

系统设计

本章主要介绍设计人机交互系统，网络流量可视化系统和后台管理系统的前期准备，在需求分析中将介绍项目背景和设计目标；本章还将介绍设计原则和技术要求；在系统架构一节中我们将介绍系统架构以及数据库设计。在之后的几章中将分开逐步实现该系统的具体功能。

需求分析

项目背景

对于人机交互系统，通过分成不同模块实现多种交互功能，充分提供便利友好的交互体验是关键。对于数据可视化系统，结合具体的平台环境，有针对性地对网络流量数据进行全面的数据可视化模拟与分析具有十分重要的意义。现有的人机交互系统和网络数据可视化技术发展成果中很少有一个较为综合的平台架构同时为人机交互和数据可视化提供更多的施展空间。本文提出的基于树莓派的人机交互和可视化系统，提出了一个软硬件深度结合，兼具视觉效果与高性能操控感的多功能系统平台。系统结合了人机交互功能，数据可视化功能、后台管理功能以及各种附加功能，集成了人机交互体验，数据集研究、真实数据反馈、跨平台数据交互等需求。

设计目标

基于嵌入式设备树莓派的人机交互及可视化系统处理的数据信息一方面是来自用户，通过获取用户信息转化成为操作命令，用更直观有效的方式进行数据的双向传递。第二个方面是来自运行在树莓派上的后台管理系统，也就是本文所配套的后台管理平台，对于真实的流量数据和资源数据有更深层次的把握，可以在实际的生活工作环境中了解数据资源以及网络流量数据的方方面面，也可以让网络安全人员更准确地了解网络攻击和网络威胁，为网络安全提供便利。

本系统的目标是在充分结合人机交互系统和网络流量数据可视化系统的基础上，为了满足管理便捷，数据交互安全可靠，多平台全面涵盖的要求，搭建合

理有效的多功能系统，抽取有效数据，记录和处理相关数据，有效监控，满足日常管理等。

设计原则与技术要求

设计原则

为充分发挥树莓派人机交互及可视化系统的整体效能，整个平台设计及后期的工程建设遵循以下原则：

- （1） 多平台访问系统。包括设计系统的嵌入式设备本地功能模块、移动端页面和电脑端页面，移动端包括数据可视化微信小程序，电脑端包括涉及具体功能的网站系统。基于树莓派的人机交互，运行需稳定，操作需流畅。
- （2） 数据可靠。系统内的数据交互需要实现数据统一和数据安全，用于交互的数据是各自所需要的数据。
- （3） 稳定性，可靠性高。
- （4） 易于维护和二次开发。

技术要求

嵌入式设备树莓派人机交互及可视化系统融合所有设想的系统框架，增强系统的更迭力度，整体化管理，优化细节。

选用不同的技术框架完成不同的子系统功能，针对不同的功能点进行各自的框架流程分析，实现技术方面的多元化。

系统设计

硬件架构

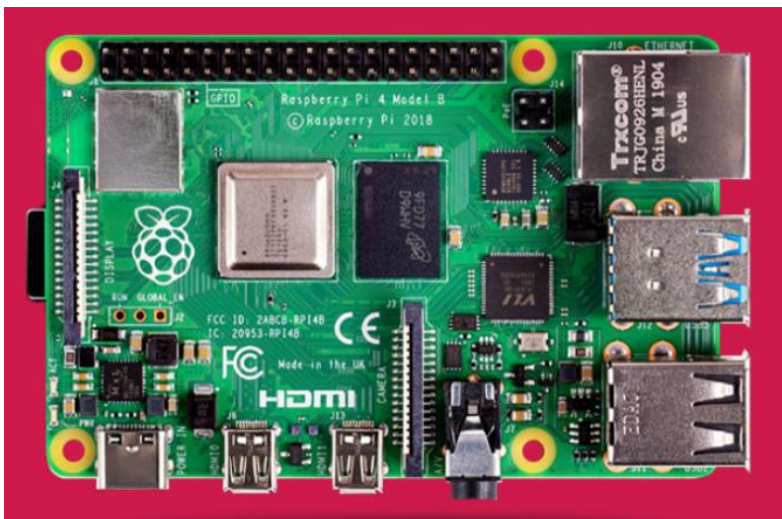


图 3-1 树莓派 4B



图 3-2 设备示意图

硬件配置：如图 3-1 所示，树莓派 4b 主板一块，15G 的 TF 卡一张，5v/3A 充电头一个，type 数据线一根，TF 读卡器一个，一块树莓派专用 3.5 寸触摸屏，USB 音箱一对，USB 录音器一只，CSI 摄像头一只。图 3-2 是设备的整体状态。

系统架构

关于嵌入式设备人机交互及可视化系统的架构，我们将在之后章节中介绍其各个部分的内部实现。首先我们将设计基于嵌入式设备树莓派的人机交互功能，接受用户操作信息并执行相应功能；其次我们提出人机交互及可视化系统的移动端扩展——Web 系统部分，扩展了人机交互的承载形态；同时我们提供了微信小程序和入口门户的后台服务系统；最后我们将介绍嵌入式设备树莓派人机交互及可视化系统的可视化管理部分，用于管理资源数据并提供可视化服务。因此，我们设计出本系统的系统架构框架以满足所有模块功能和流程的构思，如图 3-3 所示。

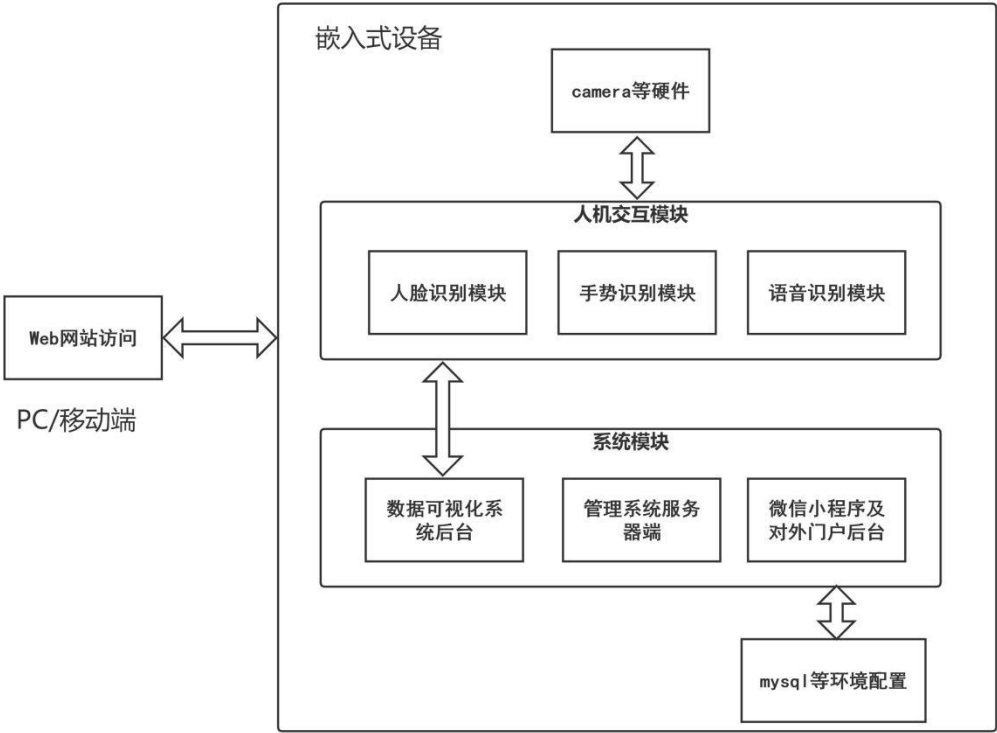
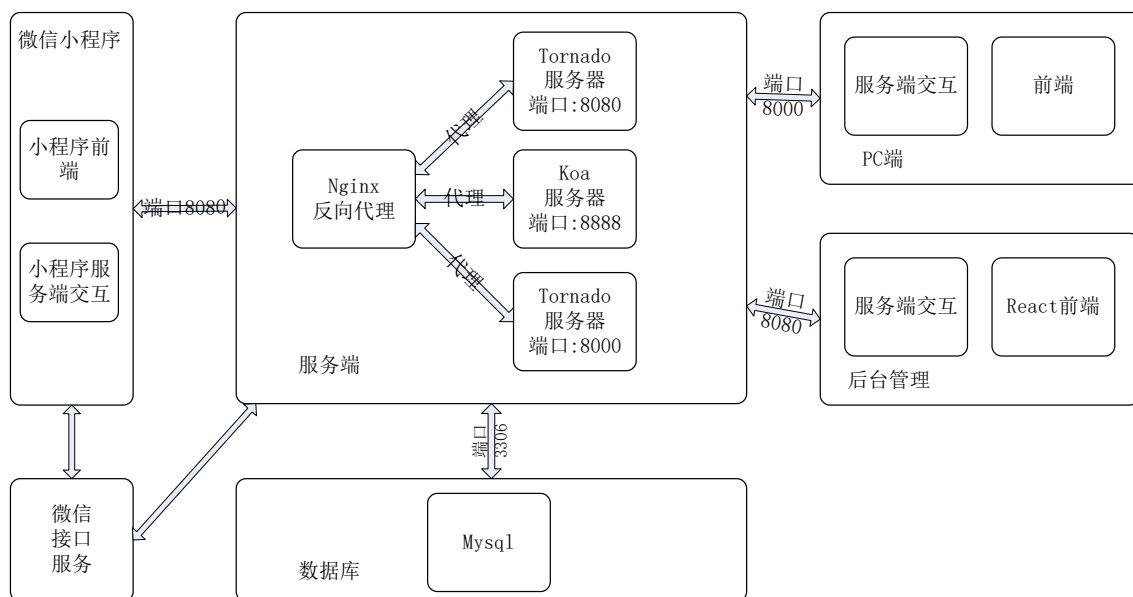


图 3-3 嵌入式设备人机交互及可视化系统架构图

树莓派人机交互及可视化系统框架可以将本文所介绍的功能进行优化整合，既保证了合理性，也保证了功能的完备性。用户交互端可以通过 Web 端界面与嵌入式设备进行交互。调用摄像头或者语音模块实现树莓派和用户的数据交互。PC/Phone 端数据可视化功能可以提供数据输入与可视化输出；管理系统功能可以提供全平台的资源管理，以实现全方位的交互与数据研究；微信小程序作为新生的更为方便快捷的数据传播媒介为人机交互和数据可视化系统提供多平台可视化的可能性。系统的 Web 部分架构图如下所示。



数据库设计

(1) PC 端数据库表设计见图 3-4

PC 端数据可视化部分由于上传文件比较大，如果使用数据库存储耗时会比较大，所以只设计了登录的用户表。用户属性主要有用户名，登陆密码，注册邮箱以及 id（唯一标识）。

