ_i 和 θ_i 分别为 y_{t-1} 和 e_{t-1} 的系数[3]。

ARIMA 模型运行流程如下:

1) 首先对态势变量时间序列进行平稳性检验并持续差分操作,直到若干次差分后序列满足平稳性检验,进入 2);

2) 确定模型参数 p,d,q,以 AIC 信息准则限定 p,q 范围,遍历 (p,q) 组合并找出具有最小 AIC 值的 (p,q) 值;

3) 模型残差检验,确保模型服从正态分布的白噪声序列,当模型残差符合白噪声序列即可将序列的信息充分提取到模型中;

4) 将 2) 中确定的最优参数应用于 ARIMA 模型进行预测,得到置信区间。

1.2 DBN 模型

动态贝叶斯网络包含两个设定:一是 1 阶马尔可夫假设,即各节点之间不能够跨越时间片,那么 t 时刻节点的取值仅受到 $t-1$ 时刻及以外节点取值的影响;二是同构性假设,即长度为 T 的时间序列限定于同一条马尔可夫链生成的稳定分布,其模型结构及参数无法随时间推移而改变[4]。

DBN 模型在处理离散不确定变量时较有优势,它由两个贝叶斯网络构成,可定义为 $<B_0, B\rightarrow>$,其中,B_0 为先验网络,包含定义在初始状态 y_0 的联合概率分布,$B\rightarrow$ 为转移网络,包含定义在变量 y_t 与 y_{t+1} 上的转移概率 $P(y_{t+1}/y_t)$,可得 $<B_0, B\rightarrow>$ 组成的 DBN 模型是对应于动态变量上的有初始值而半无限的网络结构。若只考虑一个有限的时间段 $1,2,\cdots,T$,可将 DBN 模型限定在 $y_1, y_2, \cdots, y_t$ 上的一个网络结构[5]。

2 基于转换点生成的 NH-DBNs 算法

2.1 NH-DBNs 基本思想

文献[6 - 10]中均有突破贝叶斯网络同构的设想。在此基础上,提出将多转换点及转换点跳转与生成引入 DBN 建模过程。

NH-DBNs 模型的基本思想是在传统同构 DBN 模型的基础上,将时序数据 T 划分为若干个不相交的时段 $T_1, T_2, \cdots, T_k$,k 为划分的时段数。在设定的若干个时段内,均用一个特定的 DBN 模型来仿真时序数据。图 1 为具有 k 个转换点的非同构动态贝叶斯网络,图中圆节点及连接线代表各变量因子的相互作用关系。

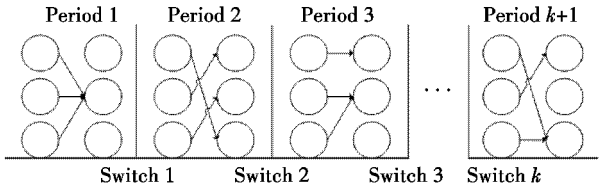

图 1　转换点分割的非同构动态贝叶斯网络

Fig. 1　The NH-DBNs of intersected switches

该模型的优势是,在整个时序内可以忽略各时段之内的指标联系,仅需考虑时段之间的联系,即默认 t 时刻态势变量的值仅受 $t-1$ 时刻节点取值的影响,而每个时段之内的时序数据均建立 DBN 模型,且均满足 1 阶马尔可夫假设。

2.2 NH-DBNs 模型的贝叶斯回归模型

利用贝叶斯回归模型,对 NH-DBNs 模型的每个时段内的时序数据进行建模。在时段 $h(h=1,2,\cdots,k+1)$ 的某一个节点的预测指标值 $y(t)$ 由 $t-1$ 时刻的指标值和回归模型决定,其表达式为

$$y_i(t) = a_{i0}^h + \sum_{j\in M_i^h} a_{ij}^h y_j(t-1) + \varepsilon_i(t) \tag{4}$$

