首先要说明的是"什么是 Google hack？"

前言

googlehacking 其实并算不上什么新东西,在早几年我在一些国外站点上就看见过相关的介绍,但是由于当时并没有重视这种技术,认为最多就只是用来找找未改名的 mdb 或者别人留下的 webshell 什么的,并无太大实际用途.但是前段时间仔细啃了些资料才猛然发觉 googlehacking 其实并非如此简单...

googlehacking 的简单实现

记得以前看见过一篇文章写的就是简单的通过用 www.google.com 来搜索 dvbbs6.mdb 或 conn.inc 来获得一些站点的敏感信息.其实使用 google 中的一些语法可以提供给我们更多的信息(当然也提供给那些习惯攻击的人更多他们所想要的.),下面就来介绍一些常用的语法.

intext:

这个就是把网页中的正文内容中的某个字符做为搜索条件.例如在 google 里输入:intext:动网.将返回所有在网页正文部分包含"动网"的网页.allintext:使用方法和 intext 类似.

intitle:

和上面那个 intext 差不多,搜索网页标题中是否有我们所要找的字符.例如搜索:intitle:安全天使.将返回所有网页标题中包含"安全天使"的网页.同理 allintitle:也同 intitle 类似.

cache:

搜索 google 里关于某些内容的缓存,有时候也许能找到一些好东西哦.

define:

搜索某个词语的定义,搜索:define:hacker,将返回关于 hacker 的定义.

filetype:

这个我要重点推荐一下,无论是撒网式攻击还是我们后面要说的对特定目标进行信息收集都需要用到这个.搜索指定类型的文件.例如输入:filetype:doc.将返回所有以 doc 结尾的文件 URL.当然如果你找.bak、.mdb 或.inc 也是可以的,获得的信息也许会更丰富:)

info:

查找指定站点的一些基本信息.

inurl:

搜索我们指定的字符是否存在于 URL 中.例如输入:inurl:admin,将返回 N 个类似于这样的连接:http://www.xxx.com/xxx/admin,用来找管理员登陆的 URL 不错.allinurl 也同 inurl 类似,可指定多个字符.

link:
例如搜索:inurl:www.4ngel.net 可以返回所有和 www.4ngel.net 做了链接的 URL.

