文章编号:2095-3046(2016)01-0068-06 DOI:10.13265/j.cnki.jxlgdxxb.2016.01.013

面向 IaaS 云平台的用户异常行为检测方法

郑剑, 周艳丽, 刘聪
(江西理工大学信息工程学院,江西 赣州 341000)

摘 要:针对 IaaS(Infrastructure as a Service)云平台中用户异常行为的检测问题,提出了一种基于用户行为模型和神经网络相结合的异常检测方法. 该方法通过构造一种基于时间、地点和事件的用户行为模型,在此基础上建立用户的正常行为模式,并与神经网络算法相结合,将用户当前行为网络输出值与给定阈值进行比较,以此来判断用户的行为是否异常,从而实现用户行为的异常检测. 实验结果表明,相比其它类似的用户行为检测方法,该方法能更有效发现用户的异常行为.

关键词:IaaS 云平台;用户行为模型;异常检测;神经网络
中图分类号:TP309 文献标志码:A

Detection method of users′ abnormal behaviour oriented IaaS cloud platform

ZHENG Jian, ZHOU Yanli, LIU Cong
(School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China)

Abstract: To detect users′ abnormal behavior in IaaS cloud platform, an anomaly detection method based on users′ behavior and neural network was proposed. In order to correctly detect users′ abnormal behaviour, a users′ behavioral model was constructed based on time, place and event, and on this basis, users′ normal behavior patterns were established. By combining the model and neural network, the method can detect and identify whether the behavior is normal or not by comparing the network output of the users′ current behavior with a given threshold. The experimental results show that the proposed method can detect users′ abnormal behavior more effectively than other similar methods of users′ behaviour detection.

Key words: IaaS cloud platform; users′ behavioral model; anomaly detection; neural network

0 引 言

云计算改变了传统的 IT 方式, 为人们带来便利的同时也面临着安全挑战[1]. 在云计算环境中, IaaS 云平台使用虚拟化技术实现对底层计算、网络和存储资源的封装, 并以虚拟机的形式提供给远程用户,同时向用户提供开放的访问接口,云终端用户可以直接使用和操作云服务提供商提供的各种云服务[2],因此云用户的不良行为对云资源安全的影响是非常严重的. 云用户的身份是否真实,用户是否可信是保障云资源安全的关键内容[3]. 身份认证技术已经比较成熟,例如数字证书、动态口令等,但身份认证无法阻止合法用户的恶意行为,因此对云终端用户行为的异常检测[4-6]对保障云安全具有重要的意义.

收稿日期:2015-12-05
基金项目:国家自然科学基金项目(61462034);江西省教育厅科学技术研究项目(GJJ13415)
作者简介:郑剑(1977-),男,博士,副教授,主要从事可信计算、云计算的可靠性和安全性等方面的研究,E-mail:zhengji25@163.com.

国内外学者在用户行为研究方面主要有两方面的成果. 用户行为可信评估部分成果：陈亚睿等[7]提出了一种云服务提供商和云终端用户之间的重复博弈模型，该模型通过多次观察用户博弈中用户行为，并结合其历史行为数据来分析终端用户的类型. 姜帆[8]建立了基于 Chord 算法的云用户信任模型，将云用户信任分为直接信任、间接信任和初始信任，并给出了三类信任对应的评价标准. 根据用户行为综合可信度和信任评价标准，得出用户行为的信任等级. 吕艳霞等[9]提出了一种基于三角模糊网络分析法的用户行为可信评估方法，结合动态的行为可信安全措施，分析云用户的异常行为.

用户行为异常检测部分成果：Doelitzscher 等[10-11]提出了一种云审计策略语言 CAPL 和基于规则的异常检测方法，该方法利用 CAPL 语言描述状态安全规则，通过这些规则能够很好的检测出异常行为，该方法的缺点是很难对云平台中的所有操作行为定义规则. Pannu 等[12]提出了一种基于支持向量机的异常检测方法，该方法不需要先前的异常历史行为数据，它能够实时学习异常行为. Fu 等[13]提出了一种云环境中自主异常检测的性能指标体系，它是基于决策树识别的性能指标，用于异常检测系统. Doelitzscher 等[14]提出了一种基于时间行为模型的检测方法，并在该模型下定义异常场景，用神经网络分析和学习用户正常行为，根据网络输出值判断行为是否异常，该方法的缺点是仅考虑时间因素很难区分正常开机和关机时间内使用虚拟机的行为是否异常. 文中提出的基于时间、地点及事件的行为模型检测方法能弥补这个缺点. 实验结果表明，用户模型能很好地表示用户的行为，且基于该模型的方法能有效检测出用户异常行为.

