

基于动态转移概率的网络态势预测方法研究

李剑蓝

(中国石油大学(华东) 计算机与通信工程学院, 山东 青岛 266580)

摘要:网络态势预测是网络态势感知领域的重要组成部分。鉴于目前网络状态变化莫测,网络攻击形式层出不穷,而现有的网络态势预测手段具有实时性差或者计算量繁重的特点,难以对网络态势进行准确有效的预测。对此,文中通过分析不同形式网络攻击下的网络状态变化特点,引入网络波动率的概念并用其描述网络波动状况,结合马尔可夫链提出一种基于动态转移概率的网络态势预测方法。该方法通过实时计算影响力衰减周期来计算转移概率,根据实际网络波动情况自适应更新转移概率的计算方式,从而动态地对网络态势进行预测,得到更准确有效的预测效果。对基于动态转移概率的网络态势预测模型进行了形式化描述并进行了原型实验,实验结果显示,该网络态势预测方法具有较高的准确性和可行性。

关键词:网络预测;动态转移概率;网络态势感知;马尔可夫链;网络安全

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2019)11-0062-05

doi:10.3969/j.issn.1673-629X.2019.11.013

Research on Method of Network Situation Prediction Based on Dynamic Transition Probability

LI Jian-lan

(School of Computer and Communication Engineering, China University of Petroleum (East China),
Qingdao 266580, China)

Abstract: Network situation prediction is an important part of network situation awareness. In view of the unpredictable changes of network state and the emergence of network attack forms, the existing network situation prediction methods have the characteristics of poor real-time performance or heavy computation, which makes it difficult to predict the network situation accurately and effectively. For this, we introduce the concept of network volatility to describe the network fluctuation by analyzing the characteristics of network state changes under different forms of network attacks. Combining with Markov chain, a network situation prediction method based on dynamic transition probability is proposed. This method calculates the transition probability by calculating the attenuation period of influence in real time, and adaptively updates the calculation method of the transition probability according to the actual network fluctuation, so as to dynamically predict the network situation and achieve more accurate and effective prediction effect. In this paper, the network situation prediction model based on dynamic transition probability is formally described and prototype experiments are carried out. The experiment shows that the network situation prediction method has high accuracy and feasibility.

Key words: network prediction; dynamic transition probability; network situation awareness; Markov chain; network security

0 引言

近年来随着互联网的发展,网络安全问题得到了高度关注。在这样的背景之下,网络态势感知成为了当前研究热点之一。

网络态势感知技术源于空中交通监管态势感知,并在1999年由Tim Bass提出^[1]。网络态势感知是指在现实网络环境下,获取、理解能够引起网络态势发生

变化的安全要素,并预测未来的发展趋势。网络态势预测作为网络态势感知的一部分,起着越发关键的作用,但是由于网络状态的变幻莫测和随机性,网络预测的有效性和准确性一直是一个关键问题。

结合实际网络状态的变化特点,文中提出一种基于动态转移概率的网络态势预测方法,使其能够更符合网络变化情况,自适应地对网络态势进行预测,以达

收稿日期:2018-12-16

修回日期:2019-04-17

网络出版时间:2019-06-26

基金项目:国家自然科学基金(61772551)

作者简介:李剑蓝(1993-),男,研究生,研究方向为信息安全和计算机网络。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190626.0840.066.html>

到更高的准确性。

1 相关工作

网络态势感知的概念由 Tim Bass 提出,他随即在 2000 年提出了基于多传感器数据融合的入侵检测框架,该框架能够实现入侵行为检测、入侵率计算、入侵者身份和入侵者行为识别、态势评估以及威胁评估等功能。Cristina Abad 等^[2]利用安全态势感知数据库系统 UCLog+设计了一个网络安全态势感知系统,该系统将当前采集的安全事件信息与数据库中的历史安全事件进行规则匹配,获取下一时刻安全事件出现的概率。文献[3-4]将聚类算法和隐马尔可夫模型(HMM)相结合,以分析最有可能的攻击序列和识别攻击的类型。

文献[5]提出了数据挖掘方法挖掘警报之间的因果关系。这些方法既无需预定义知识库也无需网络的配置信息,就可以发现攻击行为之间的因果关系和识别未知的攻击行为。Tang 等^[6]从被保护对象的角度出发分析攻击者的目的,提出了使用动态后向传播神经网络和协方差相结合的方法,基于当前每个主机的服务、攻击活动、服务重要性等分析可能要遭受攻击的服务。

