第 5 章　信息收集与信息泄露

信息收集是 Web 安全攻防的基础，攻击者通过收集目标系统的公开信息、社交资料、网络信息等，能够掌握目标系统的资产、人员、业务等情况，从而为后续的漏洞挖掘和攻击行动提供思路。信息泄露是 Web 安全攻防中的一个重要风险点，攻击者通过发现和利用目标系统的信息泄露漏洞，可以获取敏感信息，甚至控制目标系统。本章将带领读者深入理解信息收集与信息泄露的相关知识，并学习如何在 Web 安全攻防实践中应用这些知识。

5.1　信息收集概述

在 Web 安全攻防中，信息收集（也称为资产收集）是前期的一项主要工作，主要收集目标的情报信息或资产信息。信息收集的过程涉及多种方法和工具的运用，旨在全面搜集有关目标网络应用及其所处环境的详尽信息，这些信息包括但不限于目标域名、子域名、IP 地址、网络拓扑结构、Web 服务器类型和版本、开放端口、运行的服务、目录结构、Web 应用程序框架、漏洞信息、敏感文件等。

信息收集的目的是帮助攻击者或防御者了解目标系统，从而更全面地评估系统的安全性。通过收集足够的上下文信息，攻击者能够洞察系统潜在的漏洞、弱点和攻击面，从而提高攻击的质量和精确度。与此同时，防御者也可以利用信息收集识别和修复漏洞，从而提升系统的整体安全性。

在信息收集的过程中，攻击者可能会运用多种方法，包括 WHOIS 查询、域名枚举、DNS 扫描、端口扫描、服务指纹识别、漏洞扫描、Web 应用程序爬虫、社会工程学等。综合应用这些方法有助于攻击者勾勒出目标系统的全貌，并为后续攻击活动提供有力支撑。

信息收集方式可以分为主动方式和被动方式。

主动方式是指攻击者直接与目标系统进行交互，通过扫描、查询、测试等主动措施获取目标信息。其优点在于能够获取深入、全面、实时的信息；缺点在于可能会在目标系统的日志中留下痕迹，容易被目标系统的安全监测机制发现，甚至可能导致系统中断或服务中断等问题。

主动方式的信息收集举例如下。

（1）端口扫描：使用 Nmap、Masscan 等工具扫描目标系统的开放端口，了解其网络服务和 Web 应用程序的部署情况。

（2）服务指纹识别：通过发送特定的请求，判别目标系统上运行的具体服务类型和版本。

（3）漏洞扫描：使用 Nessus、OpenVAS 等工具检测目标系统可能存在的漏洞。

（4）Web 应用程序爬虫：使用 Burp Suite、ZAP 等工具模拟爬取目标 Web 应用程序，了解网站的结构和功能。

被动方式是指攻击者在不直接与目标系统进行交互的情况下，通过观察、查询公开可用的信息源收集有关目标系统的信息。其优点在于对目标系统的影响较小，不容易被目标系统日志记录或监测系统追踪；缺点在于可能受限于公开可用的信息源，无法对目标系统进行全面了解，且信息可能缺乏实时性。

被动方式的信息收集举例如下。

（1）在线 WHOIS 查询：获取目标域名的注册信息，包括域名所有者、注册日期、过期日期、DNS 服务器信息等，可利用的工具包括 VirusTotal、DNSDumpster、DNSDB 等。

（2）搜索引擎查询：使用百度、Bing 等搜索引擎搜索目标域名或相关关键词，收集公开可用的信息。

（3）社会工程学分析：通过分析目标组织在社交媒体的活动，了解员工、组织结构和可能的安全风险。

（4）第三方工具：利用站长之家、爱站网、微步在线、云悉指纹、潮汐指纹等工具收集目标系统的端口信息、子域名信息、C 段信息、敏感信息、指纹信息等。

5.2　信息收集的常用方法

根据不同的收集目标和对象，信息收集通常采用多种方法，每种方法大都专注于收集特定类型的信息。

5.2.1　基于搜索引擎

搜索引擎常被用于信息收集。攻击者通常会利用搜索引擎的高级搜索功能及一些特定的语法指令发现和访问互联网中的敏感信息、漏洞以及未经授权的资源。本节将以百度搜索引擎为例，介绍基于搜索引擎进行信息收集的常用语法和示例。

