第二次提交物文档

Luckycat小组

目录

[第二次提交物文档 1](#_Toc51190247)

[一、人员名单及分工 3](#_Toc51190248)

[二、工作计划及执行情况 3](#_Toc51190249)

[一）吴焘宇 3](#_Toc51190250)

[二）周志豪 3](#_Toc51190251)

[三）田希宇 4](#_Toc51190252)

[四）张修平 4](#_Toc51190253)

[五）孙欢 4](#_Toc51190254)

[六）彭元驰 4](#_Toc51190255)

[七）执行情况 5](#_Toc51190256)

[三、项目管理 5](#_Toc51190257)

[一）项目管理工具 5](#_Toc51190258)

[二）开发环境 5](#_Toc51190259)

[三）开发配置 5](#_Toc51190260)

[1.笔记本硬件 5](#_Toc51190261)

[2.虚拟机配置 5](#_Toc51190262)

[四）项目甘特图 6](#_Toc51190263)

[五）项目风险预测及对策分析 6](#_Toc51190264)

[1.需求变更风险 6](#_Toc51190265)

[2.沟通不良风险 6](#_Toc51190266)

[3.进度风险 6](#_Toc51190267)

[4.质量风险 7](#_Toc51190268)

[5.系统性能风险 7](#_Toc51190269)

[6.工具风险 7](#_Toc51190270)

[7.技术风险 7](#_Toc51190271)

[8.团队成员能力和素质风险 7](#_Toc51190272)

[9.团队成员协作风险 7](#_Toc51190273)

[四、项目经济 8](#_Toc51190274)

[一）预算 8](#_Toc51190275)

[五、需求分析 8](#_Toc51190276)

[一）引言 8](#_Toc51190277)

[1.目的 8](#_Toc51190278)

[2.系统用户需求 8](#_Toc51190279)

[二）功能需求 9](#_Toc51190280)

[1.系统功能列表 9](#_Toc51190281)

[三）非功能需求 9](#_Toc51190282)

[四）概要设计 10](#_Toc51190283)

[六、会议记录 10](#_Toc51190284)

[一）第一次会议 10](#_Toc51190285)

[二）第二次会议 11](#_Toc51190286)

[三）第三次会议 12](#_Toc51190287)

[四）第四次会议 13](#_Toc51190288)

[五）第五次会议 14](#_Toc51190289)

# 一、人员名单及分工

|  |  |  |
| --- | --- | --- |
| 姓名 | 职责1 | 职责2 |
| 吴焘宇 | 组长、数据处理负责人 | 未知协议分类算法 |
| 周志豪 | 项目管理负责人 | 数据处理、测试 |
| 田希宇 | 已知协议分类算法负责人 | 数据处理 |
| 张修平 | 未知协议算法分类负责人 | 已知协议分类算法 |
| 孙欢 | 文档负责人、数据库负责人 | 已知协议分类算法 |
| 彭元驰 | 测试负责人 | 未知协议分类算法 |

# 二、工作计划及执行情况

## 一）吴焘宇

1、 设计pcap包数据获取子系统，监察网口流量， 使用wireshark捕获pcap包，将获得的pcap包进行预处理，提取出载荷信息。

2、 设计数据处理子系统，将已有的载荷信息进行裁剪，并其转化成为灰度图格式，传至深度学习模块。

3、 获取深度学习系统传入的信息，并将数据转换成k-means方法所需要的数据格式，传入k-means分析模块中。

4、 调整数据格式与裁剪方法，提高算法正确率。

## 二）周志豪

1、使用project进行项目计划制定

2、使用git进行项目代码与文档的版本管理

3、制定项目的财务预算

4、对项目的进度进行把控

5、制定PPT，对每个阶段的项目结果进行记录与展示

6、总体代码可分为两个部分，第一部分为数据处理，第二部分为深度算法实现。

7、设计测试用例并对两个不同的模块分别进行白盒以及黑盒实验。

8、模块接口测试

9、设计测试用例对整体进行黑盒测试  
10、测试环境：linux系统

## 三）田希宇

1.训练样本预处理。将pcap包载荷部分处理加工成图像，并对图像进行预处理操作。

2.cnn网络选择。cnn有多种，如LeNet-5,AlexNet,ZFNet,VGG等。我们要基于特征和层的设计，选择一种适合的cnn网络。

3.参数设计。如卷积层的卷积核，激活函数，优化函数和损失函数的设计和选择。

4.模型训练。将处理好的数据导入cnn网络进行训练，得到识别模型。

5.识别测试。测试模型对原始数据的识别情况。

## 四）张修平

1、查阅相关文献，了解有关未知协议分类的算法原理和实现方法，这里主要是K-means聚类算法。

2、使用python具体实现K-means聚类算法，调试运行

3、用K-means聚类算法分析解析好的数据，记录分类结果

4、分析聚类结果，修改相关参数，改进算法效率。

## 五）孙欢

1、用git进行版本控制

2、撰写或整理需求分析、开发过程、软件测试、使用说明、项目总结、其他文档、经济预算、人员名单及分工、详细开发过程、测试方案、测试开发过程、个人总结、团队总结、项目总结等文档

