doi :10.3969/j.issn.1671-1122.2020.09.019

基于图像和机器学习的虚拟化平台异常检测

王湘懿 [1,2]，张健 [1,2]

（1. 南开大学网络空间安全学院，天津 300350；2. 天津市网络与数据安全技术重点实验室，天津 300350）

摘　要：文章提出一种基于机器学习的虚拟化平台异常行为动态检测方法，该方法依托虚拟化平台，提取正常程序和恶意软件运行过程中的系统内存并转储为文件，将其中的部分信息经 SimHash 提取形成灰度图像并采用局部二值模式（LBP）进行描述，得到图像的纹理特征，再利用图像的纹理特征训练构建的卷积神经网络，通过生成的模型判断虚拟化平台是否存在异常行为。实验表明，虚拟化平台异常检测率可以达到 97.5%，能够有效发现云攻击事件。

关键词：云计算；虚拟化；卷积神经网络；图像特征；异常行为检测

中图分类号：TP309　**文献标志码：**A　**文章编号：**1671-1122（2020）09-0092-05

中文引用格式：王湘懿，张健 . 基于图像和机器学习的虚拟化平台异常检测 [J]. 信息网络安全，2020，20（9）：92-96.

英文引用格式：WANG Xiangyi, ZHANG Jian. Abnormal Behavior Detection of Virtualization Platform Based on Image and Machine Learning[J]. Netinfo Security, 2020, 20(9): 92-96.

Abnormal Behavior Detection of Virtualization Platform Based on Image and Machine Learning

WANG Xiangyi[1,2], ZHANG Jian[1,2]

(1. *College of Cyber Science*, *Nankai University*, *Tianjin* 300350, *China*; 2. *Tianjin Key Laboratory of Network and Data Security Technology*, *Tianjin* 300350, *China*)

Abstract: This paper proposes a method for dynamically detecting abnormal behavior of a virtualization platform based on machine learning. This method relies on the virtualization platform, extracted the system memory during normal program and malware running and dumps it into a file, extracted part of the information through SimHash to form a grayscale image and used local binary mode(LBP) to describe the texture features of the image. The features of image are used to train the constructed convolutional neural network, and the generated model determines whether the virtualization platform has abnormal behavior. Experiments show that the detection rate of virtualization platform can reach 97.5%, which can effectively detect cloud attack events.

Key words: cloud computing; virtualization; convolutional neural network; image feature; abnormal behavior detection

收稿日期：2020-7-16

基金项目：天津市重点研发计划 [20YFZCGX00680]；天津市科技重大专项与工程 [19ZXZNGX00090]

作者简介：王湘懿（1999—），女，辽宁，硕士研究生，主要研究方向为云安全、网络安全、系统安全；张健（1968—），男，天津，正高级工程师，博士，主要研究方向为云安全、网络安全、系统安全。

通信作者：张健 jeffersonzj@qq.com

0 引言

云计算是近年来新兴的一种互联网分布式计算技术，可为用户带来高效便捷的互联网服务。但针对云平台的入侵事件频发且攻击方式复杂多变，云平台存在的安全隐患日益凸显[1]。云平台的基础是虚拟化技术，而虚拟机自省技术（VMI）作为有效提高云环境安全性的重要技术之一，在针对云平台的网络攻击、黑客入侵等异常行为检测中均得到了广泛的研究[2]。

本文使用VMI技术提取虚拟机平台的内存文件信息并通过图像化方式处理提取的内存特征，使用图像特征训练机器学习模型从而得到对应分类算法，进而提高虚拟化平台异常行为检测的准确率和识别效率，提升对云安全事件的监测预警能力。

1 国内外研究现状

2003年，GARFINKEL[3]等人提出了虚拟机自省技术的概念，并给出了该技术的定义：以分析虚拟机中运行的软件为目的，在虚拟机外部监测虚拟机的方法。VMI技术通过在目标虚拟机（TVM）外部获取TVM底层状态数据，能够实现安全软件与恶意软件的分离，有效应对直接攻击带来的挑战。但是在TVM外部获取的底层状态数据以二进制形式呈现，无法获悉其高层语义信息，这种语义之间的差异称为语义鸿沟[4]。语义重构是指由低级语义重构出操作系统级语义的过程[5]。研究人员根据语义重构方式的不同将虚拟机自省技术分为基于目标虚拟机的依赖型自省方法、基于安全虚拟机的依赖型自省方法、基于软件结构知识的独立型自省法和基于硬件架构知识的独立型自省法4类[6]，并发现获得信息量较丰富的方法需要依托的语义知识更多，而依托语义知识较少的方法获得的信息量不够充分。因此，如何解决语义鸿沟问题，增强VMI技术的可移植性、透明性和鲁棒性是一个重要的研究方向。

