背景&前景&是啥

当今互联网迅速发展，每个人对网站都有自己的需求，但并不是所有人都是网站开发的工程师，特别是在信息安全的门槛越来越低，各种傻瓜式黑客工具曾出不穷的今天，想要无经验零基础开发一个安全的网站可以说已经不可能了。所以各种开源建站系统，内容管理系统，建站框架渐渐火了起来。但是这些开源项目的安全性不容乐观，比如前一阵子的 strutsII 漏洞使各大平台和高校网站被入侵。

当前渗透测试人员对网站的 web 应用进行渗透测试时，有一种常用的方法——对网站应用进行 web 指纹识别，之后针对该指纹去搜索相应的漏洞详情，之后编写利用程序，若是对目标站渗透无果，还可能需要对 C 段或者旁站进行测试，这样上百个站点进行这样繁琐的手工测试是不可能实现的。

如今各大漏洞库的建立为我们提供了丰富的漏洞资源。国内有乌云网 sebug，国际上有 CVE 和 exploit-db。我们可以从这些网站和各大安全论坛获得漏洞资源，编写利用程序。

实现自动化渗透测试将大大减轻工程师的工作量，并且，漏洞的类型是不断更新的，这就要求自动化渗透软件应是可扩展的。所以我提出这样一个项目——可扩展的、针对 web 漏洞的自动化渗透测试平台。

集成了 web 指纹识别、漏洞利用、漏洞扫描、目标收集的自动化渗透测试平台。我们的目的是致力于完成国内第一款面向大众的、可扩展的、针对 web 漏洞的自动化渗透测试平台，让渗透测试更加简单，更加大众，更加方便。最大的亮点是我们提供了一系列接口和开发包来开发该平台的插件，这是一个 web 平台，是一个插件式的、模块化的平台。它的强大并不是最初的开发人员赋予的，我们提供了一个平台，一个插件式开发的框架，它真正的强大是由它的插件（模块）的数量和质量决定的。

该平台将各种功能有机的结合成一整套渗透测试流程，实现了渗透测试的自动化。其 web 平台的性质和 Python 的特性还有我们提供的开发包和接口将其可扩展性达到极致，更新迅速功能强大。

所以，我们从 web 建站系统和框架入手，研究一套能快速构造 exp 和 poc 的利用框架。

并且，增强探测模块。扩大探测漏洞的能力。

技术概述

把基本功能给建起来，先比如说一些辅助功能，C 段、旁站、搜索引擎的数

据采集。之后是一些基础功能，web 指纹识别，服务协议识别。然后是测试功能，这一块主要是渗透测试的功能，我觉得在一些方面没必要再创新自己开发，比如 SQL 注入和 XSS，完全可以是 sqlmap 和 Xsser 这些经典工具的复写。当然，检测功能要分为多个版本，比如 sql 注入我觉得应该根据 sqlmap 做一个简化检测和一个深度检测，因为要考虑到效率问题。

在测试功能中，最核心的应该是插件模块，web 指纹识别插件、扫描插件、漏洞利用插件，我们需要提供一系列简洁的接口、强大而方便的开发包、高效的插件载入程序架构。

我认为这应该是个云平台——因为我们要做成一个单机版的软件的话，这样不利于广大安全人员对插件的开发，因为要是做成一个单机版的话，大家写了插件只会自己用，并且插件都在客户端，很容易被破解。软后，考虑到一些漏洞类型的检测，云平台有绝对的优势。所以我建议建成一个云平台，卖服务，并建立开发者社区和奖励机制。

云平台的话，不可能就一台服务器计算和测试流量和计算资源都跑不动，必须做成分布式的，

建议把每一个接入扫描的用户都作为一个扫描的节点。

Web2.0 黑客的战场在 Web 前端（实际上就是浏览器层面上），Web2.0 黑客的攻击往往是被动式攻击，黑客得准备好陷阱，诱惑目标踏入，而不是拿起「刀斧」直接攻击。这样说似乎很抽象，按照我的理解，XSS 就是这种攻击手法。目前各种扫描器对储存型 XSS 的检测都不尽人意，这就是测试的时间等待问题，云平台在处理这种问题上有绝对的优势。

首先，web 漏洞主要还是 http 协议的基础上来说的，那么 http，准确来说是 html、css 和 javascript 是以浏览器呈现的，如果说自动化渗透测试的话我认为要实现一个浏览器内核。这个可以使用 chrome 的内核。

加入文本挖掘技术，主要是判断返回信息的价值并进行有效的抓取。比如我注入报错爆出了物理路径，这个就要能自动抓出来，再比如我上传绕过了，之后返回了 shell 地址，这个也要自动抓出来。这些都是简单的例子，往复杂了说，比如我要挖一个 XSS，我就要从 JS 和数据包中分析出我们可控并有效的部分。