3、总体代码可分为两个部分，第一部分为数据处理，第二部分为深度算法实现。

4、设计测试用例并对两个不同的模块分别进行白盒以及黑盒实验。

5、模块接口测试

6、设计测试用例对整体进行黑盒测试

7、测试环境：linux系统

## 六）彭元驰

1、在linux环境下装数据库

2、将程序的原始数据、训练集、测试集、输出结果全存在数据库里

## 七）执行情况

见ppt

# 三、项目管理

## 一）项目管理工具

Project/Git

## 二）开发环境

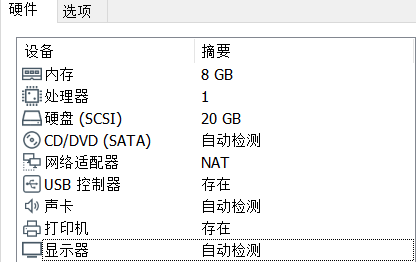
Linux/Python

## 三）开发配置

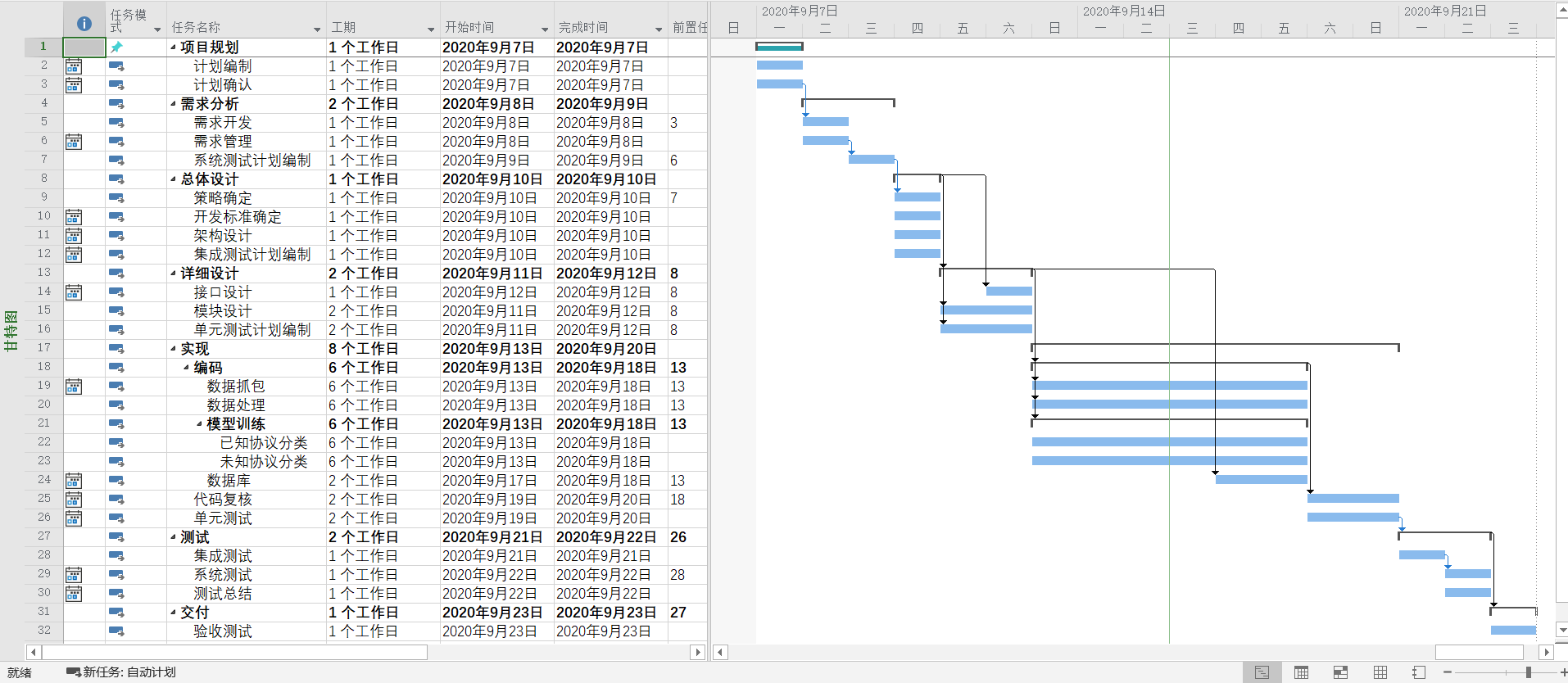
### 1.笔记本硬件

cpu i5 7300H，显卡 1050Ti

### 2.虚拟机配置



## 四）项目甘特图



## 五）项目风险预测及对策分析

在项目的建设过程中，风险几乎无处不在。有的风险可以预料，有的风险不可预料，但无论如何，提前预测风险及制定对策，对项目的成功起着至关重要的作用。

以下是此次项目中可能遇到的风险及相应对策：

### 1.需求变更风险

需求变更是软件开发项目中经常发生的事情。开发人员往往由于无限度的需求变更而苦不堪言甚至导致项目的失败。

预防这种风险的办法是项目建设之初就仔细分析项目需求，确定需求变更控制流程、记录并归档用户的需求变更申请。.

### 2.沟通不良风险

项目组与项目各干系方沟通不良是影响项目顺利进展的一个非常重要的因素。

预防这种风险的办法是项目建设之初就和项目各干系方约定好沟通的渠道和方式、项目建设过程中多和项目各干系方交流和沟通，并注意培养和锻炼自身的沟通技巧。

### 3.进度风险

有些项目对进度要求非常苛刻，项目进度的延迟意味着违约或市场机会的错失。尽管此项目对进度要求不高，我们同样要考虑该风险

预防这种风险的办法是分阶段交付产品、增加项目监控的频度和力度、多运用可行的办法保证工作质量避免返工。.

### 4.质量风险

此项目对软件质量有较高的要求,如果项目组成员同类型项目的开发经验不足，则需要密切关注项目的质量风险。

预防这种风险的办法一般是经常和用户交流工作成果、采用符合要求的开发流程、认真组织对产出物的检查和评审、计划和组织严格的独立测试等。

