关于云环境下共存攻击多目标响应策略机器学习模式的研究

Research on Machine Learning Model for Multi-target Response Strategy to Co-resident Attacks in Cloud Environment

吴志祥，曲　谱，李　耕（中国联通黑龙江省分公司，黑龙江 哈尔滨 150001）
Wu Zhixiang，Qu Pu，Li Geng（China United Network Communication Co.，Ltd.，Heilongjiang Branch，Harbin 150001，China）

摘　要：
云环境可以通过虚拟化技术共享软硬件资源，但是使用者在使用虚拟化平台时可能会面临额外的安全威胁。共存攻击是指攻击者利用共享基础设备的特性，来攻击共存在同一实体机上的其他虚拟机。使用机器学习来训练云环境共存攻击多目标响应系统，以最小成本和最小威胁检测目标共存攻击。实验结果显示，该模型具有2 000倍加速比的效率提升，同时获得87.9%高准确度的解答，在响应策略与时间效率上取得平衡。

关键词：
云环境；共存攻击；机器学习；云环境入侵检测系统
doi：10.12045/j.issn.1007-3043.2022.09.014
文章编号：1007-3043（2022）09-0065-06
中图分类号：TN915.08
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
Hardware and software resources can be shared through virtualization technology in cloud environment，while the users may face additional security threats when using virtualization platforms. Co-resident attack refers to an attacker taking advantage of the characteristics of the shared infrastructure to attack other virtual machines co-located on the same physical machine.It uses machine learning to train the multi-target response system for co-resident attack in cloud environment to detect co-resident attacks with the minimum cost and minimum threat. The experimental results show that the model has an efficiency improvement of 2 000 times the acceleration ratio and obtains the solution with high accuracy of 87.9%，and strikes a balance between response strategy and time efficiency.

Keywords：
Cloud computing；Co-resident attack；Machine learning；Cloud intrusion response system

引用格式：吴志祥，曲谱，李耕. 关于云环境下共存攻击多目标响应策略机器学习模式的研究[J]. 邮电设计技术，2022(9)：65-70.

0 前言

云环境可按照用户的需求，以低成本向使用者提供软硬件设备，使用者可以方便地使用服务，并且无需了解云环境的细节，也不需进行设备维护。虚拟化技术可以按照需求共享资源，提高了资源的使用率，图1给出了云环境虚拟化示意图。由于云环境的虚拟化特性，多个虚拟机会在同一主机上同时运行，如果虚拟机之间没有适当的隔离，攻击者可以从同一主机上取得其他虚拟机的相关信息。

2017年，Abazari等人提出了云环境共存攻击多目标检测系统，整理了虚拟机共存攻击和检测对策的类型，并提出一套模型来计算云环境整体的最佳响应对策。他们提出的模型考虑到虚拟机之间的共存时间长短，并在威胁和成本这2个冲突的目标中取得帕雷托最优解得到最佳响应对策。图2为帕雷托最优解示意图，其中F_1代表总威胁，F_2代表总成本。若有一个解x，当无法在不提升总成本的情况下再降低总威胁，且无法在不提升总威胁的情况下再降低总成本，那解x就为帕雷托最优解。在2个互相冲突的目标的情况

