



网络安全态势的云预测

朱 殊,刘丽霞,邱晓华

(武警工程大学信息工程系 西安 710086)

摘 要:针对现有网络安全态势预测算法存在的主观因素影响较大、预测结果与选取数据关联程度较高的情况,以云的网络安全态势预测规则挖掘算法为基础,提出了一个改进算法。该改进算法不仅拥有原挖掘算法无需对初始数据进行预先处理的优点,同时避免了预测规则与实际情况相孤立造成预测错误的情况,并通过实验验证了算法的可行性和有效性。基于云的网络安全态势预测改进算法,有效地提高了网络安全态势预测的准确性。

关键词:云模型;网络安全;态势预测

doi: 10.3969/j.issn.1000-0801.2013.12.013

Cloud Prediction of Network Security Situation

Zhu Shu, Liu Lixia, Qiu Xiaohua

(Department of Information Engineering, Engineering College of Chinese Armed Police Force, Xi'an 710086, China)

Abstract: The existing algorithms of network security situation prediction were affected with subjective factors. The results of algorithms were highly relevant to selection data. An improved algorithm was proposed based on the algorithm of network security situation prediction based on cloud. It not only have the advantage of original algorithm that don't need to process the initial data in advance, but also solve the problem that errors caused by the results of the original algorithm isolate from the actual situation. Experiment results show that the algorithm is feasibility and effectiveness. The improved algorithm of network security situation prediction based on cloud can improve the accuracy of the results of network security situation prediction.

Key words: cloud model, network security, situation prediction

1 引言

随着互联网技术的不断发展、网络应用的日益普及,计算机网络已经逐渐成为人们生活中不可缺少的一部分。但随之而来的,以网络病毒、DoS/DDoS攻击为主的网络攻击所造成的危害和损失也越来越大,并且向着组织严密化、目标直接化和行为趋利化的方向发展。传统的网络安全防护方式诸如防火墙、防病毒、IDS等单一的被动式网络安全防护技术在面对海量的攻击警告信息时,很难使管理员了解系统的真实安全状态并及时采取合理的响应措

施。因此,如何形象、准确地描绘当前网络安全态势,已经成为网络安全领域一个新的研究热点。网络安全态势感知技术,是用于对当前以及未来一段时间内的网络安全状态及形势进行定量和定性的评价,并对网络进行实时监测和预警的新的安全技术。作为网络安全态势感知的一项重要部分,态势预测能够使系统管理员及时掌握网络系统的变化趋势,针对可能到来的攻击做出切实有效的防御,从而提升系统的安全水平。

目前,常用的网络安全态势预测的方法主要有基于RBF(radial basis function,径向基函数)神经网络的态势预

测方法、基于支持向量机的网络攻击态势预测技术、基于 Verhulst 模型的态势预测方法和 Brynielsson J 和 Arnborg 提出的基于贝叶斯网络的态势感知与预测模型等方法^[1]。

基于 RBF 神经网络的态势预测方法^[2],是通过训练 RBF 神经网络找出态势值的前 N 个数据和随后 M 个数据的非线性映射关系,进而利用该关系进行态势值预测。但是, RBF 算法还存在着一些不足:预测结果受所选取的数据影响较大,容易带有个人主观看法;预测结果需要专家进行解读,这要求管理员具有专门的知识。

为尽可能消除人为因素的影响,提出了根据云的网络安全态势预测方法:通过对历史数据记录进行挖掘,从而得到系统的当前状况及变化趋势,通过当前状态对未来状态进行预测,从而消除人为初始数据的选取对预测结果所造成的影响,增强预测结果的客观性。

在基于云的网络安全态势预测算法中,预测规则的好坏对预测结果有着重要的影响,根据基于云的网络安全态势预测规则挖掘方法,在解决区间划分导致游离在多区间之间的数据失去模糊特性和人为设置最小支持度和置信度阈值对预测规则造成主观影响的问题的基础上,结合当前系统趋势,防止算法因预测规则孤立而出现预测错误,从而得出更为精确的预测结果。

