基于灰度图纹理指纹的恶意软件分类

张晨斌 张云春 郑 杨 张鹏程 林 森
(云南大学软件学院 昆明 650095)

摘 要 随着安卓恶意软件数量的快速增长,传统的恶意软件检测与分类机制存在检测率低、训练模型复杂度高等问题。为解决上述问题,结合图像纹理特征提取技术和机器学习分类器,提出基于灰度图纹理特征的恶意软件分类方法。该方法首先将恶意软件样本生成灰度图,设计并集成了包含 GIST 和 Tamura 特征提取算法在内的 4 种特征提取方法;然后将所得纹理特征集合作为源数据,基于 Caffe 高性能处理架构构造了 5 种分类学习模型,最终实现对恶意软件的检测和分类。实验结果表明,基于图像纹理特征的恶意软件分类具有较高的准确率,且 Caffe 架构能有效缩短学习时间,降低复杂度。

关键词 恶意软件,灰度图,纹理特征,分类学习
中图法分类号 TP399 **文献标识码** A

Malware Classification Based on Texture Fingerprint of Gray-scale Images

ZHANG Chen-bin ZHANG Yun-chun ZHENG Yang ZHANG Peng-cheng LIN Sen
(School of Software, Yunnan University, Kunming 650095, China)

Abstract With the rapid increment of the number of Android malwares, the traditional malware detection and classification methods were proved to be with low detection rate, highly complex training model and so on. To solve above problems, the texture feature of gray-scale image-based malware classification method was proposed by combining the image texture feature abstraction and machine learning classifiers. The proposed method starts with converting the malware samples into grayscale images. Four feature abstraction methods were designed including GIST and Tamura-based feature abstraction algorithm. By taking the texture feature as the source data, 5 kinds of classification learning models were constructed by using high performance architecture Caffe. Finally, the detection and classification of malwares were done. The experimental results show that the image texture feature-based malware classification achieves high accuracy, and the Caffe architecture can effectively improve the learning time which further reduces the complexity.

Keywords Malwares, Gray-scale images, Texture feature, Classification learning

1 引言

移动互联网在给人们带来便利的同时,其中的恶意软件也带来了不容小觑的安全隐患。腾讯移动安全实验室的数据显示,2017 年第一季度新增病毒数同比增加 21.42%,总数达到了 465 万,是 2014 年的 33 倍。随着物联网、大数据、云计算平台等技术的最新发展和广泛使用,安全问题更加突出,尤其是大量恶意软件的肆意传播。面临如此严峻的形势,开发能够检测和识别恶意软件的机制是保障网络安全的重要措施。

恶意软件的检测和分类研究由来已久,但始终未出现有效且完整的解决方案。人工检测恶意软件的方式效率低下,效果不理想。采用机器学习技术来实现恶意软件的检测和分类具有重要的研究意义和应用价值。传统恶意软件检测技术主要分为动态检测和静态检测两种类型[1]。其中,静态检测采用特征匹配、广谱特征码和启发式扫描的方式[2]。静态检测技术实现简单,但借助程序的静态特征进行检测的技术往往会遭到加壳、变形、变种等技术的干扰,逆向分析难度增大,检测效果不佳。动态检测技术通过在虚拟机环境中运行恶意代码来检测其行为,通过运行时的行为特征进行检测。动态方法准确度高,但分析复杂,一个恶意程序的分析时间最短也要几分钟。

针对现有方法存在的不足,本文通过采用机器学习的最新技术,以高性能机器和分析平台为依托,设计和实现了基于图像纹理特征的恶意软件分类机制。本文的主要创新点包括:1)在恶意软件生成灰度图的基础上,集成了 GIST 纹理特征提取等 3 种纹理提取算法,设计并实现了基于 Tamura 的纹理特征提取算法,生成恶意软件纹理特征库;2)以纹理特征数据为基础,利用高性能 Caffe 框架进行纹理特征数据的处理,减小了分析数据的规模;3)使用数据挖掘分类算法实现了