收稿日期：2022-08-08

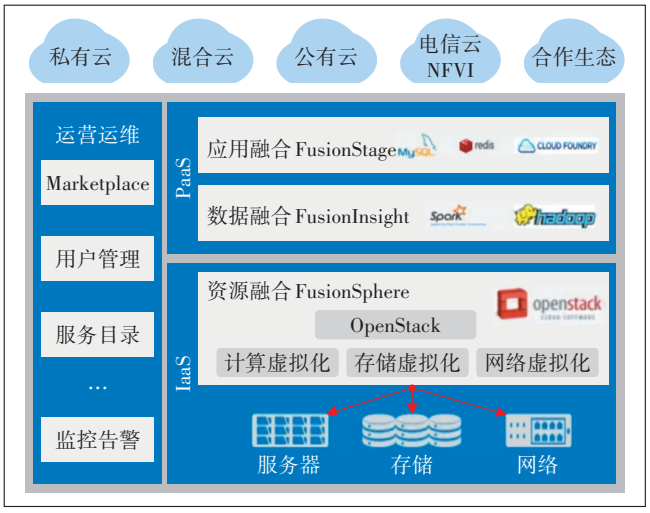

图1　云环境虚拟化示意图

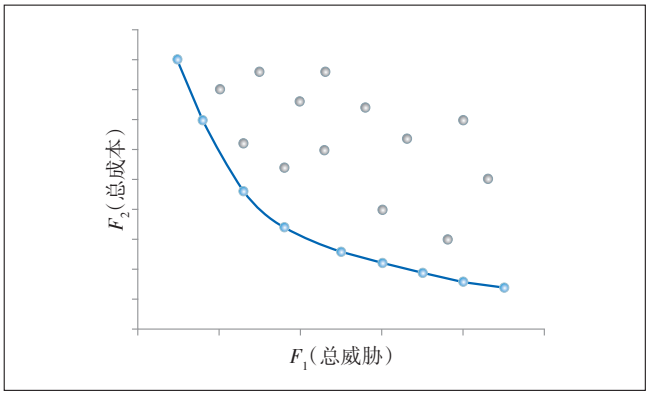

图2　帕雷托最优解示意图

下，会有多个帕雷托最优解。

实验发现，Abazari 等人的模型在具有大量虚拟机的云环境中计算检测对策需要耗费相当长的时间。因此，本文提出一个自动入侵检测系统的新模型，其使用机器学习来产生检测对策，并进行一系列实验来证明模型的效果。从实验结果可以看出机器学习可以以良好的准确度近似帕雷托最优解，并且相较于 Abazari 等人的模型在检测速度方面有 2 000 倍的提升。

1 背景知识

1.1 云环境概念

云环境计算根据服务类型可以分为 3 类 SPI：软件即服务（SaaS）、平台即服务（PaaS）和基础架构即服务（IaaS）。其中 IaaS 是云环境供应商通过虚拟化技术在一部大型的实体主机上建立多个虚拟机，根据使用者需求动态分配资源，使用者可以根据自己的需求租借适量的服务。本文讨论的云环境共存攻击就是在 IaaS 架构下隐藏的威胁。

1.2 云环境共存攻击类型

共存攻击可分为 3 种类型，第 1 种类型包含信息窃取和侧通道攻击，使受害者隐私泄露，造成不可逆的伤害；第 2 种类型是攻击者利用虚拟机之间的通信来传播恶意程序；第 3 种类型，攻击者的目标是得到不公平的共享资源，导致同一实体机上的其他虚拟机受到阻断服务攻击。

1.3 虚拟机共存时间关系

根据研究文献，虚拟机受到共存攻击的机率与共存时间呈正相关，通过虚拟机共存时间关系图可以获得虚拟机之间的共存时间，并用来计算云环境中所有虚拟机的威胁等级。虚拟机共存时间关系图如图 3 所示，图 3 中的节点代表虚拟机（VM），其中虚拟机可分为 2 种类型：集合 M 表示恶意 VM，由虚线方框表示；集合 S 表示一般 VM，由实线方框表示。节点之间的连线代表虚拟机曾经或正在同一实体机中共存，连线上的数值代表共存时间长短，时间会经过正规化变成 0~1 的数值，由 $w_{i,j}$ 表示 VM_i 和 VM_j 的共存时间，举例来说，图 3 中 VM_2 和 VM_3 的共存时间 $w_{2,3} = 0.8$，VM_8 和 VM_9 的共存时间 $w_{8,9} = 0.2$。

