个人简历

侯捷 | (+86) 131-7624-4325 | h.j.13.new@gmail.com

基本信息

- 侯捷 男 22岁 汉族
- 手机 : 13176244325
- Emial : h.j.13.new@gmail.com / z.g.13@163.com
- QQ : 450943084
- 微信 : houjie_13
- 技术博客 : houjie13.com
- Github : h-j-13

实习经历

哈尔滨工业大学 网络与信息安全技术研究中心

2015.10 — 至今 域名体系安全技术研究室 恶意域名组

作为研究室核心成员，主要负责域名安全相关工作，包括域名基础数据获取，恶意域名挖掘，基础数据统计分析；另外，负责研究室数据库的维护、设计与优化等工作。

教育经历

哈尔滨工业大学
本科(大三) - 信息安全

项目经历

非法域名监测系统

2017.9 - 至今
旨在对于非法域名进行检测、数据挖掘与关联分析

个人负责部分及特性 :
系统架构、数据库架构、域名WHOIS数据获取与处理、数据中间层与后台

- 基于水平、垂直拆分的MySQL数据库与同步ElasticSearch服务器作为存储架构
- 基于tornado构建的HTTPS的api数据中间层
- 基于token的身份认证机制
- RESUful 风格api
- 优化的数据库操作及数据缓存机制

系统目前收录及监控了约**2600万**条域名及其相关的WHOIS、WHOWAS及其他相关数据

域名WHOIS数据获取系统

2015.11 - 至今
旨在尽可能多的获取域名及WHOIS信息

个人负责部分及特性 :
基础数据获取, WHOIS数据获取，数据库维护优化，网站后台

- 基于AMQP协议下的RabbitMQ框架通过Python实现了一个高性能、高效的分布式域名WHOIS数据获取引擎
- 通过底层SOCKS代理、核心数据多源轮询更新、基于RFC协议等方式深度优化了WHOIS获取过程，核心代码完全原创，不依赖任何第三方库及第三方网站数据
- 采用动态解析模板，可将形式各异的WHOIS数据解析出统一的特征项。并基于文本分类，编辑距离等算法对地理位置和电话、邮箱等相关字段进行了深度解析
- 采用水平、垂直拆分的MySQL数据库设计，保证了数据的高可用性
- 额外的基于日志的检测系统可以获取系统详细情况、监测系统异常变动并即使发送邮件提醒、以便日常维护与开发

系统获取了超过**1.2亿的域名数据**（超过全球域名总量的1/3）与**1.5亿条域名WHOIS\WHOWAS数据**，覆盖超过1500个顶级域，1800个WHOIS服务器。数据支撑实验室其他相关项目

非法域名挖掘与画像系统

2016.12 - 2017.07
第十届全国大学生信息安全竞赛 作品赛 作品

个人负责部分及特性 :
系统、数据库架构设计，数据获取与分析，部分机器学习与后台

基于域名与页面关联关系的非法域名分析与挖掘系统。旨在通过域名多元信息对域名进行精准画像，进而分析抽象获取非法域名相关特征，使用基于决策树的机器学习模型进行域名性质判定。通过多过程多方式进行恶意域名挖掘。

- 采集与挖掘的域名相关信息包括：域名WHOIS/WHOWAS信息，IP、ICP、地理位置、CDN情况、页面解析情况、页面关联关系。
- 通过WHOIS注册信息、站点间连接关系及关键词搜索进行数据反查与挖掘。
- 通过基尼指数与决策树对域名进行非法可疑值量化与性质判别。

总决赛 **全国三等奖**

开源项目 - WHOISpy

一个纯Python的智能WHOIS客户端

- 覆盖95%的已知顶级域(1444 / 1529)
- 基于RFC 3912的域名WHOIS获取方式,不依赖任何第三方接口
- 自动分析抽取WHOIS数据关键字段
- 高效的数据获取:
  - 自动处理二级WHOIS服务器的情况
  - 支持SOCKS4,SOCKS5 代理
  - 额外使用转发WHOIS服务器提高数据质量
- 基础WHOIS数据分析:
  - 基于电话及邮编的地理位置解析
  - WHOIS数据中不同格式时间数据标准化
  - 验证WHOIS注册邮箱是否真实存在

开源项目 - Racoon

一个基于qt的简单源文本代码编辑器

- 一键注释/自动格式化/代码高亮/撤销与重做/查找并替换

专业技能

熟练掌握Python，有部分并发、分布式(AMQP)、数据库、socket、MySQLdb、Scrapy及pika开发经验。
有丰富的Linux使用经验，熟练掌握git、vim、linux基本命令与配置经验。熟悉shell脚本。
有tornado,flask(Python WEB框架)后台及RESTful风格API开发经验。
熟悉C/C++，了解常见的设计模式、了解cmake，有简单的qt及socket开发经验。
熟悉常用的算法、数据结构。
本科在实验室期间维护优化大型分布式数据库，对MySQL数据库略有理解与心得及较多的实践及优化经历。有ElasticSearch、Redis开发经历。

自我描述

之前主要方向为域名安全，数据处理与分析。对于新技术及开源社区兴趣浓厚。较为善于沟通，有过多次带领团队与主导项目开发的经历。 自学能力较强，解决问题的能力强，技术尚可，追求优雅的代码，踏实肯干，有较好的数学、英语基础。

预期实习时间 3-6 个月，每周可工作 4-5 天，时间比较充足，可以尽快到岗。
期望实习工作：研发工程师（Python） / 研发工程师（C++） / 数据研发实习生