本文受云南省应用基础研究计划青年项目(2012FD004),国家自然科学基金项目(61363084,61363021),云南大学软件学院教育创新基金项目(2012EI07)资助。

张晨斌(1994-),男,主要研究方向为机器学习、网络安全;**张云春**(1981-),男,博士,讲师,主要研究方向为无线网络,E-mail:yunchunzhang@hotmail.com(通信作者);**郑 杨**(1996-),男,主要研究方向为网络安全;**张鹏程**(1996-),男,主要研究方向为网络安全;**林 森**(1994-),男,主要研究方向为机器学习。

基于纹理特征的恶意软件分类，与同类方法相比，在保证分类准确率的同时，降低了分类的复杂度；4）基于现有的 4 种图像纹理特征提取方法，利用 Caffe 框架，设计和开发了一个高效的恶意软件分类平台。

本文第 1 节介绍恶意软件的相关背景，并阐述要解决的关键问题；第 2 节对恶意软件检测和分类有关的国内外研究工作进行介绍；第 3 节阐述基于灰度图的 4 种纹理特征提取算法的原理及具体实现；第 4 节实现基于 Caffe 框架的分析平台，并通过数据样本进行实验分析，测试了恶意软件分类的效果；最后进行总结，并对下一步可行的研究方向进行展望。

2 国内外研究现状

在恶意代码的分类和检测领域，目前有两个方向[3]：一是基于恶意代码对应的二进制内容的静态检测方法，另一种则是基于恶意代码运行时其行为的动态检测。其中，静态检测方式首先通过分析恶意样本的二进制文件、反汇编生成的文件和文件组织来建立各样本的特征，包括程序的控制流、编码错误或漏洞等；然后对抽取的特征使用分类或聚类算法进行分析，从而实现检测或分类的目的。静态分析方式具有对程序覆盖面较全、处理速度较快的优点；但是容易受到混淆技术的干扰，反汇编的逆向分析难度增大，从而存在一定程度的误报和漏报。动态检测方式主要是将需要检测的未知程序放入隔离沙箱中运行，借助其在运行过程中产生的行为来判断其是否具有恶意。常见的行为包括：对系统关键注册表相关内容的修改，以及对网络通信资源、文件和互斥锁资源的使用等。与静态方式相比，动态方法检测的准确率更高，无需对文件进行解密或解压缩，并且是观察恶意文件动态行为的唯一方式；但由于需要实际运行程序，因此所需时间更长；同时高维特征向量增加了分析算法的复杂度，导致其可扩展性较差；因在沙箱中缺少必要的触发条件，部分恶意行为未能及时捕获。总之，不管是动态方法还是静态方法，面对规模庞大且增长迅速的恶意代码库，很难对所有的未知程序进行甄别。

作为静态方法中的重要分支，基于恶意软件灰度图的研究取得了显著的进步。早在 2011 年，Nataraj 等[1]就首次提出了将待分类恶意代码的可执行文件转换为对应的灰度图像，并基于图像纹理特征生成特征向量，利用 KNN 算法实现对恶意文件的分类。该工作开创了基于图像分析的恶意软件分析方向的先河，随后大量方法被提出。国内具有代表性的是韩晓光等[4]进行的研究，其将图片映射为无压缩的灰度图像，通过灰度共生矩阵算法进行特征抽取，之后借助特征选择算法计算出 6 个贡献最大的特征，经过规范化之后建立纹理特征库，同时取得了不错的准确率。

在基于图像技术和纹理指纹技术进行恶意软件的检测和分类方面，国内外均取得了显著的进展。然而，该类研究由于刚刚起步，还存在一些不足，如：在研究时存在使用的特征提取算法与人类视觉感受差异较大；使用的特征提取算法的时间复杂度过高，计算量过大；提取的特征规模过大且纬度高，致使算法可扩展性差等。因此，本文在设计与实现动态检测的基础上，结合最新的处理框架来降低分析的时间复杂度。