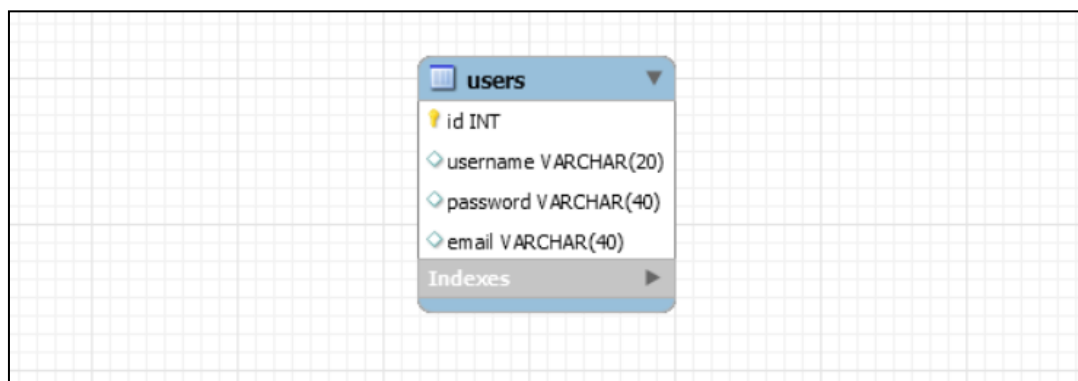


图 3-4 PC 端数据库用户表

(2) 管理系统部分数据库表设计见图 3-5

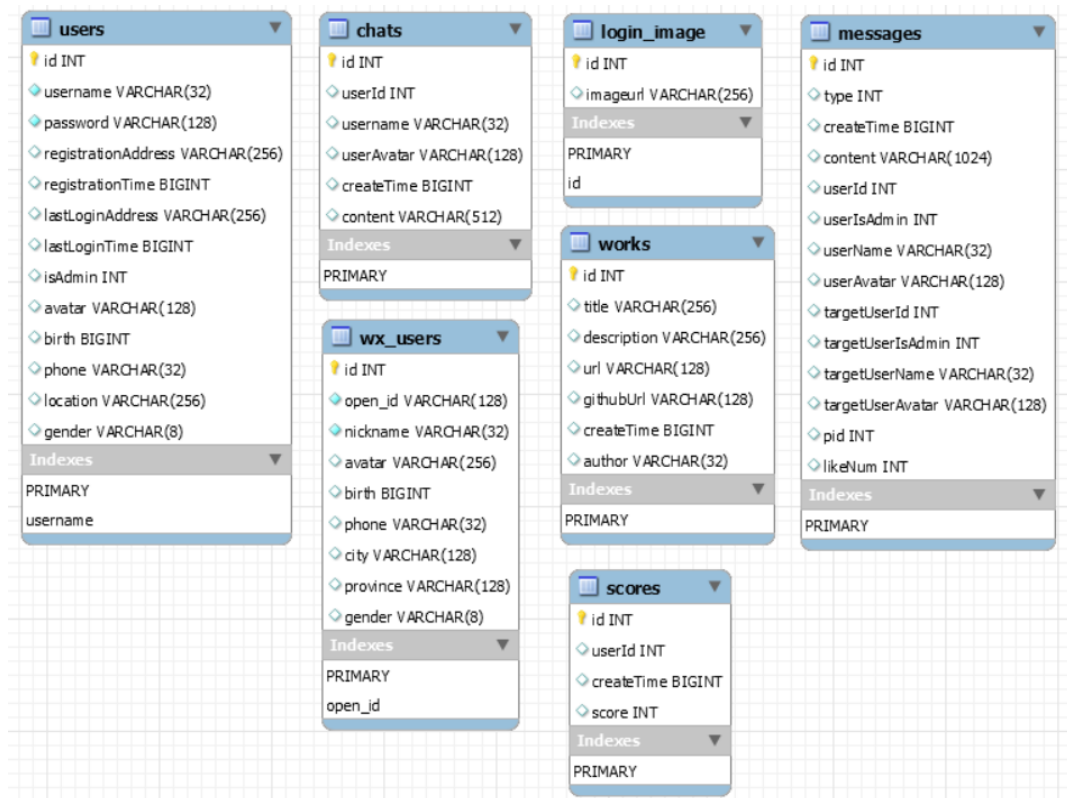


图 3-5 后台管理数据表

还有一些其他的数据库表涉及管理系统中与数据可视化不相关的功能，在此不作介绍。

本章小结

本章主要介绍了网络流量可视化系统需求分析中的项目背景和设计目标，介绍了网络流量可视化系统的设计原则和技术要求，还介绍了本系统的系统架构以及数据库设计。并说明在之后的几章中将逐步实现该系统的具体功能。

系统实现

本章将介绍嵌入式设备树莓派人机交互及可视化系统的具体实现，其中包括人机交互部分，可视化管理部分等。本章将用几个小节分别对每个部分包括：人机交互部分、可视化管理部分进行介绍。

人机交互

技术选用

本文在实现人机交互功能时，主要编程语言选择了在人工智能领域比较火的 Python。针对人脸识别和手势识别使用的机器视觉库是 OpenCV；针对语音识别使用的是 wukong-robot 开源框架，调用了百度语音 API，使用了开源轻量级语音唤醒引擎 snowboy。

(1) OpenCV

OpenCV 是一个基于 Apache2.0 许可（开源）发行的跨平台计算机视觉和机器学习软件库，可以运行在 Linux、Windows、Android 和 Mac OS 操作系统上。OpenCV 轻量级而且高效——由一系列 C 函数和少量 C++ 类构成，同时提供了 Python、Ruby、MATLAB 等语言的接口，实现了图像处理和计算机视觉方面的很多通用算法。

(2) Wukong-robot

wukong-robot 是一个简单、灵活、优雅的中文语音对话机器人/智能音箱项目。wukong-robot 有模块化的特点。功能插件、语音识别、语音合成、对话机器人都做到了高度模块化，第三方插件单独维护，方便继承和开发自己的插件。

(3) Snowboy

Snowboy 是 KITT.AI 开发的人工智能软件工具包。通过 Snowboy 软件，可以在一些硬件设备上添加“语音热词探测”功能。

功能流程与功能模块

人机交互部分主要基础是机器视觉和语音识别的相关技术，核心是机器学习，开发目的主要是为了提供基于树莓派的人机互动体系。

以下介绍人机交互部分的具体功能流程：用户可以直接登入树莓派或者使用

VNC 进入系统，系统进行初始化，在此之后用户可以根据需要查看相关功能模块，并进行运行或者调用。具体的功能流程如图 4-1

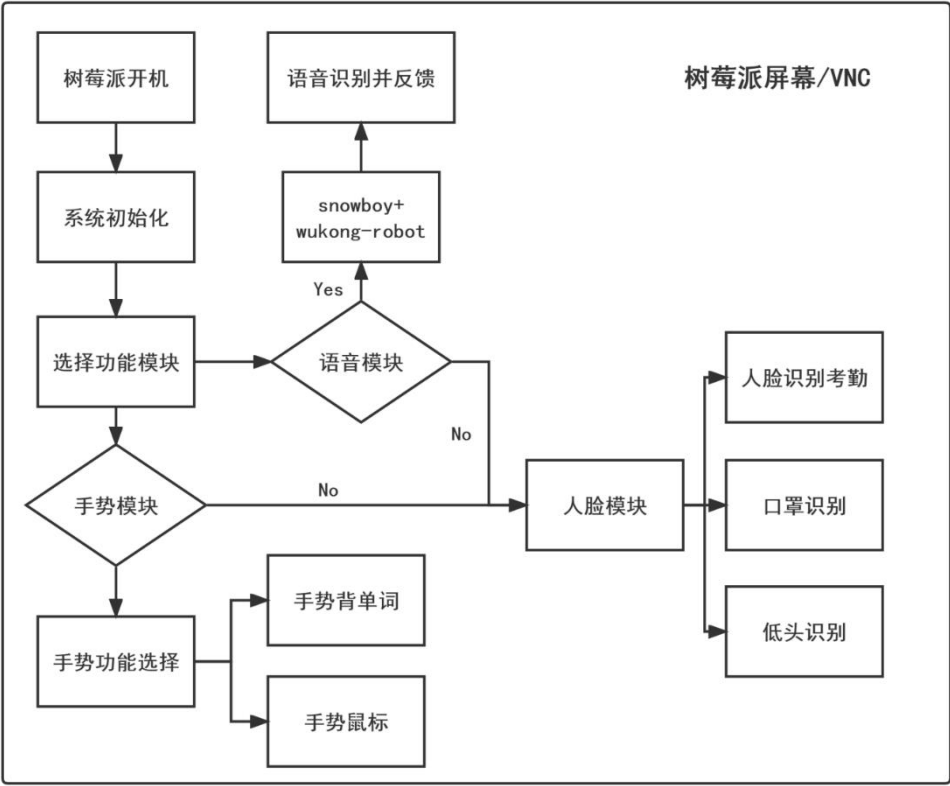


图 4-1 功能流程图

人机交互部分包括以下三个模块：语音识别，人脸识别和手势识别。以下图 4-2 给出了功能模块图。

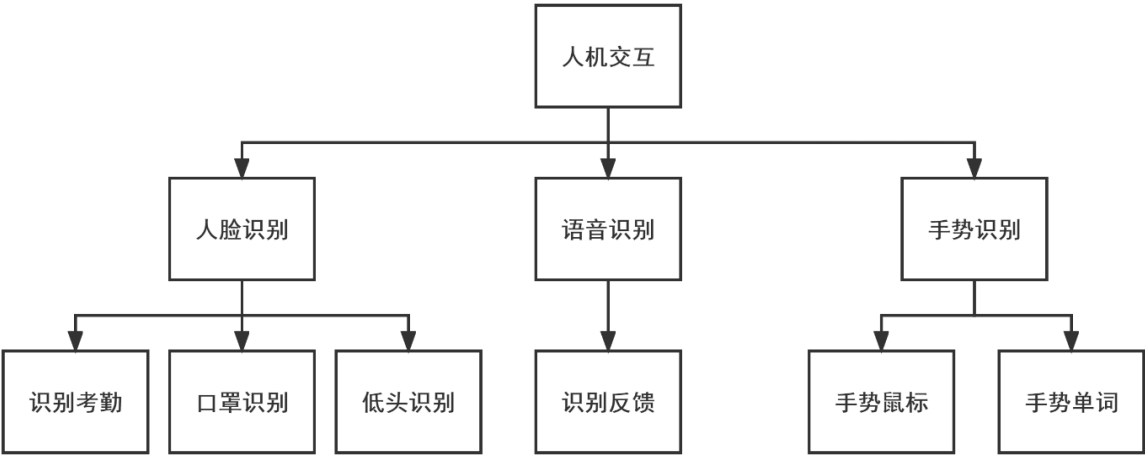


图 4-2 功能模块图

核心框架和解释

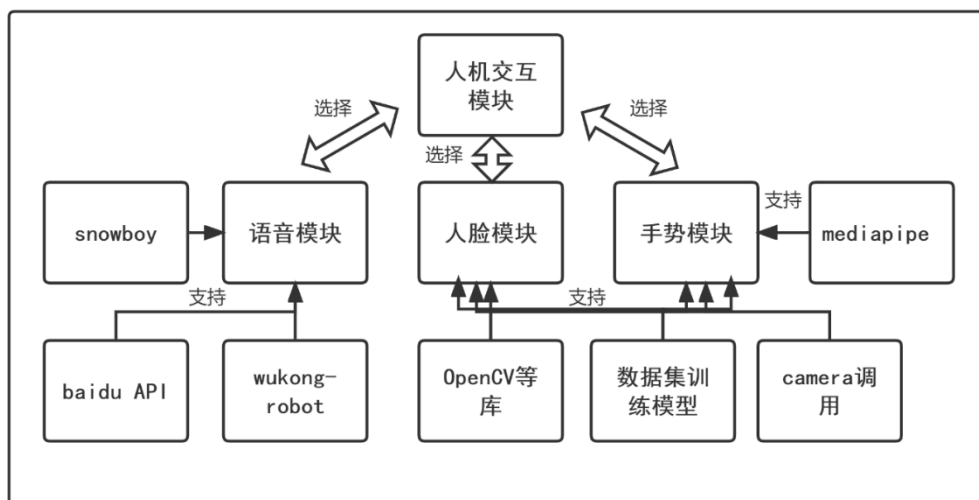


图 4-3 人机交互系统

在人机交互系统中，主要调用了硬件中的摄像头，录音器和音箱。使用基于 Python3.7 的 OpenCV 作为机器视觉库，调用了其中的很多功能函数，并生成自己的数据集训练合适的模型。在实现手势识别的过程中使用了 mediapipe。在实现语音识别模块时使用了 snowboy, wukong-robot 这些开源框架，并调用了百度的语音 APIs。

MediaPipe 是一个基于图形的跨平台框架，用于构建多模式（视频，音频和传感器）应用的机器学习管道。MediaPipe 可在移动设备、工作站和服务端上跨平台运行，并支持移动 GPU 加速。使用 MediaPipe，可以将应用的机器学习管道构建为模块化组件的图形。

数据可视化平台

技术选用

本文在实现数据可视化功能时使用的可视化基础是 D3.js，服务器端选用 Python 的 Tornado 框架，数据库选用 MySQL8.0。

（1）D3.js

D3.js 是一个可以通过数据来操作文档的 JavaScript 库。D3.js 可以通过使用 HTML、SVG 和 CSS 实现数据可视化。D3.js 严格遵循 Web 标准，使其能够方便地与现代主流浏览器兼容，避免对某些框架的依赖。项目中需要一些特殊的图表，如 IP、数据库拓扑等，它可以帮助快速实现这些图表效果。

D3.js 测试了当下主流的各个浏览器，其大部分组件可以在旧的浏览器运行。

(2) Tornado

Tornado 是非阻塞式的轻量级的 Web 服务器，使用 Python 编写。其非阻塞的方式和对 `epoll` 的运用使其速度比较快。Tornado 可以解决高并发，每秒可以处理数以千计的连接，超过了很多主流的 Web 框架，因此 Tornado 对于实时 Web 服务来说是一个较为理想的 Web 框架。

(3) MySQL8.0

MySQL 是一个常见的在 Web 应用方面很好用的关系型数据库，在 MySQL 的各个版本中，MySQL 8 更快。MySQL8.0 版本相比之前版本的一些特性如：默认编码 UTF-8，降序索引，设置持久化等。