近年来，随着机器学习技术的发展，基于机器学习的入侵行为和恶意软件检测引起了研究人员的广泛关注。2001年，SCHULTZ[7]等人使用机器学习方法，通过分析恶意软件和正常软件存在的差异特征实现了恶意软件的识别检测。2010年，CONTI[8]等人提出将二进制文本转化为灰度图片的方法。VASAN[9]等人基于文献[8]的思想，提取虚拟机中的恶意可执行文件并生成RGB三通道图片，通过卷积神经网络对这些图片进行分类，判断恶意软件所属家族的准确率可达到97%以上。NI[10]等人提出一种名为MCSC（恶意软件使用SimHash和CNN进行分类）的恶意软件分类算法，可转换反汇编的恶意软件为基于SimHash的灰度图像，通过卷积神经网络识别其族网络。DAI[11]等人将恶意的EFI文件转化为灰度图像，使用双三次插值法将图像缩放成等高等宽的图像，并结合方向梯度直方图对图像进行特征提取，提高了分类的准确率。结合图像的特征提取为恶意软件的识别与分类提供了新的方向和思路，也可进一步推广应用于虚拟化平台的异常行为检测。

2 检测方法实现

本文实现了一种基于机器学习的异常行为动态检测方法，异常行为检测流程如图1所示。该方法直接从动态运行的系统内存镜像转储文件中提取有效信息，采用机器学习算法检测发现客户虚拟机中的异常行为，无需进行语义重构，不受虚拟机上层操作系统和数据存储方式的影响。

对于从虚拟机中提取出的内存文件，本文采用SimHash序列描述内存特征。SimHash算法作为基于Hash函数的相似检测算法，能够在不损失数据特征的情况下有效进行数据降维。

SimHash算法按照顺序每次读取1024 B作为一组序列，每8位字节转换成1个字符。本文使用MD5算法计算每个字符的Hash值，得到64位的Hash序列，判断Hash序列的每一位，如果是1则加上该字符的权值；反之，则减去其权值，得到新的序列。对于新序列的每一位，如果小于0，则置0；反之，则置1，将新序列相加得到初始序列的SimHash值。此时获得的SimHash值是64位，将其划分为2个8位十六进制数。

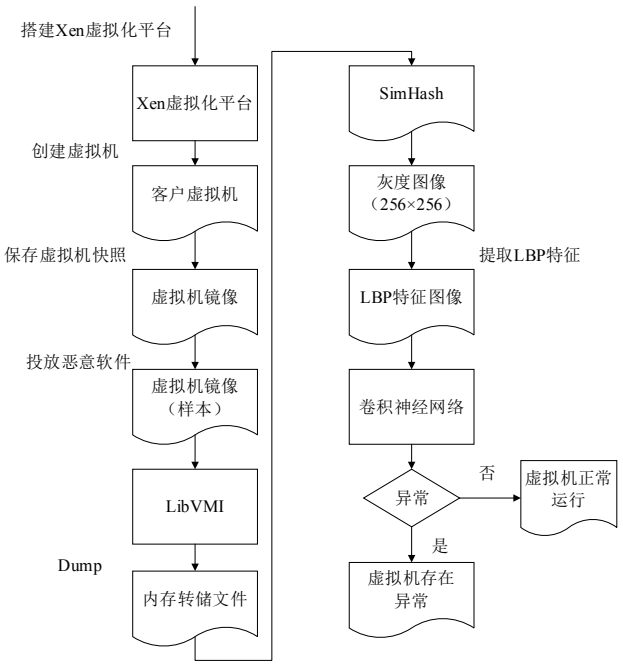

图 1 异常行为检测流程

十六进制数的每一位按照大于7记为1，小于等于7记为0的规则生成两个数值，数值的范围为0~255。将这个两个数值作为二维图像的横纵坐标映射到初始为全黑的灰度图像上，对应坐标点的灰度值增加16。将每一组数据生成的图像叠加，得到包含该文件序列特征的灰度图像。图2描述了横纵坐标的生成过程。

