中原工学院

第六届“挑战杯”大学生课外学术科技作品竞赛

作品研究报告

作品名称： 基于浏览器的网站安全事件检测系统

学院名称： 计算机学院

申报者姓名：洞见研发组

指导教师： 冯国朋

类别：

□自然科学类学术论文

□哲学社会科学类社会调查报告和学术论文

☑科技发明制作A类

□科技发明制作B类

# 摘要

互联网的发展给人们带来诸多便利的同时，许多安全问题也日益突显，做好信息安全防护，成为国家、企业乃至个人重要的工作。如果没有做好信息安全防护工作，不管是对于个人，还是企业和国家，所造成的危害都将是致命的。基于浏览器的网站安全事件检测系统能够突破传统扫描器不能适应新型的网站框架的弱项，通过调用Google开源框架Chromium进行模拟浏览器访问网站，包括加载网站操作Dom树后的全部内容。对网站进行是否存在劫持、被黑、暗链、挖矿、反共关键字等多方面高效检测。并对与存在的安全事件与威胁做快照存储。第一时间通知网站管理员，降低网站被黑后带来的负面影响。此检测系统采用Docker部署，同时对部署的服务器性能要求极低。

**关键词：**网站安全；扫描器； Chromium；

目录

[摘要 2](#_Toc531203457)

[第 1 章 绪 论 4](#_Toc531203458)

[1.1 研究的背景以及目的 4](#_Toc531203459)

[1.2 国内形式分析 4](#_Toc531203460)

[第 2 章 系统分析与设计 6](#_Toc531203461)

[2.1 需求分析 6](#_Toc531203462)

[2.2 开发环境 6](#_Toc531203463)

[2.3 引擎模块架构设计 6](#_Toc531203464)

[2.3.1 系统架构图 6](#_Toc531203465)

[2.3.2 任务调度模块 7](#_Toc531203466)

[2.3.3 启发式爬虫模块 7](#_Toc531203467)

[2.3.4 搜索引擎模块 8](#_Toc531203468)

[2.3.5 安全事件模块 8](#_Toc531203469)

[2.3.6 漏洞扫描模块 8](#_Toc531203470)

[2.4 Web模块架构设计 9](#_Toc531203471)

[2.4.1 系统架构图 9](#_Toc531203472)

[2.4.2 平台用户设计 10](#_Toc531203473)

[2.4.3 个人信息模块 10](#_Toc531203474)

[2.4.4 个人信息修改 10](#_Toc531203475)

[2.4.5 个人信息展示模块 11](#_Toc531203476)

[2.4.6 安全事件审核模块 12](#_Toc531203477)

[2.4.7 信息记录模块 12](#_Toc531203478)

[2.4.8 信息记录模块 13](#_Toc531203479)

[2.4.9 事件通报模块 13](#_Toc531203480)

[第 3 章 关键技术及创新点 14](#_Toc531203481)

[3.1 关键技术 14](#_Toc531203482)

[3.2 技术创新点 14](#_Toc531203483)

[第 4 章 前景分析 15](#_Toc531203484)

# 绪 论

## 研究的背景以及目的

Intemet为代表的全球性信息化浪潮日益深刻，网络信息技术的应用正日益普及，应用层次逐渐深入，应用领域从传统的、小型业务系统逐渐向大型、关键业务系统扩展。随着网络的普及，安全日益成为影响网络效能的重要问题，而Internet所具有的开放性、国际性和自由性在增加应用自由度的同时，对安全提出了更高的要求，这主要表现在以下几个方面：

开放性的网络，导致网络的技术是全开放的，任何个人、团体都可能获得网络上的信息，因而网络所面临的破坏和攻击可能是多方面的。例如，可以对物理传输线路实旋攻击，也可以对网络通信协议和实现实施攻击，可以对软件系统实施攻击，也可以对硬件系统实施攻击。

国际性的网络，意味着网络的攻击不仅仅来自本地网络的用户，它可以来自Internet上的任何一台机器。也就是说，网络安全所面临的是一个国际化的挑战。自由性的网络，意味着网络最初对用户的使用并没有提供任何的技术约束。用户可以自由地访问网络，自由地使用和发布各种类型的信息。

虽然信息技术的使用给人们生活、工作的方方面面带来了数不尽的便捷和好处。然而，计算机信息技术也和其他科学技术一样是一把双刃剑。大部分人使用信息技术提高工作效率，为社会创造更多的财富，而另外一些人却利用信息技术却做着相反的事情。比如，非法侵入他人的计算机系统，窃取机密信息，篡改和破坏数据，给社会造成难以估量的巨大损失