功能流程与功能模块

数据可视化部分主要基础是 Web 技术，核心是数据可视化，开发目的主要是为了方便研究人员直观清晰地研究某些网络数据集。

以下介绍数据可视化部分的具体功能流程：用户选择注册及登录进入系统，系统分为系统首页与实验案例两部分，在系统首页用户可以根据需要查看基于某些数据集的网络攻击数据展示，用户也可以根据自己的需要上传自己的数据进行可视化操作，具体的功能流程如图 4-4。

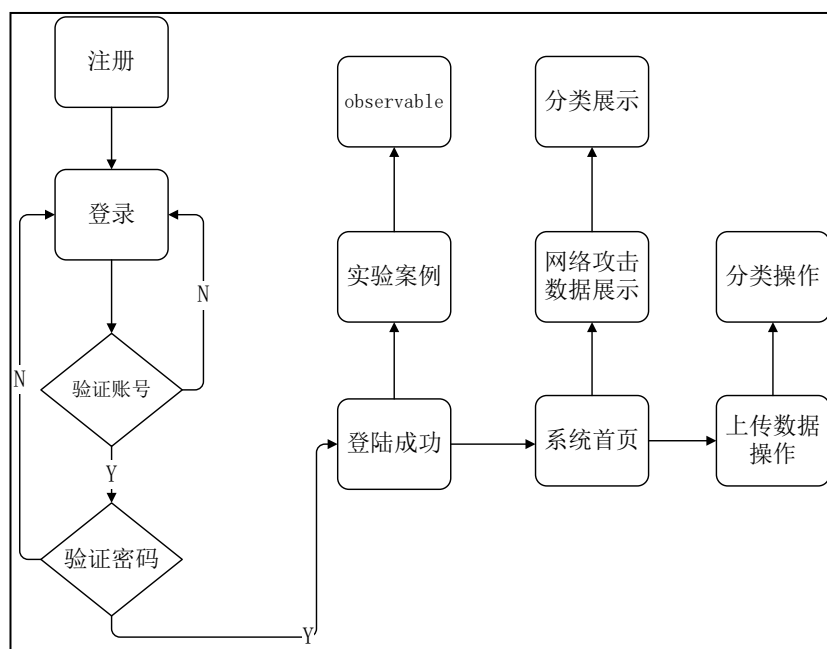


图 4-4 功能流程图

数据可视化部分包括以下两个模块：网络攻击数据展示和上传数据处理操作，以下图 4-5 给出了功能模块图。

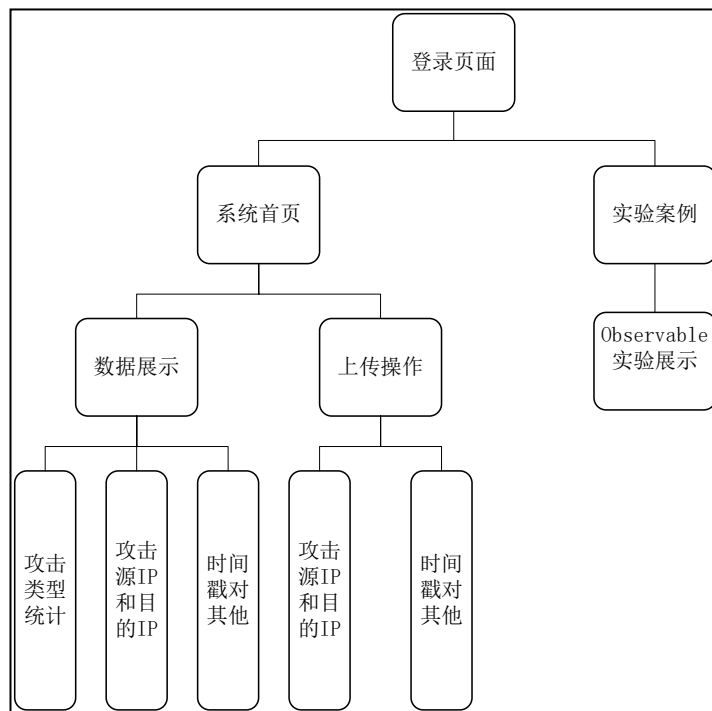


图 4-5 功能模块图

核心框架和解释

数据可视化平台架构如图 4-6 所示。

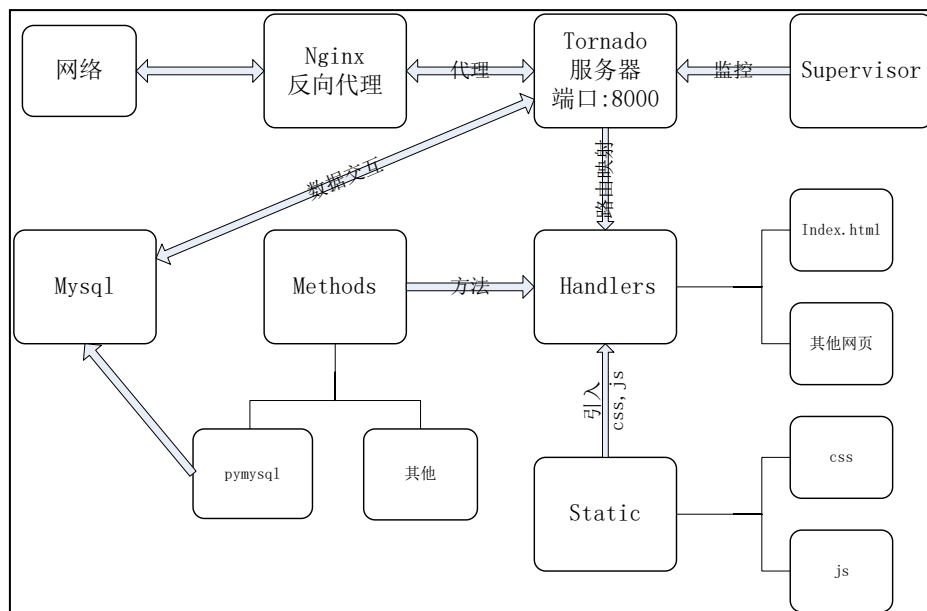


图 4-6 数据可视化平台

在本系统中，使用 Python 的 Tornado 作为 Web 服务器兼 Web 应用框架，使用 Nginx 分发 Web 请求并负载均衡，并用 Supervisor 进行监控，Handlers 中是系统中的所有页面，Methods 中使用 Python 中的 pymysql 配置并使用 MySQL。

Nginx 是一款开源的、高性能的、基于 REST 架构风格的 HTTP 服务器和反向代理服务器。代理服务器可以中转分发客户端请求到规定的服务器。

Supervisor 是一套通用的进程管理程序，是使用 Python 开发的，它可以将正常的命令行进程更改为后台守护进程，监视进程状态，如果异常则自动重新启动。Supervisor 是通过 fork/exec 的方式把被管理的进程当作其子进程来启动，这样只要把被管理进程的可执行文件路径写进 Supervisor 的配置文件中即可。

用于连接 MySQL 服务器的库在 Python3.x 版本中是 pymysql，而 Python2 中则是使用 mysqldb。

微信小程序

本章前一节介绍了数据集数据可视化平台，而单一功能的数据可视化系统满足不了现在的网络流量数据分析需求，因此本文提出了基于微信小程序的业务拓展平台，以满足网络流量数据可视化研究过程中的多元化需求。

技术选用

本系统使用 Tornado 为小程序端提供数据交互和路由，使用微信小程序扩展系统功能，使用 MySQL8 作为数据库支持。分发 Web 请求和负载均衡使用的是 Nginx。

（1）微信小程序开发介绍

微信小程序是一种新的开放能力，使用者无需下载安装，开发者也可以根据开发文档配合小程序开发工具快速地开发一个可以广泛传播的小程序。开发体验与用户体验都很出色。

（2）微信小程序目录介绍

一个微信小程序项目的目录结构如下：app.json（配置文件），app.js（逻辑文件），app.wxss（样式文件）。小程序各个页面的目录放置在文件夹 pages 中，pages 中每一个目录代表一个页面。各个页面文件夹中有四个文件：js（逻辑文件），json（配置文件），xml（结构文件），wxss（样式文件）。

（3）微信小程序开发流程

注册微信小程序账号；获取 APPID；绑定开发者；创建微信小程序项目；编写微信小程序代码（包括脚本文件，配置文件，样式表文件等）；手机预览；上

传小程序代码；提交微信审核及小程序发布。

功能流程与功能模块

本系统中微信小程序的主要功能流程如下：用户选择微信登录进入微信小程序，系统展示大部分功能，其中关键的功能是 Echarts 和网站导航。在系统的各个功能模块，用户可以根据需要自行更改微信小程序中所需的素材，或是查看并下载可利用 PC 端数据可视化子系统操作的相应文件。另外小程序与后台的数据交互由微信服务端负责。具体的功能流程如图 4-7。

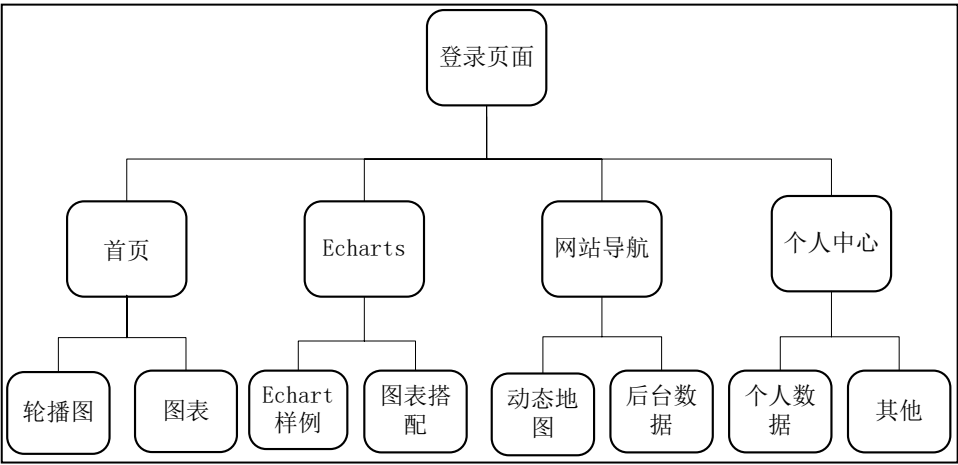
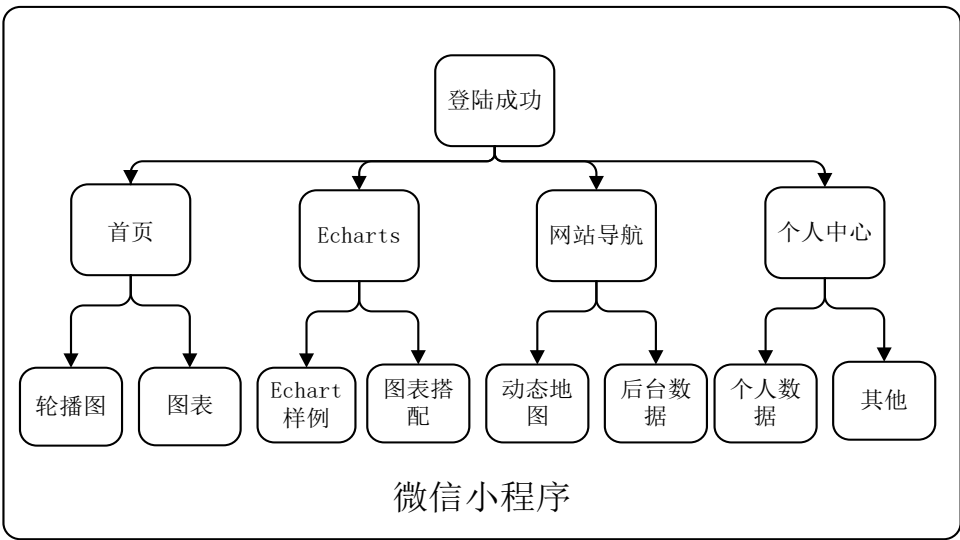