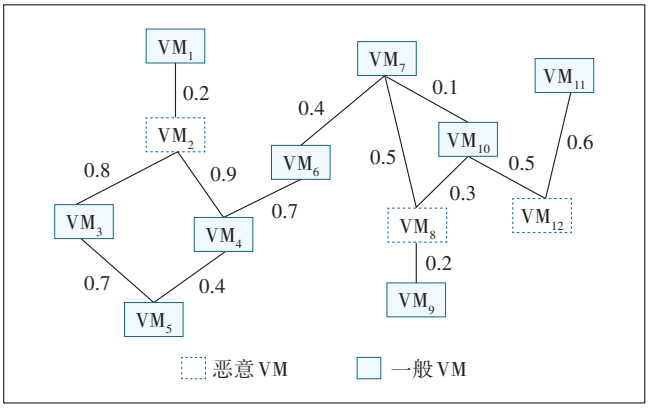

图3　虚拟机共存时间关系图

1.4 响应对策表

每种攻击都可以通过特定的对策来防御。响应对策表如表 1 所示，其中 SC 表示侧通道攻击，MP 表示恶意软体传播，DoS 表示阻断服务攻击，Y 表示该对策可防御此种攻击，N 表示该对策无法防御此种攻击。表 1 中列出了每个共存攻击的对策，总共可分为 5 个对策，对策 6 表示不进行任何动作，以下对每个对策进行详细说明。

表1　响应对策表

编号	VM属性	防御方法	SC	MP	DoS	成本
1	M/S	虚拟机迁移	Y	Y	Y	增加功耗且性能下降，成本高
2	M	更改管理程序	Y	N	N	增加能耗，成本中
	S	更改数据的存储模式	Y	N	N	增加能耗，成本中
3	M/S	通信隔离	N	Y	N	网络重新配置，成本低
4	M/S	限制资源分配	N	N	Y	增加能耗，成本中
5	M/S	限制网络流量速率	N	Y	Y	网络重新配置，成本低
6	M/S	不进行任何动作	N	N	N	无成本

a）第1种对策是虚拟机迁移，其能够防御所有共存攻击。

b）第2种对策能防御侧通道攻击，针对攻击者的虚拟机，增加其恶意行动的延迟。如果是受害虚拟机，则可以将硬盘、CPU和内存等共享资源的存取方式改为一致的模式，防止攻击者从受害者的行为中获取信息。

c）第3种对策可以防止恶意软体传播，利用虚拟交换机的功能重新配置虚拟网络。

d）第4种对策可以防御阻断服务攻击，主要是通过更改管理程序的配置降低最大计算负载或内存容量，限制攻击者的资源使用。

e）第5种对策是通过限制相关虚拟机允许的最大流量速率，防御恶意软体传播和阻断服务攻击。

f）第6种对策是不进行任何动作。

1.5 机器学习

机器学习属于人工智能的范畴，是一种让机器从学习到推理的过程，机器学习的演算法结合了各种学科，如概率学、统计学和逼近论等，通过组合这些数学模型设计出让计算机可以自我学习的演算法。从训练资料中截取重要特征，训练时通过演算法将特征进行分析获得规律得到模型，最后就可以利用模型对新资料进行预测。

机器学习的概念如图4所示，Abazari等人的模型是使用Matlab的fgoalattain函数来计算多目标优化问题，目标是指总响应成本和总威胁，本文的研究在降低威胁的同时也降低成本。而在机器学习演算法方面，本文对Lasso、Elastic Net、Ridge Regression这3种机器学习演算法进行了测试，先制造200笔资料，并随机选取80%作为训练资料，20%作为测试资料，接着分别使用3种机器学习演算法，平均准确度如表2所示。由表2可知，Ridge Regression的准确度较高，所以

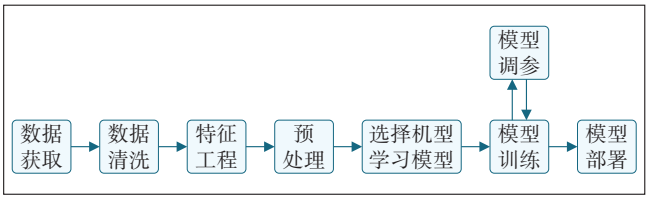

图4　机器学习概念图

表2　机器学习演算法准确度比较

机器学习算法	Lasso	Elastic Net	Ride Regression
平均准确度/%	63.7	65.6	70.2

本文选择Ridge Regression作为机器学习的演算法。

1.6 Scikit-learn

Scikit-learn简称SKlearn，是Python机器学习和资料分析的开源套件，其中包含许多知名的机器学习演算法，也内建许多知名的资料集。在演算法方面，依照演算法功能分为6个类别，如图5所示，简易流程图可以让使用者根据资料集的形态来挑选适合的演算法。