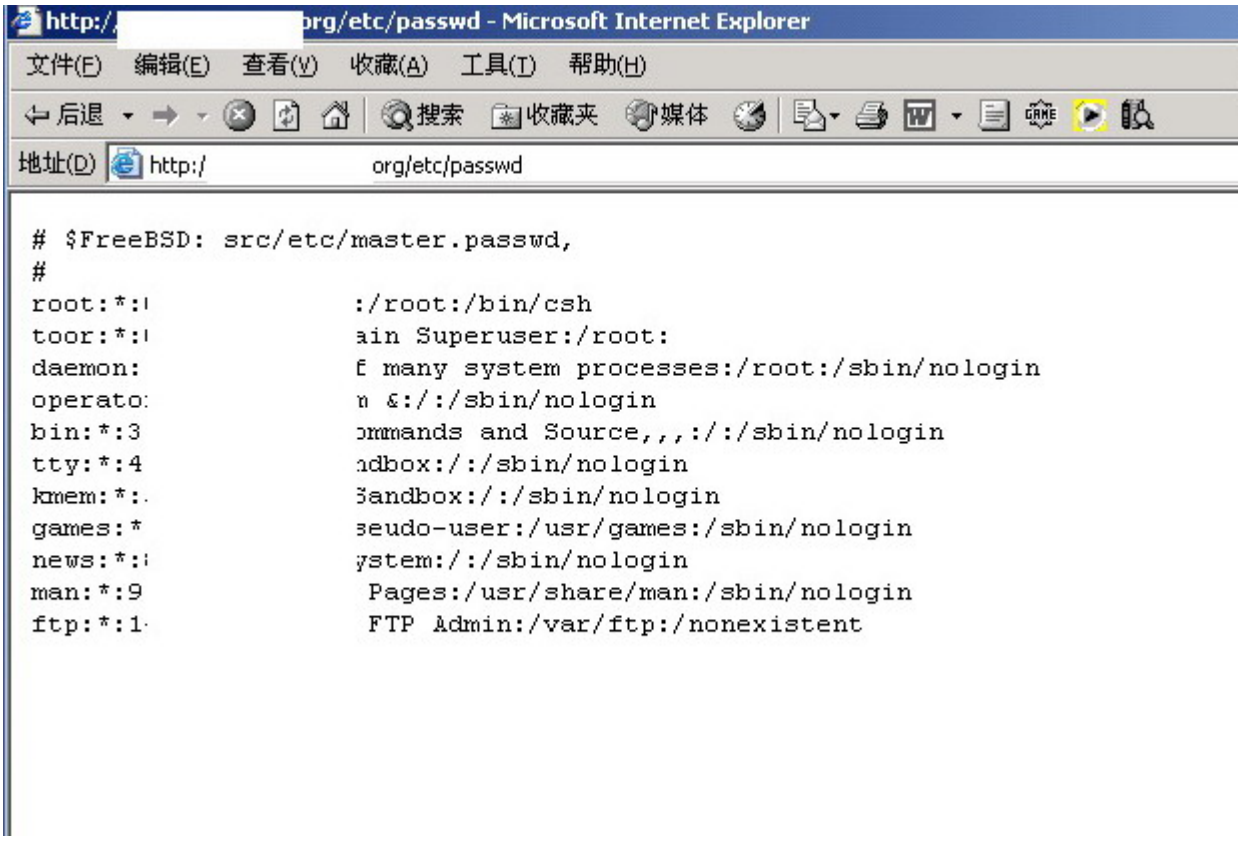
site:
这个也很有用,例如:site:www.4ngel.net.将返回所有和 4ngel.net 这个站有关的 URL.

对了还有一些操作符也是很有用的:
+把 google 可能忽略的字列如查询范围
-把某个字忽略
~同意词
.单一的通配符
*通配符，可代表多个字母
""精确查询

下面开始说说实际应用(我个人还是比较习惯用 google.com,以下内容均在 google 上搜索),对于一个居心叵测的攻击者来说,可能他最感兴趣的就是密码文件了.而 google 正因为其强大的搜索能力往往会把一些敏感信息透露给他们.用 google 搜索以下内容:

intitle:"indexof"etc
intitle:"Indexof".sh_history
intitle:"Indexof".bash_history
intitle:"indexof"passwd
intitle:"indexof"people.lst
intitle:"indexof"pwd.db
intitle:"indexof"etc/shadow
intitle:"indexof"spwd
intitle:"indexof"master.passwd
intitle:"indexof"htpasswd
"#-FrontPage-"inurl:service.pwd

有时候因为各种各样的原因一些重要的密码文件被毫无保护的暴露在网络上,如果被别有用心的人获得,那么危害是很大的.下面是我找到的一个 FreeBSD 系统的 passwd 文件(我已做过处理):

