ID	邓召群
Work	百度 3.5 年, 爱奇艺 2 年
Blog	http://secfree.github.io/blog/
Stackoverflow	https://stackoverflow.com/users/4164722/secfree
Email	zzd7zzd@gmail.com
Phone	15801905401
Basic	男, 1989-06, 湖南

技能列表

item	level	comment
大数据处理	擅长	百度 3.5 年主要工作
机器学习	入门	理解常见算法原理, 在一个项目中应用并取得巨大收益
后端开发	擅长	早期的主要工作, 现在的日常工作
安全行业和业务理解	熟悉	5 年工作所在的业务方向
项目管理和团队协作	入门	作为部门内部数据方向负责人一年以上, 管理三人团队半年
Java	擅长	大数据处理工作中主要使用语言
Python	擅长	数据处理和后端开发主要使用语言
Scala	入门	基于它实现过两个小项目
C	入门	学生时代/早期工作 使用语言

工作经历

百度 | **内部安全** | **数据处理+数据分析** | 2015/12 -- 至今

- 职责: 负责内部安全所有相关数据处理和数据分析, 为扫描和威胁监测业务提供支持
- 项目-安全资产库(负责): 解析公司 PB 级别的 Web 日志得到 URL, 基于机器学习进行 URL 相似去重, 使百度 Web 扫描业务扫描成本有数量级的下降.
  (*Detail: 百度主要以 Web 的方式提供服务, Web 可以是认为受攻击最广的点. 攻击的入口是 URL, 同样, 自己做扫描的入口也是 URL. 我们通过解析 accesslog 得到一个基础的 URL 集, 但百度每天的 accesslog 是 PB 级别, 同时扫描一个 URL 的成本很大, 因此对相似的 URL 做去重就很有必要. 我用 machine learning 的方式, 找了一个比较合适的切入点, 取得了很好的效果, 使需要扫描的 URL 有数量级的下降.*)
- 项目-用户数据使用行为审计系统(参与): 个人实现对离职用户的异常下载代码行为进行预警的功能, 是目前公司内对代码安全保障的重要一部分.
  (*More: 技术实现比较简单, 但重点在于建立一套闭环的流程.*)
- 项目-业务访问日志备份(参与): 推动全公司对外业务备份 accesslog 六个月, 以符合<<网络安全法>>要求. 个人负责方案制定和评审, 流程完善以及参与推动.
- 项目-数据治理(负责): 梳理和维护超过 20 种安全相关的数据流, 存储并提供检索.

百度 | **商业安全部** | **数据研发+后端研发** | 2014/06 -- 2015/12

- 产品-云加速(参与): 主要业务是为站点做加速和安全防护, 个人负责云加速部分日志的 ETL 以及大数据平台的运营.
- 产品-云观测(参与): 基本原理是用分布在全国各地的节点来探测网站的速度. 个人主要负责后端研发, 做数据和任务处理.
- 工具-`distcp-ex`(负责): 实现百度版 HDFS 到社区版 HDFS 的 distcp 功能.

爱奇艺 | **网络安全工程师** | **后端开发** | 2012/06 -- 2014/06

- (负责) 研发 Unix Shell 行为审计系统
- (负责) 研发基于 FreeRADIUS, MOTP 的双因素认证系统
- (负责) 开发 webshell 检测程序 (实时 + 非实时)
- (负责) 开发基于 Qt 的 web 漏洞扫描器

教育背景

- 2008/09-- 2012/06 | 天津商业大学 | 计算机科学与技术 | 本科 | 英语六级