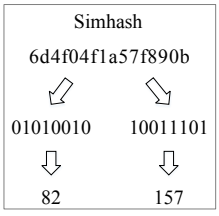

图 2 横纵坐标值生成过程

特征提取是图像处理和识别的关键步骤。1996年，OJALA[12]等人提出局部二值模式（LBP）用于描述图像局部纹理特征。原始LBP算子定义在3×3的窗口内，窗口中心像素作为中心，相邻8像素的灰度值与其进行比较。若周围像素值大于中心像素值，则该像素点的位置标记为1；否则，标记为0。这样，3×3邻域内的8个点经过比较可产生8位二进制数，即该窗口中心像素点的LBP值，并用此值反映该区域的纹理信息。使用数学模型来描述如公式（1）、公式（2）所示。

$$LBP(x_c, y_c) = \sum_{p=1}^{8} s(I(p) - I(c)) \times 2^p \quad (1)$$

$$s(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & \text{其他} \end{cases} \quad (2)$$

其中，p 表示 3×3 窗口中除中心像素点外的第 p 个像素点；$I(c)$ 表示中心像素点的灰度值，$I(p)$ 表示领域内第 p 个像素点的灰度值。

初始生成的灰度图像如图3所示。将初始灰度图像经过LBP操作后生成新的图像，该图像中的信息是原来图像中每个像素点的LBP值，如图4所示。

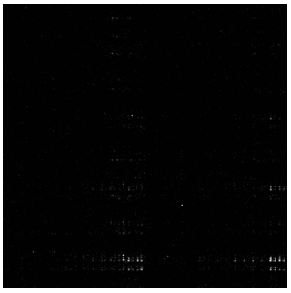

图 3 SimHash 灰度图像

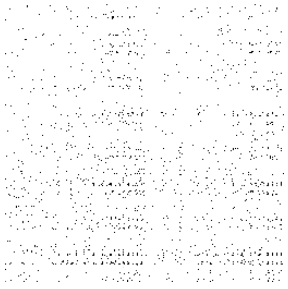

图 4 灰度图像对应的 LBP 特征图像

3 实验及分析

3.1 数据集与参数设置

搭建ubuntu16.04操作系统的虚拟机作为安全虚拟机（Dmo0），搭建Xen虚拟化平台，客户虚拟机（DmoU）安装Windows 7操作系统，模拟云计算平台基础实验环境。将内存转储工具LibVMI安装在Dmo0上，监测DmoU的行为并及时转储文件。

实验所需的恶意软件来自Virus Share，包括木马、后门软件、蠕虫、广告软件等，如表1所示。选取可

在 Windows 7 系统上运行的软件。

表 1 样本种类与数量

种类	数量
木马	15
后门软件	30
广告软件	38
蠕虫	17

Windows 7 版本客户虚拟机安装完成后即刻保存快照。将恶意软件投放到 DmoU 中，运行 30 s 后转储到当前内存形成一份异常内存文件，并在运行下一个软件之前将系统恢复到快照所保存的状态，保证每次试验过程中系统的一致性。从内存文件中提取受到攻击的异常内存文件 100 个，运行正常内存文件 100 个，共 200 个内存文件。因文件数据量过大（4.3 GB/ 个）且冗余信息较多，本文仅选择了文件起始部分（10 MB）作为实验对象，每次读取 1 KB 数据作为一个 SimHash 序列，共 10240 组。将 80% 的试验数据作为训练集，其余的作为验证集。将图片大小统一设置为 256×256。本文使用 TensorFlow 框架与 Inception V3 网络结构，该网络结构可将大卷积分解成小卷积，降低参数和计算量，减轻过拟合问题。本文依据模型在数据集训练中的表现对模型中部分参数进行调整，批处理参数 batch size 设置为 16，起始学习率为 0.01，模型初始化函数为 tf.glorot_normal() 函数。