2 基于云的网络安全态势预测规则改进挖掘算法

作为网络安全态势预测的一项基础,网络安全态势评估通常是根据一定算法对感应到的态势指标数据进行处理,得出网络安全态势值。由于网络威胁时序的混沌性和自相似性^[3],当前网络态势指标数据与态势数据之间存在着某种联系,历史数据也与变化趋势之间有着一定的关系。因此,如何从态势指标数据中挖掘出有用的信息,找出历史数据记录同变化趋势之间的关系,以提供尽可能精确的预测结果,是基于云的网络安全态势预测方法中亟待解决的问题。同时,如何使算法尽可能符合变化趋势,在当前状态相同而变化方向不同的情况下也能得出正确的结论,是主要探讨的问题。因此,在一种基于云的预测规则挖掘算法中进行了改进,使预测结果更加贴近实际情况。

2.1 云的相关理论

1995 年,我国的李德毅院士在传统模糊数学理论和概率统计的基础上提出了定性定量不确定性转换模型——云模型。云的定义如下。

定义 1 设 U 是一个用精确数值表示的定量论域, C 是 U 上的定性概念,若定量值 $x \in [0, 1]$ 是具有稳定倾向的随机数,则 x 在论域 U 上的分布称为云,每一个 x 称为云滴^[4]。

$$u: U \rightarrow [0, 1], \forall x \in U, x \rightarrow u(x) \quad (1)$$

云的数字特征可以用 3 个数值表示:期望值 Ex 、熵 En 和超熵 He 。

- 期望 Ex : 云滴在论域空间分布的期望,是最能代表定性概念的点,也是这个概念量化的最典型样本。
- 熵 En : 定性概念的不确定性度量,由概念的随机性和模糊性共同决定。一方面, En 是定性概念随机性的度量,反映了能够代表这个定性概念的云滴的离散程度;另一方面,又是定性概念亦此亦彼性的度量,反映了论域空间中可被概念接受的云滴的取值范围。
- 超熵 He : 熵的不确定性度量,即熵的熵。超熵的大小间接地反映了云的厚度。超熵越大,云滴离散度越大,隶属度的随机性越大,云的厚度越大。超熵由熵的随机性和模糊性共同决定。

云模型是定性和定量之间的转换模型,一般通过云发生器实现。云发生器一般可以分为正向云发生器和逆向云发生器。正向云发生器是一个前向的、直接的过程,根据云的期望值 Ex 、熵 En 和超熵 He 以及所需的云滴数,实现将定性概念转换为定量数值;逆向云发生器则是正向云发生器的逆过程,通过输入已知的精确数据,将其转换为定性的概念,并得到其数字特征,完成从定量数值到定性概念的转换。

2.2 云的网络安全态势属性区间划分

因为网络安全态势评估涉及多个指标,本文选取攻击的威胁程度这一指标进行研究,设 $A(t) = \{A_1, A_2, A_3, \dots, A_n\}$ 表示 IDS 检测到的 n 种事件的数量, $B(t) = \{B_1, B_2, B_3, \dots, B_n\}$ 表示 n 种事件所对应的威胁等级,则当前网络安全态势值 $F(A_i, B_i)$ 可以由式(2)计算:

$$F(A_i, B_i) = \frac{1}{n} \sum_{i=1}^n A_i \cdot B_i \quad (2)$$

计算出的网络安全态势值 $F(A_i, B_i)$ 结果越大,说明网络越不安全。

因为网络安全态势感知中获取的数据都是连续的数据,为便于计算,需要将它们转化为离散值,将连续的数据值按照性质划分到不同的区间之内,完成定量到定性的转换。设 $T = \{T_1, T_2, T_3, \dots, T_m\}$ 是一组时序态势数据值, T_i 表示 i 时刻系统感知到的态势数据,而 $K = \{K_1, K_2, K_3, \dots, K_m\}$ 则表示