3 基于灰度图纹理指纹特征的安卓恶意软件分析

鉴于安卓系统在市场上被广泛使用，同时其在处理器、存储资源等方面存在局限性，该文设计的恶意软件检测方法主要针对安卓系统。以恶意软件生成的灰度图作为分析的对象，基于高性能数据处理框架和优化流程，本文设计了主要工作流程如图 1 所示的恶意软件分类系统。文中重点设计和实现了灰度图纹理特征抽取算法、特征选择和降维方法，最后采用 Caffe 框架实现了高效的学习，进而确定各恶意样本所属的“家族”，完成分类。

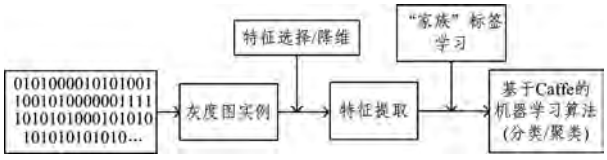

图 1 基于灰度图纹理特征的恶意软件检测流程

3.1 恶意软件灰度图的生成技术

灰度图的生成以恶意软件的二进制文件为输入，通过如下步骤将其转换为灰度图：1）以二进制的形式读取样本中任意一个十六进制的恶意代码文件，每读取 8 bits 作为一个基本单元，并执行后续步骤。2）将该基本单元中的二进制序列转换为一个无符号的整型变量，则该变量的范围为[0，255]，该值映射为图像中任意一个像素的灰阶值。具体实现中定义 255 表示纯白，0 表示纯黑。3）将读入的每一个像素对应的灰阶值按照固定宽度（该文采用 128 bits）存入新矩阵中。重复上述步骤，直到文件全部读取完毕。为了不影响最终分析结果，对于最后一次读取的内容不足 8 bits 的情况，用 0 进行填充。4）将该恶意文件对应的矩阵保存为无压缩的 PNG 图片，作为后续分析的源数据。

使用上述方法将所有样本中的实例全部转换为灰度图，后续采用图像识别方法对其进行分类，在本项目中所有样本的分类标签已知。对于分类标签未知的样本，采用与文献[5]和文献[6]中同样的方法，使用 VirusTotal 工具进行标注。

3.2 灰度图纹理特征的提取

3.2.1 图像识别与纹理特征

将恶意软件样本转换为灰度图后，可使用图像识别技术对其进行分类。分类的主要依据在于，恶意软件的不同变种在文件的指令、结构等方面存在一定的差异，而从同一个原始文件变异出的所有“变种”都保持了某种相似性，进而在其对应的灰度图中体现出某种“特征”。因此，本文将图像识别技术应用于灰度图识别，从而实现对恶意软件的分类。

灰度图识别的关键在于提取纹理指纹特征。纹理是物体表面最本质的属性，是所有实物都拥有的自然属性，一般是指实物上所展现的表面纹路[7]。在图像处理技术中，纹理通常代表图像的灰度或色彩在空间上的变化和重复。纹理一般被分为两个大类：天然纹理和人工纹理。天然纹理是指自然界原本存在的、没有人工痕迹的纹理，具有不确定性高、分布不具有规律性的特点。人工纹理是指非天然形成的、经过人类处理的纹理，纹理分布的规律性很强。本文恶意软件灰度图的纹理主要为第二种。图像纹理是图像特征的关键组成部分，在恶意软件灰度图像的分类中，图像纹理是实现同类“变种”恶意样本分类的关键。在恶意软件灰度图的生成算法中，其输出是灰度图，而不是色彩，因此色彩不作为该文恶意软件灰度图分类的依据。

3.2.2 图像纹理特征的提取

提取灰度图特征向量时，以恶意软件灰度图为输入，通过

图像纹理特征提取方法建立特征向量，然后基于特征向量，运用分类模型识别出每个恶意软件的“家族”。到目前为止，在图像纹理特征提取领域，常见的方法包括：统计方法、信号处理方法、模型方法以及几何方法。上述 4 种方法分别作用于灰度图特征的提取。

1)统计方法

图像纹理的分布可能是随机的，但基于像素点及其邻域内灰度特征的统计学方法，仍然可以对图像的纹理特征进行描述。具有代表性的方法包括灰度共生矩阵(Gray Level Co-occurrence Matrix，GLCM)[8]或半方差法。

