文章编号:1001 - 9081(2007)10 - 2443 - 03

基于 PC-LINMAP 耦合赋权及云理论的入侵检测系统

张秋余,孙 磊
(兰州理工大学 计算机与通信学院,兰州 730050)
(andy@ mail2. lut. cn)

摘 要:提出一种基于 PC-LINMAP 耦合赋权、云理论来判断系统入侵发生可能性大小的新方法。首先运用 PC-LINMAP 耦合赋权法计算系统主要性能指标的权值,并将得到的权值与理想状态下各个性能指标的数值做加权融合,从而得到理想状态下的综合评价结果。将任意时刻通过加权融合得到的数值与理想状态下得到的综合评价结果相比较得出偏差值,最后基于云理论构造定性评测云发生器并结合偏差的大小对当前入侵发生的可能性进行定性描述。实验结果表明了该方法的有效性。

关键词:PC-LINMAP 耦合赋权;云理论;入侵检测;定性评测云发生器
中图分类号: TP309 **文献标志码**:A

Intrusion detection system based on PC-LINMAP and cloud theory

ZHANG Qiu-yu, SUN Lei
(*College of Computer and Communication, Lanzhou University of Technology, Lanzhou Gansu* 730050, *China*)

Abstract: A new decision method to judge the damage degree of system intrusion possibility was proposed, which was based on PC-LINMAP and cloud theory. Firstly, the index of system performance will be given weighted values by using PC-LINMAP Theory, and these computed weighted values will be compared and fused with the values out from index of system performance under ideal conditions. Then this result of comprehensive assessment under ideal conditions could be obtained. Following next, the variation could be obtained by the comparison of real time's weighted fusion values with comprehensive assessment under ideal conditions at each specific moment. Finally, the Cloud Generator which is based on Cloud Theory combined with the above variation can achieve the qualitative description about the damage degree of current system intrusion possibility. The result of experiment shows that this method is effective.

Key words: PC-LINMAP; cloud theory; intrusion detection; qualitative evaluation cloud generator

0 引言

随着电子计算机和网络应用的不断深入,人们对系统的安全要求也越来越高,安全技术和安全产品也在不断地发展和成熟,然而绝对安全的系统是很难实现的。人们更关心的不是系统是否发生入侵,而是系统入侵发生的可能性大小,以此来判断系统的主要服务或者关键服务等能否正常工作。本文基于云理论[1]以及 PC-LINMAP 耦合赋权法[2]建立了一个能够判断入侵发生可能性大小的入侵检测系统,对系统入侵发生的可能性大小进行了定性描述,从而进一步完善了现有的入侵检测系统。

当入侵进入系统并发起恶意的修改、攻击等行为时,必然会引起系统主要性能指标的异常变化[3],而这些变化的幅度决定着系统受入侵危害的程度。入侵危害越深,系统异常变化的幅度也就越大,异常变化持续时间也就越长。因此,首先需要确定出能够影响整个系统性能的主要性能指标,然后基于 PC-LINMAP 耦合赋权法对多个正常样本(每个样本含多个性能指标的状态参数)进行分析得到各个性能指标的权重,然后计算理想状态下性能指标的综合评价结果 θ'。基于云理论构造五个评语的评语集 $V = (v_1, v_2, \cdots, v_5)$[4],分别表示系统遭受入侵可能性的大小,即很小、小、一般、大、很大五个评语。将某一时刻主要性能指标的综合评价结果 θ 与 θ' 作比较得出一个偏差值 $a = |\theta - \theta'| / \max(\theta, \theta')$,最后基于定性评测云发生器和偏差的大小对当前系统发生入侵的可能性大小进行定性描述。

1 确定系统的主要性能指标

对于不同的工作环境,其主要性能指标也有所不同。可以通过统计分析、专家知识等方法来确定系统的主要性能指标。本文以主机系统的主要性能指标为例,一般来说,主机系统的性能指标主要体现在[5]:

