安全小课堂第五十期【分布式扫描器的实现思路 】

京东安全应急响应中心 2017-03-26

作为一个安全从业人员，在平常的工作中会用到一些web扫描器对一些web系统做一些安全扫描和漏洞检测，从而确保在系统上线前尽可能多的解决了已知的安全问题，更好地保护我们的系统免受外部的入侵和攻击。而传统的web安全检测和扫描大多基于web扫描器，但在传统扫描器中由于爬虫无法获取到一些隐藏地很深的系统资源容易导致漏报。现在流行的分布式web扫描器或许能web扫描器的这一缺点弥补，JSRC **安全小课堂第五十期**，我们邀请到了**李长歌**、**安全小飞侠**师傅分享一下他们在分布式扫描器中的实现思路及方法。 以及 JSRC 白帽子花开，沦沦，feng、 胖猴粉、1ce、ziwen、虾米的讨论。

讲师：李长歌

讲师简介：

唯品会高级信息安全工程师，多年的安全测试、安全开发经验,擅长于代码审计和漏洞分析,目前关注于漏洞检测及企业级安全建设，shadow7团队核心成员。

讲师：安全小飞侠

讲师简介：

安全研究员，代码狂热分子，白帽子,专注安全开发与渗透测试。

分布式扫描器能实现对什么类型的漏洞进行检测?

京安小妹

分布式扫描器适用于常见漏洞的扫描，其对漏洞类型没有要求,只要是通过发送数据包来验证漏洞是否存在的都可以，分布式只是将原来的一台机器扫描改成了多台同时扫描而已。
扫描器能扫什么，它就能扫描什么，分布式不过是一种扫描的处理模式而已。以web扫描器为例，基本上来说凡是具有请求应答的漏洞都可以检测，所谓的请求应答就是说对于发送的请求会有直接或者间接response，比如 XSS，sql注入，文件包含，远程代码/命令执行等等。

讲师：李长歌、安全小飞侠

实现一个分布式扫描器的思路是什么样的?

京安小妹

首先需要有爬虫程序，如果没有爬虫也可以使用流量镜像或proxy代理方式获取扫描链接存入数据库，然后要有一个可以用的扫描程序,将这个扫描程序分别部署几个节点，再此的基础上新增一个mastet节点，其作用是任务调度和规则下发模块，mastet节点通过读取数据库里面的待扫描链接，然后分别下发任务到各个节点进行扫描，扫描的结果再分别存储到数据库中。

讲师：李长歌、安全小飞侠

分布式扫描器必备模块有哪些？哪一个模块是最重要的？

京安小妹

中央存储数据库、中央任务和资源调度系统、请求收集模块（如爬虫和代理），扫描引擎以及web管理控制端。其中，中央任务和资源调度系统最为重要，因为这是负责控制各个子模块的中枢系统，就像人的大脑一样。

必备的模块我觉得应该要有：爬虫模块，数据解析模块，数据存储模块，漏洞扫描模块，任务调度模块，这些模块都挺重要的，每个模块都有其用途，相互结合才能完成好整个扫描过程。

讲师：李长歌、安全小飞侠

白帽子提问:不知道你们架构是什么样的?

爬虫模块可以是py的requests，phantomjs，burp scanner;
代理模块可以是burp，mitmproxy;
数据存储模块可以是mysql，mangodb，sqlite都可以;
web管理模块可以是py的flask等等。

讲师：李长歌、安全小飞侠

白帽子提问:能否推荐几个分布式扫描器?

我个人写了一个成品的叫proxyscanner

讲师：李长歌、安全小飞侠

在实现分布式扫描器的过程中遇到过什么问题？如何解决的?

京安小妹

我个人觉得最复杂也是最棘手的问题应该就是中央任务和资源调度系统的设计，如何让各个子扫描系统彼此分离且不重复扫描，如何合理的分配扫描任务给多个字扫描系统。
这里我的解决方法是，设置一个主队列存放所有搜集到的请求，然后让各个子扫描系统各自设置自己的从队列来接收来自主队列的扫描请求和任务，同时让字扫描节点定期回传自己的队列状态至中央任务和资源调度系统，在由其判断存在哪些空闲的扫描节点并分配更多的扫描任务。
比方说爬虫获取的连接过少问题，这个模块可能影漏洞的检测率，前期可以多部署几个爬虫来扫描链接，也可以和其他的爬虫进行对比，然后进行逐步改进。

讲师：李长歌、安全小飞侠

白帽子提问:怎么解决点击交互产生的请求?

交互产生的请求可以模拟搜索页面中的click，button之类的使用类似于phantomjs之类的无界面浏览器引擎去解析页面中的js。
流量镜像或者proxy代理也是一种有效的获取连接的方式。

讲师：李长歌、安全小飞侠

白帽子提问:爬虫爬多深，这个是怎么设计的?

一般设可以定一个默认的深度3-5层目录。
我个人不太喜欢爬虫的检测方式，我比较倾向于代理镜像的方式。

讲师：李长歌、安全小飞侠

分布式扫描器的优点?

京安小妹

分布式扫描器的优点自然是多节点多任务同时处理和扫描，大大地节省了扫描时间和周期。

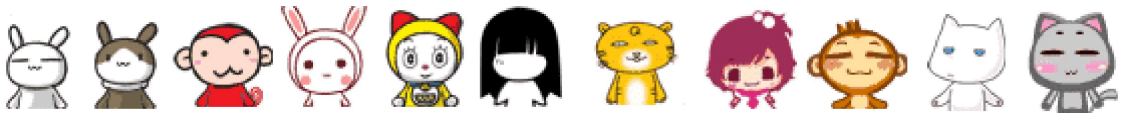
使用分布式扫描的主要优点是其降低了服务器负载和节省扫描时间，比如说之前对全网域名做一次漏洞扫描需要10个小时，使用分布式扫描加入10多个节点同时进行时，扫描时间减少到1个小时。这对于范围性的进行漏洞验证效果还是挺显著的，而爬虫分布式部署也可以尽可能多的获取到连接，这对后面的漏洞检出有帮助。

讲师：李长歌、安全小飞侠

白帽子提问:你们用的什么做定时任务?

定时任务可以直接在Linux下面做crontab。
其实简单的写个while循环都行。

讲师：李长歌、安全小飞侠



本期JSRC 安全小课堂到此结束。更多内容请期待下期安全小课堂如果还有你希望出现在安全小课堂内容暂时未出现，也欢迎留言告诉我们。

安全小课堂的往期内容开通了自助查询，回复“安全小课堂”或者点击阅读原文进行查看。

最后，广告时间，京东安全招人，安全开发、运营、风控、安全研究等多个职位虚位以待，招聘内容具体信息请扫描二维码了解。

简历请发送：cv-security@jd.com

微信公众号：jsrc_team

新浪官方微博：京东安全应急响应中心