网络态势预测作为网络态势感知的重要部分,近年来也引起了学者的广泛关注。文献[7]设计了一个基于 D-S 证据理论的态势评估架构,融合不确定信息进行不确定性推理,量化网络的安全态势。Wang 等^[8]通过统计的方法,统计攻击者需要多少个不同 0-day 漏洞才能对系统造成破坏,以评估网络的安全状况。Juan 等^[9]提出了将无偏灰度理论(unbiased grey theory)和马尔可夫理论相结合的方法,分析网络的风险变化情况。

层次决策分析(analytic hierarchy process)^[10]由 Satty 等提出,该方法将定性分析与定量分析相结合,是一种无结构的多准则决策方法,通过思维过程的层次化和数量化达到分析复杂问题的目的。文献[11]使用贝叶斯网络对网络中的“不确定因素”建模,计算攻击成功的概率,实时地评估攻击的严重程度。Aguilar 等^[12]结合模糊逻辑与神经网络技术,在认知图的基础上提出了 FCM 的概念,利用它获取网络中重要资产的依赖关系进行危害程度评估。文献[13-15]将时间序列分析技术与概率模型、数据挖掘等技术相结合来分析 DDos 攻击特征及行为变化。文献[16]在贝叶斯网络方法的基础上提高了方法的适应性,提出了伪贝叶斯网络方法对攻击行为进行识别预测。

在以上研究的基础上,文中结合网络的变化特点,利用动态改变转移概率的计算方式来自适应地调整网

络态势的预测值,达到快捷准确预测网络态势的目的。

2 基于动态转移概率的网络态势预测模型

2.1 马尔可夫链

对于随机变量序列 $X = \{X(t), t \in T\}, T = 0, 1, 2, \dots$, 其状态空间为 $S = \{0, 1, 2, \dots\}$ 。如果对于任意的正整数 m, n, k , 以及任意的状态值 $\{x_{n+k}, x_n, \dots, x_2, x_1\}$, 满足:

$$P\{X(n+k) = x_{n+k} | X(n) = x_n, \dots, X(2) = x_2, X(1) = x_1\} = P\{X(n+k) = x_{n+k} | X(n) = x_n\}$$

则称 X_r 为马尔可夫链。

其中 $P\{X(n) = x_n\}$ 表示系统在 n 时处于状态 x_n 的概率。若记系统在 n 时所在的状态为 i , 在 $n+k$ 时所在的状态为 j , 那么系统从状态 i 转移到状态 j 的条件概率 $P\{X(n+k)=j | X(n)=i\}$ 称为马尔可夫链的 k 步转移概率, 记为 $p_{ij}^{(k)}(n)$ 。特别地, 当 $k=1$ 时, 称之为马尔可夫链的转移概率, 通常记为 p_{ij} 。

由转移概率组成的矩阵称为马尔可夫链转移概率矩阵, 其形式可描述为:

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

此矩阵的每一行元素之和为 1。当转移概率只与状态变量 i, j 和时间间隔 n 有关, 称此转移概率具有平稳性, 同时称此链是时齐的。马尔可夫链一般被认为是时齐的。

马尔可夫链具有“无后效性”的特征, 即下一时刻的状态只与当前时刻状态有关, 与之前的时刻状态无关。

2.2 网络安全环境变化特点

在分析网络安全环境的变化特点之前, 将网络安全程度划分为 k 个等级, 即 $\{S_1, S_2, \dots, S_k\}$, 这 n 个等级分别代表不同程度的安全问题, 其中 S_1 表示没有安全风险, 并且安全风险 $S_1 < S_2 < \dots < S_k$ 。

网络攻击在实施之前一般都存在探测, 踩点, 扫描等过程, 这些异常行为会引起轻度的安全风险, 并且这些轻度的安全风险往往会意味着更严重的安全问题。以 web 攻击为例, 在最终获取后台权限之前, 会有探测主机 ip, 开放端口, 尝试注入等操作, 这些异常行为在当前周期可能会引起轻度的安全问题, 在下一周期可能会导致更加严重的安全问题。当攻击结束的时候, 风险程度又会慢慢降低, 所以网络攻击态势常常会出现如图 1 所示的状态。

图 1 是在实际 SQL 注入攻击实验下, 通过收集到的流量信息和入侵检测系统的警报信息, 将网络状态分为 3 个等级, 即 $k=3, S_1=0, S_2=1, S_3=2$, 并由专家

