**基于机器学习的DDoS入侵检测算法**

选题原因：

在一个网络安全公众号上看到一条新闻，讲的是一个黑客组织经常攻击小型互联网公司新发布的产品，然后勒索大量赎金，使用的方法就是DDoS攻击。听到这个消息很震惊也很气愤，就像试试自己能不能为这个问题尽一份力。

最近，乌克兰国防部等网站遭大规模DDoS攻击而关闭。不久后，俄罗斯也遭到了黑客组织“匿名者”的DDoS攻击

基础知识：

攻击类型：

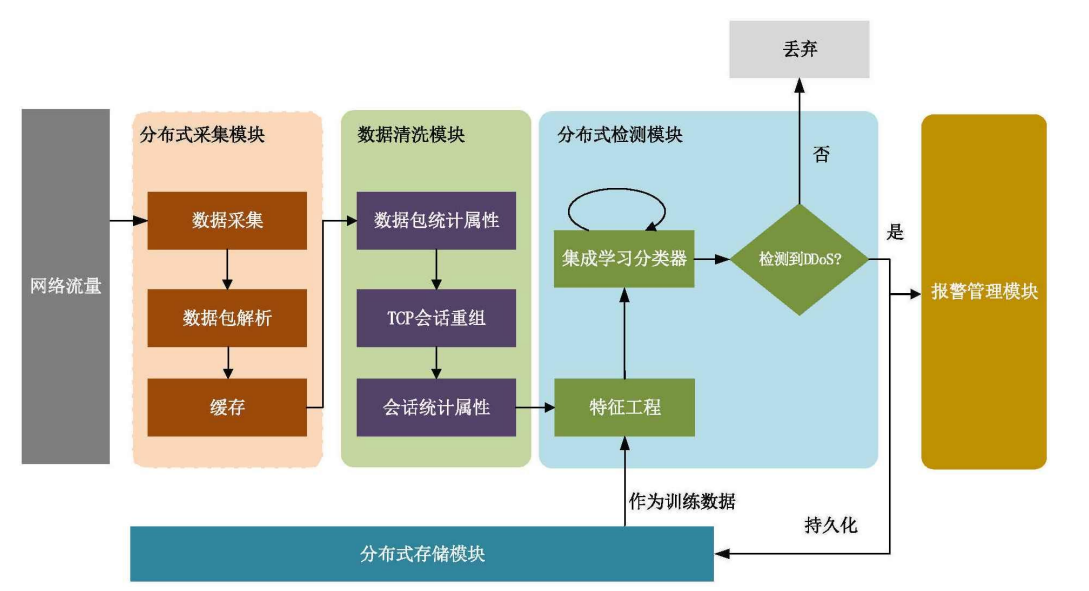
SYN-Flood 攻击：利用TCP三次握手的缺陷，在完成第二次握手后，不再进行下一次握手，使服务端处于忙等的状态，并占满缓存区。

Smurf ping 放大攻击、DNS放大攻击：发送源地址为被害主机的请求，中间节点收到后，会向被害主机进行相应，从而沾满带宽。

畸形报文攻击：攻击方向被害主机发送畸形的报文，如源地址与目的地址一致或分片重叠等，导致服务器不知如何处理而导致崩溃。

Slowloris慢速攻击：缓慢的发送多个请求头信息，达到最大连接数

检测系统主要分为四个模块类似于下图。首先是流量捕获模块，使用scapy库获取每个发送到服务器指定端口的数据包P，并进行缓存。然后是特征提取模块，提取出数据包P的特征属性，如源IP地址，目的IP地址等。将提取的基础属性数据进一步的计算，从中构建出更多维度的特征属性，构成此数据包的特征向量x。第三是攻击检测模块，这也是整个算法最核心的地方。将特征提取模块得到的特征向量送入假设函数h(x)【注】中，如果结果为1则判定改流量为攻击流量，不再交付给上层，并将此向量存储到数据训练集中。反之则正常交付，并存储到数据训练集中。



【注】

|  |  |
| --- | --- |
| 逻辑函数 |  |
| 假设函数 |  |
| 读入训练集，特征映射 | |
| 代价函数 |  |
| 计算梯度(j = 0单独计算) |  |
| 梯度下降 | 选用工具库求解，不再需要选择alpha和迭代  scipy.optimize.fmin\_tnc() |
| 得出使代价函数最小的theta，带入假设函数 | |