1)CPU 的利用率;
2)内存的利用率;
3)进程的服务时间;
4)物理内存和虚拟内存的大小;
5)I/O 的使用情况等。

引起一个性能指标数值异常变化的可能性有很多,有些属于正常的,有些则属于入侵或者错误产生的,只凭借一个性能指标数值的异常变化无法准确判断系统发生入侵的可能性大小,所以必须综合考虑各个性能指标数值的异常变化。因

收稿日期:2007 - 04 - 20;**修回日期**:2007 - 06 - 22。 **基金项目**:甘肃省自然科学基金资助项目(ZS021-A25-018-G)。
作者简介:张秋余(1966 -),男,河北辛集人,副研究员,主要研究方向:信息安全、软件工程、多媒体通信; 孙磊(1981 -),男,山东淄博人,硕士研究生,主要研究方向:信息安全。

此影响系统性能的主要指标的确定对于判断系统发生入侵的可能性大小是非常重要的。

2 PC-LINMAP 耦合赋权法

在不同的工作环境下,系统的各个性能指标所占的权重是不同的,但在一个相对稳定的工作条件下,各个性能指标所占的权重几乎不变。PC-LINMAP 耦合赋权模型将主成分分析和多维偏好线性规划进行了有机的结合,具有原始信息客观、系统分析科学、结论明确、实用性强的优点。PC-LINMAP 耦合模型分为两部分,首先,应用 PC 从原始决策矩阵求取样品的优劣排序,然后应用 LINMAP 基于求得的样品优劣有序对确定每个指标的权重,本文分以下四个步骤来完成。

1)原始数据标准化

假设参加评价的系统状态有 n 个,每个系统状态下有 p 个性能指标,于是原始数据就可以用决策矩阵 $(x_{ij})_{n\times p}$ 来表示,为了消除原指标量纲、数量级的不同,对原始数据进行标准化处理:

$$y_{ij} = (x_{ij} - x_j')/\delta_j;\ i = 1,2,\cdots,n, j = 1,2,\cdots,p$$

其中 $x_j' = \sum_{i=1}^{n} x_{ij}/n, \delta_j = \left[\sum_{i=1}^{n}(x_{ij} - x_j')^2/(n-1)\right]^{1/2};\ \forall i,j$

2)计算性能指标的相关矩阵 $\boldsymbol{R}$ 及其相关矩阵的特征值和特征向量

$$\boldsymbol{R} = (r_{jk})_{p\times p}$$

其中 $r_{jk} = [\sum_{i=1}^{n} y_{ij}\cdot y_{ik}]/(n-1);\ j,k = 1,2,\cdots,p$

用雅可比方法求相关矩阵的特征值 $\lambda_i (i = 1,2,\cdots,p)$ 并记作 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$,同时求得相应的特征向量 $\boldsymbol{\beta}_e = (\beta_{e1},\beta_{e2},\cdots,\beta_{ep}{}^T)$,其中 $e = 1,2,\cdots,p$。

3)确定主成分并计算主成分的总得分值和样品的主成分总得分值

选取前 m 个主成分,这里正整数 m 必须满足下式的最小值:$\sum_{i=1}^{m}\lambda_i/\sum_{i=1}^{p}\lambda_i \geqslant 85\%$,于是第 i 个系统状态在前 m 个主成分方向上的得分值 $z_1^i, z_2^i, \cdots, z_m^i$ 为:

$$\begin{bmatrix} z_1^i \\ z_2^i \\ \vdots \\ z_m^i \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2p} \\ \vdots & & & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mp} \end{bmatrix} \begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{pi} \end{bmatrix}$$

第 i 个样品的总得分值 $F_i = \sum_{j=1}^{m} p_j \mid z_j^i \mid$,其中 $i = 1,2,\cdots,n$;p_j 为第 j 个主成分保持原始数据总信息量的比重,即 $p_j = \lambda_i/\sum_{e=1}^{p}\lambda_e$;方案的优劣顺序按照总得分值 F_i 由大到小排列。

