基于机器学习的SQL注入检测系统的研究与实现

计算机与软件学院软件工程专业　洪楚唐

学号：2014150192

【摘要】SQL注入攻击是主要的网络攻击技术。目前为止，对SQL注入攻击的主要防御手段存在诸如过度依赖后台、开发周期长、部署难度高等问题。因此，本文提出一种工作在HTTP应用层，不依赖后台，仅需通过分析用户输入就能检测到潜在威胁的SQL注入检测技术。本文提出的SQL注入检测系统通过对用户输入进行预处理、词法分析、机器学习分析等步骤进行检测，进而执行不同的决策手段并最终输出威胁级别。该SQL检测系统的设计特点为：仅需拦截和提取用户输入，无需依赖其他后台信息；通过预处理和词法分析提取payload，并转化特征向量；通过对三种机器学习方法的训练和对比，筛选出表现最佳的机器学习模型。实验表明，相比于其他的SQL注入检测模型，该模型在仅需要采集用户输入的情况下具有较高的准确度和检测效率，能够有效地检测和防御SQL注入攻击。

【关键词】SQL注入；机器学习；朴素贝叶斯；Scikit-Learn