图 4-7 功能流程图

以下图 4-8 给出了微信小程序的系统模块图。

图 4-8 功能模块图



核心框架和代码解释

微信小程序平台的架构模式如图 4-9 所示。

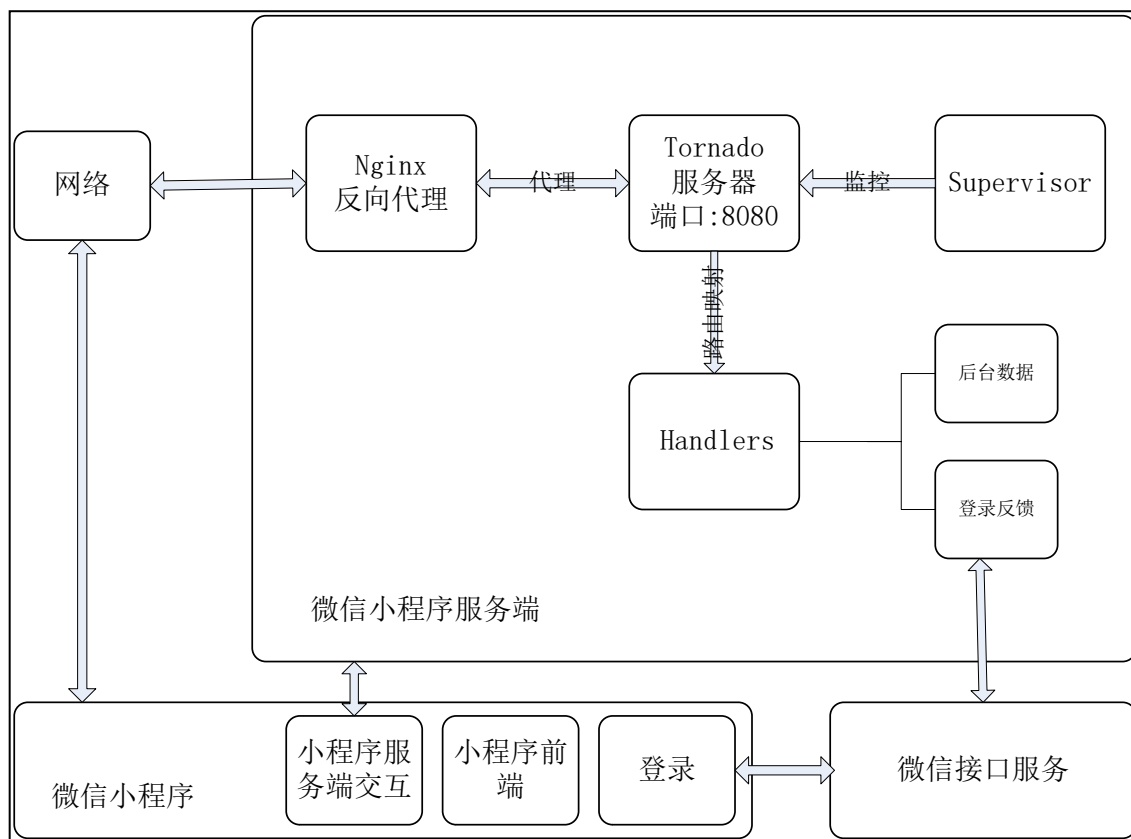


图 4-9 系统架构图

在本系统中,微信小程序端的服务器兼应用框架使用的是 Python 的 Tornado,使用 Nginx 反向代理分发 Web 请求并负载均衡,并用 Supervisor 进行监控,Handlers 中包含反馈后台数据的所有路由。

在本实验的 Nginx 配置中，主机名分别为 tornadoes、wx_tornadoes 以区分是 PC 端数据可视化系统还是微信小程序的服务器端。

微信小程序有专业的开发文档指导辅助,根据微信小程序开发手册的指导开发者可以进行个性化的小程序制作。本文关注于微信小程序对于 ECharts 的扩展,为数据可视化做了平台扩充。图 4-10 展示了微信小程序的登录流程时序^[12]。

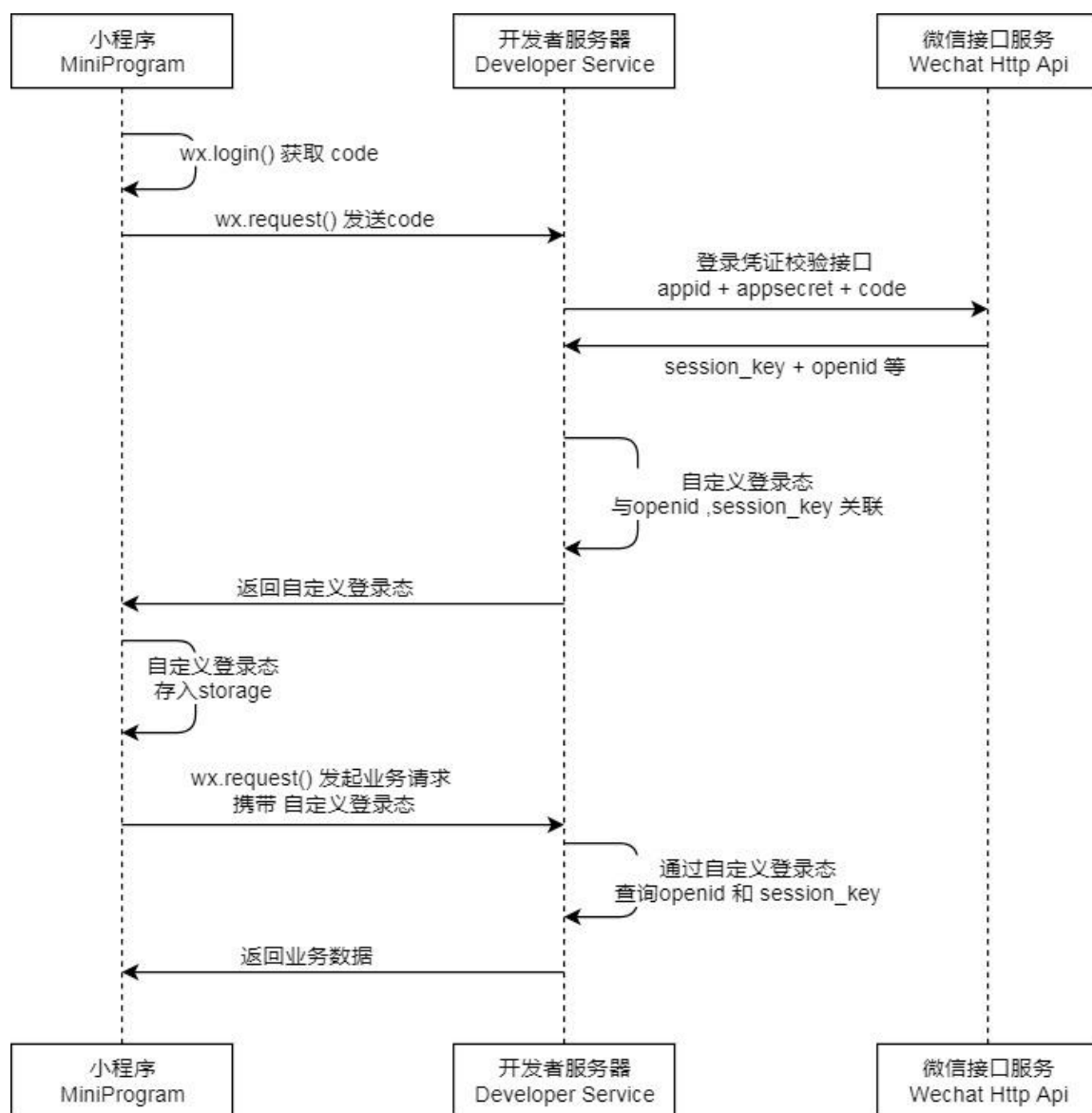


图 4-10 微信小程序登录流程时序^[12]

ECharts 的微信小程序版本是由 ECharts 和微信小程序官方团队合作所提供的。ECharts 是一个 JavaScript 图表库，其主要基于 HTML5 Canvas，可以提供交互式，直观易懂，自由化的数据可视化图表。在用户体验方面为了提升了用户对大量数据进行提取、分析、整合的能力而创新增添了数据视图、拖拽重计算等特性。ZRender 是 ECharts 的底层基础，用以创建图例，坐标系，工具箱，提示等基础组件，并在此基础上设计出众多实用性较高的图表，如柱状图、力导向图等。

后台管理部分

本章前三节分别介绍了人机交互部分，数据可视化部分和 Android 小程序部分的设计和架构，这些部分是可以相互联系的。本节将介绍本系统中的后台管理

部分，并通过分析 Tcpdump、Slowhttptest、CICFlowMeter 等工具在系统中的嵌入方式，论述整合网络攻击模拟、网络流量数据整理、数据预处理的方案，并完成了与前两节介绍的几部分相匹配的融合方式。这一部分的工作最终归纳为一套较为成熟的后台管理平台。

技术选用

本系统将使用 React 构建后台管理部分的用户界面，使用 Koa 作为后台管理系统的 Web 框架；使用 MySQL8 作为数据库支持。使用 PM2 进行监控。在系统还嵌入了 Tcpdump、Slowhttptest、CICFlowMeter 等工具，以整合网络攻击模拟、网络流量数据整理、数据预处理等功能。

（1） React 及 React-Antd-Admin 介绍

React 是一个 JavaScript 库，用于构建用户界面。其主要用于构建 UI，相当于是 MVC 中的 V（视图）。React 起源于用来架设 Instagram 的网站的 Facebook 的内部项目，于 2013 年 5 月开源。React 代码逻辑简单，拥有较高的性能，因此目前被广泛运用。

React 特点主要有：

声明式设计。React 采用声明范式，可以轻松描述应用。

高效。通过对 DOM 的模拟，React 可以减少与 DOM 的交互。

灵活。React 可以很好地与已知的库或框架配合。

JSX。JSX 是 JavaScript 语法的扩展。

组件。为了使得代码更加容易得到复用，可以在 React 中构建并使用组件，组件在大项目的开发中能够得到很好的应用。

React-Antd-Admin 是一个通用管理后台，是使用 React 和 Ant Design 搭建的 JavaScript 应用。React 搭配 Antd 易用易迁移的高质量组件，使用方便，搭建快捷，非常适合后台产品，可以简化后端人员的前端开发。

（2） Koa 介绍

Koa 是基于 Node.js 平台的 Web 开发框架，提供了一个更轻量级、更稳定的 Web 应用和 API 的开发基础。Koa 弃用回调函数，而利用 async 函数，并有力地提高了错误处理能力。Koa 没有捆绑任何中间件，使用起来干净利落，并且体积小、编程方式较为干净。

（3） Tcpdump

Tcpdump 是一种截获网络包并输出其内容的工具。具有强大的功能和灵活的捕获策略，是 Linux 系统下进行网络分析和纠错的首选工具。Tcpdump 支持对一些标签特征进行过滤。

一个在项目中使用的 Tcpdump 例子如下：

```
Tcpdump -i any -vv -w filename -s 0
```

该例子表示抓取所有网络界面的数据包，更详细地显示指令执行过程，并将数据包数据写入 filename 代表的文件中，并防止包截断。

（4） Slowhttptest

Slowhttptest 是依赖 HTTP 协议的慢速攻击 DoS 攻击工具，其设计的基本原理是服务器在请求完全接收后才会进行处理，如果客户端的发送速度缓慢或者发送不完整，服务端为其保留连接资源池占用，大量此类请求并发将导致 DoS。

Slowhttptest 攻击模式包括 Slowloris，Slow http post，Slow read attack。以下分别解读：

Slowloris：完整的 http 请求是以\r\n\r\n 结尾，攻击时仅发送\r\n，少发送一个\r\n，服务器认为请求还未发完，就会一直等待直至超时。

Slow http post：通过声明一个较大的 content-length 后，body 缓慢发送，导致服务器一直等待。

Slow read attack：发送一个正常合法的 read 请求到服务器来请求一个很大的文件，把 TCP 滑动窗口设置得很小，服务器就会以滑动窗口的大小切割文件，然后发送。文件长期滞留在内存中导致资源消耗。

功能流程与功能模块

（1） 后台管理部分

后台管理部分的主要功能流程如下：用户选择注册及登录进入后台管理平台，系统展示大部分功能，其中关键的功能是素材管理，文件下载和检测操作。在系统的各个功能模块，用户可以根据需要自行所需的素材，或是查看并下载可利用数据可视化子系统操作的相应文件。具体的功能流程如图 4-11。

