# MUTI研究工作进展与规划

目录

[一．工作概述 1](#_Toc4380)

[1.蓝盾数据中心 1](#_Toc14272)

[(1)系统基础架构 1](#_Toc3030)

[(2) 数据中心自动化 1](#_Toc3471)

[(3) 为合作方软件基因团队提供病毒样本 2](#_Toc2917)

[(4) 数据中心的维护以及升级 2](#_Toc21009)

[2. 公安部重点实验申请书 2](#_Toc8690)

[(1)写申请材料以申请公安部资金，“软件脆弱性智能分析技术” 2](#_Toc22864)

[二．未来规划：软件脆弱性智能分析技术 2](#_Toc17954)

[1、基于深度学习的漏洞检测系统 2](#_Toc28441)

[目标 O 2](#_Toc28002)

[时间范围 2](#_Toc28843)

[关键成果 KR 2](#_Toc28499)

[简述 3](#_Toc1056)

# 一．工作概述

### 1.蓝盾数据中心

(1)系统基础架构

* 构建底层高效的大数据体系架构。目前为止,我们数据中心系统计算力(CPU 42核，内存106G，存储34T),对于分布式数据存储与容错需求，我们选择了分布式存储系统Hadoop的HDFS（http://172.16.7.46:50070,用于数据交换，视频共享）,对于大数据的计算与分析需求，我们选择专为大规模数据处理而设计的快速通用的分布式计算引擎Spark（http://172.16.7.46:8080，master为46,worker有13台)

1. 数据中心自动化

* 在此基础架构中编写大量执行脚本，以此来增加大数据体系的自动化程度,以下几点：

1. 爬虫:周期性的从各种数据源中数据源中收集（如病毒样本，病毒分析报告以及程序源码等）
2. 大规模数据分析，对数据样本作交互性分析，使用Spark/Python Notebook(http://172.16.7.46:9001)

* 目前在backblaze云端数据有千万级别，蓝盾数据中心总数据量已达到4690294,日下载量最高能达到十万以上（日下载量可以通过调节进程数以及受公司网速影响）

1. 为合作方软件基因团队提供病毒样本

* 基于以上基础，我们每周为其提供病毒样本（提供对方指定的Md5值的病毒样本进行深度学习模型试验）

1. 数据中心的维护以及升级

### 公安部重点实验申请书

(1)写申请材料以申请公安部资金，“软件脆弱性智能分析技术”

二．未来规划：软件脆弱性智能分析技术

## 1、基于深度学习的漏洞检测系统

### 目标 O

* 构建基于深度学习的漏洞检测系统

### 时间范围

* 2018

### 关键成果 KR

* 深度学习方向1位研究员，完成以下目标：
  + 基于深度学习的智能漏洞检测系统原型
  + 实施5种技术比较（Checkmarx, Flawfinder, RATS, VUDDY and VulPecker）
  + 完成每周一篇论文以及一篇阅读笔记

### 简述

软件脆弱性智能分析技术是目前重要研究课题。但是，目前解决方案主要是依靠人工自定义软件脆弱性特征，导致经常遗留下大量软件漏洞（例如出现高误报率）。

大量网络攻击来源于软件漏洞，尽管在安全编程技术方面做了努力，但是软件漏洞仍然遗留，并将继续成为一个标志性的问题。这可以通过以下事实来证明：2010年通用漏洞披露（CVE）中登记的漏洞数量约为4,600个，2016年增长至约6,500个。

有一种替代的方法是在软件程序中（或简称程序）智能检测漏洞。目前已经存在有大量的静态检测漏洞的系统以及学术研究，范围是从开源工具到商业工具，再到学术研究的项目。然而，目前漏洞检测的解决方案有两大缺点：消耗人工劳动，以及出现高漏报率。

所以我们需要将AI与漏洞检测结合，用基于深度学习的漏洞检测来减轻人工繁杂的自定义特征的工作。