随着计算机网络的发展，网站安全也逐渐引起国家和个人的重视。随着《网络安全法》的推出，从法律层面重上保护了公民的权益，但是仍然存在着网站被黑等现象。即使国家和企业不断加强信息安全建设，但是形势依然严峻。很多单位网站架设了多种网络安全设备，仍然遭到攻击。通过扫描网站漏洞增强网站安全性是每个网站必需做的安全工作。Web网站扫描器的功用就是用于扫描网站当中存在的安全漏洞，进而进行修补，使网站处于一个安全状态。

## 国内形式分析

目前，全世界每20秒就有一起黑客事件发生，仅中国每年因此造成的经济损失就高达80多亿元。网络安全问题己经成为一个关系到国家安全和主权、社会的稳定、民族文化的继承和发扬的重要问题。

但是目前很多企业单位购买网络安全产品大都采用默认配置，没有根据安全防护对其进行加固，往往通过防火墙作为安全防护。网站管理员很难全面对每个安全产品进行有效的安全配置以及应用，安排专业的安全人员负责每个单品必然正价单位的运营成本。

网站安全问题日趋严峻，几乎每天都有大量的网站被篡改，被植入木马、后门的恶意的修改。网站安全扫描产品也如同雨后春笋般的出现，虽然网站安全扫描产品不能像安全防护设备那样被动防御、提前保护的方式加固网站安全，但也能够帮助网站管理员及时了解网站内存在的安全隐患。

综上所述，网络安全问题己引起人们的广泛关注，并成为当前网络技术研究的热点。

目前，存在多种解决网络安全问题的策略，其中网络检测系统可以发现系统存在的问题，防患于未然，成为实现网络安全的重要技术之一。因此，对网络检测系统的研究具有重要的意义。

# 系统分析与设计

## 需求分析

随着网站Web2.0前端框架的流行，越来越多的政府、学校等单位摆脱传统的前端架构，更换为新的前端架构，在使得网站的界面更优美、使用更人性化的同时也给传统网站安全扫描器带来了扫描精准度下降的缺陷。传统的获取数据方式只是获取网站源码，没有将网站的内容加载展示。（基于浏览器的网站安全事件监测系统--待拟定）整体分为两部分：Api引擎模块与Web展示模块。

（基于浏览器的网站安全事件监测系统--待拟定）API整体模块采用Google的chrome内核模拟浏览器访问网站，拦截网站传输数据，对数据进行安全分析。突出网站数据完整性，与后台规则库匹配快速定位事件产生的原因。同时采用chrome内核调用第三方搜索引擎如：百度、必应、360、Google等的搜索接口。针对搜索引擎下的网站恶意SEO排名产生的劫持事件做特殊关键字搜索，对于此方式产生的网站劫持事件隐蔽性很高普通用户以及站长都很难发现。此平台对于被检测网站无攻击渗透行为，不影响网站线上功能。实现自动化安全测试，大大降低人工重复测试成本。

Web展示界面。将被检测网站与监测数据结果在监测平台展示，实现统一管理，查询，用户不仅可以获取被检测网站的安全事件还能对网站事件进行溯源。

## 开发环境

平台开发环境设计如表1所示。

|  |  |
| --- | --- |
| 系统结构 | B/S(浏览器/服务器) |
| 开发平台 | JetBrains PyCharm 2018.2.2 x64 |
| 系统数据库 | Mongodb、Redis、Memcached |
| 开发语言 | Python3.6.5 |

表1 平台开发环境设计

## 引擎模块架构设计

### 系统架构图

引擎模块分为5个小模块：任务调度模块、启发式爬虫模块、搜索引擎模块、安全事件模块、漏洞扫描模块。

整体结构采用前后端分离，模块化开发形式。数据全部采用restful-api来传输，有专门的web管理平台来管理扫描网站与展示结果数据。底层提供专门的API接口集成了启发式爬虫、搜索引擎、规则事件、漏洞检测等功能模块。 模块间相互独立且能相互调用合作，耦合性低。