### 5.系统性能风险

此项目可能用于较大数据量的处理，系统对性能要求很高，这时项目组就需要关注项目的性能风险。

预防这种风险的办法一般是在进行项目开发之前先设计和搭建出系统的基础架构并进行性能测试，确保架构符合性能指标后再进行后续工作。

### 6.工具风险

项目开发和实施过程，所必须用到的管理工具、开发工具、测试工具等是否能及时到位、到位的工具版本是否符合项目要求等，也是项目组需要考虑的风险因素。

预防这种风险的办法一般是在项目的启动阶段就落实好各项工具的来源或可能的替代工具，在这些工具需要使用之前跟踪并落实工具的到位事宜。

### 7.技术风险

在此项目开发和建设的过程中，技术因素是一个非常重要的因素。项目组一定要本着项目的实际要求，选用合适、成熟的技术，不可无视项目的实际情况而选用一些虽然先进但并非项目所必须且自己又不熟悉的技术。如果项目所要求的技术项目成员不具备或掌握不够，则需要重点关注该风险因素。

预防这种风险的办法是选用项目所必须的技术、在技术应用之前，针对相关人员开展好技术培训工作。

### 8.团队成员能力和素质风险

团队成员的能力(包括业务能力和技术能力)和素质，对项目的进展、项目的质量具有很大的影响，项目经理在项目的建设过程需要实时关注该因素。

预防这种风险的办法是在组建团队时选对人、不同方面的人才要兼具，将合适的人安排到合适的岗位上。

### 9.团队成员协作风险

团队成员是否能齐心协力为项目的共同目标服务，是影响进度和质量的关键因素。

预防这种风险的办法是项目在建设之初项目经理就需要将项目目标、工作任务等和项目成员沟通清楚，倡导团结互助共同进步。

# 四、项目经济

## 一）预算

|  |  |  |
| --- | --- | --- |
| **开支种类** | **开支名** | **预算（单位：元）** |
| 设备费 | 带有RTX2080TI显卡主机 | 20000 |
| 知识产权事务费 | 正版软件 | 5000 |
| 劳务费 | 人员工资 | 8000\*6 |
| 会议费 | 开会场地费 | 5000 |
| 专家咨询费 | 请专家 | 5000 |
| 其他费用 | 其他费用 | 10000 |
| 总计 | 无 | 92000 |

