1. 毕业设计（论文）主要研究内容、进展情况及取得成果

主要研究内容为在使用强化学习的算法上，训练智能体，对恶意PE进行扰动修改，使其绕过反病毒软件（例如360杀毒，ClamAV，火绒等）的检查。

目前对恶意程序进行修改的actions已全部完成，包括下列内容：

1.PE程序尾部新增无意义字节

2.PE程序节内空洞添加无意义字节

3.PE程序添加随机资源，例如图标

4.PE程序可选首部的校验和修改

5.PE程序添加伪造的签名

6.PE程序修改时间戳

7.PE程序的debug信息被删除

8.PE程序新增一个节，但节中包含的信息是无意义内容

9.PE程序经过加壳处理

等操作

对PE程序扫描的ClamAV服务扫描模块已经完成。

同时添加了可选日志功能，允许程序记录对PE文件的更改操作。

强化学习的算法目前成功完成SARSA算法，设置state总数为16，限制16次操作，为避免文件太大拖延智能体的训练时间，训练的样本是从卡饭论坛上下载的PE恶意程序，目前大约有1500个，分成三组进行训练。

还需要完成Q-Learning算法，以及在20号或者27号之前完成论文初稿的大部分内容，以及下载更多样本。actions可能会新增一种GitHub上的开源壳：amber，但它不如UPX壳稳定，很多情况下哪怕是无害PE文件都会因为这个壳被误杀。

目前三组PE恶意程序数据训练的结果大概如下：

杀毒软件查杀率：

360杀毒（联网，只开启云查杀引擎，不开启QVM2引擎）：668/674->22/674

ClamAV（离线病毒库已更新到最新版）：128/674-> 98/674

火绒：未测试

360杀毒（联网，开启云查杀引擎和QVM2引擎）：未测试

另外还有一些内容值得研究：

为什么ClamAV对恶意PE文件的扰动修改的抗干扰能力更强，相对360杀毒来比较，初步可以认为360杀毒更依赖云端上传使用沙箱来进行分析，同时360的病毒库很小，包含QVM和鲲鹏病毒库只有8MB左右，而ClamAV的病毒库大约有600MB以上，说明包含了更多病毒的特征。

在断网情况下，笔者之前做了一次实验，用了部分样本，使用360杀毒（断网 开启云查杀引擎和QVM2引擎）进行查杀，结果查杀率只有18/51，而在联网情况下，这部分样本全会被查杀，经过强化学习模型免杀处理后的样本查杀率只有14/51。

进展可以认为项目的代码基本完成，有效果，但是不够明显，仍然需要优化，比如更换强化学习的算法，调整reward epsilon alpha gamma等参数，以及基于文件大小变换的惩罚参数k，以及扩大样本量。

而且有时候需要注意机器负载问题，在4月12号进行采用SARSA算法的强化学习对抗性样本生成模型时，PyCharm2024.3.5出现了崩溃问题，排查原因，根据错误报告发现是和PyCharm使用的JVM虚拟机有关，崩溃时出现了内存溢出问题。同时，可能需要调整可选动作数量，目前测试只使用了10个动作，是从原始动作集合中随机抽选形成一个新的action\_list来给强化学习模型使用。State数量也需要调整，有一些样本经过足够的新增节处理是可以通过ClamAV免杀的。

目前三组样本的情况：这里采用了SARSA强化学习算法，，使用ClamAV杀毒软件进行检测

sample1（样本一）

amount（总恶意程序数）:673

before1（处理前逃逸ClamAV查杀的恶意程序数）:526

after1（处理后逃逸ClamAV查杀的恶意程序数）:551

sample2（样本二）

amount（总恶意程序数）:435

before2（处理前逃逸ClamAV查杀的恶意程序数）:313

after2（处理后逃逸ClamAV查杀的恶意程序数）:339

sample3（样本三）

amount（总恶意程序数）:439

before3（处理前逃逸ClamAV查杀的恶意程序数）:351

after3（处理后逃逸ClamAV查杀的恶意程序数）:357

这些比较大的样本暂时还没经过360杀毒和其他杀毒软件进行检验，目前笔者已经正在处理第四组训练数据，同时，上述三组训练数据，在原始样本上，笔者手动删除了大小过大的恶意程序和所有非PE恶意程序（例如HTML应用程序 JavaScript恶意脚本 VBS恶意脚本 恶意shellcode 伪装成正常office文档的office宏病毒等非PE文件），这里我是按照8MB来区分的，大于8MB的恶意程序会被删除，过大的恶意程序会导致智能体的处理时间非常漫长。

