戚名钰（男，1991.07.21）

+86 **156-9728-7888** **qimingyu.security@foxmail.com**

求职目标：安全数据分析师 **个人博客：www.qimingyu.com**

教育背景：（985+211）

硕士：中国科学技术大学

专业：信息安全 学位：工学硕士 排名：5/48 2014.9—2017.6

本科：广西大学

专业：信息安全 学位：工学学士 排名：3/46 2010.9—2014.6

辅修：市场营销 学位：管理学学士 排名：1/24 2011.9—2014.3

工作经历：（百度+苏州研究院）

百度 | 安全事业部 | X-Lab | 威胁情报项目组 **安全数据分析师** 2016.3—至今

- 搭建由 50 台服务器集群构成的大数据机器学习平台，系统整体架构为：HDFS+Yarn+Spark+H2O，实现了平台的访问权限控制、数据可视化等功能，并负责该平台的运维工作。
- 在该大数据平台上，基于 MLlib 与 Sparing Water，针对移动终端（安全 SDK）每天上报到服务器集群的 Log 数据，用 Hive/Spark SQL 进行提取，做例行化的统计分析，并对其进行数据清洗后（ETL），采用 GDBT 和随机森林进行建模，根据模型对其进行黑白预测，已达到 84%的正确率。针对每天上报的网络端数据（云 WAF 拦截、SRC 数据等）进行关联分析及处理后，例行化入库到 ElasticSearch 和 MongoDB 中，同时采用机器学习建立相应的模型做自动化判断。
- 搭建由 6 台服务器构成的 Redis 集群，采用 3 主 3 从的模式，对外提供查询服务，并负责其运维工作。
- 设计并完成威胁情报平台 SaaS 的第二代 API 接口。把两个 MongoDB 及 HDFS 上面的数据（格式不同），通过 Spark-Redis 处理后，以 HSET 方式在 Redis 集群中做全量 Cache，使每秒查询率提高 70 倍以上。
- 针对百度每天的全流量请求 Log 日志（每天 500G）进行关联分析与建模，采用多维度自动化机器学习分析，甄别出恶意请求并输出给 WAF 防火墙进行拦截。

中科大苏州研究院 | 安全实验室 | Web 安全组 **渗透测试员（实习）** 2015.10-2015.12

- 根据《OWASP 测试指南 4.0》，在 HP 的 QC 系统里面按照客户要求的 Case 对 Web 应用做逐条的安全测试
- 学习并完成了 WebGoat5.2 教程的所有共 20 多个常见网络攻击实验，并通关 Try2Hack（见博客）

项目经验：（4 个）

全国智能设计竞赛 | 基于 WIFI 的手机室内定位系统项目 **组长**（共 3 人） 2014.9-2015.8

- 该项目现已实际部署应用于苏州市博物馆，并获第五届华为杯智能设计竞赛全国一等奖，项目简介见：https://github.com/qimingyu/Indoor-Positioning
- 负责定位算法输入数据预处理模块。针对信号强度的不稳定，采用了多种平滑机制（插值、去噪等），并通过指纹比对和三点定位机制，最终将定位误差控制在 5 米范围内
- 负责检测手机基准信号强度，并在信号强度与距离之间根据衰减公式做拟合模型，消除了不同手机品牌之间基准信号强度对定位精度的影响

全国信安竞赛 | 基于 SVM 的网络入侵检测系统项目 **组长**（共 3 人） 2013.9—2014.6

- 该参赛项目获第七届信安竞赛全国三等奖，详细介绍见：https://github.com/qimingyu/IDS
- 开发并完成小型网络异常检测器，改进了支持向量机的训练及参数确定算法，使样本检测正确率提升至 90%以上
- 在 Matlab 中用 KDD99 标准数据集对异常检测器进行性能测试，结果显示误报漏报率均小于 5%
- 在此基础上进一步研究了基于 SVM 的多分类集成学习模型，根据样本分布规律，采用哈夫曼编码的方式搭建多分类检测器，最终有效检测出了 4 种网络攻击流量异常类型

解放军总参三部 | 隐匿信道通信的检测与实现项目 **组员**（共 7 人） 2012.7—2012.12

- 国防“十三五”规划立项项目，项目总金额为 200 万，属实验室导师项目
- 负责研究 TCP-IP 传输协议，以及在各个不同字段（IPID、TTL 等）插值后，对存储型隐信道的响应及效果评估
- 负责实施模拟隐信道检测试验，在隐信道的稳定性、可靠性、抗检测性、容量四个维度上对其进行评估

毕业设计 | 基于差分隐私保护的数据发布技术研究与系统实现　　独自完成　　2016.9—2017.6

- 提出了两个基于差分隐私的数据发布算法 PCA-based-PPDR 和 LDA-based-PPDR，对敏感数据集进行脱敏处理后，使发布的数据能在保护个人隐私的前提下，同时具有良好的可用性。
- 设计并实现了一套大数据环境下基于差分隐私保护的数据发布系统，整体架构分为两个方面：Web 应用和 Spark 云平台。Web 应用采用 Django 框架，搭配 Gunicorn 高性能服务器和 Nginx 反向代理服务器，使系统具有较高的并发量和数据吞吐量。

专业技能：（安全基础+大数据）

- 熟悉并能熟练使用 Spark、Hadoop、MongoDB、Redis、Hive、Elasticsearch 等大数据平台及其配套数据库
- 熟悉 SVM、决策树、LDA、神经网络等各大机器学习算法，能熟练使用第三方机器学习库（MLlib、H2O）函数
- 熟悉 XSS、CSRF、SQL 注入等一般 OWASP 包含的 Web 安全攻击与防御技术
- 熟练使用 FireBug、Burpsuite、WebInspect 等渗透测试工具

科研成果：（机器学习相关）

[1]《Privacy-preserving Naive Bayes Classification》, The 8th International Conference on Knowledge Science, Engineering and Management (KSEM2015,CCF C 类会议),（第四作者）

[2]《基于 PCA 的 SVM 入侵检测研究》,《信息网络安全》2015 年第 2 期,（第一作者）

[3]《采用成分分析的差分隐私数据发布算法》,《小型微型计算机系统》已录用，编号：20160015,（第一作者）

[4] 发明专利“一种基于支持向量机的 P2P 网络贷款风险评估模型”，编号：201410801984.3,（第一作者）

获奖情况：（国家级+省级）

- 第五届华为杯全国大学生智能设计竞赛国家级一等奖　　2015.8
- 第七届全国大学生信息安全竞赛国家级三等奖　　2014.7
- 第四届蓝桥杯 JAVA 编程大赛广西区一等奖　　2013.5
- 2012 年全国大学生数学建模竞赛广西区二等奖　　2012.9