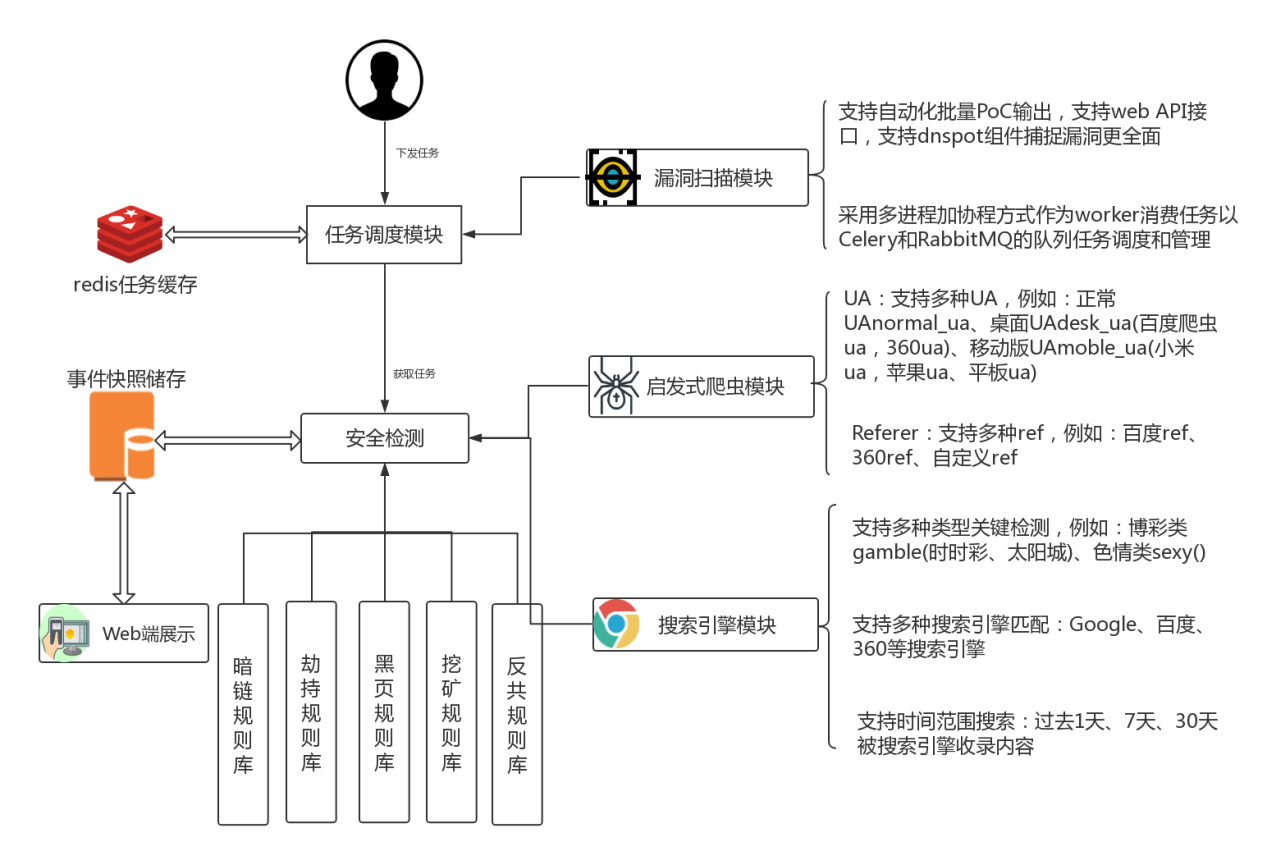


图1 引擎模块架构图

### 任务调度模块

采用redis数据库做为任务调度缓存，任务有优先级高与低之分。任务下发时，高优先级任务将先被执行。确保在碰到突发紧急安全漏洞或事件，在不影响原来任务队列的情况下，优先扫描紧急任务。

### 启发式爬虫模块

谷歌开源的puppeteer技术，主要是为了动态的获取页面内容，一般的静态爬虫例如requests只能获取到页面静态内容，获取不到页面JS加载后的内容，使用puppeteer获取的内容更准确更全面