2 研究方法

2.1 计算威胁等级

虚拟机受到共存攻击的机率与共存时间呈正相关，因此，为了计算云环境中所有虚拟机的威胁等级，首先必须拥有虚拟机共存时间关系图，还有检测到的攻击者所操控的恶意虚拟机集合M。攻击者进行攻击时，其虚拟机通常会存在资源大量使用、cache miss等异常现象，一般可通过持续的云环境监控来发现恶意虚拟机。如第1章所述，云环境共存攻击可分为3个类型：侧通道攻击、恶意软体传播和阻断服务攻击，因此设定虚拟机威胁等级为三元组（triplet）t_i，分别表示3种攻击的机率，如式（1）所示。

$$t_i = \left[t_i^{SC},\ t_i^{MP},\ t_i^{DOS}\right] \tag{1}$$

t_i是代表VM_i被攻击的机率（威胁等级），一个受害VM会因为和多个不同恶意VM共存，造成被攻击的机率提升，威胁等级必须包括任一威胁发生的机率与多个威胁同时发生的机率，因此对Abazari等人的威胁等级计算方法进行修改，实际的计算方式如式（2）所示。

$$\begin{cases} t_i = [1 \quad 1 \quad 1] - \prod_{j=1}^{n}\left([1 \quad 1 \quad 1] - t_j \times w_{i,j}\right) \\ subj.to \\ VM_j \in M \end{cases} \tag{2}$$

