基于云模型的网络攻击检测方法及其性能分析

谢立春[1,2] 张春琴[1]
(浙江工业职业技术学院 绍兴 312000)[1] (东南大学电气工程学院 南京 211189)[2]

摘 要 为了有效判断网络数据包是否存在被攻击的可能性,提出了一种新的基于云模型的检测算法DMCM(Detection Method based on Cloud Model)。该算法首先结合数据包属性的离散度和偏差定义了状态指标,并根据云模型给出了标准差分布的计算流程,以此判断数据包的异常状况。最后,通过OPNET和MATLAB进行仿真实验,深入研究了影响该算法的关键因素,同时与其它算法之间进行了性能对比,结果表明DMCM具有较好的适应性。

关键词 数据包,攻击,检测,异常,云模型
中图法分类号 TP393 **文献标识码** A

Detection Method and Performance Analysis of Network Attacks Based on Cloud Model

XIE Li-chun[1,2] ZHANG Chun-qin[1]
(Zhejiang Industry Polytechnic College, Shaoxing 312000, China)[1]
(School of Electrical Engineering, Southeast University, Nanjing 211189, China)[2]

Abstract In order to effectively determine whether network packets were attacked or not, a new detection algorithm DMCM(Detection Method based on Cloud Model) was proposed based on the cloud mode. Firstly, the state indicator of each packet was defined in terms of the discreteness and deviation of the properties of packets. Then, the process of computing the distribution of standard deviation was presented based on the cloud model. The distribution was used to determine the anomaly states of packets. Finally, a comprehensive experiments were conducted to study the performance of the DMCM algorithm through simulation using OPNET and MATLAB. Experimental results show that the proposed algorithm performs better than other algorithms in terms of adaptability.

Keywords Packet, Attack, Detection, Anomaly, Cloud model

随着Internet的迅速发展,网络信息的安全问题日渐突出,它已经成为了计算机领域的一个重大课题,引起了学者们的普遍关注。现阶段,网络攻击的手段和花样不断更新、层出不穷,根据攻击意图大致可以分为如下4大类:Denial of Service Attacks(DoS),Remote to Local Attacks(R2L),User to Root Attacks(U2R),Probes。如何有效地检测和防御攻击成为当前研究的热点和重点[1-9]。

入侵检测是防御网络攻击的重要手段。入侵检测系统(Intrusion Detection System,IDS)通过对网络数据流或访问记录的分析,识别和发现带有攻击性的网络行为。入侵检测主要分为误用检测(misuse detection)和异常检测(anomaly detection)。误用检测需要提前定义攻击的行为特征并写成规则,系统根据规则进行网络攻击行为的匹配。因此,该方法对当前最新的攻击行为无能为力。异常检测可以有效克服误用检测的缺陷,已成为目前IDS研究的主要方向。该技术建立系统或用户的正常行为模式,通过对检测系统或用户的实际行为模式和正常模式之间的比较和匹配来检测入侵行为,其特点是不需要过多有关系统缺陷的知识,具有较强的适应性,能够检测出未知的网络攻击。

目前国内外学者针对异常检测做了大量工作,文献[10]对历史数据分析建立正常的参考基线范围,一旦超出此范围就判断其为异常行为功能。文献[7]使用一般似然比方法,考虑了两个相邻时间窗口以及由这两个窗口构成的合并窗口,并计算各窗口序列残差的联合似然比,与预先设定的阈值比较,当超过预先设定的阈值时,则认为窗口边界为异常点。文献[9]针对异常网络行为会偏离正常语法规则的特点,利用改进的隐马尔可夫模型建立了一种网络攻击检测方法,其原理建立在正常行文样本的学习基础上。文献[11]根据流量在数据包头属性上的分布情况,采用定长的滑动窗口,随着窗口滑动不断计算熵值,并采用Chi-Square(x^2)假设检验比较分布差异,如果当前分布和基准分布存在较大差异,则认为出现了异常,但是该假设只能对离散且取值空间不大的分布差异有效,当取值空间较大时,误报率将增大。文献[12]利用蝴蝶突变模型刻画数据包异常行为,建立了一种突变级数的异常流量状态检测方法,但检测性能不高,且容易产生较高的误报率。