### 搜索引擎模块

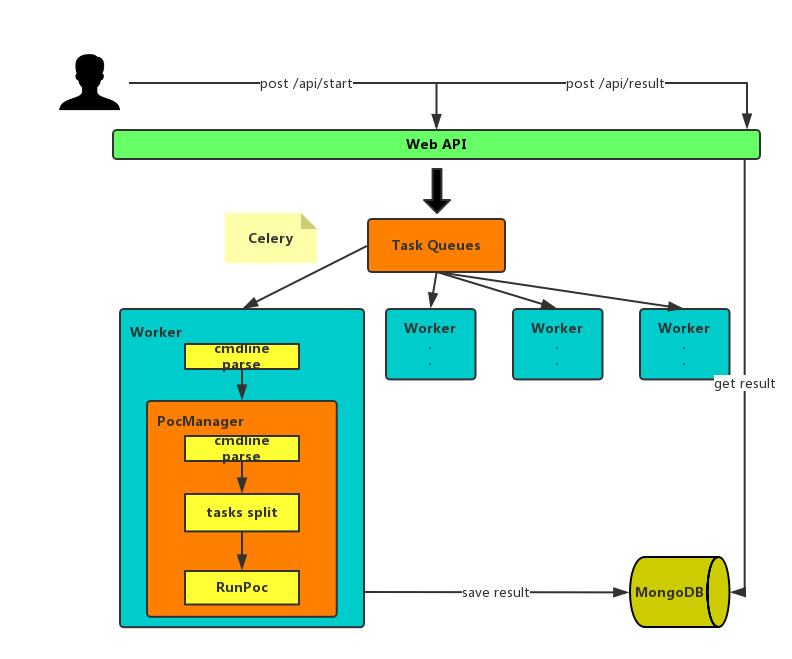
调用谷歌、百度、360等搜索引擎的接口，根据关键词(色情类，博彩类等例如澳门赌场，北京赛车)去查询指定的网站当天是否含有该关键词。此方法主要是寻找网站被劫持的新型手段。不法网站为搜索引擎SEO的提高，使用专门欺骗搜索引擎的蜘蛛池，让百度的索引到蜘蛛池内，将在被入侵的网站的url下产生随机的目录。此方法下捕捉到的安全事件，普通用户与正常的爬虫都难以发现。通过程序批量调用搜索引擎模块是一种新型的安全事件捕捉方法。

### 安全事件模块

利用启发式爬虫获取的站点资源，如：页面源码，所有加载的请求，JS源码，跳转链接等资源。通过我们自己的规则库进行规则匹配验证是否含有黑页，劫持，挖矿。规则库易于维护增加。每次命中某条规则都会给规则增加权重。进行一定时间的扫描之后，可以剔除命中率低下的规则。使得规则库保持高的命中率。

### 漏洞扫描模块

一个批量PoC自动检测框架（图2），有检测和攻击两部分。主要侧重于检测部分，通过向服务器上自动打大量的PoC文件得到服务器实时反馈来判断是否含有指定的漏洞。优点是速度快，准确性高，具有调用的Web api耦合性低。

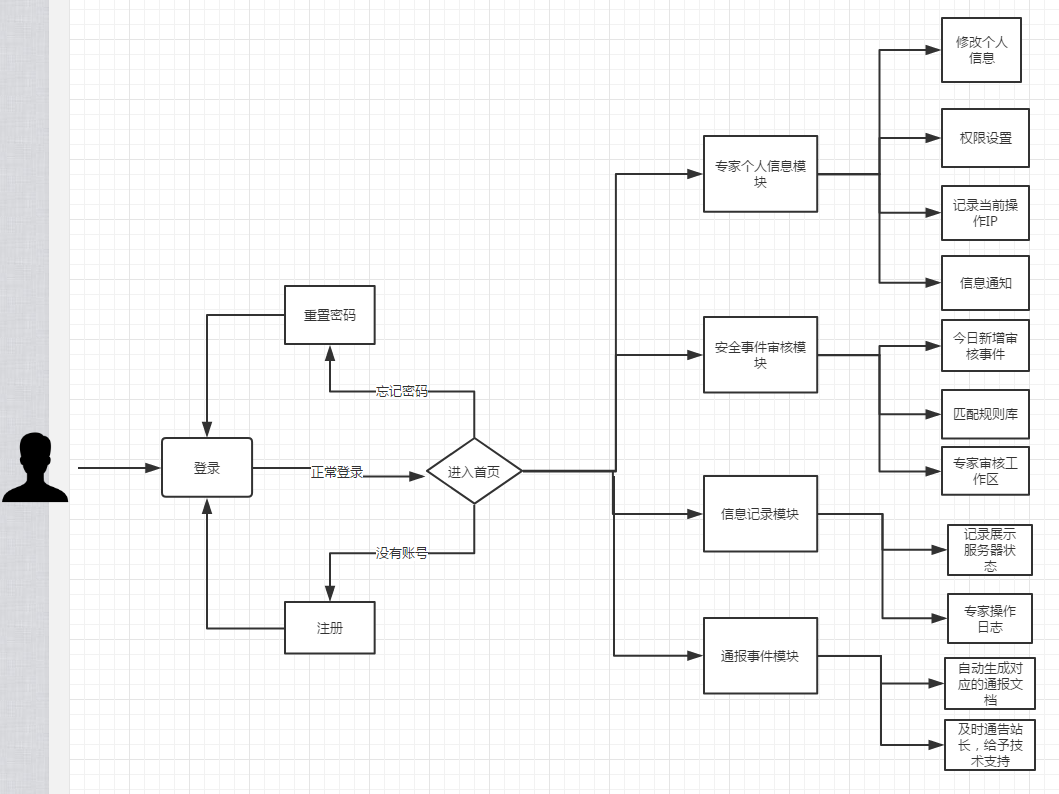
图2 PoC自动检测框架

Web api接口以轻量级框架flask开发，采用celery和RabbitMQ队列来自动调度任务，异步加多进程方式执行worker消费任务。用户操作简单只需通过接口下发任务与获取结果，celery会自动把下发任务放给worker。