半方差法是 Miranda 等[9]提出的，他详细介绍了利用该方法进行图像纹理提取的步骤。该方法使用变差函数来描述图像的纹理特征，对人工纹理能实现很好的提取和分类效果。

灰度共生矩阵法是 Gotlieb 等[10]在研究共生矩阵中各种统计特征的基础上归纳出的一种特征提取方法。韩晓光等[4]首次将 GLCM 应用于恶意软件灰度图纹理特征的提取。GLCM 方法首先计算图像中相距为 d、方向位置为 θ 的两个灰度像素同时出现的联合概率分布 $P(i,j|d,\theta)$。计算出的 GLCM 矩阵具有多种特征，如 Haralick 等[11]提取了 14 种特征。为降低后续分类学习的复杂度，通常使用特征选择算法或降维方法对其进行压缩。经过压缩后，常见的特征至少包括反差、能量、熵和相关性。

反差(Contrast)即对比度，用来衡量图像纹理中沟壑的深浅度以及图像的清楚度。反差值越大，深浅度就越大，图像也更加清晰。其计算方式为：

$$Con=\sum_i\sum_j(i-j)^2P(i,j)$$

能量(Energy)表示灰度共生矩阵中所有值的平方和，用于衡量图像纹理的变化程度是否稳定，反映人视觉感受的纹理粗细和层次的平均排列。其计算方式为：

$$Asm=\sum_i\sum_j(P(i,j))^2$$

熵(Entropy)用于权衡图像中灰度分布的随机性。图像越复杂，则熵越大。其定义为：

$$Ent=\sum_i\sum_jP(i,j)\times\log P(i,j)$$

相关性(Correlation)用于衡量图像中像素点的灰度值在行和列方向上的近似度。该值越大，表示图像在一定区域内的相关性就越高。其定义为：

$$Corr=(\sum_i\sum_jijP(i,j)-\mu_x\mu_y)/\sigma_x\sigma_y$$

2)信号处理方法

信号处理的研究经过了多年的积累，已有大量有效算法。基于信号处理方法的图像分析通过傅立叶变换将时间域或空间域的信号转换到频率域，再提取相对比较平稳的特征值，从而建立纹理特征向量。Tamura 算法[12]是信号处理领域较好的纹理特征提取算法之一，但此前并未见其在恶意软件分类领域有所应用。该文重点采用 Tamura 算法实现对恶意图像纹理特征的分析，建立基于 Tamura 纹理特征提取算法的恶意软件分类。

3)模型方法

模型方法的核心思想是假设纹理的分布由一个排列分布模型所生成，根据图像中纹理的分布和排列计算出模型的相关参数，并将该参数作为模式分类的依据。常见的模型方法包括：自回归模型、Markov 随机场方法、分形模型方法等。

4)几何方法

几何方法指基于图片纹理基元的不同特征研究纹理的提取和分析的方法。几何方法建立在图像纹理基元在排列和空间组织的分布上具有规律性的假设之上。相关研究成果表明，该方法对人工纹理图像的提取和分析具有较好效果，但对天然纹理进行处理时效果不够理想，从而限制了其应用和发展。

4　安卓恶意软件分类平台的设计

4.1　基于 Caffe 框架的平台搭建

在对比分析动态检测和静态检测方法各自的优点和缺点后可以发现，两者可相互补充，相辅相成。本文在传统的检测方式上做了一些改进，综合多种现有的检测和分类方法，设计了一个基于 Caffe 框架的恶意软件在线检测和分类平台。其中，后台采用传统的恶意软件检测分类方式，通过静态方法来匹配恶意软件的特征码和 MD5 值，完成识别。在无法识别的情况下，系统将会分离库中的恶意软件样本集的 dex 文件，进而将该文件转换为灰度图；然后，使用 Tamura 等算法提取灰度图的纹理特征；最后，利用深度学习算法对选定的恶意软件样本的纹理特征向量进行训练，构建有效且准确的分类模型，进而利用模型进行预测和识别。本平台能够解决恶意软件变种难以检测的问题，提供了便利的在线检测分类功能。