上述基于异常检测的研究都是针对某一特定的网络行为或模型进行的,且存在检测性能低、误报率及漏报率高的特

本文受浙江省自然科学基金(y1080023)资助。
谢立春(1974—),男,副教授,主要研究方向为计算机网络,E-mail:zjsxxlc@163.com;张春琴(1977—),女,副教授,主要研究方向为网络安全,E-mail:zjsxzcq@163.com。

点。近年来，模糊性、随机性及不确定性等构成的云模型已成功应用到自然语言处理、数据挖掘、决策分析、智能控制、图像处理等众多领域，基于云模型的异常检测研究也引起了学者们的广泛关注。本文基于云模型提出了一种新的检测算法，并结合数据包属性的离散度和偏差给出了检测指标，同时通过获得数据包属性的标准差分布来判断是否存在被攻击的可能性。最后，以 OPNET 和 MATLAB 进行仿真实验，对比研究了该算法与其它算法之间的性能状况。本文第 1 节给出了数据包状态评价指标；第 2 节结合云模型建立了攻击检测方法；第 3 节对该检测算法进行仿真实验；第 4 节进行了总结。

1 数据包状态评价指标

由于 HTTP 协议中数据包可以看作是遵循一定标准的字符串，由具有 k 个通用属性的数据域组成。令数据包 $Y=[y_1,y_2,\cdots,y_k]$，其中 y_k 表示 Data 头部、Host 头部、源端 IP 地址、目的端 IP 地址等。对于 n 个数据包的样本集合 $Z=[Y_1,Y_2,\cdots,Y_n]$，则可以表示为：

$$Z=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix} \tag{1}$$

那么这 n 个数据包第 k 个属性可采用序列 $Z_k=[y_{1k},y_{2k},\cdots,y_{nk}]$ 来表示。令第 k 个属性的离散度为 $\beta(Z_k)$，则整个样本集合的离散度 $\beta(Z)$ 为：

$$\beta(Z)=\frac{1}{k}\sum_{i=1}^{k}\beta(Z_i) \tag{2}$$

同时，假设某属性 k 的平均离散度为 $\overline{\beta(Z_k)}$。这里采用标准差 λ 来刻画数据包属性与整体平均离散度之间的偏差：

$$\lambda=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\beta(Z_{ik})-\overline{\beta(Z_k)})^2} \tag{3}$$

λ 越大，意味着该数据包与标准样本偏离越远，越有可能被攻击篡改信息。

针对上述判断指标，传统方法常采用隐马尔科夫模型来检测数据包是否被攻击。假设源端发送某数据包序列 $Z=[Y_1,Y_2,\cdots,Y_n]$，在某状态 t 下目的端接受到数据包 Y_n 的概率为 $p(Y_n)$，状态 t 的转移概率为 $p(t)$，利用隐马尔科夫模型获得未被攻击的数据包序列 Z 的概率为：

$$p(Z)=\sum_{n}\prod_{t}p(Y_n)p(t) \tag{4}$$

根据贝叶斯理论，其先验概率为：

$$p(\varepsilon|Z)=\frac{p(Z|\varepsilon)\prod_{\varepsilon}(1+n_\varepsilon)^{n_{\varepsilon 1}+n_{\varepsilon 2}}}{p(Z)} \tag{5}$$

其中，ε 为状态 t 时的状态，与正常状态偏差为 λ；n_ε 为隐马尔科夫模型中总状态数；$n_{\varepsilon 1}$ 为状态 ε 转移个数；$n_{\varepsilon 2}$ 为状态 ε 观察值个数。进一步，可以得到最优的评估函数为：

$$f=\prod_{\varepsilon}(1+n_\varepsilon)^{n_{\varepsilon 1}+n_{\varepsilon 2}}\prod_{Y\in Z}p(\varepsilon|Y) \tag{6}$$

但是该检测方法需要大量的先验样本信息来建立匹配模板，并且检验阈值对样本结果具有较大影响。而云模型作为一种定性和定量间转换的不确定性模型，可以有效描述模糊性和随机性，并将模糊性和随机性集成在一起，从定性信息中获得定量数据。