## Web模块架构设计

### 系统架构图

专家值守平台（图3）使用Python Flask 框架，配合Bootstrap前端框架实现前后端分离。后端采用Mongodb数据库配合Redis缓存数据库实现数据安全存储以及快速查询。前端采用Ajax实现与后端交互，最大限度优化用户体验。

图3 Web模块架构图

### 平台用户设计

根据平台功能设计，考虑平台运行，平台用户设计及相应权限如表2所示。

|  |  |
| --- | --- |
| 系统用户 | 用户权限设置 |
| 平台管理员 | 平台数据库管理，平台的运行与维护 |
| 高级专家 | 规则库的维护与更新等等权限 |
| 一般专家 | 负责安全事件的审核工作 |

表2 系统用户设计表

### 个人信息模块

此模块主要完成专家个人信息展示、修改，通知功能，主要包括修改用户名、修改密码、记录当前操作IP功能等，同时采用权限设置管理专家账号，采用通知的形式，及时告知专家事件审核情况。

### 个人信息修改

这部分主要完成专家个人信息的修改功能，主要包括专家账号邮箱修改功能、用户名修改功能、密码修改功能、以及修改头像功能。

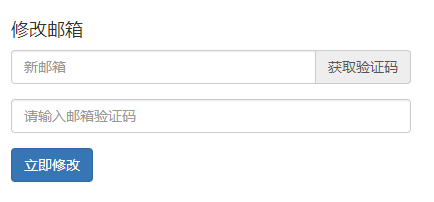
修改邮箱功能主要满足用户可以邮箱的功能，邮箱作为用户数据库的主键，通过邮箱实现识别不同用户，如图4所示。

图4 修改邮箱

修改头像功能主要满足用户可以自定义自己的头像，默认情况下随机使用系统头像。如图5所示。

图5修改头像

修改用户名功能主要满足用户可以修改自己的用户名，用户可以自己设置自己喜欢的用户名。如图6所示。

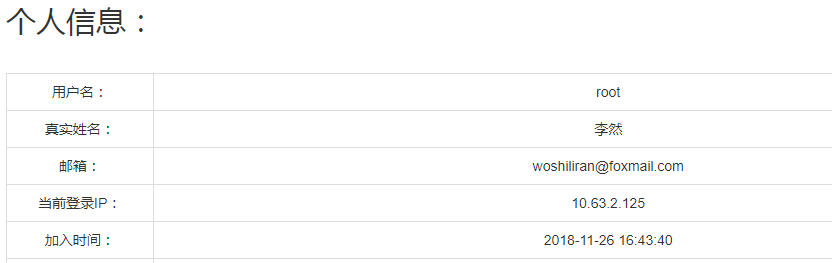
图6 修改用户名

修改密码功能主要满足用户可以在登录的情况下可以重新设置密码，用户需要填写旧密码就可以重新设置新的密码。如图7所示。

图7 修改密码

### 个人信息展示模块

这部分主要展示用户个人信息，以及用户登录、修改账号信息的记录功能。展示用户个人信息部分可以方面用户查看当前登录的信息如：用户名、真实姓名、邮箱、当前登录IP、以及加入时间。如图8所示。

图8 个人信息

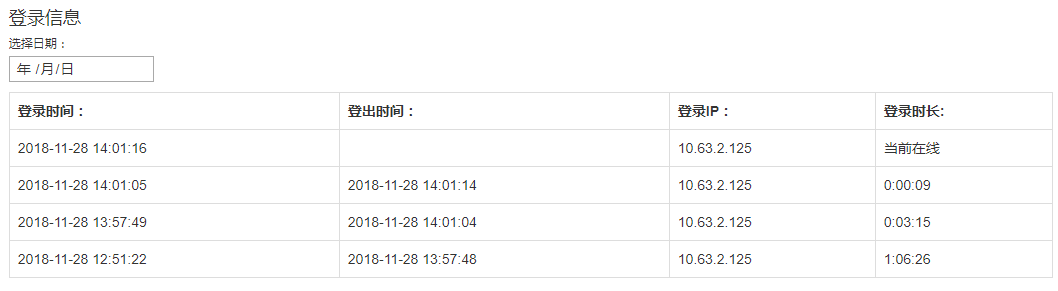
用户登录、修改账号记录功能主要记录用户登录信息，以及用户账号信息登录信息如图9所示，账号信息如图10所示。