式中:$i=1,2,\cdots,n$,n 为 h 时段内节点数;a_{i0}^h 和 a_{ij}^h 为回归系数;M_i^h 为 h 时段内节点 i 的父节点;$\varepsilon_i(t)$ 为回归模型的高斯噪声。

2.3 时序中转换点的生成

针对网络空间态势预测的实际问题,重点应用的是时间序列中转换点的生成,转换点的生成也是 NH-DBNs 算法最核心的关键环节。

基于窗口均值先验的转换点生成动作:设在某时段节点 i 左右窗口数据均值的欧氏距离为 $f(i)$,即被采样到的转换点的先验值,ξ 为先前转换点,新生成的

转换点为 ξ^+，用 $q_{bk}(\xi,\xi^+)$ 表示产生新转换点的提议概率[11]。若产生新的转换点，则新的被采样到的转换点的先验值一定正比于 $q(\xi,\xi^+)$，在不考虑消除与转换动作的情况下其表达式为[4]

$$q(\xi,\xi^+) = \frac{f(P_b)}{\sum_{P=1} f(P)} \tag{5}$$

式中，P_b 为新的被采样到的转换点。

生成动作的流程步骤为：

1）从正在采样的时段内非转换点的集合 U 中采样一个时间点作为新的转换点；

2）将选择的转换点一侧作为新的时间阶段，暂停旧阶段的模型运行，建立新的阶段采样模型；

3）由式(4)和式(5)计算此次跳转的提议概率 $q(\xi,\xi^+)$；

4）设定阈值 w，且 $w \in [0,1]$；

5）若 $w \leqslant q(\xi,\xi^+)$，接受此次跳转，生成新转换点并存储至转换点集合，否则不跳转，保持原状态并重新开始旧阶段模型。

其算法表达式为：

1）初始化转换点的位置 ξ，转换点个数 k，模型 M 以及其他参数；

2）$n=0$，$flag=0$；

3）while $(n \leqslant N \| flag==0)$ do；

4）根据现有转换点个数计算 $Birth_k$，$Birth_k$ 为生成转换点的概率；

5）采样 $w \sim Uniform[0,1]$；

6）if $\{w < Birth_k\}$ then 生成新的转换点 switch $k+1$；

7）else 更新转换点之前时段内模型；

8）end if 将采样样本加入 result 集合；

9）$n=n+1$；

10）return result。

转换点生成动作的流程如图 2 所示。

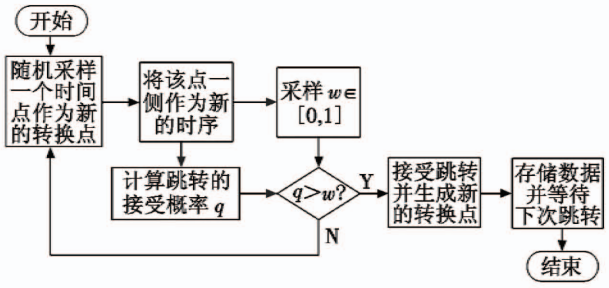

图 2 转换点生成动作的流程

Fig. 2 Flow chart of transformation point generating

3 网络空间下态势预测的 NH-DBNs 模型构建

3.1 网络态势呈现

网络空间态势是网络战场中基于网络作战过程产生影响的各个要素的运行状态和发展趋势[1]，而网络空间下态势预测与判断的前提是分析厘清各态势要素及各要素之间的影响关系。先验指标信息训练学习阶段的有效性、指导性，同时也要求各态势指标的划分清晰可靠。本文介绍两种可行的网络结构分析方式。

1）网络拓扑结构呈现。网络拓扑结构是利用信息流等传输媒介连接各网络节点的表示方法。网络拓扑模型能清晰地表示网络态势各节点间的相对位置关系、连接关系。基本的网络拓扑结构有星型网络结构、环型网络结构、风筝网络结构、树型网络结构和蜂窝网络结构等，网络拓扑结构就是由各种不同类型的基本拓扑结构组合形成的。