本文采用 Caffe(Convolutional Architecture for Fast Feature Embedding)作为深度学习框架。Caffe 是一个清晰、高效的深度学习框架，其核心语言是 C++，它支持命令行、Python 和 Matlab 接口，既可以在 CPU 上运行，也可以在 GPU 上运行。其所有的计算以 Layer 形式表示。依托 Caffe 框架，其能有效、快速地搭建卷积神经网络(CNN)。综合上述各技术，本项目设计的检测平台的系统架构如图 2 所示。

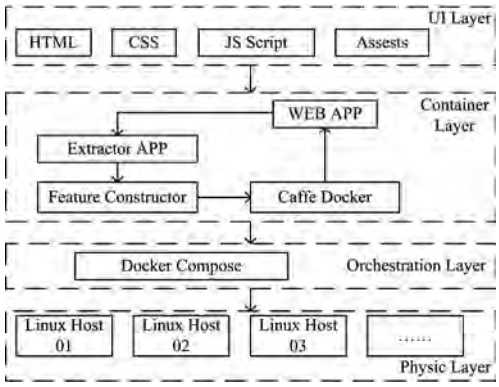

图 2　检测平台架构示意图

从图 2 中可以看出，本检测平台的架构主要分为 4 个层级。1)物理层(Physic Layer)：该层为 Linux 系统物理机组建的集群系统，同时为容器提供基础的运行环境和底层接口。2)编排层(Orchestration Layer)：主要负责为上层容器提供统一的部署和管理，并构建容器和宿主物理机之间的联系，统一分配和管理系统资源。3)容器层(Container Layer)：系统流程中的每个独立组件被封装成单个容器，可以通过多个容器并行计算来提高系统的整体效率和可用性。数据库集群为 MongoDB 主从结构组建的集群，能提高用户的访问速率与系统的整体可用性，是系统各组件数据持久化的核心。4)用户

界面层(UI Layer):该层作为 WEB 前端,负责与后端的 WEB APP SERVER 建立联系,提供友好跨平台的用户界面与用户系统的交互途径。

4.2 平台实现

4.2.1 灰度图的生成

为保证模型的有效性和准确性,本项目的输入数据采用的病毒和恶意代码样本主要来自 VxHeaven[13] 和日常收集的样本,所有样本均为可执行文件。

为便于分析,将上述每个病毒和恶意代码样本统一生成为宽度为 128 bits 的灰度图。具体实现时以 3.1 节介绍的方法最终生成无压缩的 PNG 格式图片。对文件末尾不足 128 bits 的文件,通过填充使其长度规整,填充的内容对应的区域全黑,因此不会影响最终的分析结果。

4.2.2 灰度图纹理特征的提取

平台实现时集成了多种灰度图纹理特征提取算法,其中主要采用了 GIST 纹理特征提取算法。GIST 算法[1] 最早被用于场景识别领域,结合它之前在图像检索领域的广泛运用,同时考虑到本项目中图片数据量大、图像纹理特征向量分明等特点,将该算法引入恶意代码的灰度图像的特征提取过程中。为降低分析模型的复杂度,进行特征抽取和转换,从人类心理学认知的角度,用 6 个特征对人类对图片的主观感受进行描述:人主观感受中的粗糙度(Coarseness)、对比度(Contrast)、方向度(Directionality)、线性度(Linelikeness)、规则度(Regularity)和粗细度(Roughness)。通过上述 6 个特征能更好地概括出灰度图的整体特征,进而保证后续训练模型的准确性和有效性。

4.2.3 模型的生成

获取各恶意文件的纹理特征后,为实现分类学习模型的构建,同时降低学习的时间复杂度和空间复杂度,依托 Caffe 高性能框架,集成多种深度学习算法,其中,主要使用卷积神经网络学习器。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,以发现数据的分布式特征表示。通过深度学习来训练之前的特征灰度图样本,从而生成分类预测模型。