# 五、需求分析

## 一）引言

### 1.目的

本文档是基于深度学习方法的网络流量识别与分类系统的需求说明书。主要记录了项目功能与非功能需求。

### 2.系统用户需求

连接互联网的linux用户

## 二）功能需求

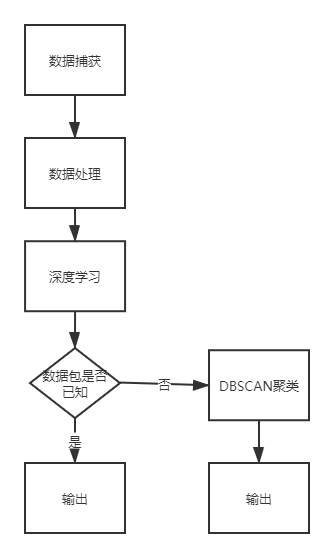
### 1.系统功能列表

|  |  |  |
| --- | --- | --- |
| 功能类别 | 功能名称 | 一般过程描述 |
| 数据获取 | 网口流量获取 | 监察网口流量，使用pypcap捕获pcap包，将获得的pcap包进行预处理，提取出载荷信息。 |
| 数据处理 | 字节流转灰度图 | 将已经处理好的字节流转换成灰度图格式，并传递深度学习模块 |
| 载荷转DBSCAN数据 | 当深度学习模型无法识别出pcap协议类型时，将pcap包标记成未知协议包，并将数据处理后传给未分类协议处理模块。 |
| 深度学习 | densenet模型训练 | 应用现有的训练集训练模型 |
| densenet数据分析 | 将训练出的模型用于分析pcap包协议类型。 |
| 未分类协议处理 | DBSCAN聚类算法 | 将深度学习算法识别不出的数据进行聚类处理 |
| 数据库 | 数据库存储 | 将处理得到的灰度图存储地址与输出的结果匹配并存入数据库 |

## 三）非功能需求

* + 处理的原始数据量不少于2GB（普通笔记本）

## 四）概要设计



# 六、会议记录

## 一）第一次会议

|  |  |  |  |
| --- | --- | --- | --- |
| **LuckyCat小组第一次会议记录** | | | |
| **时间** | 2020/9/8  19:00——19:40 | **地点** | 3号学生公寓345 |
| **主持人** | 吴焘宇 | **记录员** | 孙欢 |
| **时间调度** | 40分钟 | | |
| **参加人员** | 吴焘宇、周志豪、彭元驰、孙欢、田希宇、张修平 | | |
| **会议议题** | 系统框架设计 | | |
| **发言人及发言主要内容** | | | |
| **吴焘宇** | 主要创新点在于深度学习算法和数据特征提取  但是对具体方法的应用还没有具体的思路，希望咨询老师过后能找到合理的解决方案  提炼当前不理解的问题明天咨询老师 | | |
| **周志豪** | 了解深度学习的学习过程及如何训练模型  阅读老师的代码以及相关的文献  知悉当前进度并规划未来进展 | | |
| **彭元驰** | 大致确定两个关于深度学习算法的创新方向，其一是选择另一种更加精确的深度学习算法（较难），其二在老师的基础上对算法进行修改 | | |
| **孙欢** | 当前深度学习基础薄弱，阅读文献时有很多看不懂的地方 | | |
| **田希宇** | 要学深度学习的话要先看神经网络 | | |
| **张修平** | 对未知协议的分类，可以使用kmeans分类算法 | | |
| **会议相关照片** | | | |
| C:\Users\SunHuan\Documents\Tencent Files\2060207050\Image\Group2\R3\QM\R3QM362J)DT]G]]~BXELR%L.jpg | | | |

## 二）第二次会议

|  |  |  |  |
| --- | --- | --- | --- |
| **LuckyCat小组第二次会议记录** | | | |
| **时间** | 2020/9/9  10:50——11:40 | **地点** | 软件楼4楼 |
| **主持人** | 张继 | **记录员** | 孙欢 |
| **时间调度** | 40分钟 | | |
| **参加人员** | 吴焘宇、周志豪、彭元驰、孙欢、田希宇、张修平、张继老师 | | |
| **会议议题** | 了解老师已完成的工作，确定系统制作的大体思路 | | |
| **发言主要内容** | | | |
| 不讲创新点，先确定设计的整体思路和系统处理的范围  网络流量的分类，靶子有多大  确定系统采用将字节流转换成为灰度图并使用cnn卷积神经网络的方法分析已知协议的pcap包，并对未知协议的pcap包使用kmeans的方法进行聚类  将参考2015年对木马进行分类的获奖系统进行制作  讨论了相关财政问题  明确细化人员分工，捋清了软件的逻辑结构 | | | |
| **会议相关照片** | | | |
| 无 | | | |