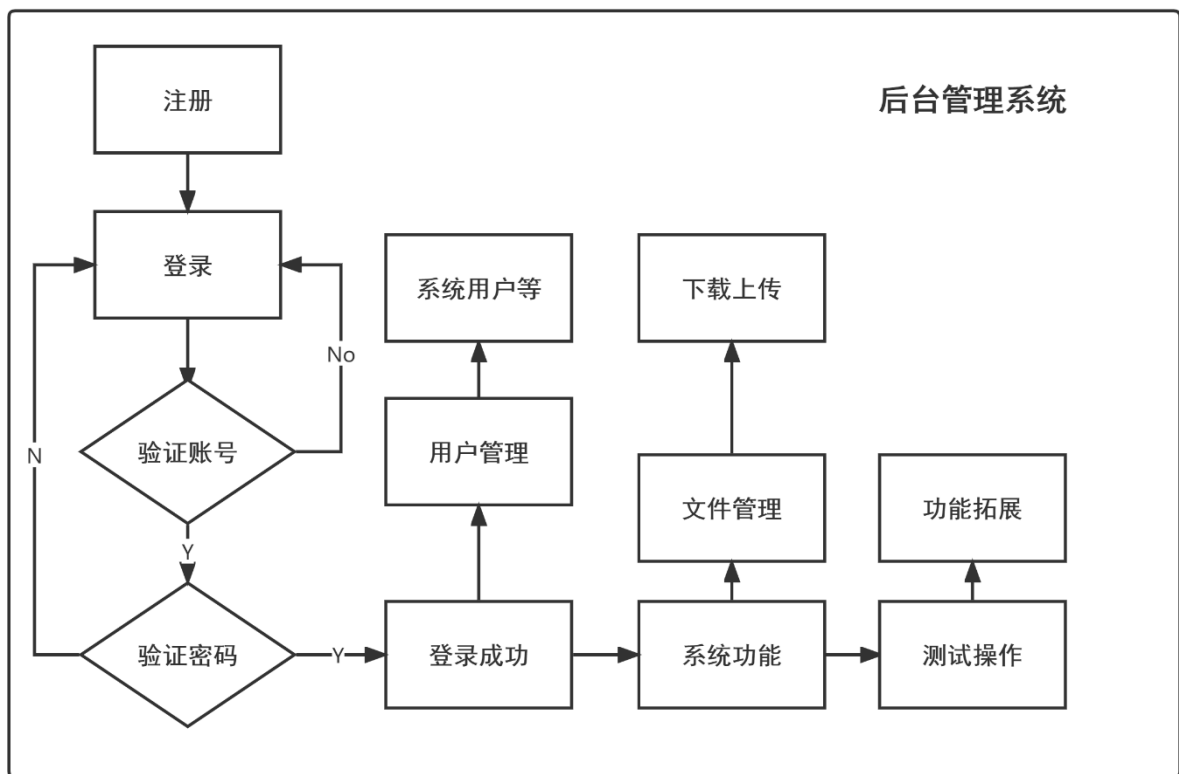


图 4-11 功能流程图

以下图 4-12 给出了后台管理部分的功能模块图。

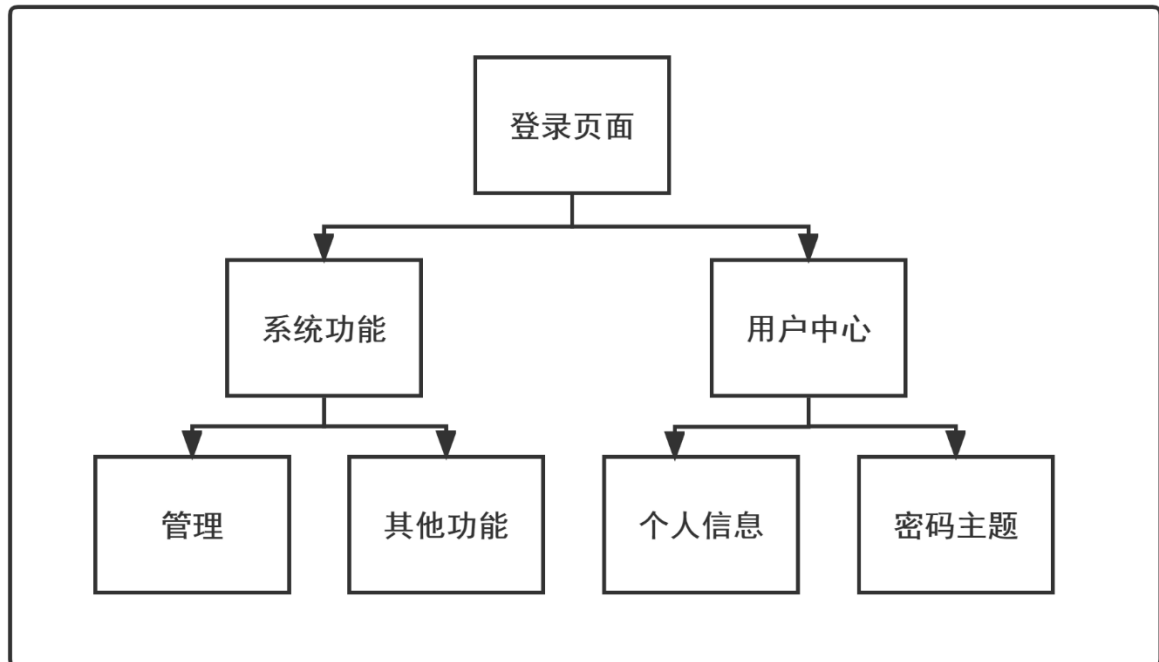


图 4-12 功能模块图

(2) Tcpdump 抓取数据包

对于 Tcpdump 抓取数据包的应用，有两种模式，分别是自动抓取和被动抓取模式。以下进行介绍：

自动抓取模式。自动抓取模式主要是为了辅助系统的文件下载功能，生成 pcap 文件与经过 CICFlowMeter 处理所得到的 csv 文件以提供真实的实验数据并通过网络流量数据可视化平台进行数据可视化操作。

被动抓取模式。被动抓取模式主要用于在实施网络攻击模拟时进行网络流量数据抓取，以实现的一段时间内的数据进行可视化。在 CIC-IDS-2017 数据集特征分析的基础上，可以根据具体攻击模式进行相应特征的数据可视化，更直观地观测数据特征的趋势和变化。

Tcpdump 嵌入功能的实现关系到后台管理系统中文件下载系统的实现。目前系统中主要利用 Tcpdump 实现数据包的自动捕获并将数据包保存到 pcap 文件中，再利用 CICFlowMeter 实现 csv 文件的获取。功能流程：在服务器端运行 Shell 脚本以运行 Tcpdump 自动捕获程序，可以自动捕捉数据包，并保存到指定位置以供系统调用，在实验中可停止捕获。Tcpdump 嵌入相应功能流程和功能模块如图 4-13 所示。

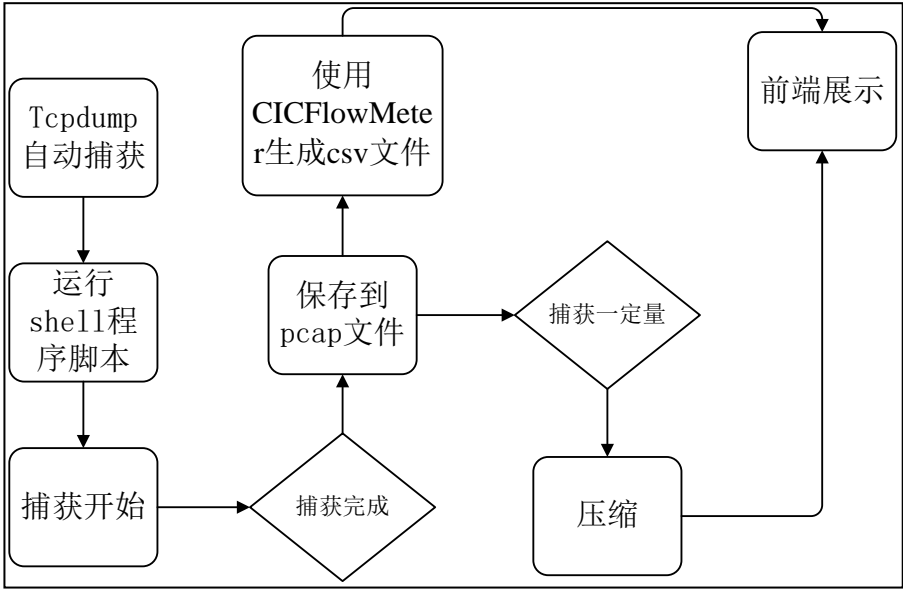


图 4-13 Tcpdump 功能流程功能模块

（3）Slowhttptest 模拟网络攻击

利用 Slowhttptest 模拟网络攻击主要是为了结合对数据集特征的研究，在了解研究各个网络攻击的最佳特征后，更真实地模拟网络攻击操作，获取真实的网络流量数据，为接下来的网络流量数据可视化技术研究提供便利。

Slowhttpstest 功能流程：在后台管理系统中选择检测操作；选择合适的网络攻击类型，触发后台的 shell 命令，对服务器的另一个端口进行网络攻击，结果将直接反馈在前端页面中。如图 4-14 所示。

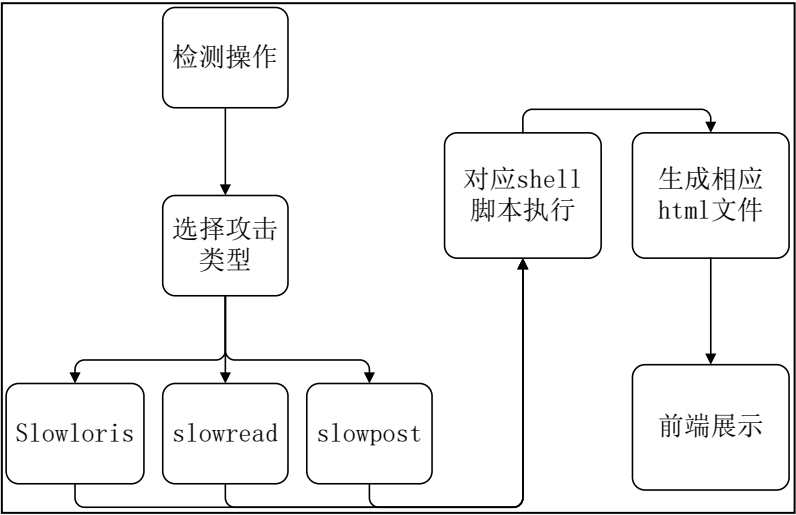


图 4-14 Slowhttpstest 功能流程

核心框架和解释

后台管理部分的总体架构模式如图 4-15 所示。

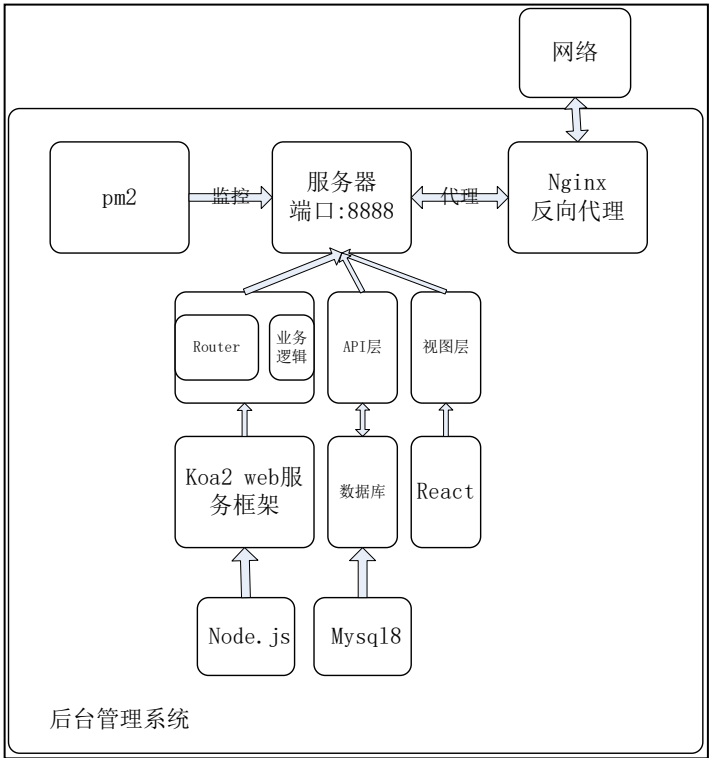


图 4-15 系统架构图

PM2 是一个 Node.js 应用进程管理工具，提供负载均衡的功能，使用 PM2 可

以使很多 Node.js 应用管理的工作得到简化，如自动重启和性能监控等，使用简单。

本章小结

本章主要介绍了树莓派人机交互及可视化系统各个功能部分的具体实现，其中包括人机交互部分，数据可视化部分，以微信小程序为主的业务拓展部分以及后台管理部分的实现，并且介绍了后台数据处理部分的实现。详细介绍了相关子系统的技术选用，功能流程与模块，核心架构与功能设计等。

系统测试及未来展望

本章将介绍对于本文所设计的嵌入式设备树莓派人机交互及可视化系统进行实验测试。


系统功能设计与测试

人机交互部分

本系统基于嵌入式设备树莓派，我们设计了基于 Web 的人机交互界面。首先本系统的人机交互界面基于数据可视化系统的支持。通过数据可视化系统的前端页面进行登录。

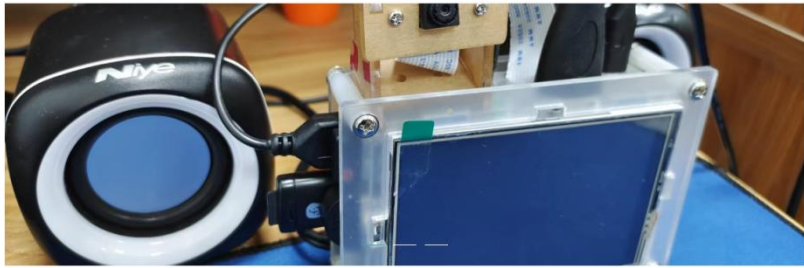


登录成功后进入可视化系统界面。我们的人机交互操作可以在数据可视化模块进行。




188-9680-90861369162653@qq.com冯波

[系统首页](#)[实验案例](#)




服务范围SERVICE SCOPE

网络流量数据可视化系统




CSV文件处理

上传csv文件，将数据导入可视化系统...