T 通过云变换所转换过来的定性的 m 个区间, K_j 表示具体的一个指标区间, $C=\{C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2), C_3(Ex_3, En_3, He_3), \dots, C_m(Ex_m, En_m, He_m)\}$ 则表示 K 所对应的云模型的概念集合, C_j 表示 K_j 属性在云模型中概念的数字特征, 每一个指标区间都对应唯一一个云模型。另外, 设 $\Delta T=\{\Delta T_1, \Delta T_2, \Delta T_3, \dots, \Delta T_n\}$ 是时序态势数值在选取的一个时间段内的变化情况, T_i 表示从 $i-\Delta t$ 到 i 时刻态势数据的变化趋势, 用两点之间直线斜率的余弦值加 1 表示, 而 $\Delta K=\{\Delta K_1, \Delta K_2, \Delta K_3, \dots, \Delta K_3\}$ 则表示 ΔT 通过云变换所转换过来的定性的概念, ΔK_i 表示具体的一个概念, $\Delta C=\{\Delta C_1(Ex_1, En_1, He_1), \Delta C_2(Ex_2, En_2, He_2), \Delta C_3(Ex_3, En_3, He_3), \dots, \Delta C_m(Ex_m, En_m, He_m)\}$ 则表示 ΔK 所对应的云模型的概念集合, ΔC_j 表示 ΔK_j 属性在云模型中概念的数字特征, 每一个指标区间也都与云模型一一对应。

在计算过程中, 时序态势数值 T 和时序态势数值变化率 ΔT 中的值可能同时属于多个云模型, 这些值被称为边界值, 边界值的归属云由它对各个云的隶属度决定, 边界值对哪个云的隶属度大, 该边界值就属于哪个云模型。这样, 可以在运算过程中减少人为干扰, 增强结果的客观性。

同时, 随着数据量的增大, 必须不断对论域进行调整, 以使得指标区间能够准确地表达出攻击威胁度的性质, 反映出网络的真实态势情况。

2.3 改进的 ORAR 算法

云模型改进优化规则关联 (optimization relation association rule, ORAR) 算法是通过建立关联矩阵 R 使整个预测过程只需进行少量次数的数据库扫描。其基本思想是: 把感知数据通过云变换转换为用云模型来表示的定性概念, 把连续的态势数据转换成离散的数据。在此基础上, 依据所建立的关系矩阵 R , 选取合适的 $k-1$ 大项集, 并由此计算出 k 项集, 通过所设定的最小支持度阈值 minsup 筛选出 $k-1$ 大项集, 然后, 根据所设定的最小置信度阈值 minconf , 输出置信度 conf 大于 minconf 的规则。本文在 ORAR 算法的基础上, 针对 ORAR 算法的结果与变化趋势相孤立, 有时无法得出正确预测结果的情况, 在生成规则后, 结合系统当前趋势对预测规则进行分析, 并对其做出合理修正, 从而获得较为精确的预测结果。改进 ORAR 算法流程如图 1 所示。

改进的 ORAR 算法的步骤如下。

(1) 划分概念区间: 设 $D=\{D_1, D_2, D_3, \dots, D_n\}$ 是攻击的威胁程度指标集合, $N=\{N_1, N_2\}$ 则是网络安全态势值的集合,

