基于机器学习APT检测

难点一：

|  |  |
| --- | --- |
| 名称 | 从原始病毒文件到可分析数据的预处理 |
| 解决方法 | 先将病毒文件放到cuckoo沙盒中运行，得到病毒文件行为的原始报告，然后用文本处理方法，提取病毒行为的关键词，形成词频统计，以矩阵形式作为模型的输入。 |
| 提出者 | 张以宁 |

难点二：

|  |  |
| --- | --- |
| 名称 | 训练模型的导出和读取 |
| 解决方法 | 使用scikit-learn的joblib模块的dump方法将模型保存到pkl文件，load方法读取文件，还原出分类器对象 |
| 提出者 | 张以宁 |

难点三：

|  |  |
| --- | --- |
| 名称 | 对用户输入的测试样例与训练出的模型进行匹配（模型种类多，不同的模型只能测某种样例） |
| 解决方法 | 暂未解决 |
| 提出者 | 杜梁楷 |

难点四：

|  |  |
| --- | --- |
| 名称 | 软件管理期间，时间较短，代码工作较多，测试时间较少 |
| 解决方法 | 由于此软件开发模型为敏捷开发，因此测试模型采用H模型，将软件测试独立开来，一旦完成某一部分的要求即开始测试。 |
| 提出者 | 罗潘亚欣 |