百度搜索引擎的常用语法指令如表 5-1 所示。

表 5-1　百度搜索引擎的常用语法指令

语 法 指 令	描　　述	示　　例
site:	搜索指定域名下的资源	site:example.cn
filetype:	搜索指定文件类型的资源	filetype:pdf
intitle:	搜索标题包含指定关键词的资源	intitle:"keyword"
inurl:	搜索 URL 包含指定关键词的资源	inurl:"keyword"
intext:	搜索包含指定关键词的资源	intext:"keyword"

百度搜索引擎的使用技巧如表 5-2 所示。

表 5-2　百度搜索引擎的使用技巧

使 用 技 巧	示　　例
使用引号确保精确匹配	"index of /admin"
使用 AND、OR 组合多种语法指令	site:example.cn AND filetype:pdf
使用减号-排除特定词语	site:example.cn -test
使用通配符 * 扩大搜索范围	site: * .example.cn
使用"Index of"和特定文件名查找目录结构	intitle:"Index of" "config.yml"

示例一：查找可能包含 FTP 信息的目录结构，输入“intitle:"Index of" inurl:ftp”，搜索结果如图 5-1 所示。

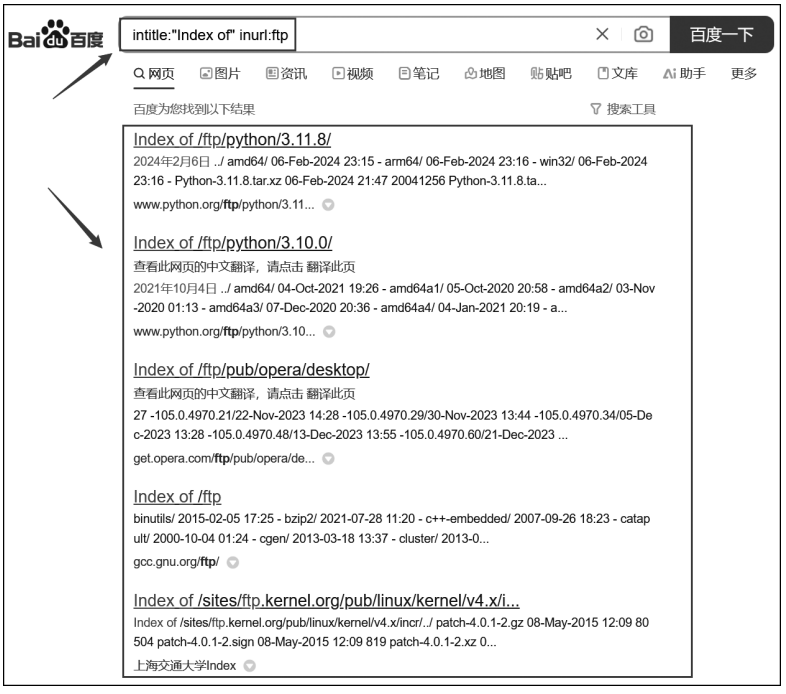

图 5-1　查找可能包含 FTP 信息的目录结构

示例二：查找可能的管理员登录页面，输入“inurl:/admin/login.php”，搜索结果如图 5-2 所示。

示例三：查找可能暴露 Git 存储库的目录，输入“intitle:"Index of" .git”，搜索结果如图 5-3 所示。

5.2.2　基于 GitHub 存储库

基于 GitHub 存储库进行信息收集是指利用 GitHub 平台上公开可见的存储库提交历史、开发者信息等资源获取有关目标系统、Web 应用程序或组织的关键信息，并对源代码、配置文件、依赖关系、提交历史等信息进行审计，以识别潜在的安全风险、漏洞和其他与目标系统有关的关键信息。