数据可视化

将数据导入可视化系统，生成图表和报告...



上传数据操作


上传个人或系统生成的数据...



相关知识

提供与系统相关的知识...

查看详情



188-9680-90861369162653@qq.com冯波

[系统首页](#)[实验案例](#)

攻击类型统计
ATTACK TYPE


攻击源IP和目的IP
SOURCE IP & DESTINATION IP

时间戳对IP
TIMESTAMP & IP

时间戳对其他
TIMESTAMP & OTHER

其他对其他
OTHER & OTHER

HTTPS is required for full functionality



连接

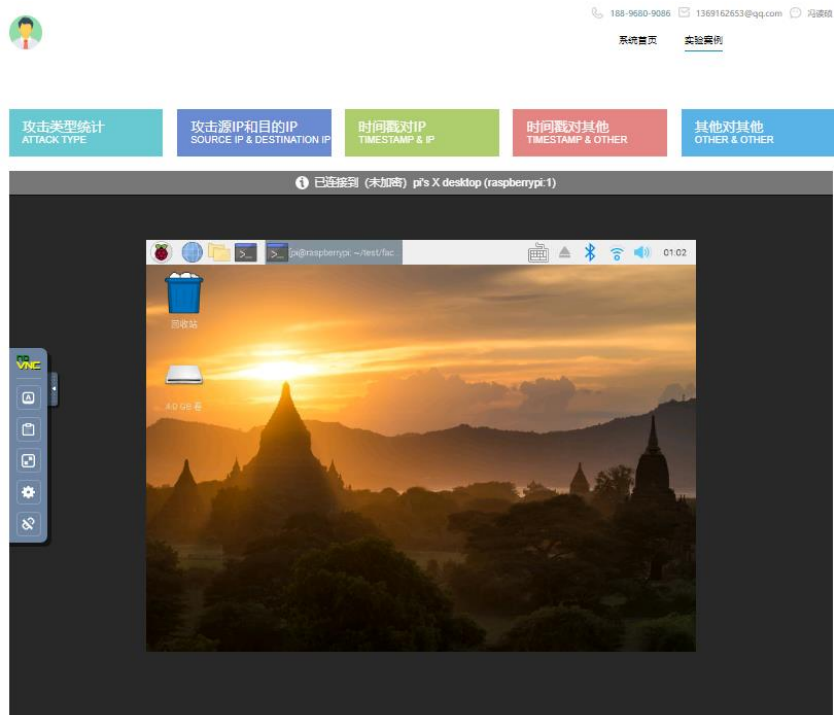


图 5-1 人机交互模块

之后我们测试人脸识别模块：

- (1) 人脸识别考勤：识别效果如图 5-2。

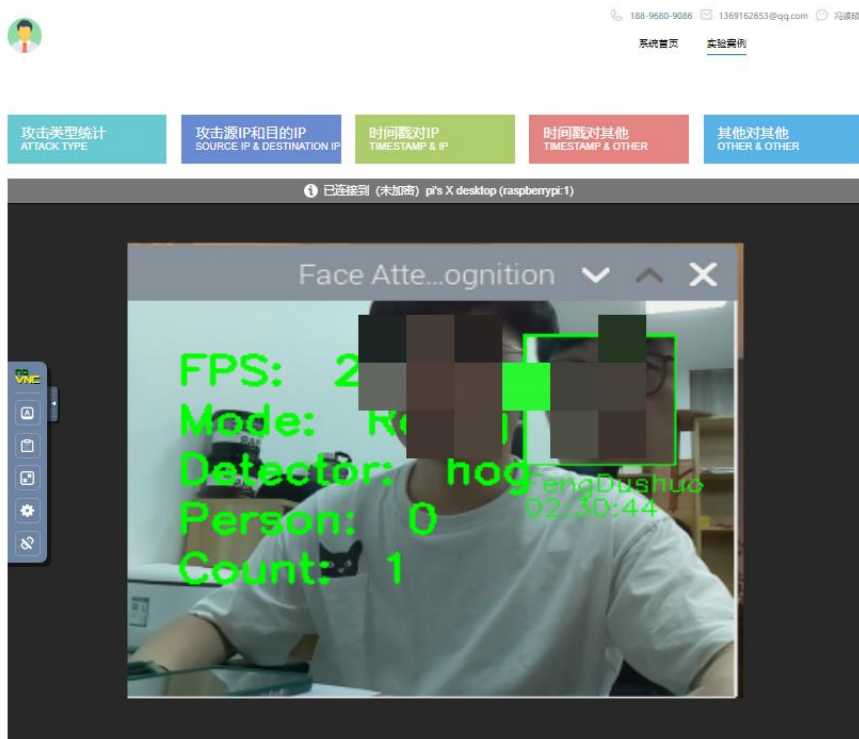


图 5-2 人脸识别考勤

- (2) 依据人脸识别口罩：识别效果如图 5-3。

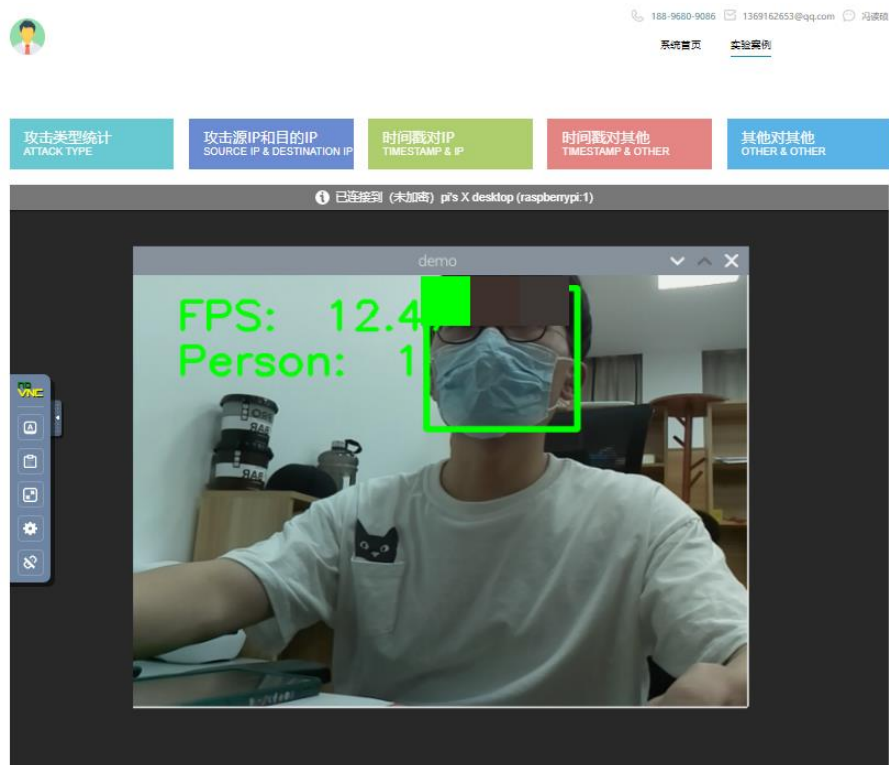
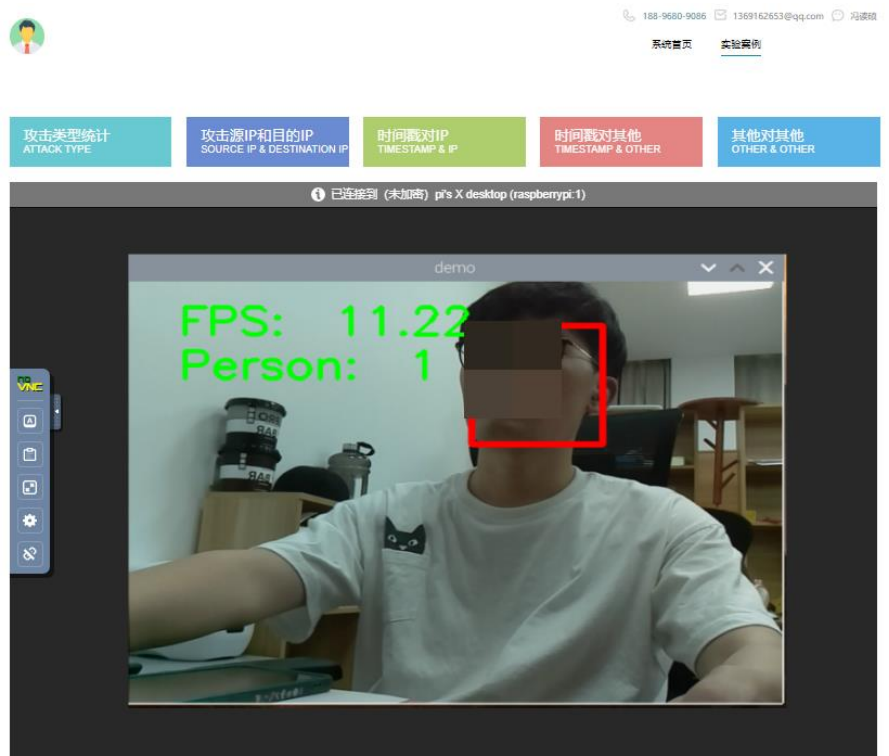


图 5-3 人脸识别口罩

(3) 依据人脸识别是否低头玩手机：识别效果如图 5-4。

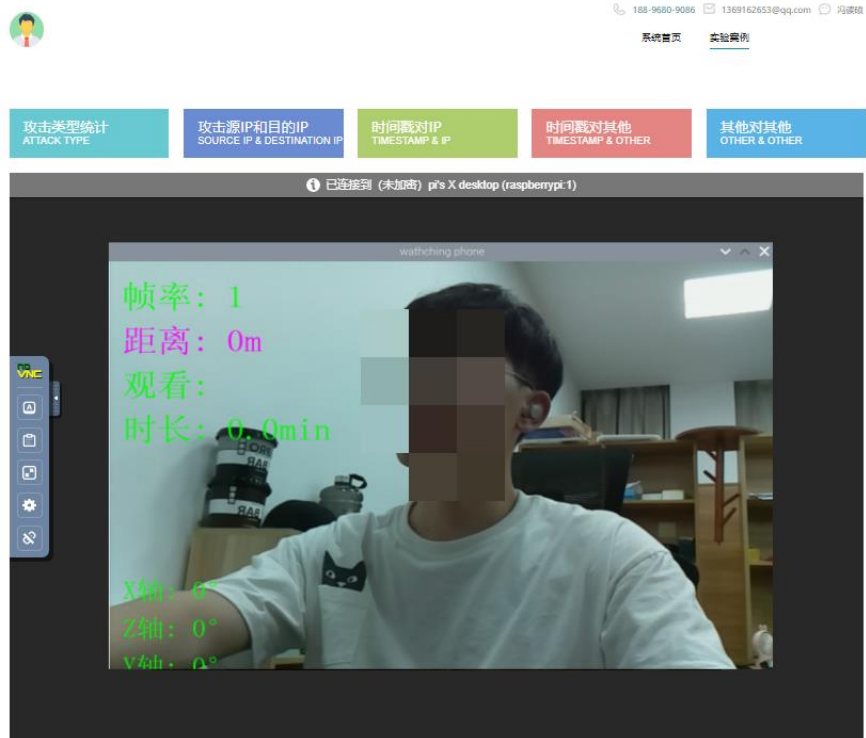


图 5-4 识别是否低头玩手机

(1) 手势背单词：识别效果如图 5-6。



图 5-6 手势背单词

最后我们测试语音识别模块：识别效果如图 5-7。

```

INFO:snowboy:Keyword 1 detected at time: 2022-05-07 02:53:16
INFO:__main__:开始录音
INFO:robot.Conversation:结束录音
INFO:robot.ASR:baidu-asr 语音识别出错了: recognition error.
INFO:robot.Conversation:命中缓存, 播放缓存语音
INFO:robot.ASR:baidu-asr 语音识别到了: ['杭州明天什么天气']
INFO:robot.AI:tuling 回答: 杭州:周日 05月08日,阴 东风转东北风,最低气温17度,最高气温26度。
INFO:robot.TTS:baidu-tts 语音合成成功, 合成路径: /tmp/tmpj9my3wqk.mp3