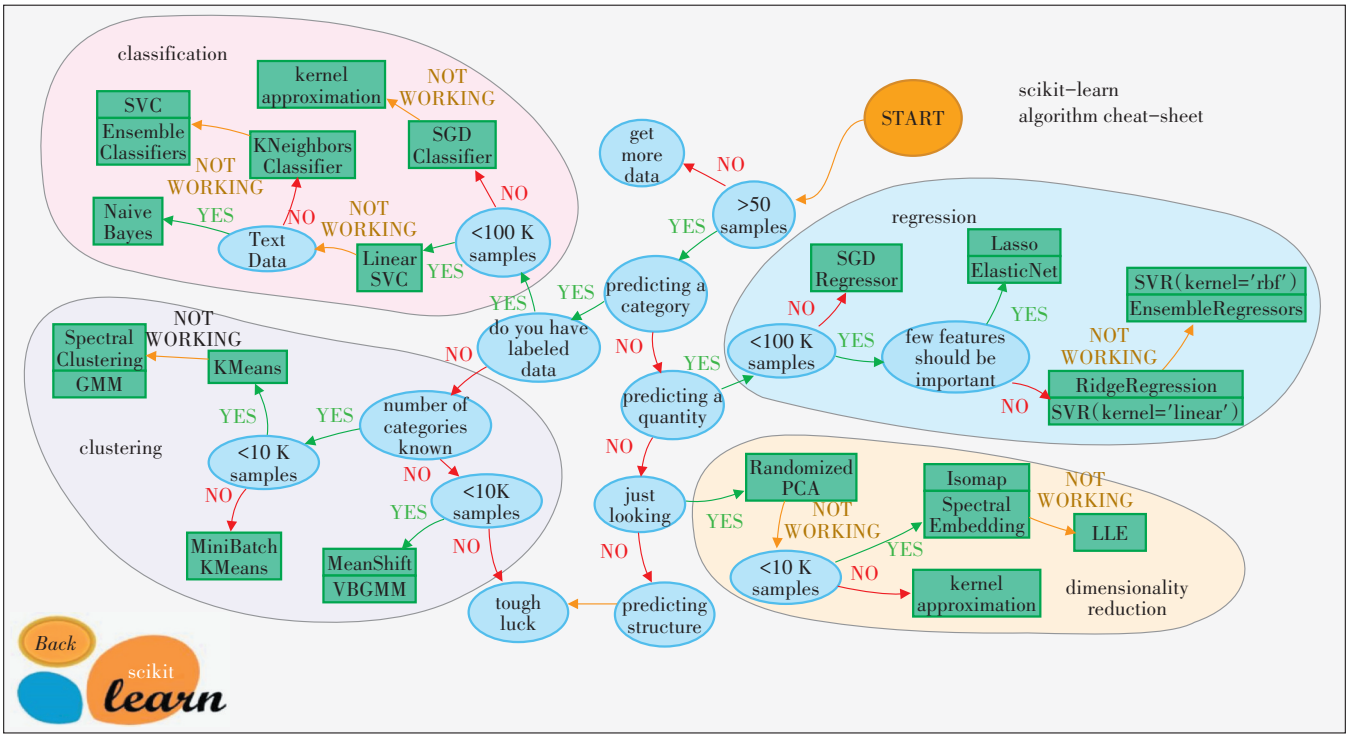

图5　Scikit-learn演算法挑选流程图

首先，VM_j要属于恶意虚拟机集合M，才需要计算其可能对VM_i造成的威胁，n为总虚拟机个数，$w_{i,j}$表示VM_i和VM_j的共存时间。$t_i \times w_{i,j}$表示VM_i被VM_j攻击的机率，$[1\ \ 1\ \ 1]-t_j \times w_{i,j}$就是不被$VM_j$攻击的机率。多个不被攻击的机率求积就等于完全不被攻击的机率，最后由$[1\ \ 1\ \ 1]$减去完全不被攻击的机率就是包括任一威胁发生的机率与多个威胁同时发生的机率。

以图3为例，给予恶意虚拟机随机的威胁等级，并根据式(2)计算一般虚拟机的威胁等级，如图6所示。假设要计算VM_4的威胁等级，VM_4和VM_2、VM_5、VM_6曾经共存或正在共存，当中只有VM_2是恶意虚拟机，所以$t_4=[1\ \ 1\ \ 1]-([1\ \ 1\ \ 1]-t_2 \times 0.9)=t_2 \times 0.9$。另外，假设要计算$VM_{10}$的威胁等级，$VM_{10}$和$VM_7$、$VM_8$、$VM_{12}$曾经共存或正在共存，当中$VM_8$和$VM_{12}$是恶意虚拟机，所以$t_{10}=[1\ \ 1\ \ 1]-([1\ \ 1\ \ 1]-t_8 \times 0.3) \times ([1\ \ 1\ \ 1]-t_{12} \times 0.5)$。