4)线性规划求权值

根据专家知识和经验,我们给定当前工作环境下指标空间中的理想点 $(y_1', y_2', \cdots, y_p')$ 则空间中任一样品点 $(y_{i1}, y_{i2}, \cdots, y_{ip})$ 到理想点 $(y_1', y_2', \cdots, y_p')$ 的加权欧几里德距离平方 s_i 为 $s_i = \sum_{j=1}^{p} w_j'(y_{ij} - y_j')^2;\ i = 1,2,\cdots,n$;其中 $w_j' (j = 1,2,\cdots,p)$ 是第 j 个指标的权重平方。

样品有序对 (k,r) 的集 Q 为 $Q = \{(k,r) \mid k$ 样品优于 r 样品$\}$,对样品有序对 (k,r) 的集 Q 中所有的有序对求和,得到总的不一致度 B 和一致度 G。

$$B = \sum_{(k,r)\in Q}(s_r - s_k)^-, \text{其中} (s_r - s_k)^- = \begin{cases} 0, & s_r \geqslant s_k \\ s_r - s_k, & s_r < s_k \end{cases}$$

$$G = \sum_{(k,r)\in Q}(s_r - s_k)^+, \text{其中} (s_r - s_k)^+ = \begin{cases} 0, & s_r < s_k \\ s_r - s_k, & s_r \geqslant s_k \end{cases}$$

那么 $G - B = h = \sum_{(k,r)\in Q}(s_r - s_k)$,于是指标权重的平方 w_j' 就可以通过线性规划得到,此时的目标函数 $\min \sum_{(k,r)\in Q}\lambda_{kr}$,约束条件:

$$\sum_{j=1}^{p} w_j'(y_{rj}^2 - y_{kj}^2) - 2\sum_{j=1}^{p} v_j(y_{rj} - y_{kj}) + \lambda_{kr} \geqslant 0$$

$$\sum_{j=1}^{p} w_j' \sum_{(k,r)\in Q}(y_{rj}^2 - y_{kj}^2) - 2\sum_{j=1}^{p} v_j \sum_{(k,r)\in Q}(y_{rj} - y_{kj}) = h$$

$w_j' > 0, j = 1,2,\cdots,p; \lambda_{kr} \geqslant 0$,对于所有的 $(k,r) \in Q; v_j$ 无非负限制($v_j = w_j y_j'$)可以得到 $w_j' (j = 1,2,\cdots,p)$ 的值,归一化之后即得各指标的权重向量 $\boldsymbol{w}_j$。

3 运用云理论实现定性评测

针对模糊理论的不彻底性,文献[6]在 1995 年提出了隶属云和隶属云发生器,突破了概率论和数理统计的不足,破除了粗集的局限性。用云模型来统一刻画语言值和数值间的随机性和模糊性,为解决定量定性转换的不确定性问题提供了一套新的技术方法。

云主要反映宇宙中事物或人类知识中概念的两种不确定性:模糊性和随机性。云模型把模糊性和随机性集成在一起,研究自然语言中最基本的语言值(又称语言原子)所蕴含的不确定性的普遍规律,使得有可能从语言值表达的定性信息中获得定量数据的范围和分布规律,也可以把精确数值有效转换为恰当的定性语言值。云的定义:

设 U 是一个用精确数值表示的定量论域,$X \subseteq U$,T 是 U 空间上的定性概念,若元素 $x(x \in X)$ 对 T 的隶属确定度 $C_T(x) \in [0,1]$ 是一有稳定趋向的随机数,则概念 T 从论域 U 到区间 $[0,1]$ 的映射在数域空间的分布,称为云(Cloud)。

$$C_T(x): U \to [0,1], \forall x \in X(X \subseteq U), x \to C_T(x)$$

这个定义可以推广到 N 维云。即如果 U 是 N 维论域,$X \subseteq U$,则 N 维元素 $x = (x_1, x_2, \cdots, x_n)(x \in X)$ 对 T 的隶属确定度 $C_T(x) \in [0,1]$ 也是一有稳定趋向的随机数。从云的基本定义出发,论域 U 上的概念 T 从论域 U 到区间 $[0,1]$ 的映射是一对多的关系,即论域中某一个元素与它对应概念 T 的隶属度之间的映射是一对多的转换,而不是传统的模糊隶属函数中的一对一的关系。