图 5-2　查找可能的管理员登录页面

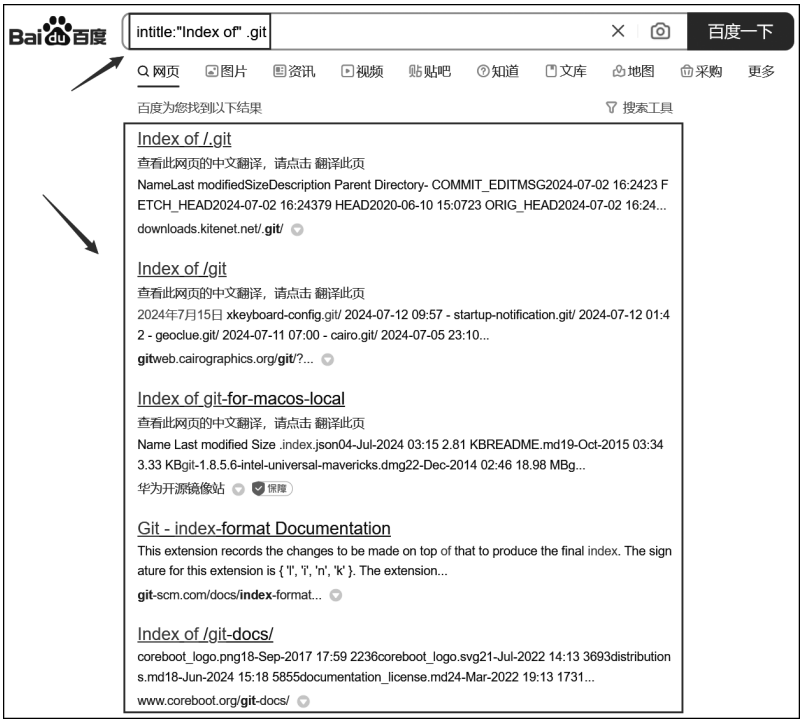

图 5-3　查找可能暴露 Git 存储库的目录

GitHub 存储库搜索常用语法如表 5-3 所示。

表 5-3　GitHub 存储库搜索常用语法

语　法	描　述	示　例
user:	搜索指定用户的存储库	user:john-doe

续表

语　法	描　述	示　例
org:	搜索指定组织的存储库	org:example-org
in:readme	搜索自述文件中包含指定内容的存储库	password in:readme
size:	搜索指定大小的存储库	size:>=1024
stars:	搜索具有指定星数的存储库	stars:>=100
forks:	搜索具有指定分支数的存储库	forks:>=50
created:	搜索指定日期创建的存储库	created:2022-01-01
pushed:	搜索指定日期范围推送的存储库	pushed:>=2022-01-01
language:	搜索使用指定编程语言的存储库	language:python

GitHub 代码搜索常用语法如表 5-4 所示。

表 5-4　GitHub 代码搜索常用语法

语　法	描　述	示　例
user:	搜索指定用户的文件	user:john-doe
org:	搜索指定组织的文件	org:example-org
path:	搜索包含指定路径的文件	path:/src/utils/
language:	搜索使用指定编程语言的文件	language:python
content:	搜索文件内容中包含指定文本的文件	content:password

GitHub 信息收集使用技巧如表 5-5 所示。

表 5-5　GitHub 信息收集使用技巧

描　述	示　例
使用双引号查询完全匹配项	"password"
使用运算符 AND、OR 和 NOT 组合搜索词	admin AND password
使用正则表达式搜索，并使用斜杠界定正则表达式的边界	/^App\/src\//
使用减号-排除限定符匹配的结果	password -language:php

示例一：搜索包含 MySQL 数据库连接关键词的敏感信息，输入“"jdbc:mysql://" AND "username" AND "password"”，如图 5-4 所示。

示例二：结合路径和编程语言搜索配置文件以寻找敏感信息或凭证，输入“path:**/config.inc language:php”，如图 5-5 所示。

注意：这种信息收集方式也可结合搜索引擎语法指令“site:github.com”使用，常用搜索语法包括“site:github.com smtp password”“site:github.com sa password”“site:github.com svn username password”等。

