

基于正态云模型的网络入侵检测

尹向东

(湖南科技学院 计算机科学系,湖南 永州 400055)

摘 要: 将正态云模型用于网络入侵检测,并提出一种基于正态云模型的网络入侵检测算法,算法首先采取特征提取并计算影响权值因子,引入云发生器计算属性特征值和差异度,入侵判断采用综合评价价值,减少了单个属性造成的局部影响.实验结果表明,该方法能够比较有效地检测真实网络数据中的未知入侵行为,在一定程度上解决了目前部分入侵检测算法存在的检测率低、误报率高的问题.

关键词: 入侵检测;云模型;正态云模型;网络安全

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2010)10-0034-03

Network Intrusion Detection Based on Normal Cloud Model

YIN Xiang-dong

(Department of Computer Science, Hunan Institute of Science and Technology, Yongzhou 400055, China)

Abstract: In this paper, the normal cloud model for network intrusion detection, and a normal cloud model is based on the network intrusion detection algorithms, to take the first feature extraction algorithm and calculate the impact of weight factor, the introduction of cloud computing attributes generators and difference eigenvalue, the value of comprehensive evaluation of the application to judge the invasion, resulted in a reduction of a single attribute of the local impact. The experimental results show that this method can detect more effective network data in real unknown intrusions, to a certain extent to solve some of the current intrusion detection algorithms detect the existence of the low rate of high false alarm rate.

Key words: intrusion detection; cloud model; the normal cloud model; network security

1 引言

入侵检测数据库中既有定量属性又有定性属性,定性属性有些还是不确定性的模糊的信息.在进行入侵检测时,通常需要专家的经验 and 知识,专家的知识 and 经验通常都是用定性语言表达的,而计算机能够处理的通常都是定量数据,因此需要将定性数据与定量数据进行转换,云模型是进行定性定量转换的工具^[1-2].云模型把随机性和模糊性结合起来.用数字特征熵,揭示随机性与模糊性的关联性.并用来表示一个定性概念的粒度.正态云模型通过期望、熵和超熵构成的特定结构发生器,生成定性概念的定量转换值,体现概念的不确定性^[3].这种特定结构不但放宽了形成正态分布的前提条件.而且把精确确定隶属函数放宽到构造正态隶属度分布的期望

函数.因而更具有普遍适用性,更简单、直接地完成了定性 with 定量之间的相互转换过程^[4].

针对云模型的研究,给出定性 with 定量之间转换的云模型形式化表示,并且将正态云模型用于计算机网络入侵检测,并提出一种基于正态云模型的网络入侵检测算法.

2 正态云模型及其算法

正态云模型是用语言值表示定量表示与定性概念之间的不确定性转换模型,它主要反映客观世界中事物或人类知识中概念的两种不确定性:随机性(发生的概率)与模糊性(边界的亦此亦彼性),并把二者集成在一起,构成定量与定性相互间的映射,研究自然语言中的最基本语言值所蕴含的不确定性的普遍规律,使得有可能把精确数值有效转换为恰当

的定性语言值,或者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.正态云模型用相互独立的一组参数共同表达一个定性概念的数字特征,反映概念的不确定性.在正态隶属函数与正态分布函数基础上,这组参数用熵 E_n ,超熵 E_c ,期望 E_x 这3个数字特征来表示:

(1) 熵 E_n 代表一个定性概念的可度量粒度,通常熵越大概念越宏观.熵还反映了定性概念的不确定性,表示在论域空间可以被定性概念接受的取值范围,是定性概念亦此亦彼性的度量.

(2) 超熵 E_c 熵的不确定性的度量,它反映代表定性概念值的样本出现的随机性,揭示了随机性与模糊性的关联.

(3) 期望 E_x 在论域空间中最能够代表这个定性概念的点,是概念量化的最典型样本点.当概念对应的论域为一维时,定性定量的不确定性转换通过正态云发生器算法实现.具体^[5]如下:

输入表示定性概念 A 的3个数字特征值熵 E_n ,超熵 E_c ,期望 E_x ,云滴数 N .

