**项目目的与意义**

Android系统以其开放性和便捷性赢得了广大用户的青睐。根据IDC发布的《全球手机市场跟踪调查季度报告》，早在2013年Android的智能手机出货量占市场份额已达到81%，同比增长51.3%。但是同传统的PC系统相比，Android系统平台无论是安全软件成熟度，还是用户安全意识方面，都不及PC系统。而随着Android智能移动终端功能上的进一步丰富，使得其中包括了很多个人信息甚至是隐私信息。这些都成为了被恶意攻击获取的目标。

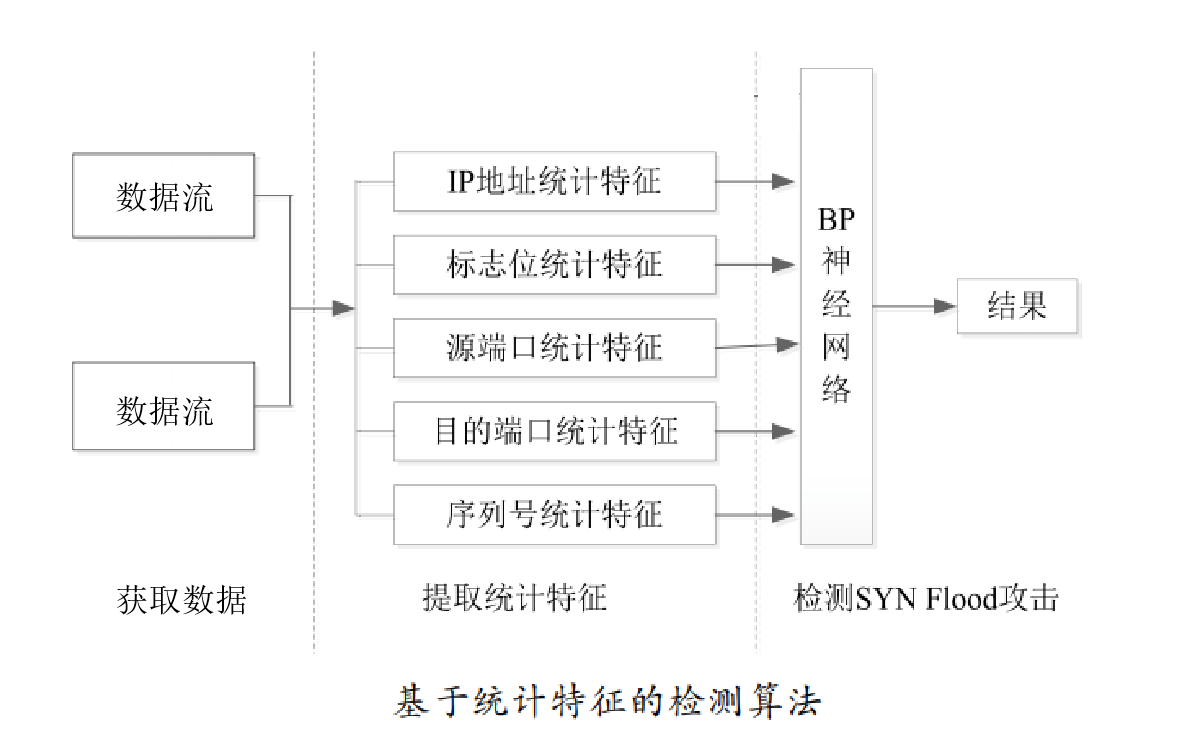
目前较为完善的Android端安全工具基本上都是针对于Android恶意应用及病毒木马的检测。这些应用都是基于对Android应用文件APK文件进行分析，判断某个APK应用文件是否为恶意应用，但是对于手机设备本身以及手机系统是否含有安全漏洞，并没有较好的解决方案，目前没有一款较为全面的Android系统漏洞检测工具。

本报告主要是提出了一个基于机器学习和统计数据匹配特征方法的针对于Android系统的入侵检测平台。这个平台主要是通过Android系统的状态参数进行实时监测，将大量的数据通过机器学习以及统计学规律的方式处理，总结在不同的攻击下Android系统呈现出的特征。利用这些特征，在面对诸如wifi，蓝牙以及其他物理接口的攻击面下的入侵可以及时的提出警告。本平台主要分为三个模块：攻击模块、监测模块、显示模块。

