**全日制本科生毕业设计开题报告**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **姓名** |  | **学号** |  | **专业** |  |
| **题目** | 基于深度学习的恶意代码检测设计与实现 | | | | |
| **设计背景与目标**  随着互联网和云计算的不断发展,人们生活的众多方面已经离不开网络和计算机。在计算机已经逐渐普遍的大数据和云计算时代中,许多不法分子也从中嗅到了商机，编写大量恶意代码窃取甚至破坏用户数据，抢占用户运算和内存资源导致客户业务无法正常运行。  随着恶意代码的种类和数量不断增加,攻击手段也在不断变化。攻击者也深知用户数据和计算量在大数据和云计算时代的重要性，从而衍生了许多攻击手段。从传统的木马后门和病毒程序到现在的勒索软件和挖矿软件，体现了攻击者的工作重心和攻击方向的转变。传统攻击方式主要以权限控制和数据窃取为主要目的，这种攻击手法不但对攻击者本身的安全素质要求很高，漏洞挖掘和恶意代码保护十分困难，不宜容易实施，攻击成本大，攻击成功率低。反观勒索和挖矿软件，它们本质上与正常程序无太大区别，没有包含非常容易识别的特征和敏感信息，攻击成本低，权限要求低，攻击范围广，可针对任何机器，加上许多运维人员对于安全意识缺失，导致非常多的未授权漏洞，攻击者有机可乘。  恶意代码数量多太多，如果使用传统基于恶意代码特征的方式进行检测，这种方法需要先使用人工静态或者动态分析之后，提取恶意代码特征，然后加入恶意代码特征库，最后将可疑程序和恶意代码特征库进行比较或者直接比较hash值，这不但不能很好地应对恶意代码的变种，而且需要花费大量时间和金钱在病毒分析人员上。传统的动态分析易受到运行环境的约束,很难获得全路径行为特征,且开投入销较大、效率较低。最近即年流行的基于机器学习的恶意代码检测方法,无法自动和有效地提取特征,比较依赖于人工提取特征,这些不太明显的特征无法准确地表示恶意代码,从而导致检测的准确率较下降等问题。针对以上问题,本文利用深度学习的思想和技术，采用tensorflow深度学习框架,从静态分析入手,对恶意代码进行检测。通过使用反编译框架capstone等，先进行传统二进制静态分析提取API和二进制代码保护的识别，将恶意代码上下文序列输入到LSTM中,结合LSTM的保存机制、遗忘机制和长时间记忆信息的特点获得长序列的操作行为信息，将其作为参数传入CNN网络进行恶意代码特征的提取，最后通过图形界面进行展示。 | | | | | |
| **设计思路、技术路线**  设计思路   1. 通过网络查找和阅读深度学习代码和论文 2. 设计深度学习模型 3. 寻找合适的数据集 4. 编写代码   技术路线  1.采用python语言设计  2.采用tensorflow或pytorch框架进行项目设计  3.采用CNN或VGG模型  4.采用微软或公开恶意代码数据集进行训练  5.采用pyqt框架进行图形界面展示 | | | | | |
| **设计进度计划**  2022.10.13-2023.1.4：完成毕业设计的选题，开题准备与开题报告  2023.1.7-2023.2.15：系统开发阶段，完成深度学习模型设计和图形界面的设计。  2023.2.16-2023.3.15：撰写论文初稿，并提交初稿和系统  2023.3.16-2023.3.25：修改论文，提交相关文档和论文终稿，修改并完善系统。  2023.3.26-2023.4：准备答辩相关事宜。 | | | | | |
| **指导教师意见**  同意开题、照此计划进行  **指导教师（签名）：罗凌** | | | | | |

说明：开题报告应在教师指导下由学生独立撰写，开题报告通过后方可做毕业设计。不能附页。