图9 登录信息

图10 账号信息

### 安全事件审核模块

此模块供安全专家审核发现的安全事件，并做出专业性修复建议指导。每日新增功能，清晰展示每天新扫描的事件，减少专家查找。事件规则库，安全专家能增加自己发现的规则。共同维护规则库。

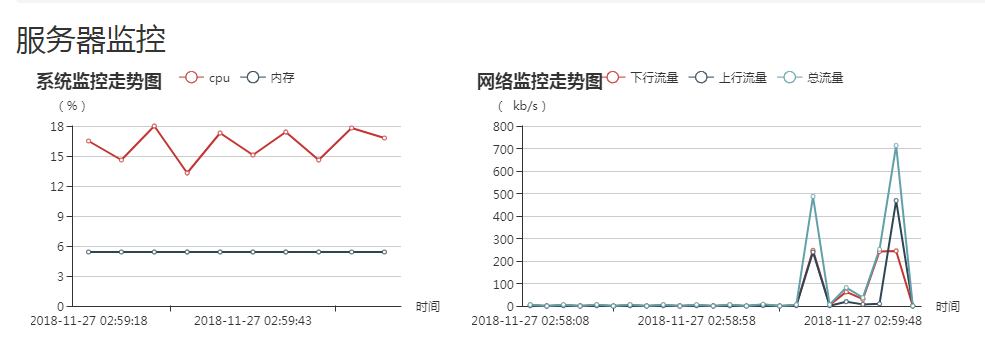
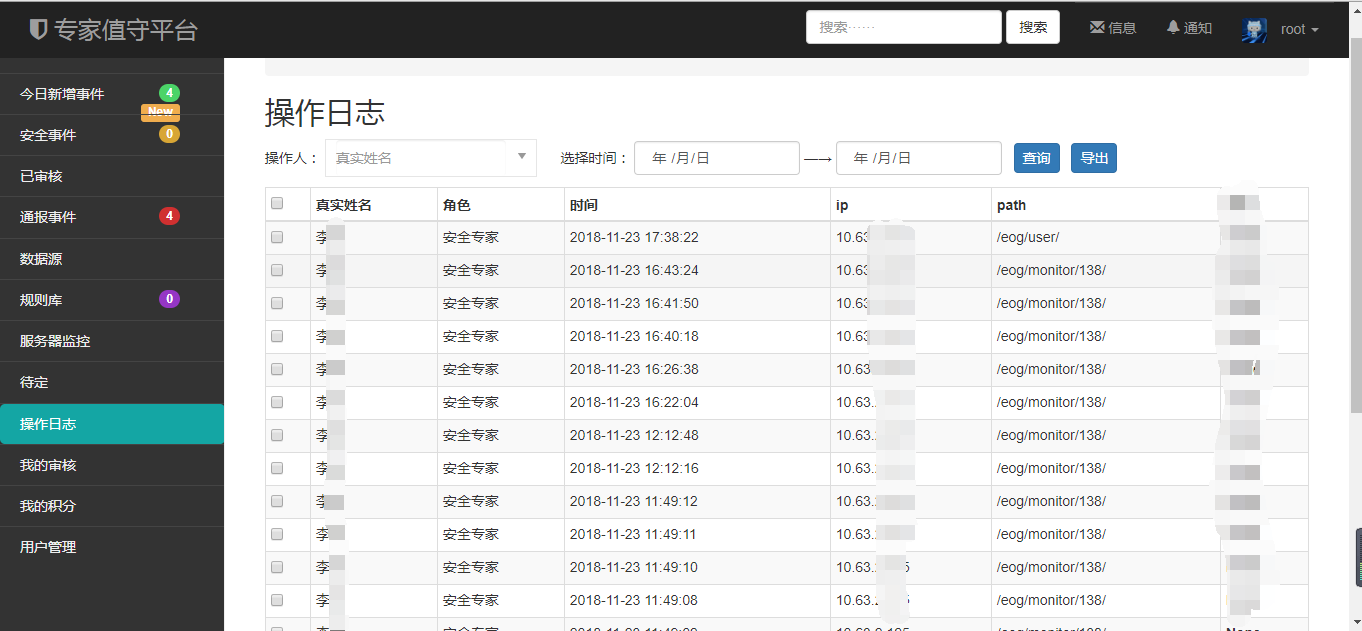
### 信息记录模块