**项目成果主要内容**

**1.算法设计**

基于统计特征的SYN Flood检测方法。



检测算法

SYN Flood（半开放攻击）是一种拒绝服务(DoS)与分布式拒绝服务(DDoS)攻击方式，其目的是通过消耗所有可用的服务器资源使服务器不可用于合法流量。通过重复发送初始连接请求（SYN）数据包，攻击者能够压倒目标服务器机器上的所有可用端口，导致目标设备根本不响应合法流量。那么SYN Flood攻击如何工作呢？

通过利用TCP连接的握手过程来进行SYN Flood攻击工作。在正常情况下，TCP连接显示三个不同的进程以进行连接。

（1）首先，客户端向服务器发送SYN数据包，以便启动连接。

（2）服务器响应该初始包与SYN / ACK包，以确认通信。

（3）最后，客户端返回ACK数据包以确认从服务器接收到的数据包。

完成这个数据包发送和接收序列后，TCP连接打开并能发送和接收数据。为了创建拒绝服务，攻击者利用这样的漏洞，即在接收到初始SYN数据包之后，服务器将用一个或多个SYN / ACK数据包进行响应，并等待握手中的最后一步。所以攻击者向目标服务器发送大量SYN数据包，通常会使用欺骗性的IP地址。然后，服务器响应每个连接请求，并留下开放端口准备好接收响应。当服务器等待从未到达的最终ACK数据包时，攻击者继续发送更多的SYN数据包。每个新的SYN数据包的到达导致服务器暂时维持新的开放端口连接一段时间，一旦所有可用端口被使用，服务器就无法正常工作。

本文从构造一个SYN攻击报文的角度分析，SYN Flood攻击会引起网络中基于IP地址、标志位、端口号、序列号的统计特征异常。据此，提出一种基于统计特征的SYN Flood攻击检测的方法。该方法分别提取基于IP地址、标志位、端口号、序列号的统计特征，最后使用BP神经网络算法得到检测结果。

1.统计特征提取：

（1）标志位

在TCP连接正常建立和终止过程中，网络中这些标志位的报文数量存在一定对称性。然而，当网络中存在SYNFlood攻击时，这种数量上的对称性将遭到破坏。

为了有效检测网络中SYN报文与SYN/ACK报文的对称性是否遭到破坏，采用下式反映对称的变化。



（2）IP地址

当SYN攻击报文中IP地址为随机值时，这时IP地址的统计特征呈现发散趋势。

为了检测IP地址是否具有发散或聚集特征，首先从数据流中分别提取IP集合，并统计每个IP出现的次数。之后采用熵的方式反映半连接IP地址的分布情况，采用下式反映。



其中，n表示此次统计的流量包数量，



（3）端口号

当构造SYN攻击报文时，源端口号/目的端口号为固定值时，使得源端口号/目的端口号统计特征呈现聚集趋势；当构造SYN攻击报文时，源端口号/目的端口号为随机值时，使得源端口号/目的端口号统计特征呈现发散趋势。

为了检测源端口号、目的端口号的统计特征，首先从数据流中分别提取源端口号集合、目的端口号集合，并统计每个元素出现的次数。之后为了检查其是否具有聚集或发散趋势，同样使用熵反映其统计特征，见下式。





其中，n表示此次统计的流量包数量，





（4）序列号

与端口号相似，当构造SYN攻击报文时，报文的序列号也存在固定值和随机值，也会导致序列号的统计特征也呈现聚集或发散趋势。

为了检测序列号的统计特征，首先从数据流中分别提取序列号集合，并统计每个元素出现的次数。之后为了检查其是否具有聚集或发散趋势，同样使用熵反映其统计特征，见下式。

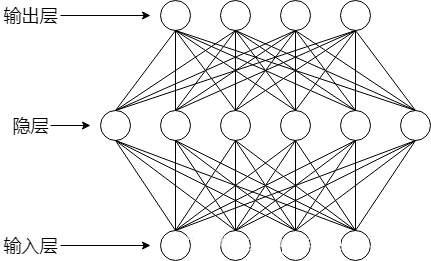


其中，n表示此次统计的流量包数量，



**2.神经网络**

对于多变量非线性系统，我们可以采用神经网络实现回归分类：



神经网络图

本次仅针对是否被洪范攻击进行学习，因此输出层可以只使用一个神经元实现二分类。

sigmoid函数可以将输出映射至0-1之间，我们只需要对数据集进行标定（1标识被攻击，0表示没有被攻击）。这样单神经元的输出即可实现二分类，接近于1的输出即为被攻击，接近于0的输出即为不被攻击。