输出 N 个云滴的定量值,以及每个云滴代表概念 A 的确定度.

步骤1:生成以 H_c 为标准差, E_n 为期望值的一个正态随机数 E'_n ;

步骤2:生成以 $\text{abs}(E'_n)$ 为标准差, E_x 为期望值的正态随机数 x ;

步骤3:令 x 为定性概念 A 的一次具体量化值,称为云滴;

步骤4:计算 $y = \exp[-(x - E_x)^2 / 2(E'_n)^2]$;

步骤5:令 y 为 x 属于定性概念 A 的确定度;

步骤6: $\{x, y\}$ 完整地反映了定性定量转换的全部内容;

步骤7:重复步骤1至步骤6,直到产生 N 个云滴.

从这个特定的生成结构序可以看出,正态云模型运用正态隶属函数与正态分布的普适性,放宽了约束条件,简化了参数,并将模糊性与随机性通过熵统一表示,构成不确定性知识表示模型.

3 基于正态云模型的入侵检测

一般基于云模型提出的入侵检测方法是基于规则发生器实现的^[6],规则往往是通过关联规则挖掘或者安全专家得到,缺乏事实标准和依据.而本文提出的算法则采用逆向云发生器从真实训练数据中得到云的数字特征,形成实际判断规则,实现正常数据

建模.数据建模首先根据服务属性分类与协议从训练数据中提取正常连接数据,其次对每类数据进行特征提取,最后利用正向云发生器计算提取属性的数字特征,完成数据建模过程;入侵检测采用正向云发生器计算正常数据模型与测试数据之间的差异度,入侵判断采用综合评价值来完成入侵检测判断过程.如图1所示.

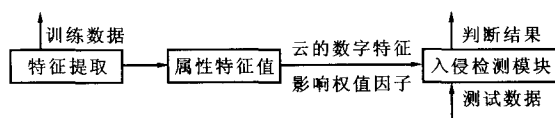


图1 基于正态云模型的入侵检测模型

3.1 特征提取

对于每种正常模型,对结果的影响程度也不同,其各个属性所占的权重不同,所以如何选取特征成为首要问题.目前特征选取主要有熵值法、卡方法、主成分分析法和均方差法等方法.无监督异常检测算法能够在最小均方误差原则下最大程度还原样本主要特征,但是这种方法不利于进行实时检测.基于 chi-square 方法在特征提取中的优点,本文利用该方法对每类正常模型提取 chi-square 值不为零的属性,得到与该类正常数据相关性较高的属性,并计算这些属性的影响权值因子.定义 n 为数据总数, m 为属性总数, i 为第 i 个数据, j 为第 j 个属性.算法如下:

(1) 输入一类正常数据,每列代表该类数据第 j 个属性值的集合,计算此列的平均值;

$$x'_j = \sum_{i=1}^n x_{ij} / n \quad (1)$$

(2) 计算每列的 chi-square 值,用于表示该属性的卡方值;公式如下:

$$CS_j = \sum_{i=1}^n (x_{ij} - x'_j)^2 / x'_j \quad (2)$$

(3) 求出各属性的影响权值因子;

$$\text{eff}_j = \text{cs}_j / \sum_{j=1}^m \text{cs}_j \quad (3)$$

3.2 计算属性特征值

一般属性特征值的计算是采用平均值法,但是该方法存在一定的误差.逆向云发生器是实现定性语言值与定量数值之间的不确定性转换模型,它将一定数量的精确数据转换为以恰当的定性语言值表示的概念.本文采用该方法计算正常数据模型的属性特征值, X 信息为正常训练数据模型中每个属性的值,每个模型可以通过多个属性云来表达.方法具体描述如下:

(1) 计算输入数据各属性的期望:

$$E_{x_j} = \sum_{i=1}^n x_{ij} / n \quad (4)$$

(2) 采用逆向云算法计算正常数据模型各属性云的超熵与熵,得到属性云的其他二个数字特征.计算公式为式(5)、(6).

$$E_{n_j} = \sqrt{\frac{\pi}{2}} * \frac{1}{n} \sum_{i=1}^n |x_{ij} - E_{x_j}| \quad (5)$$

$$H_{e_j} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - E_{x_j})^2 - E_{n_j}^2} \quad (6)$$

(3) 返回步骤(1),计算其他正常数据模型属性云的数字特征,完成数据建模过程.