此模块主要完成信息记录工作，包括服务器运行状态记录以及专家操作日志记录两个部分。服务器状态记录采用Api的形式，获取各个服务器的状态信息（包括内存、CPU以及流量状态），通过Socket技术实现与前端低延迟的实时交互。专家操作日志可以实现实时记录专家访问那些后台路由、以及审核事件情况。

### 信息记录模块

此模块主要完成信息记录工作，包括服务器运行状态记录（图11）以及专家操作日志记录（图12）两个部分。服务器状态记录采用Api的形式，获取各个服务器的状态信息（包括内存、CPU以及流量状态），通过Socket技术实现与前端低延迟的实时交互。专家操作日志可以实现实时记录专家访问那些后台路由、以及审核事件情况。

服务器监控模块主要监控运行中的服务器一些运行状态如：内存、CPU以及流量状态，此模块采用Flask\_socketio技术实现与前端的低延迟时效性高的交互，前端采用Chart.js开源框架实现图像的展示。

图11 服务器监控

操作日志模块主要记录专家审核了那些事件以及访问了那些后台重要的路由，方便管理员进行查看。如图12所示

图12 操作日志

### 事件通报模块

此模块为每个确认事件通过模板生成通告，将通过微信或邮件发送给管理员。并每周生成一次总体的周通报，并分析出河南省网站的安全状况趋势。

# 关键技术及创新点

## 关键技术

基于浏览器的网站安全事件监测系统采用Google的chrome内核模拟浏览器访问网站，拦截网站传输数据，对数据进行安全分析。突出网站数据完整性，与后台规则库匹配快速定位事件产生的原因。同时使用chrome内核调用第三方搜索引擎百度、必应、360、Google等的搜索接口。针对搜索引擎下的网站恶意SEO排名产生的劫持事件做特殊关键字搜索。

## 技术创新点

1. 调用Chrome内核拦截数据流量并做分析；
2. 实战中不断完善的规则库；
3. 平台引擎模块架构的设计，数据库设计；
4. 平台展示系统采用Python语言Flask框架运用经典MVC架构设计；
5. Nginx服务器采用负载均衡支持上万并发扫描；
6. 一键Docker弹性部署，节点模块独立运行。易于拓展，智能评估监测，对于被监控网站较少情况下，减少服务器资源；
7. 自动生成监测报告，微信，短信预警通告；

# 前景分析

随着国家相关扶植政策的陆续出台，对信息安全方面支持力度越来越大。社会迅速的进入信息化时代，随之带来巨大的网络安全隐患。对网络攻击的监测及系统漏洞的发现及修复，成为当代网络安全的首要任务。

当前没有任何一款网站安全扫描产品能让网站扫描准确率达到100%，漏报率和误报率都为0%，因此我们只有不断的完善规则库与运用更新的技术运用到扫描器中，提升扫描器的功能，减少漏报率与误报率。