经过多次调整学习率，与隐藏层神经元个数进行反向传递训练，最终可以得到接近100%的预测命中率。

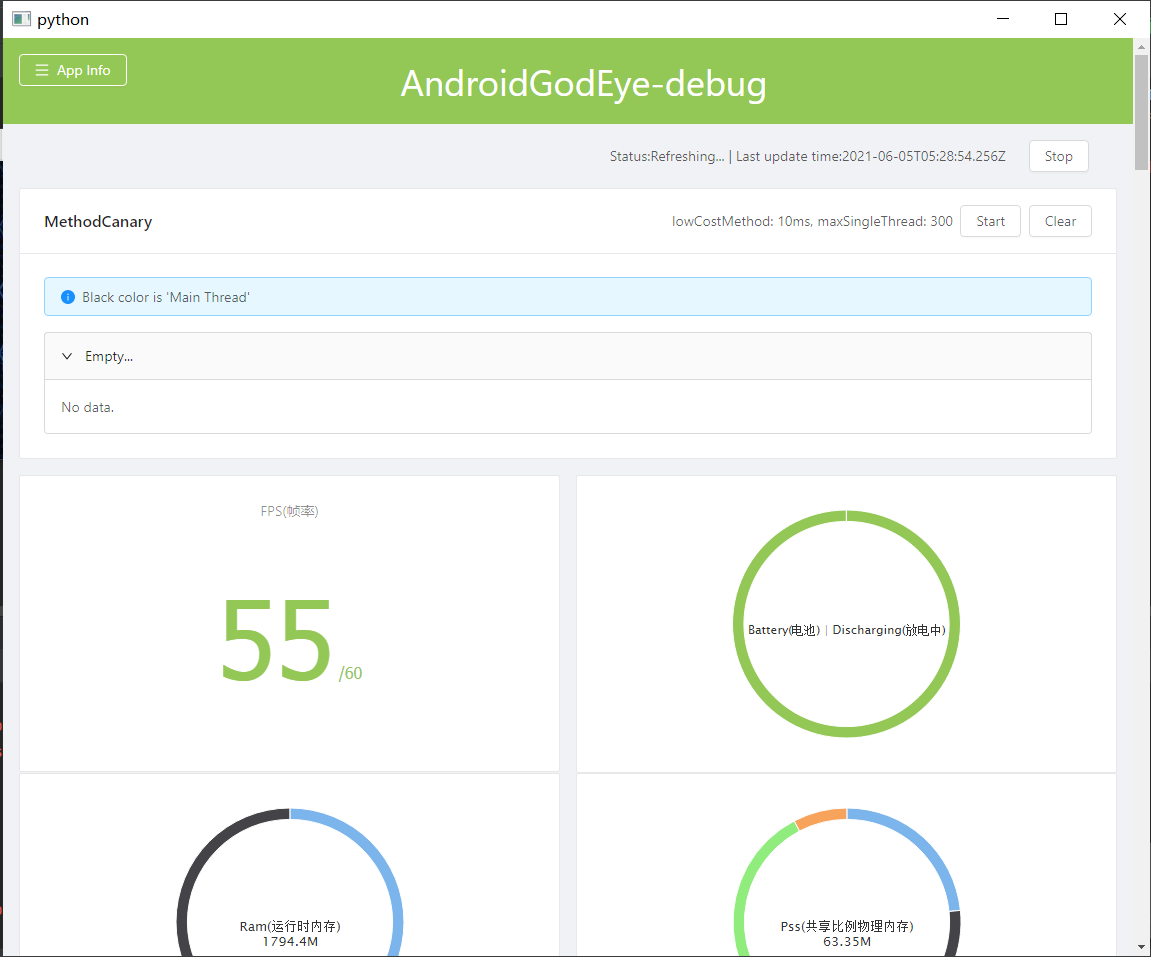
**3.实现情况**

运行代码，将会进入如下图所示的系统界面。系统由检测和统计两部分组成。



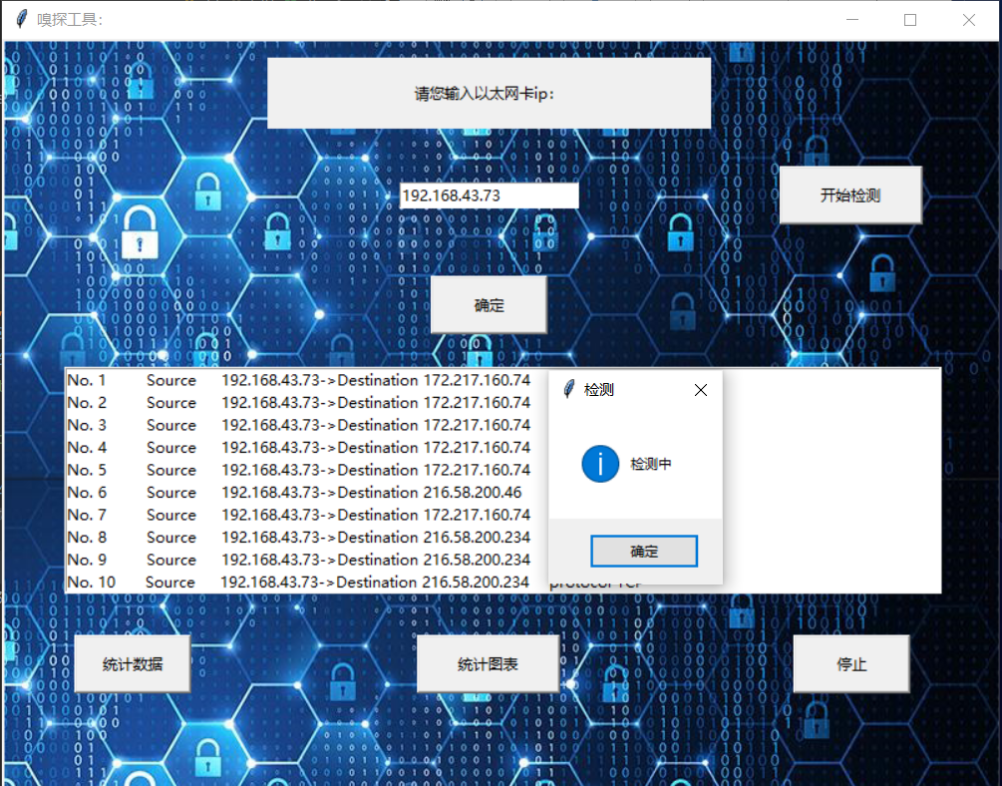
入侵检测系统

进入统计模块，可以监测手机当前的情况，包括电量，内存使用情况等。

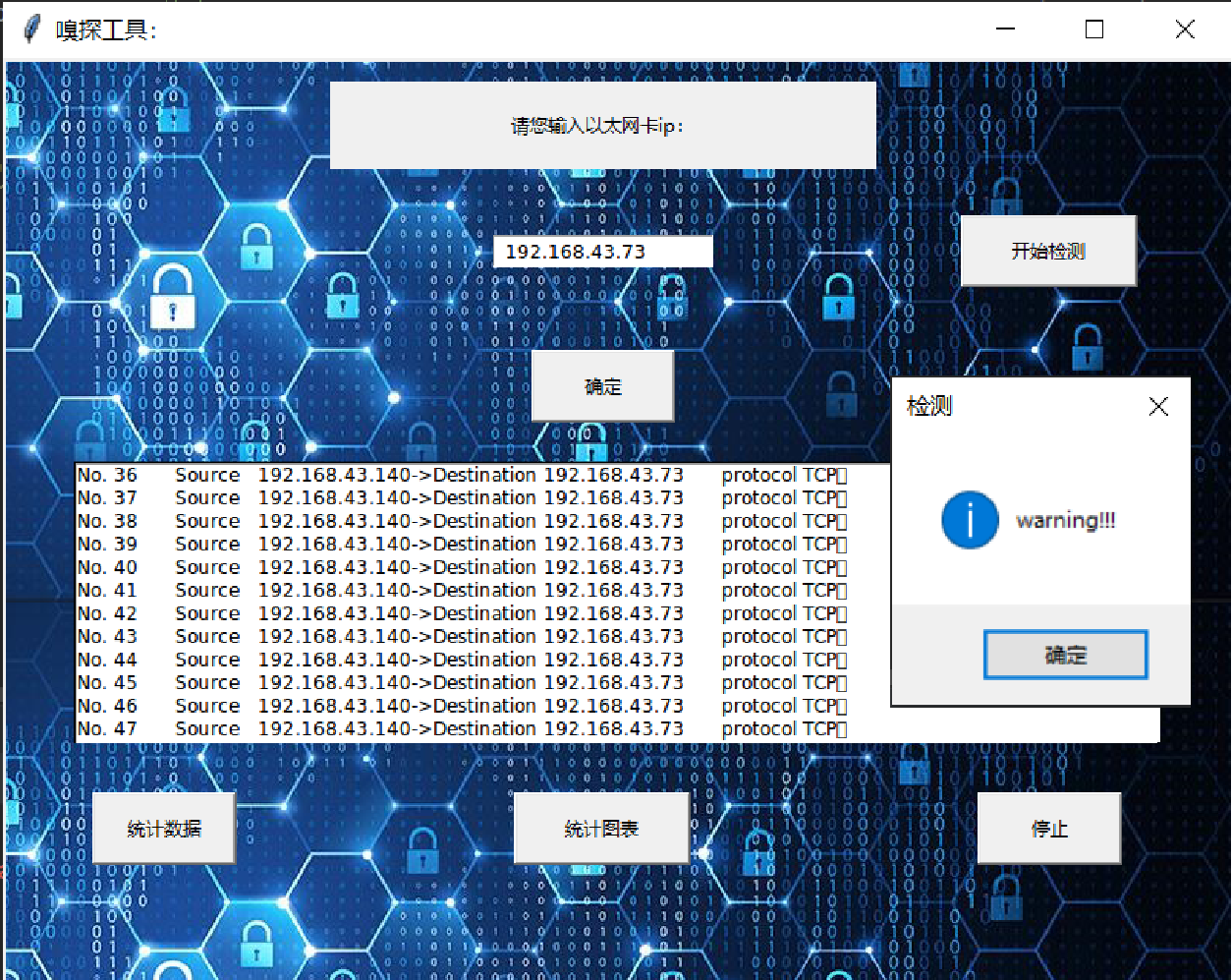