图 3 为一类网络拓扑结构，从图中可获取态势信息集合，$S=\{$网络环境，网络连通性，信息流量，传输速率，通信协议和网络服务脆弱性，端口数据稳定性，Mac 地址信息，操作系统信息，目标网络破解难度，获取目标系统权限，无线网络可探测性，目标安全策略，…$\}$。具体的态势指标要根据网络作战需求进行选择，在某些操作参数的选取上，也可以通过设定阈值进行筛选。

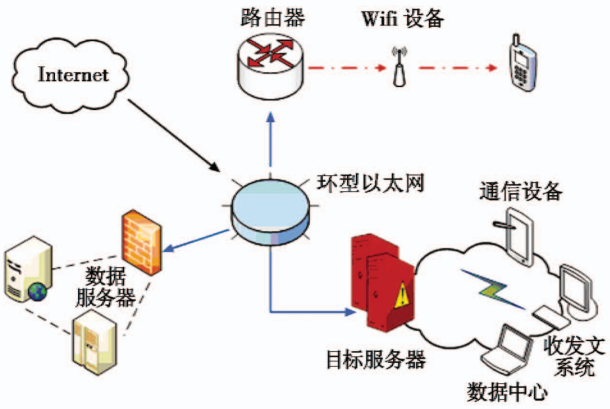

图 3 网络拓扑结构

Fig. 3 Network topology

网络拓扑结构呈现的指标参数多为并列式，如图 4 所示。图中所呈现态势为信息流量(flo)、传输速率(rat)、地址信息(mac)。

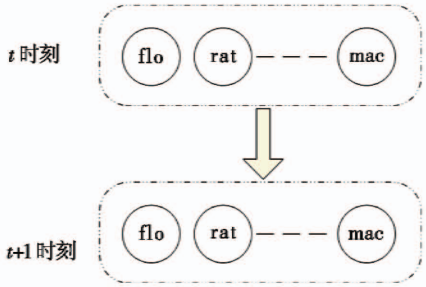

图 4 并列式态势呈现

Fig. 4 The juxtaposing state presentation

2）指标分级结构呈现。指标分级结构呈现一般

适用于具有从属性的网络态势指标，在不考虑指标间的位置和连接关系的情况下，这种方式能更加完整地表现指标间的相互作用关系，如图 5 所示。图中列举了一个网络目标防护的指标分级结构，指标信息为受威胁程度(thr)、目标安全性(sec)、网络风险性(ris)、目标隐蔽性(dis)、服务器抗毁性(sur)和防嗅探系数(sni)[5]。

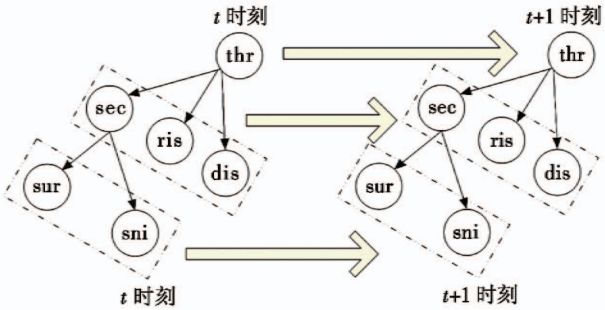

图 5　分级式态势呈现

Fig. 5　The stagewise state presentation

3.2　态势预测方法与流程

将网络空间态势预测问题抽象为一个分层问题，即输入当前 t 时刻的指标变量，输出 $t+1$ 时刻指标变量值。针对该问题，基于 NH-DBNs 模型构建网络空间下态势预测主要包括 4 个阶段：1）先验数据的分析处理阶段；2）回归模型的训练学习阶段；3）指标变量的导入阶段；4）指标变量的预测阶段。其框架如图 6 所示。