1　用户行为模型

1.1　行为模型的建立

在云计算环境下，用户行为是指用户登录云主机和使用云主机中应用程序等一系列操作. IaaS 云平台中用户行为数据包括用户 ID、登录 IP、登录时间、用户操作（运行编程软件或办公软件）、退出时间和运行时长等属性. 文章采用建模的方法将抽象的用户行为用形式化的语言进行描述，并建立基于时间、地点、事件的三元组用户行为模型，具体模型如公式（1）所示.

$$B=(T,P,E) \tag{1}$$

其中，B 代表用户行为，T 为登录或退出云主机时间，P 为用户登录云主机的地点，E 表示用户在云主机中所做的操作.

由行为模型可知，用户行为正常与否是由时间、地点和事件三者共同决定，下面将对 3 个属性正常范围进行定义.

1.1.1　时间属性

这里主要考虑两种时间：用户登录云主机时间 LIT 和 LOT 退出时间，且 LIT 和 LOT 均以秒表示[7]，例如用户在 8:00 am 登录云主机，则 LIT=28800 s. 文中假设在理想的情景下，云用户每天按时在同一个时间登录云主机进行工作，例如 8:00 am，在同一个时间退出云主机结束一天的工作，例如 6:00 pm. 很显然这不符合真实世界情景，因此，这里用核心登录时间 $CLIT$ 加上一个偏离值 ΔLIT 来表示用户真实的登录时间，同理，真实的退出时间等于核心退出时间 $CLOT$ 加上一个偏离值 ΔLOT. ΔLIT 和 ΔLOT 的范围如公式（2）所示.

$$-R\leqslant\Delta LIT,\Delta LOT\leqslant R,R\in \text{正整数} \tag{2}$$

因此，用户正常登陆或退出虚拟机的时间范围，分别如不等式（3）和不等式（4）所示.

$$CLIT-\Delta LIT\leqslant LIT\leqslant\Delta CLIT+\Delta LIT \tag{3}$$

$$CLOT-\Delta LOT\leqslant LOT\leqslant\Delta CLOT+\Delta LOT \tag{4}$$

1.1.2　地点属性

地点是指用户进入云主机管理平台的主机所处的位置而不是虚拟机本身所在位置. 在计算机领域中，通常用 IP 地址来代表主机所处的位置. 用户每天在固定的地方工作，IP 地址属于同一网络.

定义 1　正常 IP

设局域网集 $NetSet=\{N_1,N_2,\cdots,N_n\}$，如果 $\exists N\in NetSet$，使 $IP\in N$，则用户 IP 正常. 其中 $NetSet$ 为用户工作常用的网络集合，N_j 为具体网络，例如 N_i=192.168.0.0.

1.1.3　事件属性

事件指用户对云主机进行的操作，用进程表示用户在云主机上所做的操作.

定义 2　关键进程

设云主机上运行的进程列表 $PList=\{p_1,p_2,\cdots,p_m\}$，关键进程列表 $KList=\{Kp_1,Kp_2,\cdots,Kp_m\}$，如果 $\forall Kp_i\in PList$，则关键进程值 k=1，否则 k=0. $PList$ 为云主机当前运行的进程列表，$KList$ 为指定云主机必须要运行的进行列表，称之为关键进程列表. 例如假设用户主要从事编程工作，每天需要使用 eclipse 软件和 MySQL 数据库，则关键进程列表 $KList=\{elipse.exe,mysql.exe\}$.

1.2 行为定义

正常行为是指用户按照某种规定或规则对云主机进行操作，如果违反了规定的行为则为异常. 根据用户行为模型可知，用户行为是否正常与时间、地点、事件有关，如果用户的行为数据满足时间、地点、事件的规则要求，则行为正常，否则则为异常行为. 下面将用形式化语言对用户正常行为进行定义.

