**本科实验指导书**

实验名称： CAN总线入侵检测

开课学院： 计算机学院

审 核 人： 逄德明 讲师

教 辅： 高良 研究生

国防科学技术大学训练部制

一、实验目的

 实验目标是实现一种用于车载网络的高精度实时入侵检测方法，该方法可以通过较少的工作量部署在当前的车载网关节点上。为达到此目的，所提出的入侵检测方法可以部署到车辆网关平台中的软件插件或者作为要加载到现有CAN总线上的硬件节点。设计实现一种基于固定数量的消息作为滑动窗口的信息熵监测算法，以及基于模拟退火算法对滑动窗口进行优化。

二、实验内容简要描述

1．实验算法一：给定固定大小的滑动窗口在CAN总线消息中实现基于信息熵检测的DoS攻击检测算法

2．实验算法二：使用模拟退火算法改进实验一中算法的滑动窗口战略，优化决策条件，使得检测精度提高、假阳性率降低、响应时间缩短

三、实验原理、方法、手段

1. 基本概念

（1）CAN总线消息：CAN总线消息中，消息的优先级由其ID字段标识。正常情况下，在一定时间(或一定窗口)内取样得到的CAN总线消息，各ID段消息占据的比例应该不会出现大的波动，取样窗口的信息熵应处于一定的范围之内。

（2）DoS攻击： 攻击者可以在总线上的短周期注入高优先级消息，影响CAN总线对于正常消息的处理，影响汽车电子系统的正常响应。

假设采样窗口的长度为w。由于DoS攻击的特性，在一个采样窗口内高优先级消息的比例会在短时间内增大，相应的低优先级的消息比例降低。这会导致采样窗口内信息熵的异常变化。实验计算正常情况下CAN总线中信息熵的平均值a，标准差α,预测DoS攻击时设置决策条件区间（a-kα,a+kα），将信息熵超出决策条件的采样窗口预测为发生异常情况（即为DoS攻击）。其中k控制决策区间敏感度的常数。

（3）信息熵：熵是信息学上代表随机变量的不确定度的度量。一个离散随机变量X的熵H(X)定义为：

H(X) =

（4）模拟退火算法，它是一种启发式搜索算法，按照预定的控制策略进行搜索，在搜索过程中获取的中间信息将用来改进控制策略。

2. 相关定义：将一个采样（滑动）窗口看作一个块，那么定义：

（1） 预测准确率Ra=正确检测为攻击块的数目Da/攻击块总数Ta

（2） 假阳性率 Rn=正常信息块被检测为攻击块的数目Dn/正常信息块的数目Tn

（3） 响应时间 Rt=攻击开始时间At-攻击检测时间Dt

四、实验条件

1. 实验所用数据集：

DoS\_attack\_dataset\_no\_zero.csv是未加入DoS攻击的正常CAN总线数据集

Add\_DoS\_attack\_dataset1.csv是加了DoS攻击块的CAN总线数据集，DoS攻击块由ID=0标识

在excel中打开数据可能出现排版错乱，可直接在代码编辑器中打开。

2. 实验条件：建议使用python语言，

可能需要的python包：pandas，math，matplotlib，numpy

五、实验步骤

通过下述两个算法分别计算入侵检测的三个性能指标：**预测准确率、假阳性率、响应时间**

1. 算法一的实验步骤

设定采样窗口长度为w，决策敏感度为k

•第1步：从正常数据集中提取CAN消息，使用滑动窗口长度w作为信息熵采样窗口，分析在滑动窗口内捕获的正常数据信息熵。计算熵的标准差θ和均值a。

•第2步：从测试数据集中提取CAN消息IDs以获取新的IDs集。

•第3步：使用滑动窗口作为信息熵采样窗口，以分析在滑动窗口内捕获的测试数据集消息IDs的信息熵。

·第4步：将第3步中获得的信息熵与正常信息熵的决策范围进行比较，以确定是否存在攻击块。

•第5步：将步骤4中获得的攻击结果与添加标记的攻击块进行比较，得到初始检测精度数据。

2. 算法二的实验步骤

•按照一定的规则迭代选择新的采样窗口长度为w，决策敏感度为k。整个训练过程采用模拟退火算法，每次循环并重复上述步骤1到步骤5，计算当前的代价函数值。以最高的代价函数来确定最佳的窗口尺寸和决策条件。优化采样窗口长度w，决策敏感度k的取值。