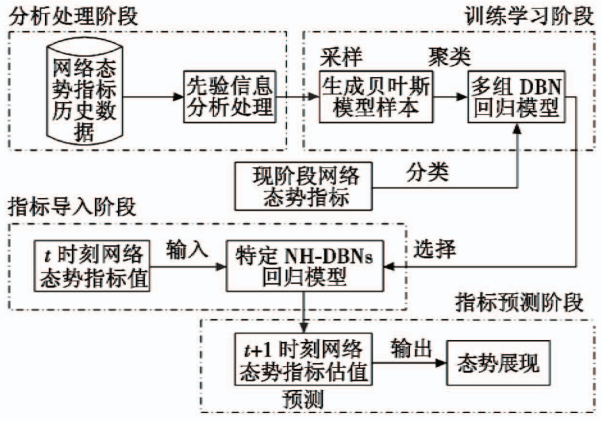

图 6　基于 NH-DBNs 的网络态势预测方法框架图

Fig. 6　Network state prediction frame chart based on NH-DBNs

4　实例仿真

为验证算法的有效性与精确性，假设在图 3 所示的网络拓扑结构中无线局域网下预测带宽数据流量、脆弱性系数(安全系数)、数据冗余度 3 类相互之间具有作用关系的网络态势。给出采样数据[8,12]：

数据流量 = [4,19,23.5,13,12,6.5,4.5,2.5,6.5,18,2,1,16,20.5,5,15.5,12,23,22.5,23.5,3.5,21,15.5,16.5,20,7,21.5,1]；

脆弱性系数 = [0.1,0.23,0.24,0.09,0.6]；

数据冗余度 = [0.028 571 4,0.557 143,0.5,0.2,0.557 143,0.185 714,0.385 714,0.4,0.671 429,0.1,0.442 857,0.442 857,0.142 857,0.042 857 1,0.542 857,0.042 857 1,0.3,0.7,0.171 429,0.571 429,0.414 286,0.614 286,0.3,0.6,0.5,0.257 143,0.157 143,0.6,0.314 286,0.285 714,0.7,0.542 857,0.157 143,0.671 429,0.271 429,0.157 143,0.528 571,0.057 142 9,0.342 857,0.7,0.414 286]；

设定以上采样数据均在周期为 100 的时间序列内，且稳定按时段分布。

4.1　实验结果展示

采用 DBN 模型、ARIMA 模型和 NH-DBNs 模型在同一个环境下(Matlab2010)进行对比仿真实验，针对上述 3 类网络态势变量的预测效果如图 7 所示。

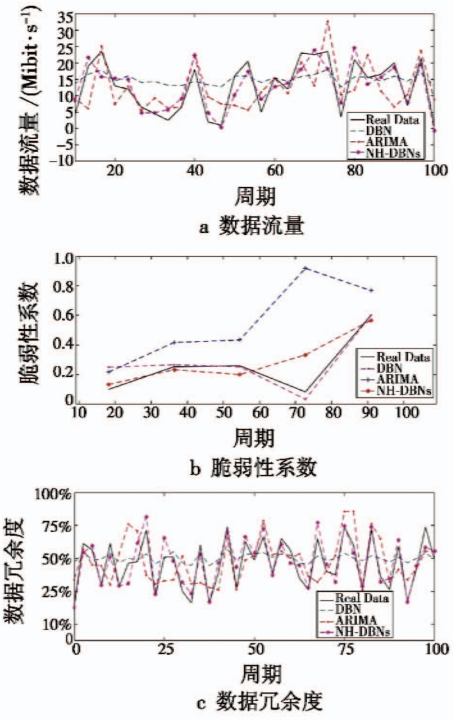

图 7　预测效果

Fig. 7　Prediction effect

由于对数据流量态势预测的对比相对明显，给出对该态势变量的平均绝对误差、准确率、精确率和召回率的对比，如表 1 所示。

表 1　预测结果对比

Table 1　The comparison of predicted results

模型	平均绝对误差	准确率	精确率	召回率
DBN	229.53	0.57	0.56	0.41
ARIMA	109.89	0.71	0.80	0.43
NH-DBNs	102.57	0.79	0.87	0.45