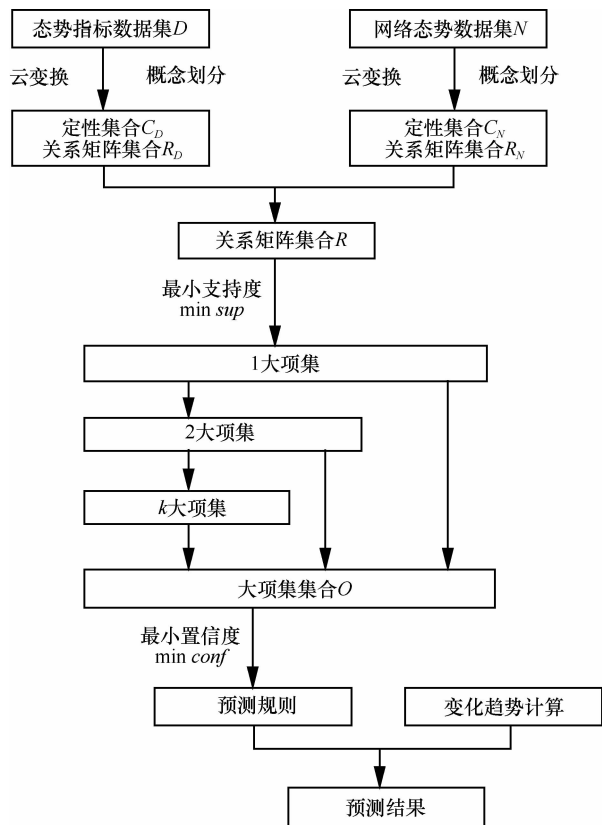


图1 改进 ORAR 算法流程

N_1 表示当前网络安全态势值的集合, 而 N_2 表示所要预测的、 N_1 下一时刻的网络安全态势值的集合。将 D 通过云变换转化为云模型的定性集 $C_D=\{C_{D1}, C_{D2}, C_{D3}, \dots, C_{Dn}\}$ 。将 N 也通过云变换转化为云模型的定性集合 $C_N=\{C_1, C_2, C_3, \dots, C_n\}$, 从而完成定量到定性的概念变换。

(2) 建立关系矩阵: 通过云的极大值判别法计算各数值对云模型概念的隶属度, 将数值归属到合适的云概念中, 建立关系矩阵 R_{Dk} 。

(3) 循环进行步骤(1)和步骤(2), 进行概念区间的划分和建立关系矩阵, 得到与指标集合 D 相关的云模型的概念集合 $C_D=\{C_{D1}, C_{D2}, C_{D3}, \dots, C_{Dn}\}$ 和关系矩阵的集合 $R_D=\{R_{D1}, R_{D2}, R_{D3}, \dots, R_{Dn}\}$; 态势值集合 N 的云模型概念集合 $C_N=\{C_1, C_2, C_3, \dots, C_n\}$ 和关系矩阵的集合 $R_N=\{R_1, R_2\}$ 。

(4) 将关系矩阵 R_D 和 R_N 合并, 得到关系矩阵集合 $R=\{R_{D1}, R_{D2}, R_{D3}, \dots, R_{Dn}, R_1, R_2\}$ 。

(5) 1 大项集的生成: 输入关系矩阵 R 和最小支持度阈值 minsup , 对关系矩阵 R 进行搜索, 运用 ORAR1 算法筛选出支持度 $\text{sup} \geq \text{minsup}$ 的项集, 从而获得 1 大项集, 并将 1 大项集纳入大项集的集合 O 中。

(6)2大项集的生成:对关系矩阵 R 进行搜索,运用ORAR2算法在步骤(5)中得出的1大项集中筛选出支持度 $sup \geq minsup$ 的项集,从而得到2大项集,并将2大项集纳入大项集的集合 O 中。

(7) k 大项集的生成:当 $k > 2$ 时,对关系矩阵 R 进行搜索,运用ORAR k 算法在 $k-1$ 大项集中筛选出支持度 $sup \geq minsup$ 的项集,从而生成 k 大项集,并将 k 大项集纳入大项集的集合 O 中。

(8)生成规则:输入最小置信度阈值 $minconf$,计算集合 O 中大项集中产生的预测规则的置信度 $conf$,筛选出置信度 $conf \geq minconf$ 的预测规则。