统计模块：浏览器监控结果

进入检测模块，将会对指定IP地址的设备进行洪泛攻击检测。下方的ip地址为安卓手机的ip地址，在检测到攻击行为时，会弹出警告弹窗，如下图所示。

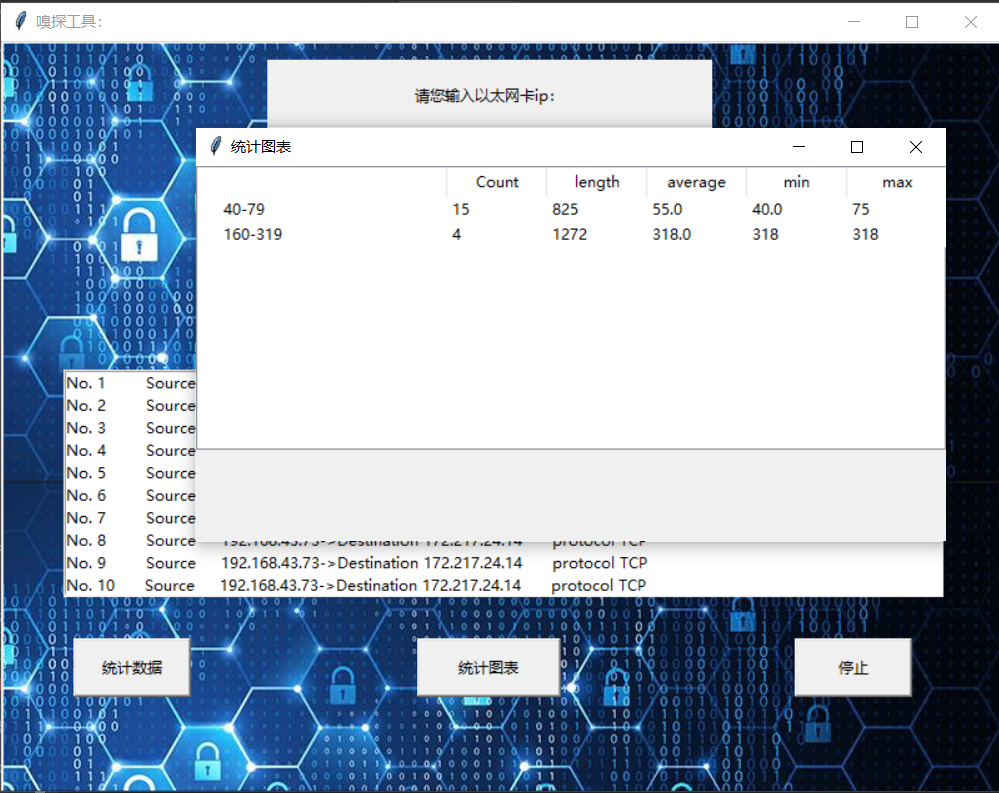


检测模块：入侵检测中

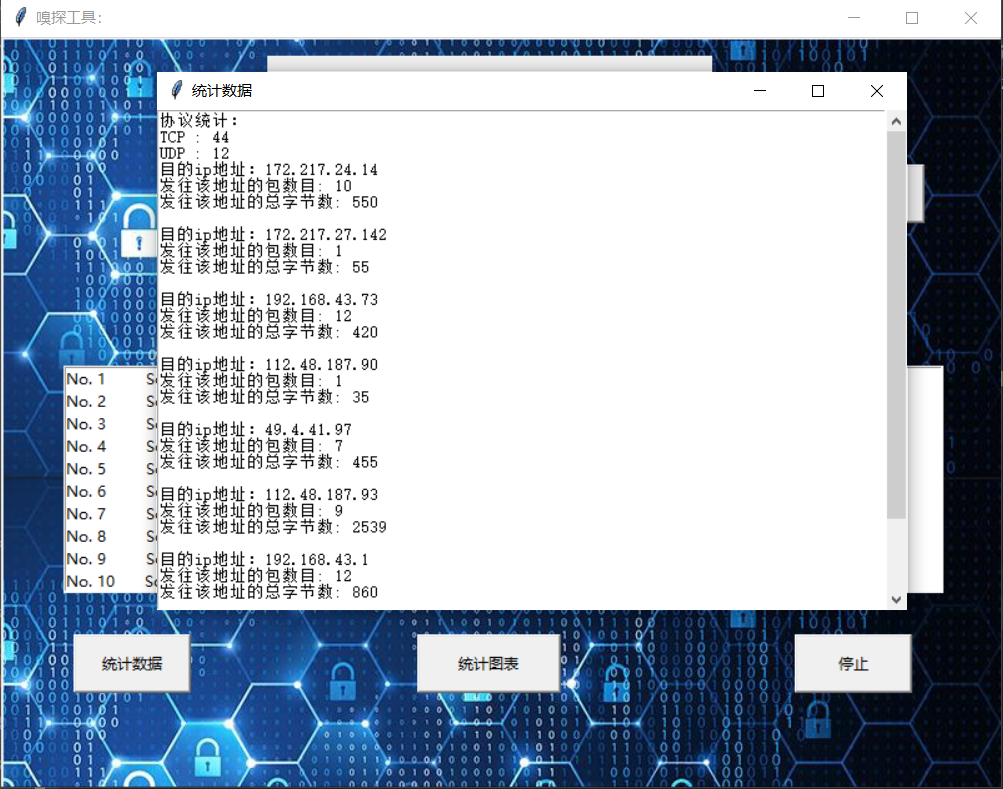


检测模块：入侵检测抓包和报警

检测模块还提供了查看统计数据和统计图表的功能。



检测模块：统计图表



检测模块：统计数据

**5 总结**

项目组实现了一个基于机器学习和统计数据匹配特征方法的针对于Android系统的入侵检测平台。这个平台主要是通过Android系统的状态参数进行实时监测，将大量的数据通过机器学习以及统计学规律的方式处理，总结在不同的攻击下Android系统呈现出的特征。

**创新特色与实践意义**

实践意义在于利用这些特征，在面对诸如wifi，蓝牙以及其他物理接口的攻击面下的入侵可以及时的提出警告。主要创新特色为应用创新，采用神经网络的方法，实现洪泛攻击的报警系统。

**财务执行情况与成员分工合作情况**

项目经费用于PCB板制作、电子元器件和zed双目摄像头的采购，已完成报销。

团队成员分工分别为算法研究、嵌入式系统设计和系统架构的研究，能共同较好合作完成该项目，实现项目预期成果。

**研究过程安排表**

阶段一（2019.1-2019.6）：项目应用与算法调研

阶段二（2019.7-2019.9）：数据集准备

阶段三（2019.10-2020.3）：搭建实验环境，学习pyQt，scrapy框架相关技术

阶段四（2020.4-2020.9）：实现网络数据时时抓取，搭建神经网络模型

阶段五（2020.9-2020.12）：训练模型，实验测试，项目总结