### 整体流程

### Vae donut www18

### 

### 异常流量数据集

KDD99数据集：特征数据集

https://kdd.ics.uci.edu/databases/kddcup99/task.html

NSL-KDD数据：特征数据集和pcap

https://www.unb.ca/cic/datasets/nsl.html

CICIDS2017数据集：特征数据集和pcap

https://www.unb.ca/cic/datasets/ids-2017.html

USTC-TFC2016数据集：pcap

https://github.com/yungshenglu/USTC-TFC2016

### 异常流量检测

仅对训练集中标签为正常的正常网络流量进行训练，也只需要识别出样本是否为异常样本，并不关心不同异常类型样本的数据分布，将所有非正常的流量视作异常处理。

对正常类的数据进行训练，使其能够准确重建这些正常类的数据。

当遇到异常数据时，由于模型未见过这些数据，其重建误差会显著增大，从而识别出异常数据。

特征选择：

基于最基本的NetFlow的几个特征维度，不需要额外的计算



MTS

<https://dl.acm.org/doi/abs/10.1145/3611643.3613896>

### 异常流量分类

检测出异常流量后进一步分析，识别是哪种类型的异常流量

基于异常流量数据集的标签，有监督的训练异常分类模型。

特征选择：

基于Clickhouse的数据查询能力去统计数据，提供更多维度的特征



分类标签：

以KDDCPU99数据集的异常流量为例：

|  |  |
| --- | --- |
| 流量类型 | 介绍 |
| DDoS | 分布式拒绝服务攻击 |
| R2L | 来自远程机器的未经授权的访问 |
| U2R | 未经授权访问本地超级用户（root）权限 |
| probing | 监视和其他探测，例如端口扫描 |

以NSL-KDD数据集的异常流量为例：

|  |  |
| --- | --- |
| 流量类型 | 介绍 |
| DDoS | 分布式拒绝服务攻击 |
| Brute Force | 暴力破解攻击 |
| Heartbleed | 利用Heartbleed漏洞进行的攻击 |
| Botnet | 受控的僵尸网络流量 |
| SQL Injection | SQL注入攻击 |
| Brute Force | 针对Web登录页面的暴力破解攻击 |
| Infiltration | 内部网络的渗透攻击 |
| PortScan | 端口扫描攻击 |
| … | … |

以USTC-TFC2016数据集的异常流量为例

|  |  |
| --- | --- |
| 流量类型 | 介绍 |
| Cridex | 银行木马 |
| Geodo | 恶意程序 |
| Htbot | 恶意程序 |
| Miuref | 恶意程序 |
| Neris | 恶意程序 |
| Nsis-ay | 挖矿木马 |
| Shifu | 银行木马 |
| Tinba | 银行木马 |
| Virut | 恶意程序 |
| Zeus | 银行木马 |

### 后续需要解决的问题

**1.开源的异常流量数据集与NetFlow数据格式有较大差异**

需要设计预处理方法把开源的异常流量数据集与目前的NetFlow数据格式统一到一起

目前开源的异常流量数据集大致可以分为两种格式：

1. 人工特征数据集。由人工手动设计的流量特征组成，已经提前计算好。
2. 原始流量数据集。采集自原始流量，格式为PCAP

**2.异常检测和异常分类的算法设计**