2 基于云模型的攻击检测方法

云模型[13,14]采用期望 Ex、熵 En 以及超熵 He 来描述某一定性对象。期望 Ex 表示代表该对象的个体样本，熵 En 表示一个个体样本的可度量粒度（熵越大，粒度越大，范围将越宏观），代表可接受的取值范围；超熵 He 是熵 He 的不确定性度量，代表定性对象样本出现的随机性，用来阐述模糊性和随机性之间的关系。在云模型中，云团由一系列云滴构成，某个云滴代表定性对象的一次实现，并且云滴之间的顺序是无关的。云滴根据定义的规则发生器，在域内产生一个随机确定度，通过这一随机确定度来激活后件云发生器，以此产生新的云滴。

针对式(3)定义的标准差 λ，这里结合云模型建立新的算法 DMCM(Detection Method based of Cloud Model)，以此判断数据包是否被攻击。具体算法步骤如下所述。

(1)在某时刻 t，初始化网络参数。

(2)将待检测某数据包 $Y=[y_1,y_2,\cdots,y_k]$ 视作云滴，对待检测数据包 Y 的属性中的每一个取值 y_k，计算其分布函数 $g(y_k)$。

$$g(y_k)=\sum_{i=1}^{k}(v\eta(y_i)+z_i) \tag{7}$$

$$|g(y_k)-\sum_{i=1}^{k}(v\eta(y_i)+z_i)|<\theta \tag{8}$$

其中，$\eta(y_i)$ 为第 i 个属性的数学期望，θ 为一随机数，$z_i=\min(|g(y_i)|,|g(y_{i-1})|)$，$v=|g(y_i)-g(y_{i-1})|$ $(g(y_k)=0)$。

(3)根据数据分布函数 $g(y_k)$ 的波峰所在位置，将其属性值取为期望 Ex，如果满足 $g(y_k)>g(y_{k-1})$，则 $Ey_k=y_k$。

(4)同时，基于 $g(y_k)$ 来计算以 Ex 为期望的熵 He，并计算其数据分布函数 $G(He)$ 和离散度 $\beta(Z_k)$。

(5)重复执行上述步骤(2)－(4)，直至所有云滴完成，计算云团熵 He 的平均值 H 和平均离散度 $\overline{\beta(Z_k)}$。

(6)对于某云滴 k，定义其离散差度 $\Delta\beta=\beta(Z_k)-\overline{\beta(Z_k)}$，并根据式(9)对其执行变异操作，产生新的云滴 j：

$$j=\Delta\beta p_1+k(1-rand()) \tag{9}$$

其中，p_1 为变异概率，$rand()$ 为(0,1)之间的随机数。

(7)判断交叉概率 p_2 与随机数 $rand()$ 之间的关系，如果 $p_2>rand()$，则由式(10)对云滴 k 和 j 执行交叉操作，获得新云滴 k_1 和 j_1：

$$\begin{cases} k_1=jp_2+\dfrac{k}{1+e^{1-rand()}} \\ j_1=kp_2+\dfrac{j}{1+e^{1-rand()}} \end{cases} \tag{10}$$

(8)根据已知的 $G(k_1)$ 和 $G(j_1)$ 以及式(3)来计算标准差分布函数 $G'(k_1)$ 和 $G'(j_1)$，并得出基于云模型的特征量熵 $En(k_1)$、$En(j_1)$ 和超熵 $He(k_1)$、$He(j_1)$：

$$En(k_1)=\frac{1}{2}(G'(k)rand()+G'(k_1)(1-rand())) \tag{11}$$

$$En(j_1)=\frac{1}{2}(G'(j)rand()+G'(j_1)(1-rand())) \tag{12}$$

$$He(k_1)=\frac{1}{5\max(G'(k),G'(k_1))} \tag{13}$$

$$He(j_1)=\frac{1}{5\max(G'(j),G'(j_1))} \tag{14}$$

(9)获得云滴对应标准差 $\lambda(i)$ 的云模型曲线：

$$\lambda(i)=\exp(-\frac{(En(k_1)-En(j_1))^2}{|He(k_1)^2-He(j_1)^2|}) \tag{15}$$

(10)令 $i=i+1$，跳转到步骤(6)，重复计算标准差 $\lambda(i)$，直到完成所有云滴的标准差计算，获得标准差分布 $\lambda=$

$[\lambda(1),\lambda(2),\cdots,\lambda(k)]$，并判断每个 $\lambda(i)$ 是否超出规定阈值，如果超出，则存在被攻击的可能性。