云环境的总威胁等级T，就是将所有虚拟机的威胁等级做加总，计算方式如下。

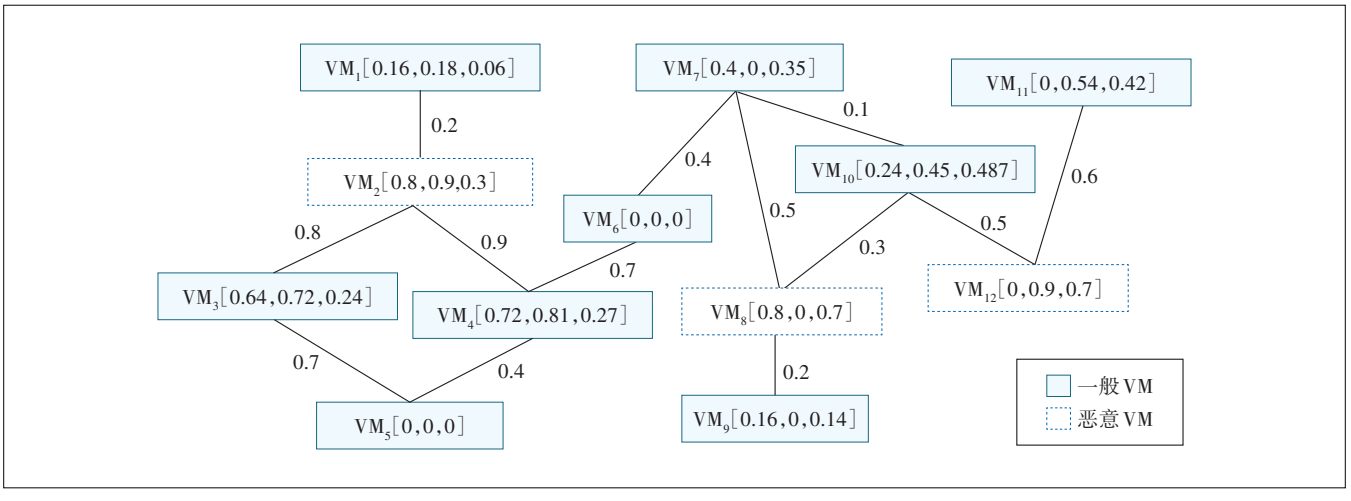

图6　虚拟机威胁等级示意图

$$T = \sum_{i=1}^{n} t_i \tag{3}$$

2.2 响应对策矩阵和成本向量

首先，定义矩阵 $C = \left[c_1, c_2, \cdots, c_q\right]^T$ 表示响应对策，如式(4)所示，其中，$c_i = \left[c_i^{SC}, c_i^{MP}, c_i^{DOS}\right]$，$q$ 代表对策总数量，本文中 $q = 6$。式(5)的 R_C 是响应成本向量，r_{c_i} 代表对策 c_i 耗费的成本。

$$C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_q \end{bmatrix} = \begin{bmatrix} C_1^{SC} & C_1^{MP} & C_1^{DOS} \\ C_2^{SC} & C_2^{MP} & C_2^{DOS} \\ \vdots & \vdots & \vdots \\ C_q^{SC} & C_q^{MP} & C_q^{DOS} \end{bmatrix} \tag{4}$$

$$R_C = \begin{bmatrix} r_{c_1} \\ r_{c_2} \\ \vdots \\ r_{c_q} \end{bmatrix} \tag{5}$$

每种对策能响应的攻击类型不同，所耗费的成本也不同。根据式(4)，由表1的SC、MP和DOS 3个栏位产生响应对策矩阵 C，如式(6)。举例来说，$C(3,2) = 1$ 代表对策3可以应对恶软件传播攻击，$C(4,3) = 1$ 代表对策4可以应对拒绝服务服务攻击。由表1的成本栏位产生成本向量 R_C，如式(7)。$R_C(3) = 0.2$ 代表对策3拥有低的耗费成本。

$$C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{6}$$

$$R_C = \begin{bmatrix} 1 \\ 0.5 \\ 0.2 \\ 0.4 \\ 0.3 \\ 0 \end{bmatrix} \tag{7}$$

2.3 资料生成

Abazari等人提出的云环境共存攻击多目标响应系统，已经能针对云环境整体选择出最适合的响应对策，但是存在响应时间过久的问题。因此，本文的训练数据就由Abazari等人提出的模型来产生，步骤如下。

a）产生一个随机的VM共存时间关系图。

b）随机挑选约30%的VM作为恶意VM。

c）随机产生恶意VM威胁等级。

d）计算每个VM的威胁等级。

e）使用Matlab的fgoalattain函数产生解答。

数据生成的部分使用Matlab完成。资料集产生之后，随机选取80%作为机器学习的训练数据，再随机选取20%数据作为测试数据。机器学习使用Scikit-learn的Ridge regression。Matlab的fgoalattain函数用来计算多目标优化的问题，包括线性和非线性，完整程序定义如式(8)所示。其中，weight、goal、b 和 b_{eq} 是向量；A 和 A_{eq} 是矩阵；$F(x)$、$c(x)$ 和 $c_{eq}(x)$ 是回传向量的函数，可以为非线性函数；x、lb和ub可以是向量或矩阵。

$$\text{minimize}_{x,y}\ \gamma\ such\ that \begin{cases} F(x) - \text{weight} \cdot \gamma \leqslant \text{goal} \\ c(x) \leqslant 0 \\ c_{eq}(x) = 0 \\ A \cdot x = b \\ A_{eq} \cdot x = b_{eq} \\ \text{lb} \leqslant x \leqslant \text{ub} \end{cases} \tag{8}$$