基于浏览器的网站安全检测系统，拥有维护性极其方便的规则库。能在以后不断的完善，增强扫描准确率。基于浏览器的应用方式极大的降低了漏报率。

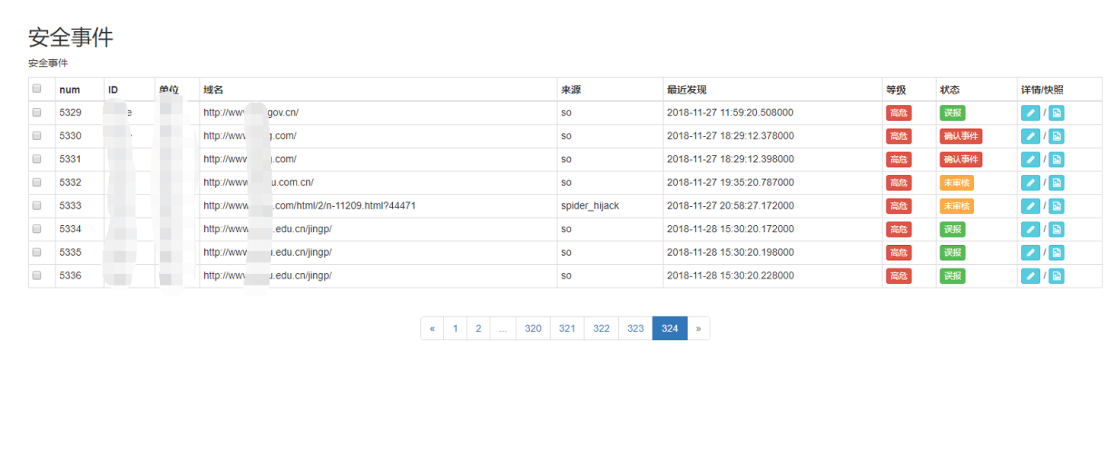
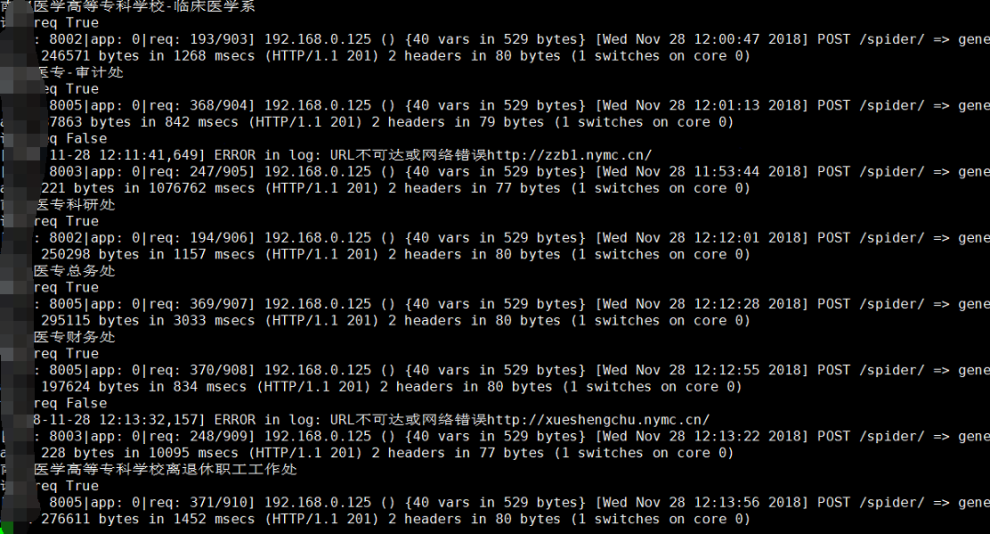
在当今市面上，高度集成的安全扫描系统服务费用是很高的。但许多商业性网站，服务器及一些机构单位对于信息安全也是必要的开销,安全产品亦会带来巨大的经济效益。

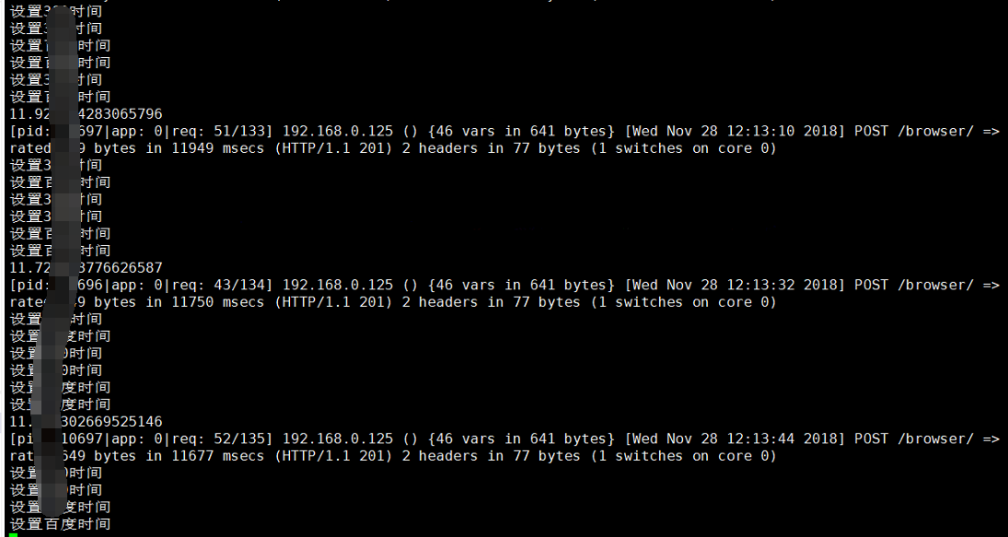
目前已将我们发现被劫持的网站通报给各网站管理员。通过及时修复并有效的降低了网站被劫持带来的负面影响，保证了网站的良好运营。

附件一

##### 扫描成果展示

安全事件截图

安全扫描截图



附件二

##### 安全通报模板

详细见文档

某某单位安全通报.docx