3 性能测试与分析

为了验证所提 DMCM 算法的有效性，本文结合 OPNET 和 MATLAB 进行仿真实验。考虑到网络通信的复杂性及攻击的不确定性，本研究提出的 DMCM 算法可以采用一个统一的云模型实现定性概念与定量描述之间的不确定转换。本研究在 OPNET 中构建了大量拓扑节点进行了实验，以验证实验结果的有效性。通过大量实验发现，DMCM 算法针对不同的网络攻击技术都具有较好的检测效果，且误报率相对比较低。

为便于说明，本研究采用如图 1 所示的简单网络仿真拓扑结构进行描述：每个网络节点缓冲区为 1024kB，链路带宽为 20M，每个数据包大小为 512B，延时 10ms。其中，节点 S 作为数据源端(IP 地址设为 192.168.1.1)，接收端节点 D(IP 地址设为 192.168.1.100)接收数据，攻击端(IP 地址设为 192.168.1.2)为节点 f，不限时向图 1 网络发起攻击(包括 DoS、Probe、R2L 和 U2R 等)，其它节点为网络中转点(从点 a 到点 g 的 IP 地址分别设为 192.168.1.3 到 192.168.1.9)。令节点 S 处发送 $n=1000$/s 个数据包到节点 D，数据包表示为 $Y=[y_1,y_2,y_3]$，y_1 为数据长度，y_2 为数据时间戳，y_3 为数据源地址。在节点 D 处监听，对从节点 S 发送的数据包并进行分析，同时节点 f 发动 DoS 攻击，对比 DMCM 算法获取的结果，其数据包长度状态 y_1 的变化情况如图 2 所示。从图 2 可以看出，DMCM 算法检测的数据包长度状态 y_1 与监听收集的实际数据样本比较吻合。对其进行数据分析，其误差为 4.57%。

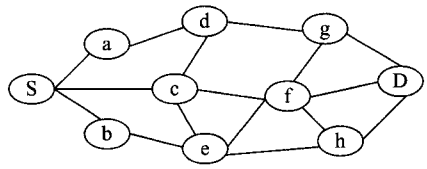

图 1 网络仿真拓扑结构

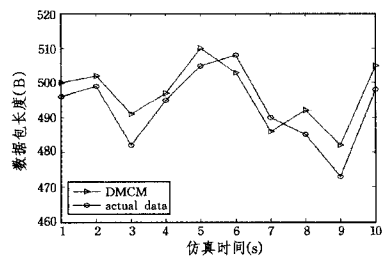

图 2 数据包长度检测结果比较

与本文同样关注网络异常检测的研究很多，本研究将与文献[15]提出的基于相对熵理论的多测度网络异常检测方法 RETMMAD(Based on Relative Entropy Theory Multi-Measure Anomaly Detection)进行对比。原因在于文献[15]已经与同类的 PHAD、ALAD、NETAD、FAD、EMERALD 等著名异常检测方法进行了比较，且 RETMMAD 算法针对不同的攻击均有较高的检测率和较低的误报率。表 1 中显示了当节点 f 发动 DoS、Probe、R2L 和 U2R 攻击时，RETMMAD 和 DMCM 算法检测成功率、漏报率以及误报率的比较情况。从表 1 可以看出，本文提出的 DMCM 算法较 RETMMAD 方法整体平均检测成功率提高了 4.69%。DMCM 算法针对 R2L 攻击存在 3.71%的误报率，其原因主要在于 R2L 攻击手段的多样性、攻击方式的复杂性，而相较于其它攻击种类，DMCM 算法具有具有一定的适应性，表现出了较好的攻击检测效果。

表 1 不同攻击种类下检测结果的比较

攻击种类 \ 算法	DMCM			RETMMAD		
	成功率%	漏报率%	误报率%	成功率%	漏报率%	误报率%
DoS	95.47	4.53	0	95.22	4.78	0
Probe	89.21	10.79	0	80.15	19.85	0
R2L	86.75	9.54	3.71	82.46	12.01	5.53
U2R	82.69	17.31	0	73.52	26.48	0
平均值	87.03	12.04		82.34	16.28	