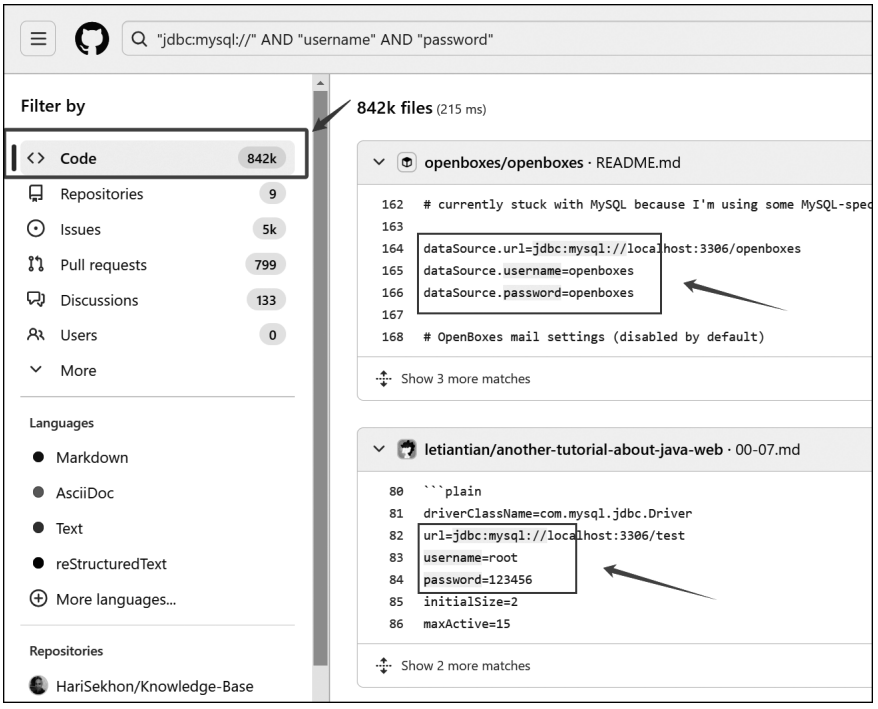

图 5-4　搜索包含 MySQL 数据库连接关键词的敏感信息

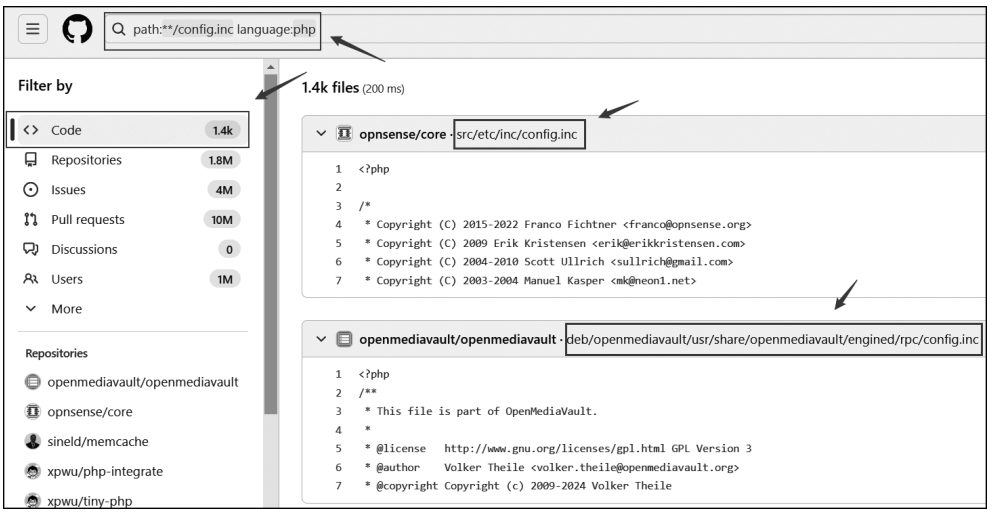

图 5-5　结合路径和编程语言搜索配置文件以寻找敏感信息或凭证

5.2.3　端口扫描

在网络通信中，端口是分配给特定服务或 Web 应用程序进程的数字标识符。每种网络服务都监听特定的端口号，以便正确接收来自客户端的请求。

端口扫描指通过专门设计的工具或技术，对目标系统的网络端口进行系统化的扫描，以识别开放的端口、运行的服务和 Web 应用程序。端口扫描的基本原理是攻击者向目标系统

的各个端口发送请求，并通过系统的响应判断这些端口是否开放。如果某个端口开放，通常意味着有一个服务正在监听该端口，并等待接收来自客户端的连接请求；如果端口关闭，任何发送到这个端口的连接请求都会被拒绝或不予响应。