定义 3 正常行为

用 $behavior=(time,ip,event)$ 表示用户一条行为，$time$ 表示用户登录或退出时间，ip 表示用户使用网络地址，$event$ 表示当前运行着的关键进程. 如果用户行为同时满足公式(5)，则 $behavior$ 为正常行为，否则为异常行为.

$$\begin{cases} time \in LIT \vee time \in LOT \\ ip \in N \wedge N \in NetSet \\ event \in KList \end{cases} \tag{5}$$

1.3 用户行为异常场景

Doelitzscher 等在文献[7]中提出了 3 种具有代表性的用户行为异常场景，具体如下.

异常 1：用户登录云主机时间和退出云主机时间在正常主机运行时间范围之外，如图 1(a)所示；

异常 2：用户登录云主机和退出云主机时间在正常主机运行时间范围之内，如图 1(b)所示；

异常 3：用户在正常登录时间范围内登录云主机，并且在正常退出时间范围内退出云主机，如图 1(c)所示.

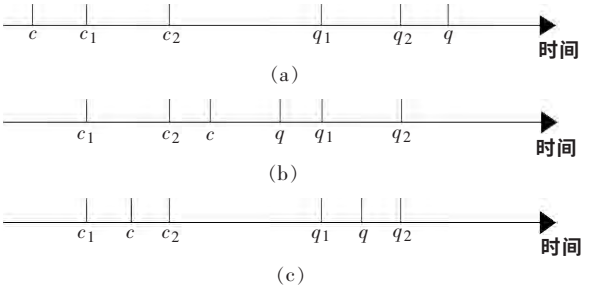

图 1 3 种异常场景

图 1 中，c 为用户登录云主机时间，q 为用户退出云主机时间，c_1,c_2 分别为正常登录云主机时间范围的下限和上限，q_1,q_2 分别为正常退出云主机时间范围的下限和上限.

上述 3 个异常场景仅考虑了用户使用虚拟机的时间，然而文中提出的用户行为模型综合考虑时间、地点和事件 3 个因素，因此，文中在上述异常场景的基础上进行修改，并给出与之对应的 3 种异常场景，具体如下定义.

定义 4 第一类异常

把用户登录云主机的时间(或退出云主机的时间)$time$ 在正常运行时间范围之外，不管用户的 ip 地址是否正常，用户在使用云主机时是否运行指定关键进程 $event$，即 $time$,ip,$event$ 同时满足公式(6)的用户行为统称为第一类异常.

$$\begin{cases} time \notin RT=LOT-LIT \\ ip \in IPSet \vee ip \notin IPSet \\ event \in KList \vee event \notin KList \end{cases} \tag{6}$$

定义 5 第二类异常

把用户登录云主机的时间(或退出云主机的时间)$time$ 在正常运行时间范围之内，不管用户的 ip 地址是否正常，用户在使用云主机时是否运行指定关键进程 $event$，即 $time$,ip,$event$ 同时满足公式(7)的用户行为统称为第二类异常.

$$\begin{cases} time \notin RT=LOT-LIT \\ ip \in IPSet \vee ip \notin IPSet \\ event \in KList \vee event \notin KList \end{cases} \tag{7}$$

定义 6 第三类异常

把满足用户登录云主机的时间 $time$ 在正常登录时间范围之内，或退出云主机的时间 $time$ 在正常退出时间范围之内，用户的 ip 地址异常，或用户在使用云主机时未运行指定关键进程 $event$，即 $time$,ip,$event$ 同时满足公式(8)的用户行为统称为第三类异常.

$$\begin{cases} time \in LOT \vee time \in LIT \\ ip \in IPSet \vee event \notin KList \end{cases} \tag{8}$$

其中 RT 为云主机正常运行时间范围，$IPSet$ 为正常 ip 地址集合.

2 实验方案的设计

2.1 数据生成算法

为了能够评估行为模型的可行性和合理性，需要大量用户历史行为数据来进行验证. 目前，获取用户行为数据的主要方法[14]有 2 种：由云服务提供商提供和使用模拟环境模拟用户行为数据. 由于涉及云用户隐私问题，云服务提供商一般拒绝提供用户数据. 因此，文中采用算法 1 生成用户行为数据.