(9)计算变化趋势:计算出 i 时刻系统状态在 $i-\Delta t$ 到 i 时刻内的变化趋势,并观察属于哪一变化区间。

(10)将变化区间与步骤(8)中的预测规则相结合,分析出最终的预测结果。

3 实验与分析

3.1 实验数据和定义概念

实验数据来源于某局域网中IDS所采集到的威胁警报数据,从中选取威胁警报数 A 和攻击警报总数 I 进行分析,根据式(2)计算出网络安全态势值 F 。网络安全态势如图2所示,点A表示 $i=335$ 时的网络安全态势,点B表示 $i=421$ 时的态势。

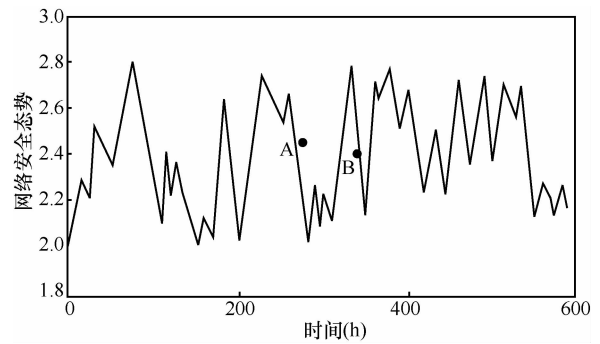


图2 网络安全态势

攻击对系统的威胁程度由攻击警报数决定,攻击警报数越大,表明攻击对系统的威胁越大,威胁等级越高。利用云的概念区间划分方法将攻击威胁分为3个等级:轻微威胁、一般威胁、严重威胁;网络安全态势分为3个等级:安全程度高、安全程度中、安全程度低;网络安全趋势分为3个等级:形势好转、形势稳定、形势恶化。表1给出了云模型中各概念的数字特征。

表1 各概念的数字特征

定性概念	期望 E_x	熵 E_n	超熵 He
扫描威胁轻微	523	87	0.7
扫描威胁一般	1 064	117	3.45
扫描威胁严重	1 823	187	7.21
总体威胁轻微	14 042	1 310	10.98
总体威胁一般	20 406	1 519	13.22
总体威胁严重	28 632	2 053	15.01
安全程度高	1.563	0.125	0.003
安全程度中	2.091	0.126	0.004
安全程度低	2.673	0.128	0.008
形势好转	0.479	0.094	0.002
形势稳定	1.011	0.012	0.001
形势恶化	1.587	0.091	0.002

3.2 实验结果

根据改进的ORAR算法对数据进行挖掘,选择的最小支持度阈值 $minsup=0.034\ 2$,最小置信度阈值 $minconf=0.616\ 4$,得出如下6个预测规则:

- 扫描威胁轻微,总体威胁一般,安全程度低→安全程度中, $\{sup=0.060\ 9, conf=0.642\ 1\}$;
- 扫描威胁轻微,总体威胁严重,安全程度中→安全程度低, $\{sup=0.037\ 3, conf=0.636\ 7\}$;
- 扫描威胁轻微,总体威胁严重,安全程度低→安全程度低, $\{sup=0.060\ 6, conf=0.762\ 5\}$;
- 扫描威胁一般,总体威胁轻微,安全程度低→安全程度中, $\{sup=0.038\ 2, conf=0.603\ 3\}$;
- 扫描威胁严重,总体威胁严重,安全程度中→安全程度低, $\{sup=0.030\ 6, conf=0.813\ 1\}$;
- 扫描威胁严重,总体威胁严重,安全程度低→安全程度低, $\{sup=0.051\ 4, conf=0.771\ 6\}$ 。

