<点击这里添加中文题目>

<点击此处添加作者信息>

<作者单位>

2.02.02.02.02.02.02.02.02.02.02.02.02.02.02.02.02.0

**摘要**：（1）对于综述文章：以背景、研究现状、研究用途的结构书写，篇幅以150-300字左右为宜，不用第一人称做主语，不与正文语句重复。（2）对于一般研究性文章：以摘录要点的形式按目的、方法、结果、结论的结构报道出作者的主要研究成果，字数在200-400字左右为宜，不用第一人称做主语，不与正文语句重复。（正式稿件中请删除上述说明性文字）

**关键词**：关键词3~6个，用分号隔开；要能反映文章的基本观点，避免广义词。第一个关键词为该文内容所属二级学科名称（正式稿件中请删除上述说明性文字）。

**中图分类号**：请查阅《中国图书馆分类法》

<The Title of your paper>

<Author's Name>

<Author's Address>

**Abstract:** In this paper……….. (10 Points, Times New Roman)

Key words: key word 1; key word 2; key word 3

1. 引言

随着互联网的普及，恶意代码的危害也变得越来越来越难以控制。2017年5 月12 日晚上20 时左右，全球爆发大规模蠕虫勒索软件感染事件，仅仅几个小时内，该勒索软件已经攻击了99个国家近万台电脑。英国、美国、俄罗斯、德国、土耳其、意大利、中国、菲律宾等国家都已中招，且攻击仍在蔓延。据报道，勒索攻击导致16家英国医院业务瘫痪，西班牙某电信公司有85%的电脑感染该恶意程序。至少1600家美国组织，11200家俄罗斯组织和6500家中国组织和企业都受到了攻击。国内也有大量教学系统瘫痪，包括校园一卡通系统。根据国内知名互联网安全厂商奇虎360在2017年2月发布的安全报告，2016年360互联网安全中心共截获PC端新增恶意程序样本1.9亿个。敲诈者病毒在国内发生两次大规模传播，全国至少有497多万台用户电脑遭到了敲诈者病毒攻击。通过对受害者调研，42.6%的受害者不知道感染病毒的原因。预计在2017年敲诈者会增长10倍，且利用挂马攻击也将再次爆发。各种病毒、木马、蠕虫等恶意代码在网络间广泛传播，已经给个人、企业甚至政府带来了难以估量的损失， 因此研究更有效的恶意代码检测技术是十分有意义的。

恶意代码的各种混淆和变形手段是其难以检测的主要原因，传统的基于特征匹配的检测方法对于各种变形的恶意代码显得无能为力。随着机器学习和数据挖掘技术的快速发展，已经有许多研究将这些技术用于恶意代码的检测，并且取的了不错的效果。其中，大部分研究都是首先对操作码进行特征的提取，此时特征维度会特别高，然后再从中选择特征进行降维处理，最后对选择出的特征应用机器学习算法。这种方法的检测结果依赖于特征选择算法的有效性，而且由于操作码的局部性，对于程序总体流程的反映有着严重的局限性。针对以上存在的问题，本文通过对大量恶意代码在汇编层次的统计，提出对汇编代码进行粗粒度划分的方法，对恶意代码在汇编层次进行简化；此时既能很好的反映了程序的总体流程，又不会维度太高而需要进行降维处理；最后，基于随机森林算法对提取出的特征进行快速精准的分类。主要成果：1）提出一种对汇编指令进行集合映射的方法，建立一种改进的n-gram模型；2）基于hadoop分布式环境，实现了分布式汇编类特征模型训练；3）有助于研究人员了解恶意代码对汇编指令的使用情况，可用于进一步恶意代码家族分类。

1. 恶意代码检测方法整体设计

本文所述的恶意代码检测技术是一种静态检测技术。检测过程主要分为四步，分别是反汇编、汇编指令分类、提取特征、使用机器学习算法对样本进行训练和测试。

本文所采用的恶意程序样本来自于卡饭论坛以及360截获的恶意软件，非恶意程序样本在360软件库中随机下载。

接下来对收集的软件进行反汇编，以便从中提取所需的静态汇编特征。反汇编的工具有很多，为了便于脚本编程，本文采用objdump工具对所有PE文件进行反汇编。首先，将反汇编的结果以文件形式保存；接下来对所有文件进行汇编指令到集合的映射操作，根据原有汇编指令的顺序获得集合的顺序；最后根据保存集合顺序信息的文件提取n-gram信息。至此，特征提取工作结束。

在提取了样本特征后，将这些特征作为机器学习算法的输入，从而获得训练检测模型。其总体设计如图1所示：

图恶意代码检测模型

1. 恶意代码检测方法详细设计与实现
   1. 恶意代码检测技术概述
      1. 三级标题

图表的标注：图片标注在下方，表格标注在表格上方，居中，汉字用宋体，数字英文用Times New Roman, 罗马字用Symbol。字号：小五。

表1 名称……（小五号宋体）

Tab. 1 英文名称

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x*/cm | *I*/mA | *v*/(m⋅s-1) | *h*/m | *p*/MPa |
| 10 | 30 | 2.5 | 4 | 110 |
| 12 | 34 | 3.0 | 5 | 111 |

注：

文献出处标注：以参考文献的序号在正文中以[1]标注出，字体：Times New Roman，字号：五号，上标，颜色：蓝色。例：。。。年产量居全国之首[1]。。。。

图1 名称……（小五号宋体）

Fig. 1 英文名称

1. 一级标题
2. 一级标题
3. 结论

本文给出了………

致谢（可选）

应向对论文有帮助的有关人士或单位表示谢意。

[参考文献] (References)

<点击这里添加参考文献〉