算法 1:generateData()

```
1  record=[];
2  anlcount=0;
3  flag=true;
```

```
4 fileilename=createfilename(filepath);
5 while(i<=TotalItems){
6 if(anlcount< AnormalItems){
7 if(i%3==1)
8 随机为 TStmp 一个 RT 范围之外的值;
9 else if(i%3==2)
10 TStmp←Randrange(RT);
11 else
12 TStmp←Randrange(LIT)
13 或 TStmp←Randrange(LOT);
14 IsKey←Randrange(0,1);
15 为 ip 随机分配一个 ip 地址;
16 if(IsKey==0||ip ∉ NetSet)
17 flag=false;
18 Anomaly=1;
19 if(flag)
20 anlcount←anlcount+1;
21 Anomaly=0;}
22 else{
23 TStmp←Randrange(RT)
   或 TStmp←Randrange(LIT)
   或 TStmp←Randrange(LOT);
24 IsKey←Randrange(0,1);
25 从 NetSet 中随机挑选一个值赋给 ip;
26 Anomaly=1;}
27 将 TStmp,ip,IsKey,Anomaly 等因素构成一条行为记录，并添加到 record;
28 Write(record,filename)
29 i←i+1;}
30 Retrun filename;
```

2.2　网络结构设计

文中采用神经网络机器学习方法检测用户异常行为，并选用 BP 算法作为网络监督算法.

对于第一类异常的检测，设计如图 2 所示的网络结构，它包含 4 个输入，2 个隐含层和 1 个输出. 输入是 4 个元素构成的向量 v，如公式(9)所示. 4 个元素分别是云管理系统事件收到云主机开机或关机的时间($TStmp$)、云主机运行状态(VMS)、云用户登录云主机的 IP 地址(IP)，云主机是否运行了关键进程 ($IsKey$). 网络结构的输出值介于 0 到 1 之间.

$$v=[TStmp,VMS,IP,IsKey] \tag{9}$$

对于第二类异常和第三类异常的检测，其网络结构包含 6 个输入，2 个隐含层和 1 个输出. 输入是在第一类异常网络结构的输入向量的基础上增加二个元素构成 6 个元素的向量 v_2，如公式(10)所示. LastVMCC 表示云主机在当前时间之前开机的次数；LastVMSC 表示云主机在当前时间之前关机的次数. 隐含层和输出与第一类异常网络结构的隐含层和输出相同.

$$v_2=[TStmp,VMS,IP,IsKey,LastVMCC,LastVMSC] \tag{10}$$

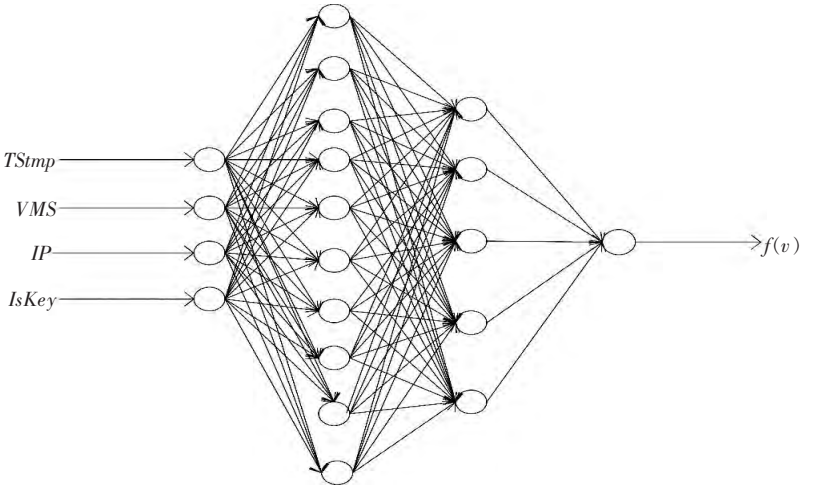

图 2　第一类异常网络结构图

3　实验与结果分析

整个实验过程包括：数据集的生成，用户正常行为的分析与学习，异常行为的检测与判断 3 个阶段. 首先利用算法 1 生成整个实验的数据集(如表 1 所示)，然后选择异常场景，并加载对应的网络，通过网络学习和分析正常行为，最后检测和判断异常行为. 表 1 中虚拟机开关机时间($TStmp$)、虚拟机运行状态(VMS)、用户进入云管理界面的 ip 地址

表 1 部分用户行为数据

行号	TStmp	VMS	LastVMCC	LastVMCS	IP	IsKey	Anomaly
1	31927	1	1	0	192.168.0.178	1	1
2	46609	0	1	1	192.168.0.189	0	1
3	52779	1	2	1	192.168.0.189	1	1
4	64999	0	2	2	192.168.0.189	0	1
5	31296	1	1	0	192.168.0.44	0	0
6	36325	0	1	1	192.168.0.10	0	1
7	53126	1	2	1	192.168.0.10	1	1
8	63094	0	2	2	192.168.0.10	0	1
9	19515	1	1	0	192.168.0.126	1	0
10	67770	0	1	1	212.211.187.150	0	0

注:Anomaly=1 为正常行为,Anomaly=0 为异常行为.