```

图 5-7 语音识别对应的文字输出

数据可视化部分

近年来, IDS 领域的研究蓬勃发展, 研究人员目前已经可以获取到很多公开数据集用以测试和评估自己的技术或工具。目前可用的 IDS 数据集如林肯实验室的 DARPA, 加州大学尔湾分校的 KDD'99, 应用互联网数据分析中心的 CAIDA 等。为了证明 CIC-IDS-2017 数据集的独特优势并阐述其符合当前对全面可靠数据集的真实需求, 本文分析和评估了自 1998 年以来公开提供的 11 个 IDS 数据集, 并将结果汇总成表, 如表 5-1 所示。

表 5-1 IDS 数据集分析

数据集	实验室	概要	缺点
DARPA	林肯实验室	人为注入攻击, 良性流量	缺乏实际攻击数据记录
KDD'99	加州大学尔湾分校	良性和攻击流量	大量冗余记录, 数据损坏
DEFCON	Shmoo 组	端口扫描和缓冲区溢出攻击	不同于现实世界的网络流量
CAIDA	应用互联网数据分析中心	特定于特定事件或攻击	不是有效的基准数据集
LBNL	劳伦斯伯克利国家实验室	中型站点记录	没有有效载荷, 匿名化
CDX	美国军事学院	网络战争竞赛	缺乏流量多样性和容量
Kyoto	京都大学	Honeypots	没有假阳性
Twente	特温特大学	Net-flow, 蜜罐网络	缺乏多样性
UMASS	马萨诸塞州大学	跟踪文件	缺乏流量和攻击的多样性
ISCX2012	新不伦瑞克大学	全包有效负载	不是基于真实世界的统计数据
ADFA	新南威尔士大学	FTP 和 SSH 密码暴力破解	缺乏攻击的多样性

CIC-IDS-2017 数据集涵盖了所有 11 个常见攻击的必要标准, 如 DoS、DDoS、Brute Force、XSS、SQL 注入、过滤、端口扫描和 Botnet。该数据集被完全标记, 并通过使用 CICFlowMeter 软件提取和计算了所有良性和入侵性流量的 80 多个

网络流量特征，该软件在加拿大网络安全研究（网站上公开提供(Habibi Lashkari 等，2017 年)。其次，Iman Sharafaldin^[8]等分析了 CIC-IDS-2017 数据集，选择最佳的特征集来检测不同的攻击，并执行了七种常用的机器学习算法来评估此数据集。

使用 CICFlowMeter 从数据集中提取 80 个流量特征。CICFlowMeter 作为流特征提取工具，能够根据提交的 pcap 文件生成有 80 多个特征的 csv 文件，使用方法有两种：在线和离线模式。在线模式可以实时监控并产生特征，监听结束之后可以保存到本地；离线模式是提交一个 pcap 文件，得到一个包含特征的 csv 文件。

为了从 80 个提取的特征中找到检测每个攻击的最佳特征集，使用 Scikit-Learn 的随（森林回归模型类(Pedregosa)人，2011 年)。首先计算整个数据集中每个特征的重要性，然后将每个类上每个特征拆分的平均标准化平均值与相应的特征重要性值相乘，得到最终结果。

本系统首页如图 5-8 所示，首页顶部是功能导航栏和轮播图（导航到 D3.js 官网等），导航栏的分类主要有系统首页和实验案例。底部是系统所提供的功能，本论文所介绍的是其中的网络攻击数据显示和上传数据操作这两个功能。

本系统主要功能包括网络安全数据集的网络攻击数据显示和上传数据操作。在网络攻击数据显示功能页面中具体展示了数据集中的攻击类型统计，攻击源 IP 和目的 IP，时间戳对 IP（后因为数据集中时间戳过于集中而省略此功能），时间戳对其他，其他对其他（根据数据集中的特征标签来命名功能），如图 5-9 所示。上传数据操作功能页中，包括攻击源 IP 和目的 IP，时间戳对其他这两个功能，其他功能因为实用性不高被舍弃。上传数据操作功能中用户可以上传自己的 csv 文件（CICFlowMeter 处理过的输出文件），生成对应的可视化图表，如图 5-10 所示。

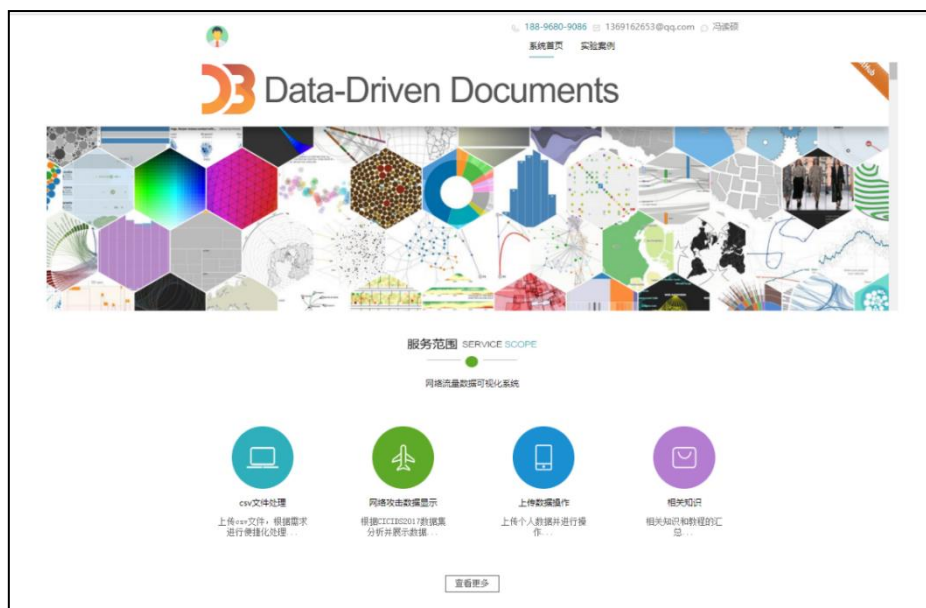


图 5-8 主页面设计图

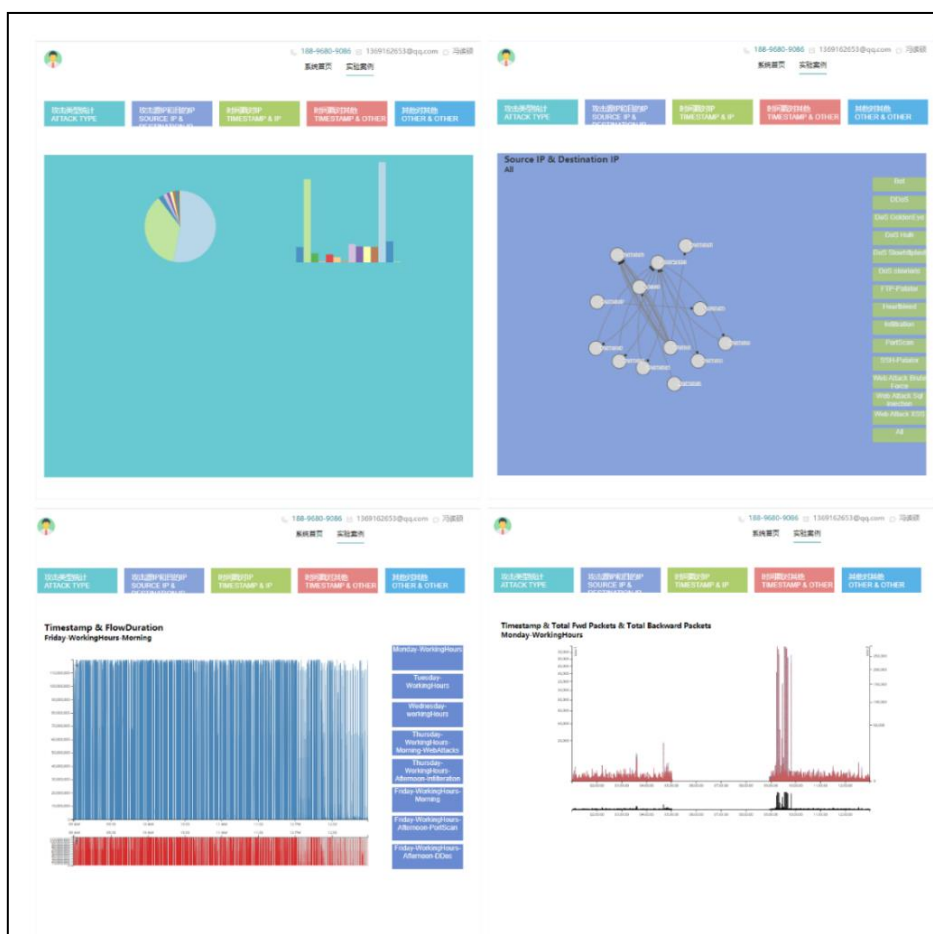


图 5-9 网络攻击数据显示功能图



图 5-10 上传数据操作功能图

微信小程序与对外门户网站

本实验中设计并开发了以后台管理系统为依托的微信小程序和对外门户网站，主要是为了扩充嵌入式设备数据分析整合平台的多样性。在此微信小程序中引入了 ECharts 的微信小程序版本，使得数据可视化技术可以在微信小程序中得到应用，满足了当前对数据可视化技术广泛应用的需求。因为微信小程序即搜即用的特点，为用户提供更好的数据可视化体验，让每个人都能参与到网络数据可视化和人机交互的研究中，为嵌入式设备人机交互和网络安全研究中各种技术的发展提供了可靠的支持。

本实验开发的微信小程序把功能侧重于数据图表的展示和多个平台之间的

交互。

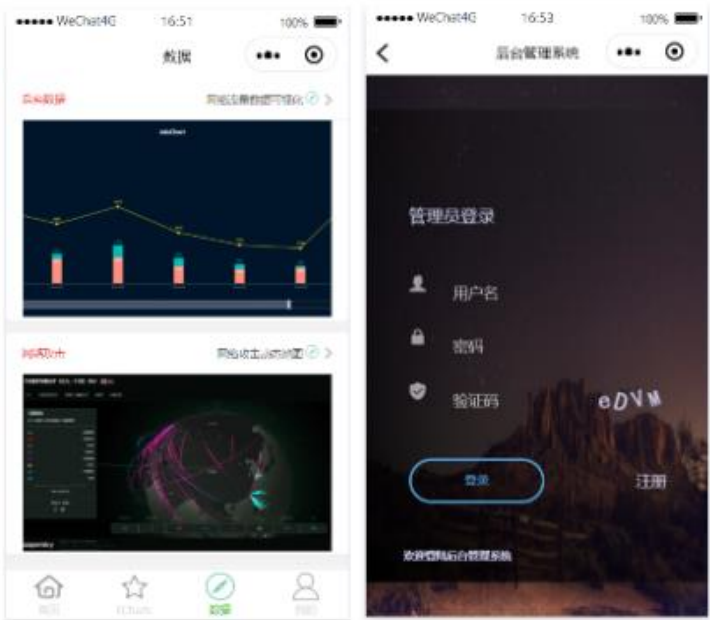
图表展示：结合微信小程序的 ECharts Demo，展示现有的 ECharts 图表，为之后的网络流量数据可视化工作做好铺垫。如图 5.11 所示。