3.2 实验结果分析

采用改进的卷积神经网络模型迭代 10000 次进行数据训练，其中，准确率参数 *accuracy* 的计算方式如公式（3）所示。

$$accuracy = \frac{TN + TP}{FN + FP + TN + TP} \tag{3}$$

召回率参数 *recall* 的计算方式如公式（4）所示。

$$recall = \frac{TP}{TP + FN} \tag{4}$$

精确率参数 *precision* 的计算方式如公式（5）所示。

$$precision = \frac{TP}{TP + FP} \tag{5}$$

其中，*TN* 为预测为负的样本中实际负样本个数，*TP* 为预测为正的样本中实际正样本个数，*FP* 为预测为正的样本中实际负样本个数（误报），*FN* 为预测为负的样本中实际正样本个数（漏报）。

准确率衡量了模型分类的正确性，召回率和精确率衡量了模型对正例的识别能力。另外，还可通过鲁棒性对卷积神经网络中的模型进行衡量。通常用损失函数来估量模型的预测值 $f(x)$ 与真实值 Y 的不一致程度。损失函数是一个非负实值函数，通常使用 $L(Y, f(x))$ 表示，损失函数越小，模型的鲁棒性就越好[13]。损失函数的表示方法公式（6）所示。

$$\theta = argmin\theta \frac{1}{N}\sum_{i=1}^{N} L\left(y_i, f\left(x_i;\theta\right)\right) + \lambda\Phi\left(\theta\right) \tag{6}$$

为了进一步验证实验结果，选取最优模型，本文从卷积神经网络模型、数据量和图像特征提取方式 3 个方面，将 Inception V3 模型与 VGG16、AlexNet、ResNet101、LeNet 等经典卷积神经网络模型进行对比。实验数据为“SimHash+ LBP”的图像，实验结果如图 5 所示。

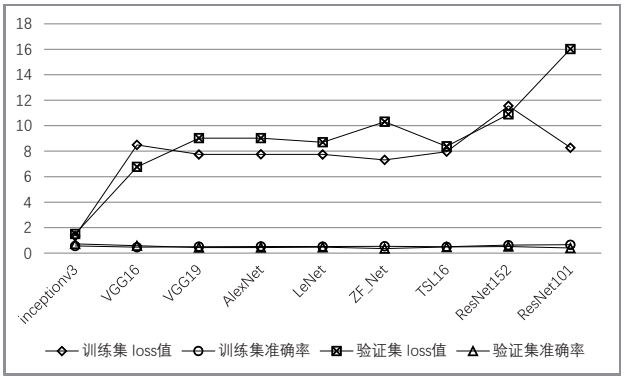

图 5 实验效果对比

由图 5 可知，Inception V3 模型的损失函数值相比其他模型有明显的优势，在训练集上的准确率与验证集上的准确率均高于其他模型。

实验将经过 LBP 特征提取的灰度图像旋转 90% 从而扩大数据量至 400 张图片，生成大数据集作为对比。实验对比集还包括生成的初始灰度图像与经过 HOG 特征提取的灰度图像。训练集图像作为训练样本，随迭代次数进行模型训练。在模型验证过程中，本文选取验证集上准确率最高的模型作为最优模型并记录对应

的结果。实验结果及对比如表2所示。

表 2 实验结果对比

方法	*accuracy*	*precision*	*recall*
Inception V3+SimHash +LBP+ 大数据集	80%	73.07%	95%
InceptionV3+SimHash	97.5%	95.24%	100%
InceptionV3+SimHash +HOG	90%	100%	80%
InceptionV3+SimHash +LBP	90%	86.36%	95%

由表2可知，使用原始灰度图像判断对应内存转储文件所属类别的整体效果较好；使用LBP纹理特征处理过的图像判断其对应的内存文件所属类别，机器学习模型的召回率较高；使用HOG特征提取处理过的图像进行内存文件分类时，模型的召回率、精确率与准确率均与原始灰度图像所形成的模型有一定的差距，推测该方法在本文实验中会损失内存文件对应的灰度图像中的可区分特征。

在大数据集的训练与验证中，模型的损失函数值增长速度快，收敛速度快，loss曲线体现出下降的趋势也早于小数据集。这说明对于本文使用的模型Inception V3，数据量的增加有利于模型整体性能提升。