等参数的真实数据,均可以从 IaaS 云平台中获取到,例如以 Openstack 为例,云用户通过管理界面开启和关闭虚拟机的时间及 IP 地址均可以通过 nova-compute 和 dashboard 模块获取到,是否运行关键进程可以利用曹立铭等[15]提出的语义重构法来判断.

为了验证用户行为模型检测方法的有效性,文中分别使用 4 种检测方法(基于时间模型的方法[1]、基于时间与进程模型的方法、基于时间与 ip 地址模型的方法以及用户行为模型的检测方法)进行异常行为检测.

为了降低实验结果的偶然性和减少实验误差,对每种异常情况均进 10 次模拟实验. 基于时间模型与基于用户行为模型检测方法的 10 次模拟实验结果分别如表 2 和表 3 所示,这里以第一类异常的检测为例. 表 2 和表 3 中,正确检测是指原本就是异常的行为,通过检测系统得出来的结果还是异常行为,这种检测称之为正确检测. 检测率定义为正确检测量与异常数据量之比.

表 2 基于时间模型下的 10 次模拟实验结果

次序	总量 /条	异常数 /条	正确检测量 /条	检测率 /%
1	10	7	7	100
2	10	3	1	33.33
3	10	2	1	50
4	10	1	0	0
5	10	4	4	100
6	10	4	0	0
7	10	7	5	71.43
8	10	1	1	100
9	10	8	5	62.50
10	10	3	3	100

从表 2 和表 3 可以看出,两种检测方法均能检测出大部分异常行为,且基于用户行为模型的方法

表 3 基于用户行为模型下的 10 次模拟实验结果

次序	总量 /条	异常数 /条	正确检测量 /条	检测率 /%
1	10	7	5	71.43
2	10	3	1	33.33
3	10	2	2	100
4	10	1	1	100
5	10	4	4	100
6	10	4	3	75
7	10	7	7	100
8	10	1	1	100
9	10	8	8	100
10	10	3	3	100

能够检测出基于时间模型检测不出的异常行为. 例如表 1 中第 5 条和第 10 条异常行为,分别使用上述两种检测方法检测这两条行为数据,基于时间模型的检测方法认为它们是正常行为,而基于用户行为模型的方法能够检测出它们为异常行为. 其原因是上述两条异常行为,只从时间因素来看,属于正常时间范围内,因此,基于时间模型的方法认为它们是正常行为,而基于用户行为模型的方法综合了时间、地点及事件 3 个因素的影响,所以能判断出它们为异常行为.

文中用 10 次模拟实验的平均检测率来评估检测方法对每种异常场景的检测效率. 通过分析计算得 4 种方法的平均检测率如表 4 所示.

图 3 展现了每种异常场景下的平均检测率. 从图 3 可直观的看出,与基于时间行为模型的方法相比,文中方法下每种异常的检测率都是最高的. 其原因是异常行为可能由时间、地点、事件三者中任一个因素引起. 文中方法综合考虑了这 3 个因素,因此,无论由哪个因素引起的异常行为,使用基于用户行为模型的方法均能够检测出.

基于时间行为模型的方法对第三类异常的检测率为 78 %左右,而文中提出的方法对第三类异

表 4　4 种方法的平均检测率

异常类型	时间模型下的检测率 /%	时间进程模型下的检测率 /%	时间 ip 地址模型下的的检测率 /%	用户行为模型下的检测率 /%
第一类异常	61.73	68.45	69.46	87.98
第二类异常	67.98	83.63	90.90	91.83
第三类异常	77.98	84.02	91.98	92.98

常的检测率约为 93 %. 由于第三类异常虚拟机开机、关机时间均在正常开关时间范围内,因此,基于时间行为模型的方法很难检测出来. 例如表 1 中第 5 和第 10 条行为数据都属于第三类异常, 时间均在正常时间范围内,利用基于时间行为模型的方法检测,检测不出这 2 条数据,而文中提出的方法却能检测出来. 通过对比分析得出, 基于时间、地点和事件的用户行为模型能够很好的描述用户行为,且基于用户行为模型的检测方法比文献[14]更容易检测出第三类异常.