3.3 入侵检测模块

测试数据通过入侵检测模块计算正常数据模型与新数据之间的差异度,并与影响权值因子求得综合评价值来判断新数据是属于入侵数据还是正常数据.采用 X 条件正向云发生器实现属性的差异度计算. X 条件正向云发生器算法如下:

(1) 生成超熵为标准差,以熵为中心的正态随机数 E'_{n_j} :

$$E'_{n_j} = \text{Nordata}(E_{n_j}, H_{e_j}) \quad (7)$$

(2) 输入每条测试数据的属性 x_{ij} ,计算数据以 E'_{n_j} 为标准差、期望 E_{x_j} 为中心的正态随机函数的隶属度 μ_{ij} :

$$\mu_{ij} = \exp \left[-\frac{(x_{ij} - E_{x_j})^2}{2E'^2_{n_j}} \right] \quad (8)$$

(3) 利用隶属度来表示正常数据与测试数据属性云的差异度,结合利用方差值计算出的影响权值因子,得到测试数据的综合评价值,计算公式为式(9).将该综合评价值与算法设置的阈值进行比较,若大于阈值,则判定为入侵数据,否则反之.

$$\sum_{j=1}^n [(1 - \mu_{ij}) * \text{eff}_j] \quad (9)$$

基于正态云模型的入侵检测方法如图 2 所示.

4 实验结果与分析

实验采用 KDD Cup 1999 标准入侵检测数据集^[7]. KDD99 CUP 数据集是由 DARPA (Defence Advanced Research Projects Agency) 和麻省理工学院的 Lincoln 实验室提供的入侵数据集采集样本.该数据集共有近 500 万条样本,每个样本包含 41 个特征属性,其中 7 个符号特征(离散属性)、34 个数值特征(连续属性).实验数据除包含正常连接数据外,

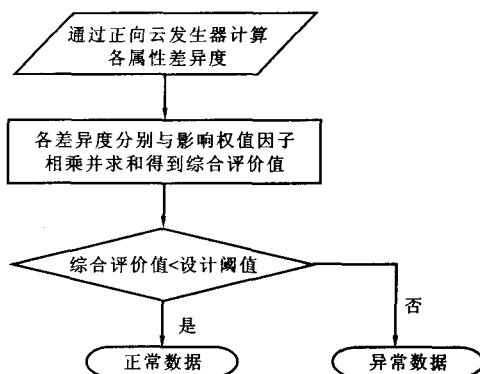


图 2 入侵检测算法流程图

还包含了四大入侵行为,分别是 R2L 远程攻击、Probe 刺探攻击、DoS 拒绝服务、U2R 获取根权限.实验从数据集中随机选取了训练数据 7 032 条,测试数据 16 408 条.训练集中 DoS 攻击 1 114 条,Probe 攻击 38 条,R2L 攻击 9 条,U2R 攻击 20 条;测试数据中 U2R 攻击 33 条,DoS 攻击 2 581 条,R2L 攻击 11 条,Probe 攻击 78 条.

实验在 512MB 内存、Intel Celeron 3.0GHz CPU、Matlab 语言编程、Windows XP 操作系统环境下进行.实验采用错误判断样本总数、检测率和误报率三个评价指标来评价算法检测效果.判断阈值是需要人为设定的参数,直接影响着入侵判断的结果.判断阈值与错误判断样本数、检测率和误报率关系.

实验表明,检测率和误报率随着判断阈值的增加而降低,为了保证较小的错误判断样本数,在下面的测试中设定判断阈值为 1.8,测试数据结果如表 1 所示.