表达概念 T 的云由许多的云滴组成,每个云滴均是这个定性概念映射到数域空间的一个点,即定性概念的语言值在数量上的一次具体样例实现。这种实现带有不确定性,模型同时给出这个点能够代表该定性概念的确定程度。

每个云滴都是随机产生的,而且每个云滴代表该定性概念语言的不确定变化,在每个云滴表现出来时,不会剧烈影响到云的整体特征,也就是说某一个特定的云滴可能无足轻重,但是一定数量的云滴的整体分布特性就体现了云映射的模糊

性和随机性,云的整体形状反映了在用定量数值表示定性概念时的不确定特性。例如“离地面100 cm左右”是一个空间概念,而“离地面 100 cm”就是一个空间数据,是该定性概念在论域中的一次具体定量实现,经过云映射,这个云滴代表该定性概念的确定程度是1。但这种实现也可能是“离地面99 cm”的数据,代表该定性概念的确定程度也可能是0.9 等。所有的这些实现累积到一定数量,经过云映射,在论域中就形成一朵云,表达“离地面 100 cm 左右”这个概念。

云用三个数字特征量表示语言值的定性含义:$A(Ex, En, D)$,其中 Ex、En 和 D 分别是云的期望值、熵和超熵。给定一组 $\{Ex, En, D\}$ 就唯一地定义了一个特定的云 A。云的期望值 Ex 是 U 中与云的重心[1] 对应的位置。论域中的数值 Ex 与语言值 A 完全相容。在实际应用中很容易确定 Ex。熵 En 反映了该语言值对数值的可覆盖测度,表示了论域中有多少数值可以被这个定性概念所接受。超熵 D 反映云滴的离散程度,仔细观察云的形状,我们可以看出云的“厚度”是不均匀的,靠近腰部最分散,而在顶部和底部汇聚性更好。通过云理论,我们可以更准确地描述当前系统所处的状态,且具有很强的直观性和易懂性。

判断任意状态$(y_{i1}, y_{i2}, \cdots, y_{ip})$下系统发生入侵的可能性大小,首先将通过 PC-LINMAP 耦合赋权法得到的各个指标的权重 w_j,代入下式得到主要性能指标的综合评价结果 θ 及理想状态下性能指标的综合评价结果 θ':

$$\theta = \sum_{j=1}^{p} w_j y_{ij}, \theta' = \sum_{j=1}^{p} w_j y_{1j}$$

其中 y_{1j} 对应为理想点$(y_1', y_2', \cdots, y_p')$。

$a = |\theta - \theta'| / \max(\theta, \theta')$,将定义的五个评语置于连续的语言值标尺上,且每个评语都用云模型来实现,构成一个定性评测的云发生器,如图 1 所示。

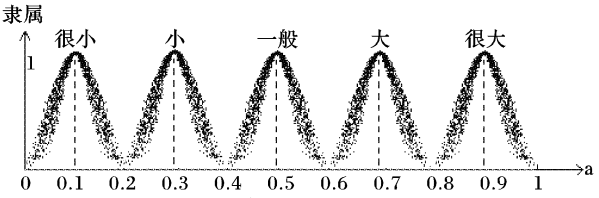

图 1　定性评测云发生器

4　判断系统发生入侵的可能性大小的实例

在一个主机系统当中,CPU 的利用率(指标 1)、内存的利用率(指标 2)、进程的服务时间(指标 3)这三个性能指标是非常敏感和重要的,通过这三个性能指标数值的异常变化来判断系统发生入侵的可能性大小是可行的。

首先假设系统正常工作下理想状态所对应的理想点$(y_1', y_2', \cdots, y_p') = (0.3, 0.3, 0.8)$,然后随机的取四个正常状态下上述三个指标的数值,构成决策矩阵 $\boldsymbol{B}$:

$$\boldsymbol{B} = \begin{array}{c} \\ \text{状态} \end{array} \begin{array}{c} \begin{array}{ccc} \text{指标 1} & \text{指标 2} & \text{指标 3} \end{array} \\ \begin{bmatrix} 0.6 & 0.5 & 1.2 \\ 0.3 & 0.4 & 1.4 \\ 0.4 & 0.5 & 1.3 \\ 0.5 & 0.3 & 1.2 \end{bmatrix} \end{array}$$