4 结束语

本文借助VMI技术提取了虚拟化平台的内存文件，使用SimHash获得了其中部分信息并转化为灰度图像，通对图像采用局部LBP进行描述，得到图像的纹理特征，并将特征用于训练构建的卷积神经网络，以训练获得的分类模型为核心，判断图像对应的虚拟机平台是否存在异常行为。该方法从虚拟机平台的内存文件中提取信息，减少了对客户虚拟机运行过程的干预，且与虚拟机当前的操作系统、存储数据结构等特征关联性不强，其可移植性强、适应范围广，但也存在操作较繁琐、运算时间较长等问题，有待进一步提升与改进。

今后将在现有研究的基础上，在不破坏特征和有效信息的前提下，进一步分析内存，获取更多有效信息，力求进一步通过图像特征分析识别更多的系统高层语义，从而判断黑客攻击的方法和系统受影响的位置。（责编 潘海洋）

参考文献：

[1] National Internet Emergency Center. China's Internet Network Security Situation in the First Half of 2019[EB/OL]. http://www.cac.gov.cn/2019-08/13/c_1124871484.htm, 2020-6-1.
国家互联网应急中心 . 2019 年上半年我国互联网网络安全态势 [EB/OL]. http://www.cac.gov.cn/2019-08/13/c_1124871484.htm, 2020-6-1.
[2] BAEK H W, SRIVASTAVA A, DER MERWE J V, et al. CloudVMI: Virtual Machine Introspection as a Cloud Service[C]//IEEE. IEEE International Conference on Cloud Engineering, March 11-14, 2014, Washington, DC, USA. New Jersey: IEEE, 2014: 153-158.
[3] GARFINKEL T, ROSENBLUM M. A Virtual Machine Introspection Based Architecture for Intrusion Detection[EB/OL]. https://suif.stanford.edu/papers/vmi-ndss03.pdf, 2020-6-1.
[4] NANCE K, BISHOP M, HAY B. Virtual Machine Introspection: Observation or Interference[J]. IEEE Security & Privacy Magazine, 2008, 6(5): 32-37.
[5] PFOH J, SCHNEIDER C, ECKERT C. A Formal Model for Virtual Machine Introspection[EB/OL]. http://www.cs.jhu.edu/~sdoshi/jhuisi650/papers/spimacs/SPIMACS_CD/vmsec/p1.pdf, 2020-6-1.
[6] LI Baohui, XU Kefu, ZHANG Peng, et al. Research and Application Progress of Virtual Machine Introspection Technology[J]. Journal of Software, 2016, 27(6): 1384-1401.
李保琿，徐克付，张鹏，等 . 虚拟机自省技术研究与应用进展 [J]. 软件学报，2016，27(6)：1384-1401.
[7] SCHULTZ M G, ESKIN E, ZADOK F, et al. Data Mining Methods for Detection of New Malicious Executables[EB/OL]. http://cseweb.ucsd.edu/~eeskin/papers/binaryeval-ieeesp01.pdf, 2020-6-1.
[8] CONTI G, BRATUS S, SHUBINA A, et al. Automated Mapping of Large Binary Objects Using Primitive Fragment Type Classification[EB/OL]. https://dl.acm.org/doi/10.1016/j.diin.2010.05.002, 2020-6-1.
[9] VASAN D, ALAZAB M, WASSAN S, et al. IMCFN: Image-based Malware Classification Using Fine-tuned Convolutional Neural Network Architecture[EB/OL]. https://www.sciencedirect.com/science/article/abs/pii/S1389128619304736, 2020-6-1.
[10] NI Sang, QIAN Quan, ZHANG Rui. Malware Identification Using Visualization Images and Deep Learning[EB/OL]. https://www.sciencedirect.com/science/article/pii/S0167404818303481, 2020-6-1.
[11] DAI Yusheng, LI Hui, QIAN Yekui, et al. A Malware Classification Method Based on Memory Dump Grayscale Image[J]. Digital Investigation, 2018, 27(12): 30-37.
[12] OJALA T, PIETIKINEN M, HARWOOD D. A Comparative Study of Texture Measures with Classification Based on Featured Distributions[J]. Pattern Recognition the Journal of the Pattern Recognition Society, 1996, 29(1): 51-59.
[13] BAMS D, LEHNERT T, WOLFF C P. Loss Functions in Option Valuation: A Framework for Selection[J]. Management Science, 2009, 55(5): 853-862.