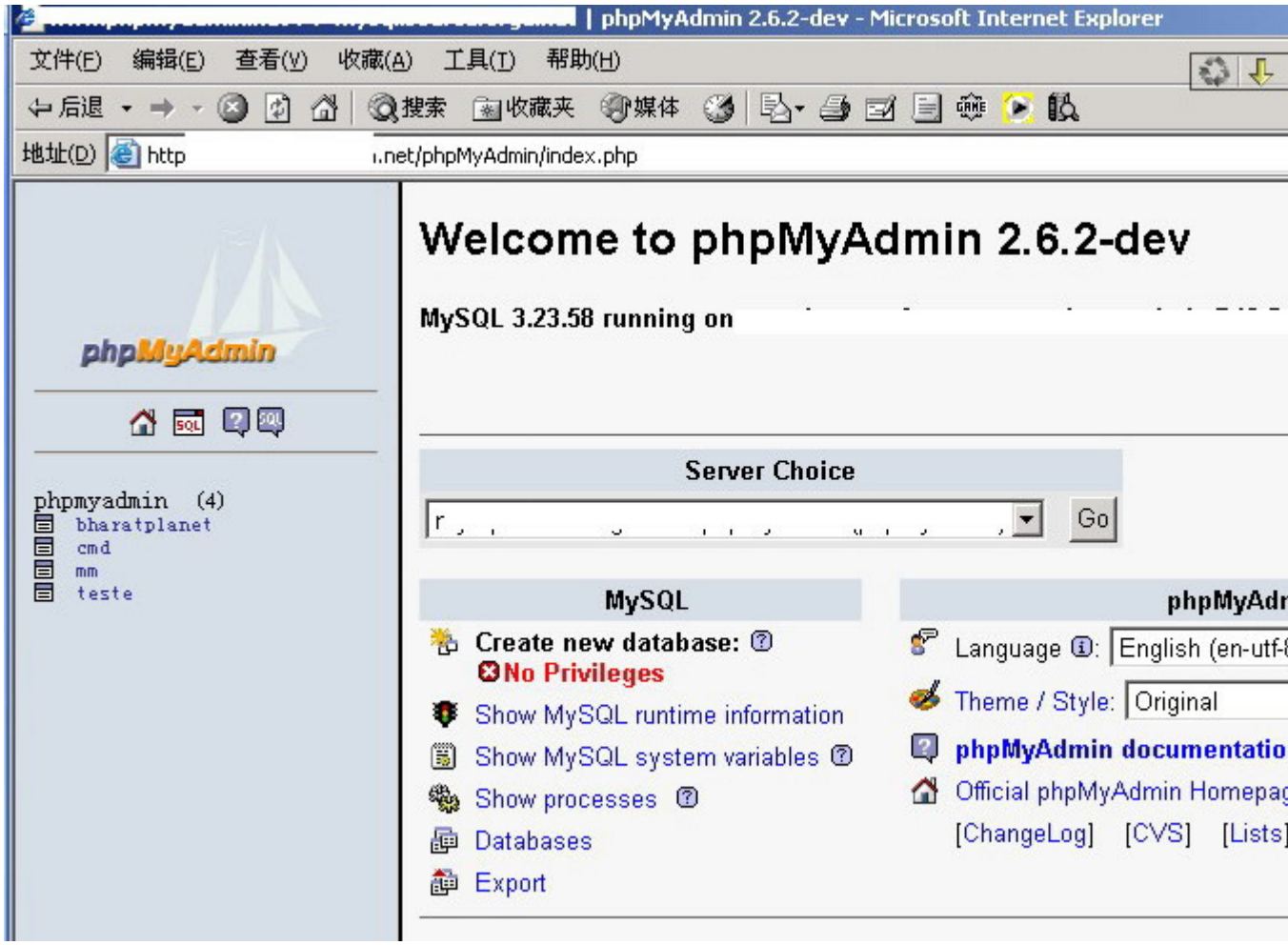


图一

同样可以用 google 来搜索一些具有漏洞的程序,例如 ZeroBoard 前段时间发现个文件代码泄露漏洞,我们可以用 google 来找网上使用这套程序的站点:
intext:ZeroBoardfiletype:php