表 1 实验测试数据表

攻击类型	攻击名称	检测率/%		误报率/%	
		本文算法	文献[5]	文献[5]	本文算法
DOS	Smurf	99.02	95.71	1.44	1.14
PROBE	Satan	99.45	93.22	2.62	1.55
R2L	Guess_passwd	99.71	91.54	1.87	1.24
U2R	Perl	98.78	85.34	2.71	1.58

通过实验数据可知文献[5]提出的入侵检测算法的检测率平均在 85%~96%之间,误报率平均在 1.47%~2.71%之间.而我们通过对不同攻击类型进行测试,检测率平均在 98%~99.7%之间,平均值为 99.247 5%,误报率平均在 1.14%~1.58%之间,平均值为 1.375%.说明本文算法新提出的算法相比于同类算法具有较高的检测率和较低的误报率,有效解决了目前入侵检测算法存在的问题.

(下转第 40 页)

5 结束语

本文分析了 Wang 等人提出的代理盲多重签名方案,指出该签名方案不满足强盲签名方案的不可追踪性,代理签名人只要保留自己的盲签名和相关数据,待消息拥有者公布签名消息后,代理签名人就可以通过计算找到自己的盲签名行为与公布的签名之间的联系,继而实现对消息拥有者的追踪.在 Wang-Li-Zhou 方案的基础上,提出了一种新的代理盲多重签名方案,安全性分析表明,该方案满足不可追踪性和其他密码学特性,在电子投票系统和电子现金系统等需要保护用户匿名性的场合具有很好的实用价值.

参考文献:

- [1] 赵泽茂,刘凤玉.广义 EL Gamal 型弱盲签名的构造方法[J].计算机工程与设计,2005,26(2):2168-2169
- [2] Mambo M, Usuda K, Okamoto E. Proxy signatures for

delegating signing operation[C]//Proc. of the 3rd ACM Conference on Computer and Computer and Communications Security. New York: ACM Press,1996:48-57.

- [3] Lin W D, Jan J K. A security personal learning tools using a proxy blind signature scheme[C]//Proc. of International Conference on Chinese Language Computing. USA: Chinese Language Computer Society Knowledge Systems Institute, 2000:273-277.
- [4] 李萍,张建中,魏峰.一种基于身份的代理盲签名方案[J].微电子学与计算机,2006,23(2):52-54.
- [5] 符景云,陈鲁生.一种新的强代理签名方案及强代理多重签名方案[J].南开大学学报,2006,39(2):35-39.
- [6] 王洪涛,李大兴,周大水.一种新的基于椭圆曲线的代理盲多重签名[J].微电子学与计算机,2007,24(2):160-163.

作者简介:

秦艳琳 女,(1980-),博士研究生,讲师.研究方向为信息安全与密码学.

吴晓平 博士,教授.

(上接第 36 页)

5 结束语

提出一种基于正态云模型的网络入侵检测算法,算法首先采取特征提取并计算影响权值因子,引入云发生器计算属性特征值和差异度,应用综合评价进行入侵判断,减少了单个属性造成的局部性影响.实验结果表明,该方法能够比较有效地检测真实网络数据中的未知入侵行为,在一定程度上解决了目前部分入侵检测算法存在的检测率低、误报率高的问题.

参考文献:

- [1] 林果园,黄皓,张永平.入侵检测系统研究进展[J].计算机科学,2008,35(2):69-74.
- [2] 唐勇,卢锡城,王勇军.攻击特征自动提取技术综述[J].

通信学报,2009,30(2):96-105.

- [3] 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):158-164.
- [4] 李德毅,刘常昱.论正态云模型的普适性[J].中国工程科学,2004,6(8):28-34.
- [5] 姜伟,高知新,李本喜.基于多维云模型的入侵检测[J].计算机工程,2006,32(24):155-156.
- [6] 刘常昱,李德毅,杜鸷,等.正态云模型的统计分析[J].信息与控制,2005,24(2):236-239.
- [7] KDD Cup 1999 Data set[EB/OL]. [2009-11-10]. <http://archive.ics.uci.edu/ml/databases/kddcup99/kddcup99.html>.

作者简介:

尹向东 男,(1976-),硕士,副教授.研究方向为网络安全、QoS.