图 5-11 图表展示

平台交互：结合后台管理系统和网络攻击地图进行网页嵌入。如图 5.12 所示。

图 5-12 平台交互

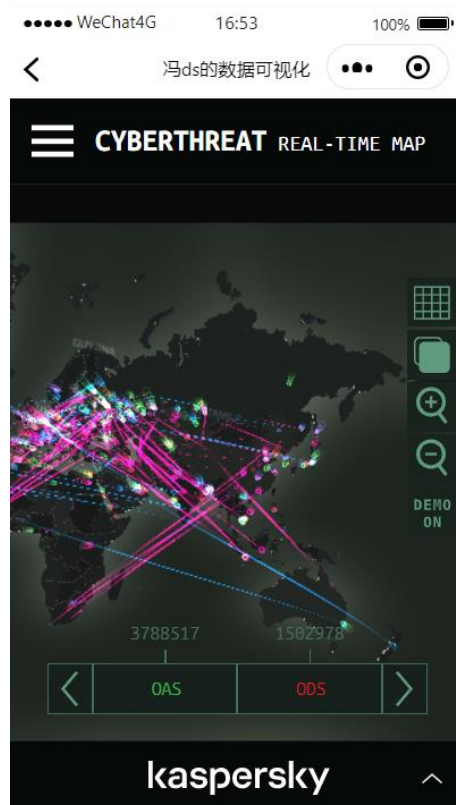


平台交互中使用小程序的 web-view 组件嵌入后台管理系统和网络攻击地图 cybermap。

web-view：web-view 是一个可以承载网页的容器。使用 HTML5 开发的网站

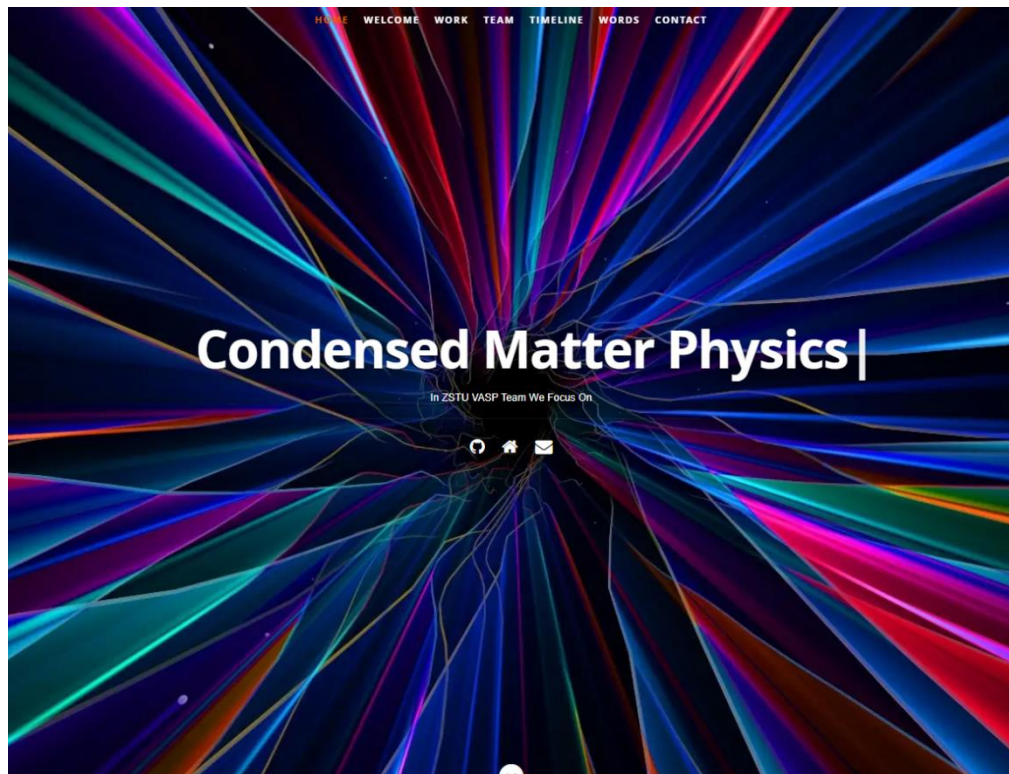
可以利用 web-view 直接迁移到微信小程序中，web-view 可以自动铺满整个小程序的页面。不过目前个人类型的小程序不支持使用，所以在线上展示中个人类型的小程序会显示不出。

图 5-13 卡巴斯基网络威胁实时地图



cybermap.kaspersky.com 网站可以实时监控全球网络威胁并通过动态地图的形式展示，虽然是商业网站，但是可视化效果很好。该网站对于网络威胁的展示很科幻绚丽，迁移到小程序中可以辅助网络数据可视化平台的多样化开发，如图 5.13 所示。

本实验开发的对外门户网站可以很好的展示一个机构的整体信息，在使用嵌入式设备的机构或组织中可以得到很好的应用。

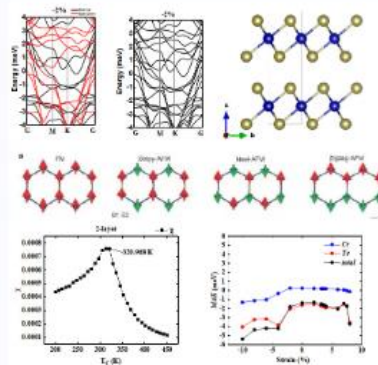


CURRENT WORK

First principle calculation

Calculation methods with our own characteristics

We have carried out characteristic first principle calculation according to our concrete needs. Our current computing work depends on many computing simulation software recognized in the industry, such as VASP, spirit, vampire and so on. In order to achieve many specific requirements, we are also trying to write our own program and apply it to the actual scene. So far we have some research results in the field of first principle calculation.



Two-dimensional materials

Research and exploration of two-dimensional materials

Among all kinds of materials, we are most interested in two-dimensional materials, which are also called two-dimensional van der Waals materials. We are now focus on structural design, electronic structure and physical property regulation of two-dimensional semiconductors and magnetic related materials, theoretical calculation of electronic state behavior at the interface of low dimensional van der Waals heterojunction.

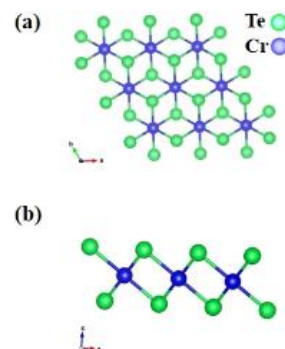




图 5-14 对外门户网站

后台管理部分

本平台中后台管理系统采用 React-Admin 框架,因为 React-Admin 框架比较适用于搭建简单通用的后台管理系统,其在开源的基础上可以引入更多的框架更新和功能扩充优化。以下将介绍网络流量数据可视化系统后台管理部分中的主要功能:

用户注册登录: 本功能为系统基础功能,用于用户的注册和登录,用户可分为管理员和普通用户,管理员可以在后台进行设置,目前为了系统的通用性,系统所有功能设置对所有用户可见。

系统首页: 系统首页设置侧边导航栏,显示具体的功能,首页也在右上角设置了用户中心以管理用户的信息。如图 5-15 所示。



图 5-15 系统首页图

用户管理：用于管理使用本系统的用户，如图 5-16 所示。

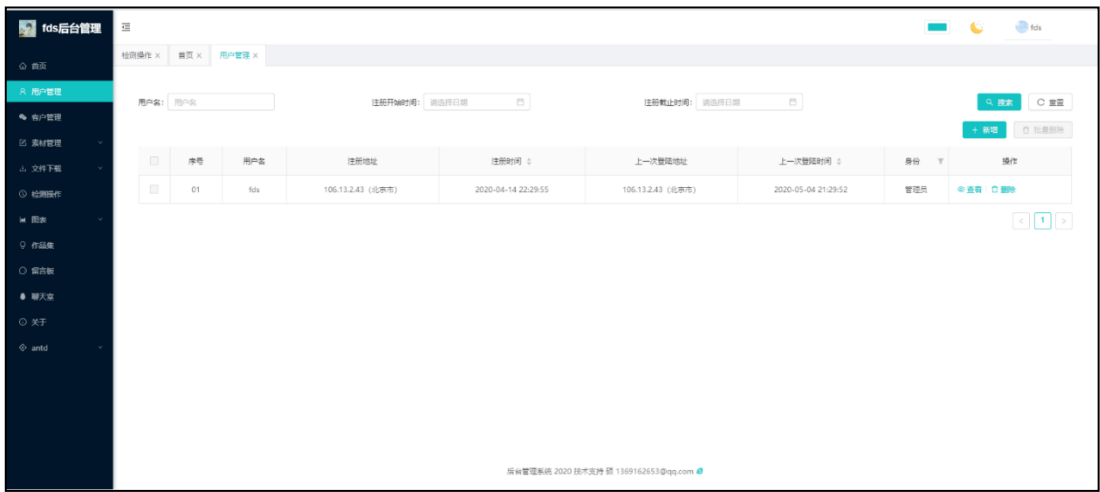


图 5-16 用户管理

素材管理：用于管理各类素材，提供更新数据的功能。如图 5-17 所示。

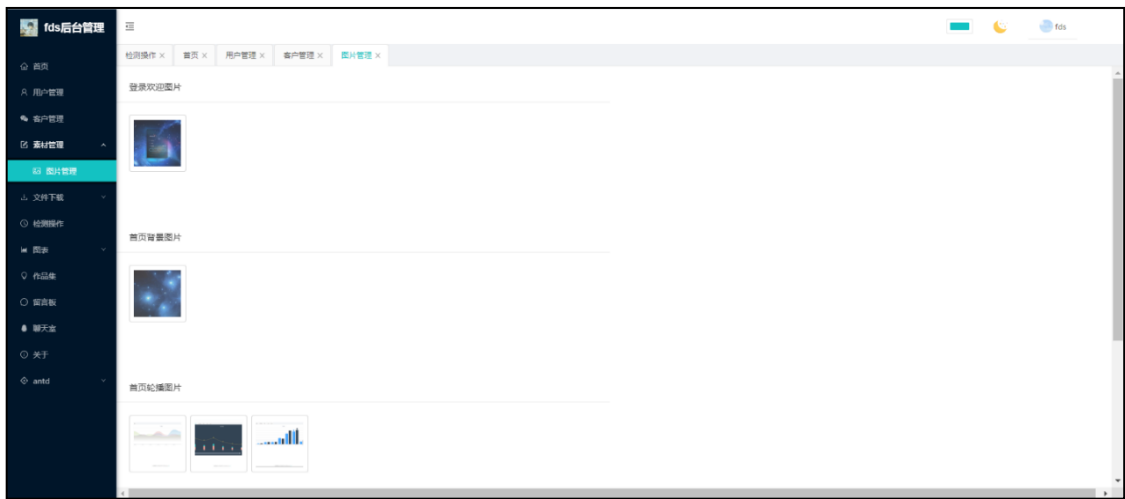


图 5-17 素材管理

文件下载：此功能为了与数据可视化子系统交互，下载适合用数据可视化系统进行操作的数据，包括 csv 文件和 pcap 文件，这些文件是由 Tcpdump 在项目所部署的服务器上所捕获的流量数据文件。具体如图 5-18 所示。

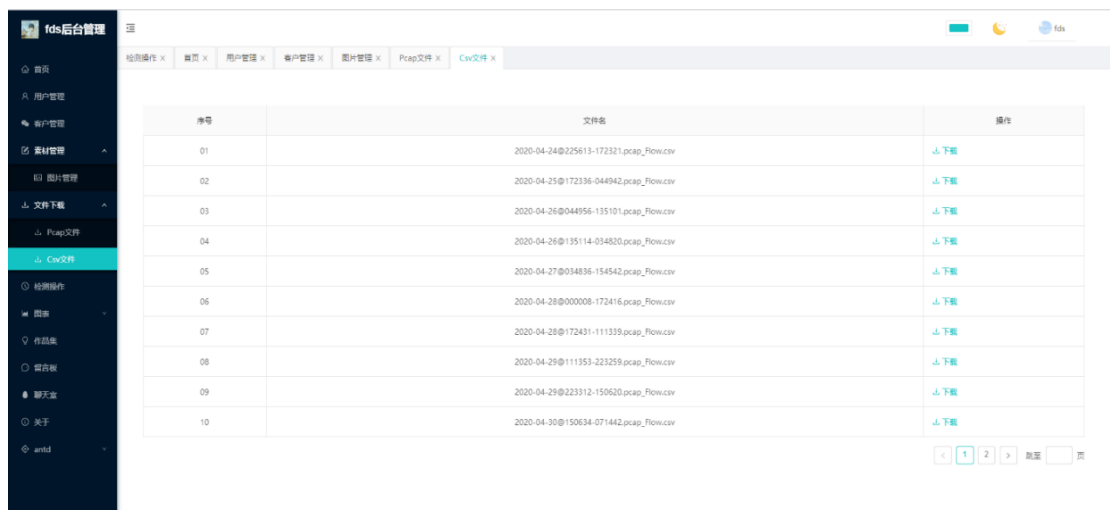


图 5-18 文件下载

其他：如图表，留言板等，这些功能借鉴开源的框架功能插件，以展示 React-Admin 框架的可扩展性。

展望

由于时间问题，还有很多功能已经在构思但是没有来得及完成，例如想多增加一些基于机器视觉或是语音识别的人机交互，实现更智能的系统整合；想要开发更专业的微信小程序拓宽与嵌入式平台交互的途径；还有基于安全考虑，希望使用更安全的硬件架构 TrustZone 进行安全开发，并引入开源的安全执行环境 OP-TEE 等。这些想法会在之后执行。

本章小结

本章首先介绍了基于 Web 和嵌入式设备的人机交互及可视化系统的主要功能设计，并通过测试。满足了我们对于人机交互和可视化平台便捷化管理的需求，并实现了平台的整体部署。