打分得到的安全风险程度图。

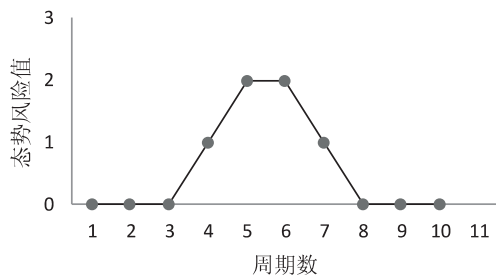


图 1 正常攻击流程下的网络态势变化

但是网络环境复杂,也不排除有风险突然爆发的情况,这种情况下,网络攻击态势常常会出现如图 2 所示的状态。

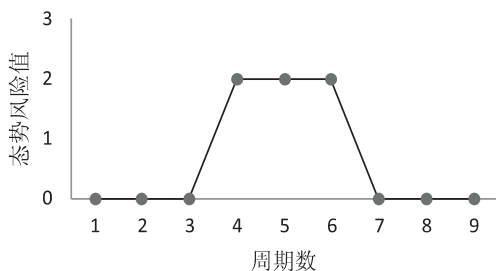


图 2 已知安全漏洞攻击下的网络态势变化

图 2 是在已知有安全漏洞的情况下,直接进行 SQL 注入攻击实验下,通过收集到的流量信息和入侵检测系统的警报信息,由专家打分得到的安全风险程度图。

由这两张图的趋势变化可以看到它们的网络态势变化情况是不一样的。文中用风险波动率 ∂ 来衡量网络风险的波动程度,风险波动率的计算公式为:

$$\partial = \frac{\text{count}_{\text{change}}}{\text{count}_{\text{all}}} \quad (1)$$

其中, $\text{count}_{\text{change}}$ 为参考周期内变化的网络风险等级的次数; $\text{count}_{\text{all}}$ 为周期内总的网络风险等级的次数。

2.3 基于动态转移概率的网络态势预测模型

马尔可夫链具有“无后效性”的特征,而网络状态的转变与当前状态是紧密关联的,这一点与马尔可夫链相符,但是网络状态的转移应该是随着网络状态的变化而动态改变的。随着网络状态的更新,转移概率也将随之自适应更新。所以文中提出一种基于动态转移概率的网络态势预测模型

假设有周期集合 $\{T_1, T_2, \dots, T_n\}$, 共有 k 个安全风险等级 $\{S_1, S_2, \dots, S_k\}$, 每个周期都有相应的网络风险等级对应 $\{I_1 \rightarrow T_1, I_2 \rightarrow T_2, \dots, I_n \rightarrow T_n\}$ ($I_1, I_2, \dots, I_n \in \{S_1, S_2, \dots, S_k\}$), 转移概率参考的周期集合为 $T_r = \{T_a, T_{a+1}, \dots, T_b\}$ ($m = b - a + 1$), 共 m 个周期。

转移概率的计算方法可描述为:

$$P_{ij} = \frac{S_{ij}}{S_{i1} + S_{i2} + \dots + S_{ik}} \quad (2)$$

其中, S_{ij} 表示参考的 m 个周期内, 状态从 S_i 到 S_j

的次数。由表达式可以看出, P_{ij} 是一个完全由参考周期集合 $\{T_a, T_{a+1}, \dots, T_b\}$ ($m = b - a + 1$) 的值来决定的, 即 $P_{ij} = f(T_r)$, 而 T_r 的值是一个动态变化的值。

文中认为以往的网络状态对下一周期的网络状态的影响是随时间衰落的, 即时间越接近的网络状态对下一网络状态的影响越大。而随时间衰落快慢的主要因素在于这一时期的网络波动情况, 波动越大, 说明网络变化幅度大, 随时间衰落的速度就越快。假设影响力因素为 μ , 波动率为 ∂ , 时间为从当前周期向前追溯, 并以周期为单位设为 t , 则变量满足方程:

$$\mu = e^{T - \&(\partial(t) + 0.1)m} - 1 \quad (3)$$

其中, T 为衡量波动率的总参考周期数; $\&$ 为一可变参数。 $\partial(t)$ 为波动率, 随时间变化。影响力因素为 $\mu = 0$ 时, 若已知 $\partial(t)$, $\&$, T 的值, 则将此时的 t 作为转移概率参考的周期数, 从而确定转移概率参考的周期集合 $T_r = \{T_a, T_{a+1}, \dots, T_b\}$ ($m = b - a + 1$)。

那么基于动态转移概率的网络态势预测模型可以描述为:

$$\begin{cases} S_d = S_c \begin{pmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \dots & p_{nn} \end{pmatrix} \\ P_{ij} = \frac{S_{ij}}{S_{i1} + S_{i2} + \dots + S_{ik}} \\ S_{ij} = \text{Count}(T_r) \end{cases} \quad (4)$$

其中, S_d 为预测的下一周期的网络态势值概率分布情况; S_c 为当前网络态势值所表示的矩阵。通过式 3 确定的 T_r 即可得到 S_d , 取概率最大的情况作为下一周期的网络态势预测值。