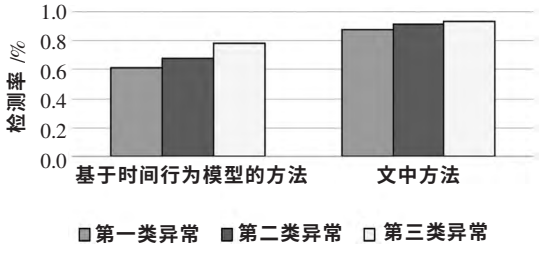

图 3　检测率趋势图

4　结束语

针对单一因素的用户异常行为检测方法存在低效的缺点,文中提出了基于时间、地点和事件的行为模型的异常检测方法. 该方法以考虑用户习惯为前提,与神经网络算法相结合,以获得好的检测效果. 通过实验结果和理论分析可知,文中提出的用户行为模型能很好的描述用户行为,且基于该模型的检测方法有效提高了检测率,并能有效的检测出正常开关机时间内使用虚拟机的行为是否异常. 由于硬件和数据采集问题, 没有搭建真正的 IaaS 云平台来获取用户行为数据,导致实验结果可能存在偏差,下一步的工作将通过 IaaS 云平台获取真实用户行为数据, 将实验进一步完善.

参考文献:

[1] Wei Y, Blake M B. Service-oriented computing and cloud computing: challenges and opportunities [J]. IEEE Internet Computing, 2010, 14 (6): 72-75.

[2] 李淑芝, 刘锋. 云环境下基于用户偏好的粒子群优化算法的 Web 服务选择 [J]. 江西理工大学学报, 2013, 34(5): 60-65.

[3] 田俊峰, 曹迅. 基于多部图的云用户行为认定模型 [J]. 计算机研究与发展, 2014, 51(10): 2308-2317.

[4] 王冬阳. 面向云计算的异常检测技术的研究与实现 [D].上海:上海交通大学, 2013.

[5] Tsai C F, Hsu Y F, Lin C Y, et al. Intrusion detection by machine learning: A review [J]. Expert Systems with Applications, 2009, 36 (10): 11994-12000.

[6] Ghosh P, Debnath C, Metia D, et al. An efficient hybrid multilevel intrusion detection system in cloud environment [J]. IOSR Journal of Computer Engineering, 2014, 16(4): 16-26.

[7] 陈亚睿, 田立勤, 杨扬. 云计算环境下基于动态博弈论的用户行为模型与分析 [J]. 电子学报, 2011, 39(8): 1818-1823.

[8] 姜帆. 基于 Chord 算法的云用户信任模型 [D].大连:辽宁师范大学, 2013.

[9] 吕艳霞,田立勤,孙珊珊. 云环境下基于 FANP 的用户行为的可信评估与控制分析 [J].计算机科学,2013,40(1):132-138.

[10] Doelitzscher F, Reich C, Knahl M, et al. An agent based business aware incident detection system for cloud environments[J]. Journal of Cloud Computing, 2012, 1(1): 1-19.

[11] Doelitzscher F, Ruebsamen T, Karbe T, et al. Sun behind clouds-on automatic cloud security audits and a cloud audit policy language [J]. International Journal on Advances in Networks and Services, 2013, 6(12): 1-16.

[12] Pannu H S, Liu J, Fu S. Aad: Adaptive anomaly detection system for cloud computing infrastructures[C]//Reliable Distributed Systems, 2012 IEEE 31st Symposium on. IEEE, 2012: 396-397.

[13] Fu S. Performance metric selection for autonomic anomaly detection on cloud computing systems [C]//Global Telecommunications Conference, 2011 IEEE. IEEE, 2011: 1-5.

[14] Doelitzscher F, Knahl M, Reich C, et al. Anomaly detection in iaas clouds [C]//Cloud Computing Technology and Science.2013 IEEE 5th International Conference on. IEEE, 2013: 387-394.

[15] 曹立铭, 赵逢禹. 私有云平台上的虚拟机进程安全检测[J]. 计算机应用研究, 2013, 30(5): 1495-1499.