本文用 Matlab 设计实现了 PC-LINMAP 耦合赋权模型,把 $\boldsymbol{B}$ 矩阵输入 Matlab 程序,求得指标的相关矩阵:

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.1348 & -0.9439 \\ 0.1348 & 1.0000 & 0.0909 \\ -0.9439 & 0.0909 & 1.0000 \end{bmatrix}$$

相关矩阵 $\boldsymbol{R}$ 的特征值为 $\lambda_1 = 1.9449, \lambda_2 = 1.0252, \lambda_3 = 0.0298$;

对应的特征向量为$[-0.7087, -0.0333, 0.7047]^T$,$[-0.0912, -0.9862, -0.1384]^T$,$[-0.6996, 0.1624, -0.6958]^T$,得到方案的优劣顺序 O 为 $O = \{$状态4,状态3,状态2,状态1$\}$ 即方案的有序对集 $Q = \{(4,3),(4,2),(4,1),(3,2),(3,1),(2,1)\}$。

求得三个指标的权重平方值 w_j' 为 $w_j' = (0.0252, 0.0433, 0.0613)$ 那么三个指标的权重向量 $w_j = (\sqrt{0.0252}, \sqrt{0.0433}, \sqrt{0.0613})$ 经归一化后得到 $w_j = (0.2583, 0.3387, 0.4030)$。

由前面的假设得知,正常工作下系统的最佳状态所对应的理想点$(y_1', y_2', \cdots, y_p') = (0.3, 0.3, 0.8)$,根据公式当前系统的最佳状态下性能指标的综合评价结果是:

$$\theta' = \sum_{j=1}^{p} w_j y_{1j} = 0.2583 \times 0.3 + 0.3387 \times 0.3 + 0.4030 \times 0.8 = 0.5015$$

并假设任一时刻 t 得到的一组样本为$(0.9, 0.8, 1.6)$,判断在这一时刻 t 系统发生入侵的可能性大小。由公式 $\theta = \sum_{j=1}^{p} w_j y_{ij}$ $= 0.2583 \times 0.9 + 0.3387 \times 0.8 + 0.4030 \times 1.6 = 1.14823$。

将上面求得的 θ' 和 θ 的值代入公式:

$a = |\theta - \theta'| / \max(\theta, \theta') = (1.1482\,3 - 0.501\,5) / 1.14823 = 0.5632$

根据定性评测云发生器,我们可以定性的判断此时的系统发生入侵的可能性是一般。

5　结语

在稳定的工作环境下,对多个状态的主要性能指标进行 PC-LINMAP 耦合赋权分析得到的权重具有原始信息的客观性、科学性、实用性强等优点。通过对比当前状态和理想状态下系统的主要性能指标的综合评价结果,结合云模型所构造的五评语定性评测云发生器,给出当前系统发生入侵可能性大小的定性语言描述,弥补了过去以判断是否发生入侵为主要目标的入侵检测系统的不足。通过简单的实例,证明了本方法的可行性和有效性,但在五评语所对应的数值属性区间划分的方法上有待于进一步的研究。

参考文献:

[1] 李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器[J]. 计算机研究和发展, 1995, 32(6): 16 - 21.

[2] 应天元. 系统综合评价的赋权新方法——PC-LINMAP 耦合模型[J]. 系统工程理论与实践, 1997, 17(2): 8 - 13.

[3] 杨鹤, 董红斌, 梁意文, 等. 人工免疫系统中危险信号的云方法定义[J]. 计算机工程与应用, 2006, 42(10): 34 - 45.

[4] 何洪成. 云模型及其在指挥控制系统可靠性分析中的应用[J]. 火力与指挥控制, 2006, 31(5): 76 - 80.

[5] 赵卫伟, 李德毅. 基于云模型的入侵检测方法[J]. 计算机工程与应用, 2003, 39(26): 158 - 160, 164.

[6] 李德毅, 史雪梅. 语言原子模型和似然推理[C]// 计算机智能接口与智能应用论文集, 1993: 272 - 277.