# 说明文档

## 1.数据集

数据集为USTC-TFC2016，包含Benign正常流量以及Malware木马流量。

文件格式为pcap

## 2.数据预处理

文件夹PcapToMnist包含数据预处理所需要的文件。

0\_Tool:包含流量切分工具

1\_Pcap:需处理的pcap文件存放位置

2\_Session:经过2\_PcapToSession处理后的会话文件存放位置

3\_ProcessedSession:经过3\_ProcessSession处理后的会话文件存放位置

4\_Png:经过4\_Session2png处理后的图像文件存放位置

5\_Mnist:经过5\_Png2Mnist处理后的mnist文件存放位置

处理流程：

1.将需要处理的pcap文件放入1\_Pcap

2.依次执行2\_PcapToSession，3\_ProcessSession，4\_Session2png，5\_Png2Mnist

3.得到mnist文件

4.将mnist文件解压后的文件与converter.m文件放在同一目录

5.在matlab执行converter.m文件得到mat文件

## 3.模型训练

文件夹tensorflow包含，模型训练的代码，代码的运行环境为MacOS 10.15.4， python 3.6.8，tensorflow 1.2.1.

模型训练流程：

1.将预处理后的mat文件放入data/images/compare文件夹

2.修改datasets/compare.py中TRAIN\_NUM和TEST\_NUM

FILES = {

'train': Datafile(os.path.join(DIR, 'compare\_train.mat'), TRAIN\_NUM),

'test': Datafile(os.path.join(DIR, 'compare\_test.mat'),TEST\_NUM)

}

TRAIN\_NUM可用matlab打开compare\_train.mat文件，查看文件中的y值得到，TEST\_NUM同理。

3.设置train\_compare.py中n\_labeled（标记样本数），training\_length训练长度，rampup\_length超参数a的变化长度

4.执行python train\_compare.py进行模型训练

## 4.对比实验

文件夹对比实验包含机器学习对比方法（ML）和论文对比方法(HTFS)：分别实现了三类对比实验

experiment1:在少标记样本下（10%、5%、1%、0.1%标记样本）的二分类（木马流量，正常流量）对比实验。

experiment2：检测未知样本的分类效果对比，即从十类样本中选取九类作为训练集，另一类作为测试集，总共重复十次。

experiment3：前述模型在少标记样本下的多分类效果对比。分别设置10类木马流量以及包含所有类的全集木马流量，标记样本占比为10%。

## 实验结果

实验结果见文件夹 “图片”