由于最近的研究发现实际网络流量具有突发性和长相关性，这里利用 FARIMA(p,d,q)模型在节点 S 处来产生带有分形特征的数据包流量，然后对比 RETMMAD 和 DMCM 算法对分形流量的检测性能。FARIMA(p,d,q)模型的参数 p 和 q 为整数，d 为实数，并且对任意流量 $Y=(Y_t:t=0,1,2,\cdots)$需满足如下形式：

$$\Phi(B)\Delta^d Y_t=\Theta(B)\omega_n \tag{16}$$

其中，$\Phi(B)=1-\Phi_1(B)-\cdots-\Phi_p(B)^p$ 为自回归项，$\Theta(B)=1-\Theta_1(B)-\cdots-\Theta_q(B)^q$ 为滑动平均项，$\Delta=1-B$ 表示差分算子，而 Δ^d 为分形差分算子。在节点 S 处产生 FARIMA(p,d,q)特征的数据包流量，首先根据 $\Delta^d Y_t=\omega$ 和 $\Delta^d=(1-B)^d=\Gamma(-d+k)/(\Gamma(-d)\Gamma(k+1))$产生分形差分噪声 X_t，然后采用 ARMA 过程来生成 FARIMA(p,d,q)流量 Y_t。在图 3 中给出了 RETMMAD 与 DMCM 算法以及实际监听收集的数据包长度状态 y_1 的对比情况，其中设置节点 f 发动 Probe 攻击。从图 3 看出，DMCM 算法检测结果与实际监听收集的结果比较接近，而 RETMMAD 方法则相差较大，DMCM 和 RETMMAD 与实际监听的误差分别为 9.93%和17.32%。

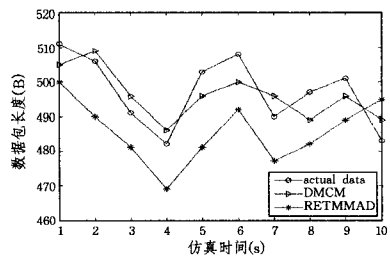

图 3 分形环境下数据包长度检测结果比较

最后，为了进一步分析 DMCM 算法性能状态，这里对关键参数变异概率 p_1 和交叉概率 p_2 进行研究。图 4 和图 5 分别给出了不同变异概率 p_1 和交叉概率 p_2 下 DMCM 算法检测误差的情况，其中节点 f 发动 DoS 攻击。从图 4 可以看出，在仿真初期，变异概率 p_1 越小，对应曲线的误差越小；而在仿真后期，变异概率 p_1 越大，对应曲线的误差越小。由于前期在节点 D 处聚集的数据包较少，采用较小的变异概率就能够有效地改善离散差度以及新的云滴性能；而在仿真后期，在节点 D 处聚集的数据包增多，此时加大变异概率对新云滴性能的改善才能起到显著效果，因此此时曲线出现了突变。而在图 5 中也存在类似现象，在仿真初期交叉概率 p_2 越大，对应曲线的误差越小；而在仿真后期交叉概率 p_2 越小，对应曲线的误差越小。在仿真初期云团整体性能较低，此时加大

(下转第 389 页)

rence on Computer Science and Service System. IEEE Computer Society, 2011

[2] 马礼,郑纬民. 信息网格环境下的综合信任度评价模型[J]. 清华大学学报(自然科学版),2009(4):599-603

[3] Rahman A A, Hailes S. Supporting Trust in Virtual Communities[C]// Proceedings of the 33rd Hawaii International Conference on System Sciences. Washington DC, USA: [s. n.], 2000: 6007-6016

[4] Josang A, Knapskog S J. A metric for trusted systems[C]// Proceedings of the 21st National Security Conference. NSA, 1998

[5] Azzedin F, Maheswaran M. Towards Trust-Aware Resource Management in Grid Computing Systems[C]// Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid. 2002: 1-6

[6] 李小勇,桂小林. 可信网络中基于多维决策属性的信任量化模型[J]. 计算机学报,2009,32(3):405-416

[7] Josang A, Hird S, Faccer E. Simulating the Effect of Reputation Systems on e-Markets[C]// the Proceedings of the First International Conference on Trust Management. Creece, 2003: 179-194

(上接第 380 页)

交叉概率有利于云团性能朝最优状态转移,同时避免陷入局部最优;而在后期随着云团整体性能的提高,云滴之间的性能差异变小,此时小的交叉概率便能优化云滴性能,并且交叉概率越小系统开销越小,对应的误差越小。