（1）确定优化目标函数

综合考虑预测准确率Ra、假阳性率 R、响应时间 Rt

max优化目标函数E=C1\*Ra-C2\*Rn-C3\*Rt

C1,C2,C3为权重参数取值C1=1,C2=0.5,C3=1

（2）初始化参数条件

窗口大小W，敏感度k， 初始温度T

（3）确定循环终止条件

外循环：（终止温度T，期望优化目标E达成，循环迭代次数N，系统熵稳定情况）

内循环：（目标函数均值，若干步变化小，迭代步数）

（4）确定状态转移函数

（W，θ）🡪(W1,θ1)

建议尽量控制W，k变化范围，降低求解空间

（5）确定冷却进度（例Tt=T0/(1+t)）

（6）确定状态转移条件，转移概率

六、实验要求与注意事项

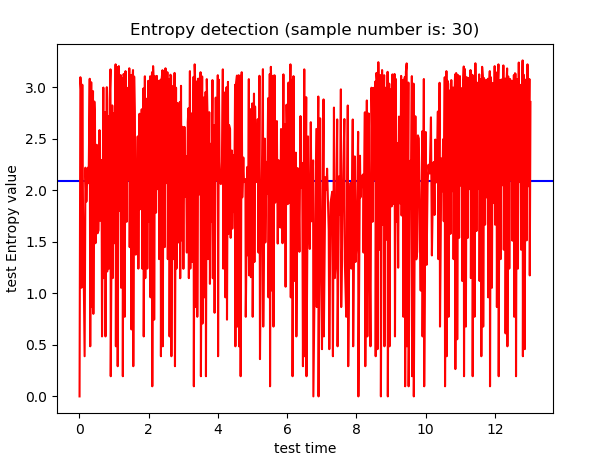
1. 实验要求：

* + 单人完成上述实验，提交实验代码并撰写实验报告，实验代码要求可以运行并添加详细注释；
  + 严禁抄袭，对于出现雷同代码的情况，最早提交代码的同学正常计分，其他后交的同学按照抄袭处理（以实验提交材料发送到我邮箱的时间计算）。
  + 每个人提交的实验材料放在一个文件中，文件夹命名方式：学号+姓名，打包发送到邮箱：[pang3724@163.com](mailto:pang3724@163.com)， [853642470@qq.com](mailto:853642470@qq.com)。

2.注意事项：

实验过程注意保留实验结果，分析自己算法的复杂度，算法原理，注意理解原论文，以上提到的终止条件，状态变化等设定可以灵活选择。欢迎提出对论文和实验相关的问题分析其优缺点分析。

需要使用绘制出类似下图的横坐标为检测时间，纵坐标为熵值的结果图。(使用matplotlib这个包很方便)



七、附录：

1. 参考论文：Sliding Window Optimized Information Entropy Analysis Method for Intrusion Detection on In-Vehicle Networks

2. 将处理数据可能用到的几个函数写在了一个小例子里，供大家参考

def example():

# 读取csv文件

df = pd.read\_csv('DoS\_attack\_dataset\_no\_zero.csv', sep='\t')

#取出窗口为0到100的数据

data = df.iloc[0:100]

#得到本窗口内第1条消息的时间戳的值

end\_time = data.iloc[0]['TimeStamp']

#得到窗口内不同ID消息出现的频率

p\_data = data['ID'].value\_counts() / data['ID'].count()

#计算不同ID消息出现的频率的标准差

arr\_std = np.std(p\_data, ddof=1)

print(arr\_std)

3. 模拟退火算法(引自http://bubuko.com/infodetail-119970.html)

（1）算法简介

模拟退火算法来源于固体退火原理，将固体加温至充分高，再让其徐徐冷却，加温时，固体内部粒子随温升变为无序状，内能增大，而徐徐冷却时粒子渐趋有序，在每个温度都达到平衡态，最后在常温时达到基态，内能减为最小。需要注意的是一定要徐徐冷却，也就是物体和外界温度相差越低，退火时间越长退火的效果就越好。这也就是这算法名字的由来。

（2）基本步骤

1. 生成初始解T0
2. 生成新解Tn+1
3. 使用评价函数评价新解Tn+1，
4. 如果新解Tn+1更优秀则接受新解。
5. 否则以一定概率p接受Tn+1
6. 重复2-5步骤，直到解足够优秀。

（3）关键点

算法中有4个部分是对算法质量有显著影响的。初始解，产生新解的方法，接受差解的概率p，以及评价函数。

通常要求初始解比较差，也就是被加热的温度很高。如果初始解就比较好，很容易坠入局部最优解。

产生的新解与原来的解差距比较要合理，也就是要徐徐的降温，降温过快就会成淬火了。

接受差解的概率p，概率不能过高也不能过低，过低会导致很容易坠入局部最优解，过高会导致收敛速度大幅下降。

通常取p=exp(-Δt′/T)。

评价函数，显而易见的评价函数能否有效表示解的优秀程度直接关系解的质量。