3 实验分析

针对基于动态转移概率的网络态势预测模型, 搭建了实验原型系统, 包括一个包含 SQL 注入漏洞的 Web 服务器和一个攻击机。攻击机在两天的时间内随机地发起攻击, 攻击类型包括扫描攻击、SQL 注入攻击和 DoS 攻击。服务器端装有抓包工具和 snort 入侵检测系统, 通过抓包工具获取 TCP, UDP, ICMP 的流量信息, 通过 snort 入侵检测系统获取警报信息, 以 10 秒为一个周期, 由专家打分得到 2 482 条网络风险评估值, 评估值分为三个等级 $\{0, 1, 2\}$, 分别代表无风险, 低风险和高风险。

(1) 参数 $\&$ 的确定。

当 $\mu = 0$ 时, $m = \lceil \frac{T}{\&(\partial(t) + 0.1)} \rceil$ 。若 T 周期内的波动率 $\partial(t)$ 已知, 那么由参数 $\&$ 决定 m 的值。文中

考虑波动率所参考的周期数 T 为 100 个周期,为了确定 $\&$ 的值,做了四组对比实验,分别是在波动率为 0, 0.3, 0.6, 0.9 的情况下, $\&$ 分别为 1, 5, 10, 20 下 50 次预测情况的准确率,见图 3。

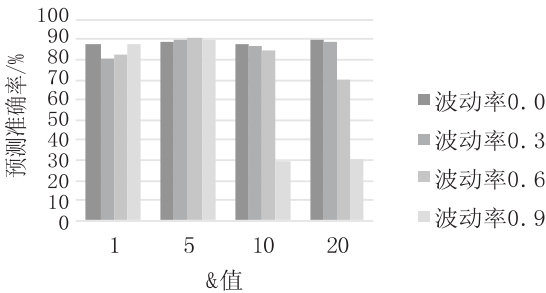


图 3 不同波动率下不同 $\&$ 值的准确率对比

由图 3 可以看到,当 $T = 100$ 的情况下, $\&$ 是不宜过大的,在波动率高的情况下, $\&$ 值过高会导致 m 的值为 1,那么转移概率计算的参考范围太小就失去了计算转移概率的意义,失去了参考的价值。

综合来看,在该实验网络环境下, $T = 100$ 时,选择 $\& = 5$ 比较合适。

(2) 网络态势预测性能分析。
针对获得的 2 482 条数据,从第 100 条开始往后进行预测,通过 Matlab 计算动态转移概率以及实际预

测值,进行了 2 382 次实验,得到准确率为 0.884,展现了不错的效果。

选取了部分周期的预测情况,将该方法与非动态转移概率的马尔可夫链预测方法和多层的 BP 神经网络预测方法进行了对比。其中非动态转移概率的方法利用所有的实验数据计算转移概率,BP 神经网络方法以预测周期的前 10 个周期作为特征,训练了 1 000 条数据,选取神经网络层数为 3,对比结果如图 4 所示。

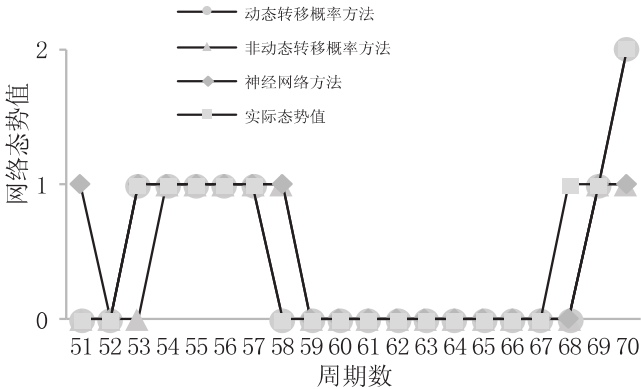


图 4 部分周期不同方法的预测值与实际值对比

由图 4 可见,在网络状态发生变化时容易出现预测错误,但是使用动态转移概率的方法相比于非动态转移概率和 BP 神经网络的方法,具有更好的预测效果,且相对于 BP 神经网络方法没有了复杂的训