3.3 结果分析

实验数据见表2。

表2表明,点A和点B的安全状态相同,点A下一时刻的预测规则1为{扫描威胁轻微,总体威胁一般,安全程度低→安全程度中},同时,点A的网络安全趋势为形势好转,与预测规则相同,因此可以判断点A的下一时刻网络安全程度为中,这与现实情况相同。而点B的下一时刻的预测规则为{扫描威胁轻微,总体威胁一般,安全程度低→安全程度中},但是点B状态下的网络安全趋势为形势恶化,综合来看,判断点B下一时刻的网络安全程度为低,这与实际结果相同。可以看到,在采用原算法的情况下,由于没有考虑网络安全趋势的变化,对点A和点B的预测规则相同,但由于二者网络安全趋势不同,造成了实际结果



表 2 实验数据

时间点 i	扫描威胁	总体威胁	网络安全态势值	网络安全趋势(ORAR 算法)	网络安全趋势(改进 ORAR 算法)
335	轻微(690 次)	一般(19 721 次)	安全程度低(2 542)	形势好转	形势好转
336	轻微(573 次)	一般(18 754 次)	安全程度中(2 231)		
421	轻微(628 次)	一般(18 401 次)	安全程度低(2 371)	形势恶化	形势好转
422	轻微(644 次)	一般(17 148 次)	安全程度低(2 393)		

的不同,使算法在预测点 B 时出现错误,而在改进算法下,因为充分考虑到网络趋势变化情况,避免了预测规则的孤立,从而得出了较为精确的结果。但是,在预测规则与趋势变化结果相冲突时,需要引入人为影响,这在一定程度上降低了算法的客观性。

4 结束语

将网络安全态势感知和云理论结合起来,针对现有的基于云的网络安全态势预测规则挖掘算法予以改进,使算法与实际结合更加紧密,结果更加精确。因为实验中验证的数据较少,因此对网络安全态势的发展预测还仅限于局部。当提供的数据量增加时,预测规则的数量将同时得到提高,预测结果也将更加精确。但是,如何将算法应用于实际以及如何解决算法中两方结论冲突时的决策问题,仍然需要做进一步的研究。

参考文献

- 1 Chang K C, Saha R K, Yin X Y. A linear predictive bandwidth conservation algorithm for situation awareness. http://volgenau.gmu.edu/~kchang/publications/conference_pdf/CON_9_A_LINEAR.pdf,1998
- 2 任伟. 网络安全态势评估智能化研究. 上海交通大学硕士学位论文,2007

- 3 宣蕾. 网络安全定量风险评估及预测技术研究. 国防科学技术大学博士学位论文,2007
- 4 李德毅, 杜鹤. 不确定性人工智能. 北京: 国防工业出版社, 2005
- 5 郝树勇. 基于云的网络安全态势预测研究. 国防科学技术大学硕士学位论文,2010
- 6 叶琼, 李绍稳, 张友华等. 云模型及应用综述. 计算机工程与设计, 2011,32(12):4198~4201
- 7 周文秀. 关联规则挖掘算法的研究与改进. 武汉理工大学硕士学位论文,2008

[作者简介]



朱殊,男,武警工程大学信息工程系硕士研究生,主要研究方向为网络安全。

刘丽霞,女,博士,武警工程大学信息工程系教授,主要研究方向为算法优化和嵌入式系统应用。

邱晓华,男,武警工程大学信息工程系助教,主要研究方向为计算机网络。

(收稿日期:2013-07-20)

(上接第 81 页)

- 13 Martin R C. Agile Software Development: Principles, and Practices. Prentice Hall PTR, 2003

[作者简介]



张小东,男,哈尔滨工业大学博士研究生,主要研究方向为云服务平台框架及其关键技术。

战德臣,男,哈尔滨工业大学教授、博士生导师,主要研究方向为服务工程与服务计算、软件工程。

聂兰顺,男,博士,哈尔滨工业大学副教授,主要研究方向为信息物理系统、资源调度、优化方法等。

孟凡超,男,博士,哈尔滨工业大学(威海)副教授,CCF 高级会员,主要研究方向为软件体系结构、云计算。

(收稿日期:2013-07-31)