或者使用:
inurl:outlogin.php?_zb_path=site:.jp

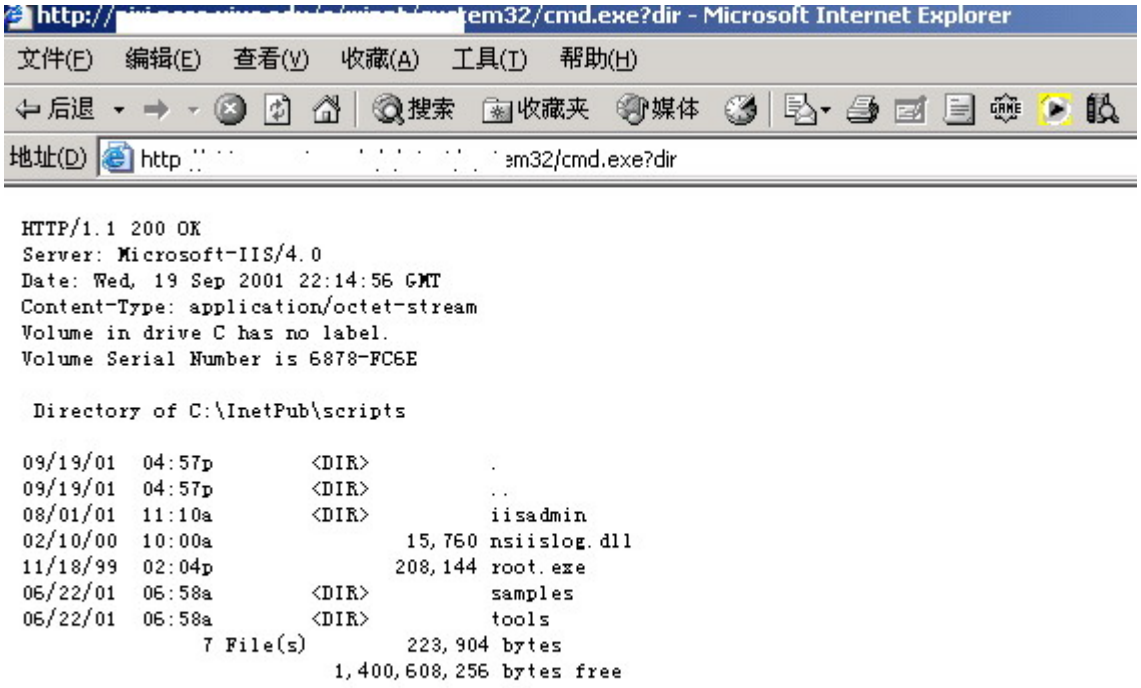
来寻找我们所需要的页面.phpmyadmin 是一套功能强大的数据库操作软件,一些站点由于配置失误,导致我们可以不使用密码直接对 phpmyadmin 进行操作.我们可以用 google 搜索存在这样漏洞的程序 URL:
intitle:phpmyadmintext:Createnewdatabase



图二

还记得 http://www.xxx.com/_vti_bin/..%5C..%5C....m32/cmd.exe?dir 吗?用 google 找找，你也许还可以找到很多古董级的机器。同样我们可以用这个找找有其他 cgi 漏洞的页面。

allinurl：winntsystem32



图三

前面我们已经简单的说过可以用 google 来搜索数据库文件,用上一些语法来精确查找能够获得更多东西(access 的数据库,mssql、mysql 的连接文件等等).举个例子示例一下:

allinurl:bbsdata
filetype:mdbinurl:database
filetype:incconn
inurl:datafiletype:mdb
intitle:"indexof"data//在一些配置不正确的 apache+win32 的服务器上经常出现这种情况

和上面的原理一样,我们还可以用 google 来找后台,方法就略了,举一反三即可,毕竟我写这篇文章的目的是让大家了解 googlehacking,而不是让你用 google 去破坏.安全是把双刃剑,关键在于你如何去运用.