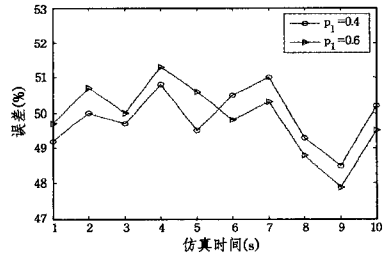

图 4 不同变异概率 p_1 下数据包长度检测结果比较

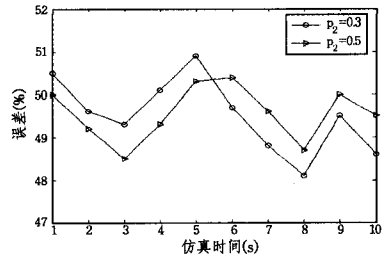

图 5 不同交叉概率 p_2 下数据包长度检测结果比较

结束语 在对已有网络异常检测方法分析的基础上,本研究结合云模型提出了一种新的检测算法 DMCM。该算法首先针对数据包属性的离散度和偏差进行定义,给出了样本判断指标,同时基于云模型定义了数据包的期望 Ex、熵 En 以及超熵 He,并通过计算属性标准差分布来判断数据包是否存在被攻击的可能。最后,以 OPNET 和 MATLAB 进行仿真实验,对比研究了该算法与 RETMMAD 算法以及实际监听的情况,结果发现 DMCM 性能较 RETMMAD 算法有大幅度提高,和实际情况比较吻合。在今后的研究中,可以考虑针对 DoS、Probe、R2L 和 U2R 等各种攻击的具体状况,来建立完善的检测方法。

参 考 文 献

[1] Varadharajan V, Tupakula U. Counteracting security attacks in virtual machines in the cloud using property based attestation [J]. Journal of Network and Computer Applications, 2014, 40(7): 31-45

[2] Younis Y A, Kifayat K, Merabti M. An access control model for cloud computing [J]. Journal of Information Security and Applications, 2014, 19(1): 45-60

[3] 杨宏宇,常媛. 基于 K 均值多重主成分分析的 App-DDoS 检测方法[J]. 通信学报,2014,35(5):16-24

[4] Yu Yong, Niu Lei, Yang Guo-min, et al. On the security of auditing mechanisms for secure cloud storage [J]. Future Generation Computer Systems, 2014, 30(1): 127-132

[5] 王会梅,鲜明,王国玉. 基于扩展网络攻击图的网络攻击策略生成算法[J]. 电子与信息学报,2011,33(12):3015-3021

[6] 谢柏林,余顺争. 基于应用层协议分析的应用层实时主动防御系统[J]. 计算机学报,2011,34(3):452-463

[7] 程宏兵,容淳铭,黄晓,等. 高效的攻击检测与数据融合算法[J]. 通信学报,2012,33(9):85-94

[8] 张玲,白中英,罗守山,等. 基于粗糙集和人工免疫的集成入侵检测模型[J]. 通信学报,2013,34(9):166-175

[9] 席荣荣,云晓春,张永铮,等. 一种改进的网络安全态势量化评估方法 [J]. 计算机学报,2014,31(3):95-101

[10] 储泽楠,李世扬. 基于节点生长马氏距离 K 均值和 HMM 的网络入侵检测方法设计[J]. 计算机测量与控制,2014,37(10):1-12

[11] 张冰涛,王小鹏. 面向 WSN 安全路由协议的自适应威胁模型[J]. 计算机应用研究,2014,31(4):1208-1211

[12] Vissers T, Somasundaram T S, Pieters L, et al. DDoS defense system for Web services in a cloud environment [J]. Future Generation Computer Systems, 2014, 37: 37-45

[13] 朱建明,宋彪,黄启发. 基于系统动力学的网络安全攻防演化博弈模型[J]. 通信学报,2014,35(1):54-61

[14] 刘禹,李德毅,张光卫,等. 云模型雾化特性及在进化算法中的应用[J]. 电子学报,2009,37(8):1651-1658

[15] 张亚玲,韩照国,任娇霞. 基于相对熵理论的多测度网络异常检测方法[J]. 计算机应用,2010,30(7):1771-1774

[16] Worku S G, Xu Chun-xiang, Zhao Ji-ning, et al. Secure and efficient privacy-preserving public auditing scheme for cloud storage [J]. Computers & Electrical Engineering, 2014, 40(5): 1703-1713