练过程。

三种方法的准确率对比如图 5 所示。可以看到在准确率上动态转移概率的方法都高于其余两者,证明了其准确性和有效性。

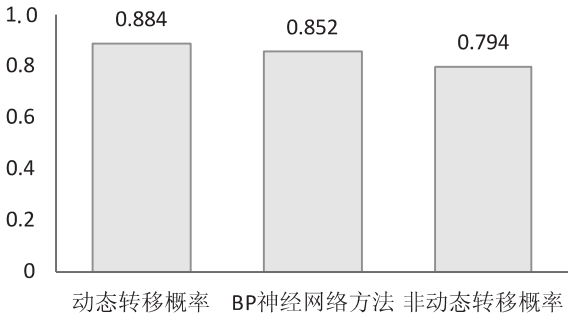


图 5 不同方法的预测准确率

4 结束语

针对网络态势预测的准确性和有效性的问题,分

析了网络变化特点,在马尔可夫链的一步转移概率基础上引入了网络波动率的概念,并在此之上提出了一种基于动态转移概率的网络态势预测方法。通过实时

计算网络的波动率来动态计算转移概率,从而预估下一周期的网络态势值。实验结果表明,相比于非动态转移概率和 BP 神经网络,该方法具有更好的准确性,并且相比于神经网络的方法省去了复杂的训练过程,证明了该方法的准确性和有效性。

参考文献:

- [1] BASS T. Multisensor data fusion for next generation distributed intrusion detect systems[C]//Proceedings of the 1999 I-RIS national symposium on sensor and data fusion. [s. l.]: [s. n.], 1999:1-6.
- [2] YURICK W, ABAD C, HASAN R, et al. UCLog+: a security data management system for correlating alerts, incidents, and raw data from remote logs[J]. Computing Research Repository, 2006, 12(3):124-129.
- [3] OURSTON D, MATZNER S, STUMP W, et al. Applications of hidden Markov models to detecting multi-stage network attacks[C]//36th annual Hawaii international conference on system sciences. Big Island, HI, USA; IEEE, 2003:73-76.
- [4] KATIPALLY R, LI Yang, LIU Anyi. Attacker behavior analysis in multi-stage attack detection system[C]//Proceedings of the 7th workshop on cyber security & information intelligence research. Oak Ridge, Tennessee, USA; ACM, 2011: 63.
- [5] KATIPALLY R, GASIOR W, CUI Xiaohui, et al. Multistage attack detection system for network administrators using data mining[C]//Proceedings of the 6th annual workshop on cyber security and information intelligence research. Oak Ridge, Tennessee, USA; ACM, 2010:51.
- [6] TANG Chenghua, WANG Xin, ZHANG Reixia, et al. Modeling and analysis of network security situation prediction based on covariance likelihood neural [C]//International conference on intelligent computing. Zhengzhou, China; Springer, 2012:71-78.
- [7] QU Zhaoyang, LI Yaying, LI Peng. A network security situation evaluation method based on D-S evidence theory[C]//2nd conference on environmental science and information application technology. Wuhan, China; IEEE, 2010:496-499.
- [8] WANG Lingyu, JAJODIA S, SINGHAL A, et al. k-Zero day safety: measuring the security risk of networks against unknown attacks [C]//European symposium on research in computer security. Athens, Greece; Springer, 2010:573-587.
- [9] JUAN Li, TAO Li, GANG Liang. A network security dynamic situation forecasting method[C]//International forum on information technology and applications. Chengdu, China; IEEE, 2009:115-118.
- [10] SATTY T L. The analytic hierarchy process [J]. Design Methods and Theories, 1980, 14(3):124-134.
- [11] XIE Peng, LI J H, OU Xinming, et al. Using Bayesian networks for cyber security analysis [C]//IEEE/IFIP international conference on dependable systems & networks. Chicago, IL, USA; IEEE, 2010:211-220.
- [12] SZWED P, SKRZYNSKI P. A new lightweight method for security risk assessment based on fuzzy cognitive maps[J]. International Journal of Applied Mathematics and Computer Science, 2014, 24(1):213-225.
- [13] FACHKHA C, BOU-HARB E, DEBBABI M. Towards a forecasting model for distributed denial of service activities [C]//IEEE 12th international symposium on network computing and applications. Cambridge, MA, USA; IEEE, 2013: 110-117.
- [14] KIM S, SHIN S, KIM H, et al. Hybrid intrusion forecasting framework for early warning system[J]. IEICE Transactions on Information and Systems, 2008, E91-D(5):1234-1241.
- [15] PONTES E, GUELF A E, KOFUJI S T, et al. Applying multi-correlation for improving forecasting in cyber security [C]//Sixth international conference on digital information management. Melbourne, QLD, Australia; IEEE, 2011:179-186.
- [16] RAMAKI A A, KHOSRAVI-FARMAD M, BAFGHI A G. Real time alert correlation and prediction using Bayesian networks[C]//12th international Iranian society of cryptology conference on information security and cryptology. Rasht, Iran; IEEE, 2016:98-103.