同时笔者发现在前三组训练数据的处理过程中，没有出现后文中的惩罚参数k异常问题，这是我在第四组训练数据中发现的问题。

1. 存在的问题和拟解决方案

目前存在的问题如下：

1.样本免杀率低，可能需要扩大训练样本数以及适量增加state数量或修改一些参数，更换更好的强化学习算法来调整。笔者认为应该鼓励智能体采用增加无害节和无害后缀字符串的行为。

2.对强化学习算法的原理掌握仍然不够熟练，尽管代码没有灾难性错误，目前在使用赵世钰老师的强化学习的数学原理这本书来进行学习，同时在GitHub上有代码样例https://github.com/MathFoundationRL/Book-Mathematical-Foundation-of-Reinforcement-Learning，以及MOOC上有完整的视频教程来进行学习。

3.样本中某些文件不是PE文件，尽管后缀名是exe，如果action没有异常处理，会直接报错导致整个程序崩溃。以及有一些原始PE文件因为自身结构问题，导致某些智能体采取某些操作会报错，这些异常如果没有被捕获对程序来说也是灾难性的。需要额外对项目中的某些模块进行异常处理捕获，防止造成程序崩溃。

4.样本筛选模块没有集合到SARSA算法智能体中，导致用户需要额外再运行一个python文件去除掉那些大于4MB的样本，没有筛选掉过大的样本，会导致处理时间太长。

5.惩罚参数k有很小的概率会为负值，这是因为k=(1/3)\*(ln(处理后文件大小/前文件大小)+1)的函数值域问题，需要更换惩罚参数k的表达式，避免意外情况的出现

可能会潜在稀疏奖励问题，可能需要鼓励智能体尽可能去做新增无害段内容或新增冗余后缀等操作，允许给予小额奖励（这里我暂时给的是1）。

5.同时需要注意恶意样本的查杀和检测有效性的操作，如果不注意，将会对计算机带来严重的危害，这些操作不应该在实体机上运行，应该换用Widnows7虚拟机中进行，同时拥有镜像可以方便更换杀毒软件种类，也能防止危害主机，虽然360杀毒开启云沙箱上传可疑文件之后能查杀几乎所有样本，但仍然需要考虑更换杀毒软件空窗期的问题。

1. 下一步研究任务与进度安排

1.翻译完指导教师和学长提供的一些英文论文，再对这些论文中的一些关键代码进行研究。

2.在卡饭论坛等网站下载更多PE恶意程序样本用于强化学习模型，完成强化学习智能体处理后的样本对火绒以及其他杀毒软件免杀的情况分析。

3.学习强化学习的数学原理这本书，补充有关强化学习算法的其他知识，已改进原有的SARSA算法模型，以及采用更好的算法

4.在四月末，基本完成论文初稿的大部分内容。

5.研究360杀毒在开启不同引擎的情况下对样本的查杀率。

6.更换惩罚函数k，减少意外情况导致原本奖励应该是正数，但是却给了负数奖励（惩罚）的情况，这通常发生于加壳处理之后，对原文件有一定压缩，导致处理后文件大小/源文件大小过于小，从而导致奖励出现意外的情况

7.更新开源协议免责声明，注明所有实验在本机封闭环境进行，且安装有杀毒软件保护，未对其他计算机造成任何影响，生成的样本仅用于学术研究探讨交流学习，禁止用于任何非法用途。所有样本均为真实威胁，可能具有严重危害性。不要在没有任何安全措施保存，打开或执行这些样本，必须在虚拟机中测试样本。