主要模块

插件式，模块化扩展

爬虫

Skadi tools

目标搜集：集成 Google、baidu、zoomeye、bing 搜索引擎，特定 URL 网站搜集、特定 cms 网站搜集、C 段、旁站、二级域名查询
cms 识别：构造特定 url，判断 http 响应；爬行页面链接，匹配特征 URL；web 指纹识别插件
漏洞扫描
漏洞利用：低配置漏洞利用；高配置漏洞利用

开发包，开发框架

总体进展

（1）平台 web 接口，即平台网站前端完成，后端基本完成。
（2）核心渗透部分，已发布第一版开发包，分为 Java 开发包和 jython 开发包两种。
（3）核心渗透部分已经完成大部分基本功能，sql 注入，反射型 xss 等基础扫描模块已经开发完毕，搜索引擎目标收集，zoomeye 目标收集，C 段查询，同服查询，等目标收集模块已经开发完毕，web 指纹识别模块已经集成了数十种建站系统和框架的指纹并有三套识别算法。
（4）扩展接口基本完成，通过反射机制调用扩展攻击模块。
（5）针对 web 建站系统和内容管理系统的攻击模块已经编写十几个。

已完成的研究工作内容以及取得的成果

（1）平台 web 接口，即平台网站前端完成，后端基本完成。我们在前端使用了 Bootstrap，jq 等框架来进行前端开发，后端使用 php 语言，并使用 thinkPHP 框架进行建站，网站主要分为用户登录注册，攻击模块、插件、字典的上传和公开，新建工程，新建策略，管理工程查看渗透测试成果等几大功能。
（2）核心渗透部分，已发布第一版开发包，分为 Java 开发包和 jython 开发包两种。开发包
（3）扩展接口基本完成，通过反射机制调用扩展攻击模块。扩展接口和开发包向联系，为平台提供强大的可扩展性：
任务调度:
task_push(servie, arg, uuid=None, target=None), 新增加一个服务，后面是参数,比如
task_push('www', 'http://www.baidu.com/')
调度器会传递所有插件的 audit 函数('www', 'http://www.baidu.com/')这样的参数
如果指定 UUID，UUID 将做为此任务的唯一标识，防止重复，如果不指定，系统将自动生成一个 UUID
如果指定 target 参数，新生的任务产生的报告所属的域名为 target 指定的值

报告函数:
通知: security_note(str)
提示: security_info(str)
警告: security_warning(str)
漏洞: security_hole(str)

util:
is_ipaddr(str) 字串是否一个 IP 地址
decode_html(head, body) 根据 HTTP 头和内容进行编码转换成 utf-8 编码
urljoin(base, ref) URL 组合，比如 urljoin('http://www.baidu.com/', 'abc/../dd.html') 将返回 http://www.baidu.com/dd.html
html2text(body, head='') HTML 转换为 txt, 如果指定 head, 尝试用 decode_html 解码后转换
get_url_host(url) 得到 URL 域名。
get_domain_root(url) 得到一个 URL 或者域名的根域名(内置 TLD 字典)
str_ration(str1, str2) 比较两个文本的相似度，会根据长度自动选择最快匹配算法
curl:
参考 curl 的一个纯 python 迷你版本，只支持 HTTP 一些协议，可代替 urllib，

（4）核心渗透部分已经完成大部分基本功能，sql 注入，反射型 xss 等基础扫描模块已经开发完毕，搜索引擎目标收集，zoomeye 目标收集，C 段查询，同服查询，等目标收集模块已经开发完毕，web 指纹识别模块已经集成了数十种建站系统和框架的指纹并有三套识别算法。
对于目标收集，我们编写了一个智能爬虫，使用 jsoup html 解析框架，其中涵盖了百度，Google，zoomeye，bing 的网站结构和请求结构，可以快速解析并抓取数据，其中为了绕过 Google 的防爬虫机制，我们伪造了 user agent 消息头，实现了绕过。
C 段和同服查询我们采用了 bing 的 IP 搜索，ip 搜索引擎需要一个微软 key 的支持。
Web 指纹识别有如下三种算法：
1：网页中发现关键字
2：特定文件的 MD5（主要是静态文件、不一定要是 MD5）
3：指定 URL 的关键字
我们搜集了数十种主流 cms 和建站系统的指纹特征，并对请求和计算进行了优化，其中准确度较高耗时较少的是关键路径识别，也就是特征 URL 识别。精度最高的是静态 MD5 比对，但是收集特征值和扫描时相对较困难。
（5）针对 web 建站系统和内容管理系统的攻击模块已经编写十几个。编写针对建站系统和内容管理系统的攻击模块时，我们主要是从各大漏洞发布机构和各大安全论坛获取漏洞详情或者相应的 exp，之后使用我们的开发包和接口进行修改，使之与平台耦合，目前我们已经编写了针对 dedecms，discuz!，Kesion 等主流系统的多个 exp 模块，可以实现获得网站管理员密码，或者直接得到 webshell 等效果。