利用 google 完全是可以对一个站点进行信息收集和渗透的，下面我们用 google 对特定站点进行一次测试。www.xxxx.com 是全国著名大学之一，一次偶然的机会我决定对其站点进行一次测试(文中所涉及该学校的信息均已经过处理，请勿对号入座:).
首先用 google 先看这个站点的一些基本情况(一些细节部分就略去了):
site:xxxx.com

从返回的信息中，找到几个该校的几个系院的域名：

http://a1.xxxx.com
http://a2.xxxx.com
http://a3.xxxx.com
http://a4.xxxx.com

顺便 ping 了一下，应该是在不同的服务器.(想想我们学校就那一台可怜的 web 服务器，大学就是有钱，汗一个)。学校一般都会有不少好的资料，先看看有什么好东西没:
site:xxxx.comfiletype:doc

得到 N 个不错的 doc。先找找网站的管理后台地址：
site:xxxx.comintext:管理
site:xxxx.cominurl:login
site:xxxx.comintitle:管理

超过获得 2 个管理后台地址：
http://a2.xxxx.com/sys/admin_login.asp
http://a3.xxxx.com:88/_admin/login_in.asp

还算不错，看看服务器上跑的是什么程序：
site:a2.xxxx.comfiletype:asp
site:a2.xxxx.comfiletype:php
site:a2.xxxx.comfiletype:aspx
site:a3.xxxx.comfiletype:asp
site:.....
.....

a2 服务器用的应该是 IIS，上面用的是 asp 的整站程序，还有一个 php 的论坛
a3 服务器也是 IIS，aspx+asp。web 程序都应该是自己开发的。有论坛那就看看能不能遇见什么公共的 FTP 帐号什么的：
site:a2.xxxx.comintext:ftp://*:*

没找到什么有价值的东西。再看看有没有上传一类的漏洞：
site:a2.xxxx.cominurl:file
site:a3.xxxx.cominurl:load

在 a2 上发现一个上传文件的页面：
http://a2.xxxx.com/sys/uploadfile.asp

用 IE 看了一下，没权限访问。试试注射，
site:a2.xxxx.comfiletype:asp

得到 N 个 asp 页面的地址，体力活就让软件做吧，这套程序明显没有对注射做什么防范，dbowner 权限，虽然不高但已足矣，backashell 我不太喜欢，而且看起来数据库的个头就不小，直接把 web 管理员的密码暴出来再说，MD5 加密过。一般学校的站点的密码都比较有规律，通常都是域名+电话一类的变形，用 google 搞定吧。

site:xxxx.com//得到 N 个二级域名
site:xxxx.comintext:*@xxxx.com//得到 N 个邮件地址，还有邮箱的主人的名字什么的
site:xxxx.comintext:电话//N 个电话

把什么的信息做个字典吧，挂上慢慢跑。过了一段时间就跑出 4 个帐号，2 个是学生会的，1 个管理员，还有一个可能是老师的帐号。登陆上去：
name：网站管理员
pass：a2xxxx7619//说了吧，就是域名+4 个数字

要再怎么提权那就不属于本文讨论访问了，呵呵，到此为止。

关于 googlehacking 的防范

以前我们站的晓风·残月写过一篇躲避 google 的文章，原理就是通过在站点根目录下建立一个 robots.txt 以避免网络机器人获得一些敏感的信息，具体大家看原文章：
http://www.4ngel.net/article/26.htm

不过这种方法我个人不推荐，有点此地无银三百两的味道。简单一点的方法就是上 google 把自己站点的一些信息删除掉，访问这个 URL：
http://www.google.com/remove.html

前几天看见又有人讨论用程序来欺骗 robot 的方法，我觉得可以试试，代码如下：

```
if(strstr($_SERVER['HTTP_USER_AGENT'],"Googlebot"))
{
    header("HTTP/1.1301");
    header("Location:http://www.google.com");
}
?>
```

后记

这段时间在国外的一些 googlehack 的研究站点看了看，其实也都差不多是一些基本语法的灵活运用，或者配合某个脚本漏洞，主要还是靠个人的灵活思维。国外对于 googlehack 方面的防范也并不是很多，所以大家还是点到为止，不要去搞破坏拉，呵呵。对于一些在 win 上跑
apache 的网管们应该多注意一下这方面，一个 intitle:indexof 就差不多都出来了：）