4.2　实验结果分析

由表 1 可看出，在数据流量指标预测中，NH-DBNs 模型较 DBN 和 ARIMA 模型在绝对误差上分别降低了

126.96 和 7.32，准确率分别提高了 0.22 和 0.08，精确率分别提高了 0.31 和 0.07，召回率分别提高了 0.04 和 0.02。从图 7 中可以看出，在脆弱性系数的变量预测上，给出的 3 类模型预测效果较为相似。DBN 模型在数据流量和冗余度的态势中，大致可预测变量变化趋势，但曲线的起伏度相对真实输出值显得过于平缓，ARIMA 模型与 NH-DBNs 模型跟真实的态势变量值较为吻合，而非同构动态贝叶斯回归模型较之 ARIMA 模型的 3 项指标变量的预测上均有更好的表现，其主要原因在于 DBN 模型回归模型是固定的，其时序不能随着时间推移而发生改变，ARIMA 模型为单变量模型，仅仅考察单变量的变化趋势，无法探测其他态势变量对预测变量的作用与影响。通过实验可以证明，采用 NH-DBNs 模型拟合效果更为精确，精度更高。

5 结束语

本文主要对网络空间下态势的呈现方式与预测进行了研究，在分析比较现有预测方法的基础上，提出基于 NH-DBNs 模型的网络态势预测方法，该方法突破传统动态贝叶斯在同一个时序下回归的方法，将多转换点引入建模过程，最后利用现有数据集运用 DBN 和 ARIMA 模型进行对比实验。结果表明，本文提出的方法能够应用于网络态势的预测，并可以提高态势的预测精确性，对于指挥者针对网络态势精确决策具有辅助作用和参考意义。

在实际应用中，该网络态势预测方法有两方面掣肘因素：1）预测结果受提取的网络态势指标参数值的限制；2）本文数据来自文献现有数据，在网络态势参数值的提取方面还未能有效解决。下一步研究将针对参数提取量化、运行速率方面加以完善，并提出随机参数态势预测的方法。

参 考 文 献

[1] 纪浩然. 网络作战态势生成和展现研究[D]. 长沙：国防科学技术大学，2011.

[2] 俞露. 基于非同构动态贝叶斯网络的研究与应用[D]. 南京：南京大学，2017.

[3] 孙建树，娄渊胜，陈裕俊. 基于 ARIMA-SVR 的水文时间序列异常值检测[J]. 计算机与数字工程，2018，46(2)：225-229.

[4] 俞露，高阳，史颖欢. 基于滑动窗口均值先验的非同构动态贝叶斯网络转换点检测算法[J]. 模式识别与人工智能，2016，29(8)：751-761.

[5] 李冯敬，姚佩阳，田晓飞，等. 动态贝叶斯网络在通信对抗目标选择中的应用[J]. 电光与控制，2012，19(8)：63-70.

[6] TALIH M, HENGARTNER N. Structural learning with time-varying components: tracking the cross-section of financial time series[J]. Journal of the Royal Statistical Society(Statistical Methodology), 2005, 67(3): 321-341.

[7] NIELSEN S H, NIELSEN T D. Adapting Bayes network structures to non-stationary domains[J]. International Journal of Approximate Reasoning, 2008, 49(2): 379-397.

[8] ROBINSON J W, HARTEMINK A J. Learning non-stationary dynamic Bayesian networks[J]. Journal of Machine Learning Research, 2010, 11: 3647-3680.

[9] GRZEGORCZYK M, HUSMEIER D. Regularization of non-homogeneous dynamic Bayesian networks with global information-coupling based on hierarchical Bayesian models[J]. Machine Learning, 2013, 91(1): 105-154.

[10] DONDELINGER F, LEBRE S, HUSMEIER D. Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure[J]. Machine Learning, 2013, 90(2): 191-230.

[11] LEBRE S. Stochastic process analysis for genomics and dynamic Bayesian networks inference[D]. Strasbourg: University of Strasbourg, 2007.

[12] 刘玉岭，冯登国，连一峰，等. 基于时空维度分析的网络安全态势预测方法[J]. 计算机研究与发展，2014，51(8)：1681-1694.