$F(x)$ 是要进行多目标优化的函数，优化的目标是云环境总威胁和总成本，如式(9)。

$$\begin{aligned} &\min_{x,y}\ \gamma\ subj.to \\ &F_1(X) - W_1\gamma \leqslant F_1^* \\ &F_2(X) - W_2\gamma \leqslant F_2^* \end{aligned} \tag{9}$$

$F_1(X)$ 是总威胁方程式，F_1^* 是其目标值；$F_2(X)$ 是总成本方程式，F_2^* 是其目标值。云环境的总威胁和总成本越低越好，因此 $(F_1^*, F_2^*) = (0,0)$，总威胁和总成本的方程式如式(10)所示。

$$\begin{cases} \text{Total Threat} = \sum_{i=1}^{n}\left[\left(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} - \sum_{j=1}^{q} x_{i,j} \times c_j\right) \times t_j\right] \\ \text{Total Cost} = X \times R_C \end{cases} \tag{10}$$

权重 W_1 和 W_2 分别代表威胁和成本的重要性，用来衡量较低的威胁或较低的成本，其中 $W = (W_1, W_2)$，$W_1 + W_2 = 1$，举例来说，当设定 W_1 接近0的时候，意味着得到的对策会有高的成本搭配极低威胁的成效。当设定的权重不同，就会得到不同的解 X，X 的定义如式(11)所示，$x = \{0,1\}$，其中 q 为对策数量，n 为虚拟机数量，解 X 会显示每个虚拟机该使用哪一个对策。

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,q} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,q} \\ & \vdots & & \\ x_{n,1} & x_{n,2} & \cdots & x_{n,q} \end{bmatrix} \tag{11}$$

最后，将权重 W、所有虚拟机的威胁等级 t、解 X 都转为一维阵列，整合成一列数据，数据表内容如表3所

示。生成多笔数据后输出成.csv文档，方便后续使用python读取进行机器学习。

表3　数据表内容

W_1	W_2	t_1^{SC}	t_1^{MP}	t_1^{DOS}	…	t_n^{SC}	t_n^{MP}	t_n^{DOS}	$X_{1,1}$	…	$X_{n,q}$

3 实验验证与结论

本文提出一个以机器学习响应云环境共存攻击的系统，为了验证提出的模型具有实用性，设计了以下实验并和Abazari等人的模型进行比较。在存在50个虚拟机的云环境模拟环境中，机器学习在响应时间方面比Matlab快2 296倍，准确度为85%，可见训练资料和虚拟机个数的增加可以提升机器学习的准确度。使用一组固定的威胁等级，并使用不同的权重，权重从$W=(0.1,0.9)$到$W=(0.9,0.1)$，每次增减0.1。分别使用Matlab和机器学习产生响应对策，然后计算使用对策后云环境总威胁和总成本，结果如图7所示。