## 三）第三次会议

|  |  |  |  |
| --- | --- | --- | --- |
| **LuckyCat小组第三次会议记录** | | | |
| **时间** | 2020/9/12  8:30——11:30 | **地点** | 3号学生公寓二楼 |
| **主持人** | 吴焘宇 | **记录员** | 孙欢 |
| **时间调度** | 3小时 | | |
| **参加人员** | 吴焘宇、周志豪、彭元驰、孙欢、田希宇、张修平 | | |
| **会议议题** | 数据库相关问题、聚类算法及经典卷积神经网络算法 | | |
| **发言主要内容** | | | |
| 1.任务再分工：加入了MongoDB数据库的使用，分配相关人员，以及其他任务的微调整  2.数据库相关问题：相关人员学习MongoDB数据库的使用，大致了解其如何储存pcap包及对应的类型与调用的接口  3.除了kmeans（划分聚类）之外的聚类算法如划分聚类（k-中心点算法、CLARANS算法）、层次聚类、模糊聚类、基于密度聚类等，挑选合适的聚类算法，希望在下次汇报前能有实质的进展  4.经典卷积神经网络算法如VGGNET、GOOGLENET、RESNET等的大致了解，挑选合适的详细学习如何代码实现  5.对老师提供的代码进行了更加细致的分析，大致读懂各部分代码的功能 | | | |
| **会议相关照片** | | | |
| C:\Users\SunHuan\Documents\Tencent Files\2060207050\Image\Group2\I0\2]\I02]JS[_C3RPLPPUA5_RLHQ.jpg | | | |

## 四）第四次会议

|  |  |  |  |
| --- | --- | --- | --- |
| **LuckyCat小组第四次会议记录** | | | |
| **时间** | 2020/9/13  19:30——20:10 | **地点** | 3号学生公寓345 |
| **主持人** | 吴焘宇 | **记录员** | 孙欢 |
| **时间调度** | 40分钟 | | |
| **参加人员** | 吴焘宇、周志豪、彭元驰、孙欢、田希宇、张修平 | | |
| **会议议题** | 交流每个人相关工作的进展 | | |
| **发言主要内容** | | | |
| 1.已知协议部分进展：通过对比各种经典CNN模型的优缺点，为了使结果更加准确，确定使用DenseNet作为训练网络  2.数据库部分进展：确定使用MongoDB非关系型数据库储存pcap包的类型  3.未知协议部分进展：确定分别使用BIRCH或Kmediode聚类算法对未知协议进行无监督聚类，根据运行效率和精确度择优选择。  4.数据处理部分进展：把数据包的内容的每一个字节转化为0—255的数字，暂取28\*28个字节 | | | |
| **会议相关照片** | | | |
| C:\Users\SunHuan\Documents\Tencent Files\2060207050\Image\Group2\61\`)\61`)J51OTQNJ{Z7_P4B%GOF.jpg | | | |

## 五）第五次会议

|  |  |  |  |
| --- | --- | --- | --- |
| **LuckyCat小组第五次会议记录** | | | |
| **时间** | 2020/9/16  19:30——20:30 | **地点** | 3号学生公寓345 |
| **主持人** | 吴焘宇 | **记录员** | 孙欢 |
| **时间调度** | 1小时 | | |
| **参加人员** | 吴焘宇、周志豪、彭元驰、孙欢、田希宇、张修平 | | |
| **会议议题** | 讨论当前的工作进展及未来规划 | | |
| **发言主要内容** | | | |
| 1.经过老师建议，我们将包头的数据也放入灰度图中进行处理，经过实际的操作发现，包头放入对训练模型无任何帮助。  2.经过比较后确认DBScan为最终的无监督聚类算法，学习配置了sklearn库相关的API，计划这周具体实现响应算法并完成调参。  3.构建了Densenet网络，采用Densenet-BC网络结构 | | | |
| **会议相关照片** | | | |
| C:\Users\SunHuan\Documents\Tencent Files\2060207050\Image\Group2\IV\V6\IVV6(XB}WG)T5XP(E$5TDPB.jpg | | | |