端口扫描可分为 TCP 全连接扫描、SYN/半开放扫描、UDP 扫描等类型，详细对比如表 5-6 所示。

表 5-6　端口扫描常用类型对比

类　型	说　明	优　缺　点
TCP 全连接扫描	通过建立完整的 TCP 连接扫描目标端口	优点：精确，确认端口真实开放；可获取服务详细信息。缺点：相对慢，对目标系统产生较大负载；容易被监测系统发现
SYN（Synchronize）/半开放扫描	通过发送 SYN 包并监测响应，不完成三次握手	优点：较快，不完全建立连接，减小影响；适用于大范围扫描。缺点：不能确认端口是否真实开放；可能被防火墙发现
UDP 扫描	通过发送 UDP 数据包并分析响应判断端口状态	优点：适用基于 UDP 协议的服务；不产生完整的连接，减小影响。缺点：不可靠，易受网络延迟影响；无法确认端口是否真实开放
NULL 扫描	通过发送无标志位的 TCP 数据包并观察响应	优点：利用目标系统对非法包的不同响应识别开放端口。缺点：需要目标系统的开放端口与非开放端口对非法包的响应不同，不适用于所有系统
FIN(Finish)扫描	通过发送带有 FIN 标志位的 TCP 数据包并观察响应	优点：利用目标系统对非法包的不同响应识别开放端口。缺点：同“NULL 扫描”
XMAS 扫描	通过发送带有 FIN、PSH(Push)和 URG（Urgent）标志位的 TCP 数据包并观察响应	优点：利用目标系统对非法包的不同响应识别开放端口。缺点：同“NULL 扫描”
ACK（Acknowledgement）扫描	通过发送带有 ACK 标志位的 TCP 数据包并检测防火墙响应	优点：可以识别目标主机中的防火墙规则和过滤策略。缺点：无法明确判断端口是否真实开放

常见服务的默认开放端口及可被利用的攻击方向如表 5-7 所示。

表 5-7　常见服务端口及其可能利用的攻击方向

Web 应用服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
HTTP 服务	80/8080	Web 攻击、暴力破解
SSL 服务	443	OpenSSL“心脏出血”漏洞
JAVA RMI 服务	1090/1099	反序列化远程命令执行
Lotus Domino 邮件服务	1352	弱口令、信息泄露、暴力破解
Weblogic 服务	7001/7002	Java 反序列化、弱口令、SSRF

续表

Web 应用服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
JDWP 服务	8000	远程命令执行
Tomcat 服务	8080	弱口令、示例目录
Jboss 服务	8080	未授权访问、反序列化
Resin 服务	8080	目录遍历、远程文件读取
Jetty 服务	8080	远程共享缓冲区泄漏
Jenkins 服务	8080	未授权访问、远程代码执行
GlassFish 服务	8080/4848	弱口令、任意文件读取
WebSphere 服务	9090	Java 反序列化、弱口令
Webmin-Web 服务	10000	弱口令
数据库服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
LDAP 服务	389	未授权访问、弱口令
MSSQL 服务	1433	注入、弱口令、暴力破解
Oracle 服务	1521	暴力破解、注入
MySQL 服务	3306	注入、暴力破解
Sybase/DB2 服务	5000	暴力破解、注入
PostgreSQL 服务	5432	暴力破解、注入、弱口令
CouchDB 服务	5984	垂直权限绕过、任意命令执行
Redis 服务	6379	未授权访问、弱口令、暴力破解
Elasticsearch 服务	9200	未授权访问、命令执行
MemCache 服务	11211	未授权访问
MongoDB 服务	27017/27018	暴力破解、未授权访问
远程连接服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
SSH 远程连接服务	22	弱口令、暴力破解、用户名枚举、SSH 隧道及内网代理转发
Telnet 远程连接服务	23	暴力破解、嗅探、弱口令
RDP 远程桌面连接	3389	Shift 后门(Windows Server 2003 以下版本)、暴力破解
VNC 远程连接服务	5900	弱口令、暴力破解
文件共享服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
FTP/TFTP 文件传输服务	21/22/69	匿名登录、弱口令、暴力破解、嗅探