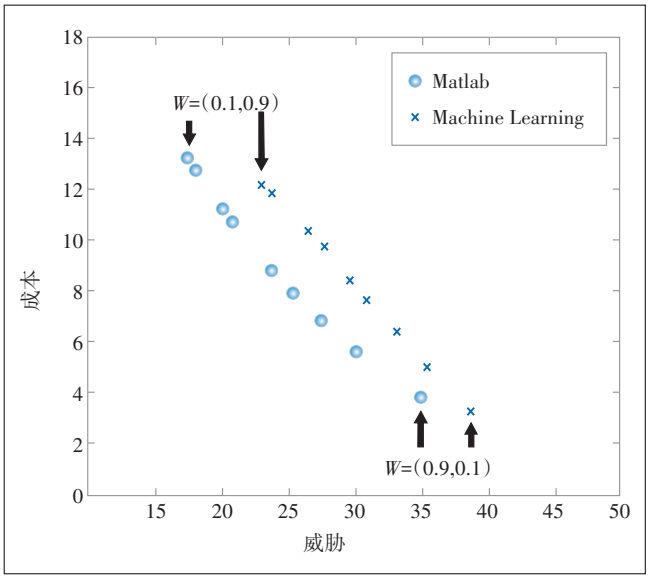

图7　不同权重下的威胁与成本比较图

图7中Matlab的点可视为帕雷托最优解，而机器学习的曲线也近似于帕雷托最优解，并不会偏离太多。在相同权重的设定下，机器学习的解有较高的威胁，但有较低的成本。

从实验结果可以看出，传统方法和本研究的机器学习模型都可以得到针对当前环境适当的响应对策，传统方法在总威胁和总成本的取舍下得到帕雷托最优解，而本研究的机器学习模型可以以高准确度得到近似的解。在响应时间方面，从实验的云环境来看传统方法需要的响应时间增长了约12倍，而机器学习模型仅增加1.23倍，资料笔数的增加和虚拟机量的增加也让机器学习的准确度有所提升，可见在更大量虚拟机的云环境中，本文的机器学习模型有更好的效果。

参考文献：

[1] ABAZARI F, ANALOUI M, TAKABI H. Multi-objective response to co-resident attacks in cloud environment[J]. International Journal of Information & Communication Technology Research, 2017, 9(3): 25-36.

[2] ALI M, KHAN S U, VASILAKOS A V. Security in cloud computing: opportunities and challenges[J]. Information Sciences, 2015, 305: 357-383.

[3] ANWAR S, ZAIN J M, ZOLKIPLI M F, et al. Response option for attacks detected by intrusion detection system[C]//2015 4th International Conference on Software Engineering and Computer Systems (ICSECS). Kuantan: IEEE, 2015: 195-200.

[4] BATES A, MOOD B, PLETCHER J, et al. On detecting co-resident cloud instances using network flow watermarking techniques[J]. International Journal of Information Security, 2014, 13(2): 171-189.

[5] NVD. Technical report 2015: CVE-2015-3456[S/OL]. [2022-04-25]. https://nvd.nist.gov/vuln/detail/CVE-2015-3456.

[6] ABAZARI F, ANALOUI M, TAKABI H. Effect of anti-malware software on infectious nodes in cloud environment[J]. Computers & Security, 2016, 58: 139-148.

[7] BALASUBRAMANIAN S, LOBBES M M, O'CONNELL B M, et al. Automated response to detection of threat to cloud virtual machine: 14495973[P]. 2014-09-25.

[8] ABAZARI F, ANALOUI M. Exploring the effects of virtual machine placement on the transmission of infections in cloud[C]//7'th International Symposium on Telecommunications (IST'2014). Tehran: IEEE, 2014: 278-282.

[9] SUNDARESWARAN S, SQUCCIARINI A C. Detecting malicious co-resident virtual machines indulging in load-based attacks[C]//Information and Communications Security. Cham: Springer International Publishing, 2013: 113-124.

[10] ALTUNAY M, LEYFFER S, LINDEROTH J T, et al. Optimal response to attacks on the open science grid[J]. Computer Networks, 2011, 55(1): 61-73.

[11] WUEEST C. Threats to virtual environments[Z]. Symantec, 2014.

作者简介：
吴志祥，毕业于吉林大学，高级工程师，硕士，主要从事网络安全技术研究、网络安全技术培训和网络安全事件应急处置工作；曲谱，毕业于哈尔滨工业大学，工程师，学士，主要从事网络安全防护，风险评估，安全事件分析等工作；李耕，毕业于黑龙江大学，工程师，学士，主要从事网络安全平台维护，语音呼叫中心平台安全防护与研究工作。