4.3 分类结果的对比

测试时,选用的分析器包括 J48、朴素贝叶斯、KNN(K-近邻)、随机森林和基于 Caffe 的卷积神经网络 5 种。在大量实验结果的基础上,上述学习算法的准确率均达到了 90% 以上。各算法的识别率和误判率如表 1 所列。

表 1 各分类算法结果的比较

算法	识别率(*TPR*)	误判率(*FPR*)
J48	0.8920	0.0050
朴素贝叶斯	0.8960	0.0050
KNN	0.8450	0.0070
随机森林	0.9300	0.0030
CNN(Caffe)	0.9288	0.2005

通过上述结果可见,依托高性能框架 Caffe 的卷积神经网络对恶意样本的分类具有较高的识别率,虽略低于随机森林,但 Caffe 框架减少了学习所需的时间,降低了时间复杂度。

结束语 该文在分析了当前恶意软件造成的严峻形势的基础上,总结了国内外现有的恶意软件检测分类技术。为进一步开发有效的检测分类工具,结合现有的图像识别的深度学习方法,提出了基于机器学习的高性能恶意软件检测分类技术架构。其中,多种灰度图纹理特征提取算法的集成、高性能 Caffe 架构的使用都具有开创性的意义。下一步的主要研究方向是:增大学习样本数量,进而完善恶意软件的特征库;集成多种学习器,完善分类效果;在现有特征的基础上,构建高效且精简的特征优化方法;针对学习器进行对抗学习等。

参 考 文 献

[1] NATARAJ L, KARTHIKEYAN S, JACOB G, et al. Malware Images: Visualization and Automatic Classification[C]// Proceedings of the 8th International Symposium on Visualization for Cyber Security (VizSec'11). New York, USA, 2011: 401-407.

[2] 王蕊,冯登国,杨轶等. 基于语义的恶意代码行为特征提取及检测方法[J]. 软件学报,2012,23(2):378-393.

[3] NARUDIN F A, FEIZOLLAH A, ANUAR N B, et al. Evaluation of machine learning classifiers for mobile malware detection [J]. Soft Computing, 2016, 20(1): 343-357.

[4] 韩晓光,曲武,姚宣霞,等. 基于纹理指纹的恶意代码变种检测方法研究[J]. 通信学报,2014,35(8):125-136.

[5] MALIK J, KAUSHAL R. CREDROID: Android malware detection by network traffic analysis[C]// Proceedings of the 1st ACM Workshop on Privacy-Aware Mobile Computing. ACM, 2016: 28-36.

[6] KOLOSNJAJI B, ZARRAS A, WEBSTER G, et al. Deep learning for classification of malware system call sequences[C]// Australasian Joint Conference on Artificial Intelligence. Springer International Publishing, 2016: 137-149.

[7] 高程程,惠晓威. 基于灰度共生矩阵的纹理特征提取[J]. 计算机系统应用,2010,19(6):195-198.

[8] MOHANAIAH P, SATHYANARAYANA P, GURUKUMAR L. Image texture feature extraction using GLCM approach[J]. International Journal of Scientific and Research Publications, 2013, 3(5): 1.

[9] CARR J R, DE MIRANDA F P. The semivariogram in comparison to the co-occurrence matrix for classification of image texture[J]. IEEE Transactions on Geoscience and Remote Sensing, 1998, 36(6): 1945-1952.

[10] GOTLIEB C C, KREYSZIG H E. Texture descriptors based on co-occurrence matrics[J]. Computer Vision, Graphics, and Image Processing, 1990, 51(1): 70-86.

[11] HARALICK R M, SHANMUGAM K, DINSTEIN IH. Textural features for image classification[J]. IEEE Transactions on Systems, Man and Cybernetics, 1973, SMC-3(6): 610-621.

[12] PATEL J M, GAMIT N C. A review on feature extraction techniques in content based image retrieval[C]// International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE Computer Society, 2016: 2259-2263.

[13] HEAVEN V X. Computer virus collection [EB/OL]. URL: http://vxheaven.org/vl.