续表

文件共享服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
Samba 文件共享服务	139	暴力破解、未授权访问、远程代码执行
LDAP 目录访问协议	389	注入、允许匿名访问、弱口令
SMB 文件共享服务	445	信息泄露、远程代码执行
NFS 网络文件系统服务	2049	配置不当
邮件服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
SMTP 邮件服务	25	匿名发送邮件、弱口令、用户枚举
POP3 邮件服务	110	暴力破解、嗅探
IMAP 邮件服务	143	暴力破解
常见网络协议端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
DNS 服务	53	域传送漏洞、欺骗、缓存投毒
DHCP 服务	67/68	劫持、欺骗
SNMP 服务	161	弱口令
特殊服务端口		
服 务 名 称	默认开放端口	可被利用的攻击方向
Linux Rexec 服务	512/513/514	暴力破解、Rlogin 登录
Rsync 服务	873	匿名访问、弱口令
Zookeeper 服务	2181	未授权访问
Docker 服务	2375	未授权访问
SVN 服务	3690	SVN 泄露、未授权访问
ActiveMQ 服务	8161	弱口令、任意文件写入、反序列化
WebSphere 服务	9043	控制台弱口令、远程代码执行
Elasticsearch 服务	9200/9300	未授权访问
Zabbix 服务	10051	远程命令执行、SQL 注入
Memcache 服务	11211	未授权访问
Hadoop 服务	50070	未授权访问

常用的端口扫描工具介绍如下。

1. Nmap

Nmap 是一款广泛应用于网络安全领域的开源工具，提供端口扫描、主机发现、操作系统识别以及服务和 Web 应用程序探测等多种功能。Nmap 强大的扫描引擎和丰富的功能

使其成为安全专业人员和系统管理员首选的工具之一，Nmap 不仅能够帮助用户全面了解网络环境，还能够发现潜在的安全漏洞，提示用户采取相应的防护措施。限于篇幅，本书不介绍 Nmap 的安装，读者可自行查阅相关资料进行安装。

格式如下。

```
nmap［选项］目标
```

常用选项如下。

- -sS：SYN 扫描，通过发送 TCP SYN 包识别开放端口。
- -sT：TCP 全连接扫描，通过建立完整的 TCP 连接识别开放端口，速度较慢。
- -sU：UDP 扫描，用于识别开放的 UDP 端口。
- -sP：Ping 扫描，用于快速扫描目标系统的存活主机。
- -sn：无端口扫描，使用 Ping 扫描发现存活主机，但不进行端口扫描。
- -A：启用操作系统检测、服务版本检测、脚本扫描等多种功能，进行全面的信息收集。
- -p <port-range>：指定端口范围，可以是单个端口、一系列连续端口（如 1-100），或以逗号分隔的多个不连续端口。
- -O：启用操作系统检测，尝试识别目标系统的操作系统类型。
- -script <script>：执行特定的 Nmap 脚本，用于进行漏洞检测和服务探测。
- -v：提供详细的信息输出。
- -T<0-5>：指定扫描速度，0 为最慢、5 为最快。
- --open：只显示开放的端口，忽略关闭的端口。

此处基于 CentOS7 攻击机和 Windows7 靶机进行演示：使用 CentOS7 攻击机中的 Nmap 扫描 Windows7 靶机（IP 地址为：192.168.1.101）的 1～1000 号端口，确定每个端口的状态。命令为“nmap -p 1-1000 192.168.1.101”，执行结果如图 5-6 所示。

```
root@websec:~# nmap -p 1-1000 192.168.1.101
Starting Nmap 7.80 ( https://nmap.org ) at 2024-06-08 19:53 EDT
Nmap scan report for 192.168.1.101
Host is up (0.00070s latency).
Not shown: 996 closed ports
PORT    STATE SERVICE
80/tcp  open  http
135/tcp open  msrpc
139/tcp open  netbios-ssn
445/tcp open  microsoft-ds
MAC Address: 00:0C:29:10:1D:8A (VMware)

Nmap done: 1 IP address (1 host up) scanned in 1.59 seconds
```

图 5-6　Nmap 使用示例

2. Zenmap

Zenmap 是 Nmap 的图形用户界面版本，提供了直观的界面用于执行扫描，具体用法参考 Nmap。限于篇幅，本书不介绍 Zenmap 的安装，读者可自行查阅相关资料进行安装。

此处基于 Windows10 攻击机和 Windows7 靶机进行演示。使用 Windows10 攻击机中的 Zenmap 快速且详细地扫描 Windows7 靶机，获取其操作系统信息、服务版本、开放端口及其对应的服务，并提供详细的扫描过程信息。命令为“nmap -T4 -A -v 192.168.1.101”，其中“-T4”表示设置扫描速度为较快的等级，“-A”表示启用操作系统检测、服务版本检测、脚