5

INFORMATION GATHERING

In this chapter we begin the information-gathering phase of penetration testing. The goal of this phase is to learn as much about our clients as we can. Does the CEO reveal way too much on Twitter? Is the system administrator writing to archived listservs, asking about how to secure a Drupal install? What software are their web servers running? Are the Internet-facing systems listening on more ports than they should? Or, if this is an internal penetration test, what is the IP address of the domain controller?

We'll also start to interact with our target systems, learning as much as we can about them without actively attacking them. We'll use the knowledge gained in this phase to move on to the threat-modeling phase where we think like attackers and develop plans of attack based on the information we've gathered. Based on the information we uncover, we'll actively search for and verify vulnerabilities using vulnerability-scanning techniques, which are covered in the next chapter.

Open Source Intelligence Gathering

We can learn a good deal about our client's organization and infrastructure before we send a single packet their way, but information gathering can still be a bit of a moving target. It isn't feasible to study the online life of every employee, and given a large amount of gathered information, it can be difficult to discern important data from noise. If the CEO tweets frequently about a favorite sports team, that team's name may be the basis for her webmail password, but it could just as easily be entirely irrelevant. Other times it will be easier to pick up on something crucial. For instance, if your client has online job postings for a system administrator who is an expert in certain software, chances are those platforms are deployed in the client's infrastructure.

As opposed to intelligence gained from covert sources such as dumpster diving, dumping website databases, and social engineering, *open source intelligence* (or *OSINT*) is gathered from legal sources like public records and social media. The success of a pentest often depends on the results of the information-gathering phase, so in this section, we will look at a few tools to obtain interesting information from these public sources.

Netcraft

Sometimes the information that web servers and web-hosting companies gather and make publicly available can tell you a lot about a website. For instance, a company called Netcraft logs the uptime and makes queries about the underlying software. (This information is made publicly available at *http://www.netcraft.com/*.) Netcraft also provides other services, and their antiphishing offerings are of particular interest to information security.

For example, Figure 5-1 shows the result when we query *http://www.netcraft.com/* for *http://www.bulbsecurity.com*. As you can see, *bulbsecurity.com* was first seen in March 2012. It was registered through GoDaddy, has an IP address of 50.63.212.1, and is running Linux with an Apache web server.

Armed with this information, when pentesting *bulbsecurity.com*, we could start by ruling out vulnerabilities that affect only Microsoft IIS servers. Or, if we wanted to try social engineering to get credentials to the website, we could write an email that appears to be from GoDaddy, asking the administrator to log in and check some security settings.

Site report for https://rourab.com

▶ Look up another site?

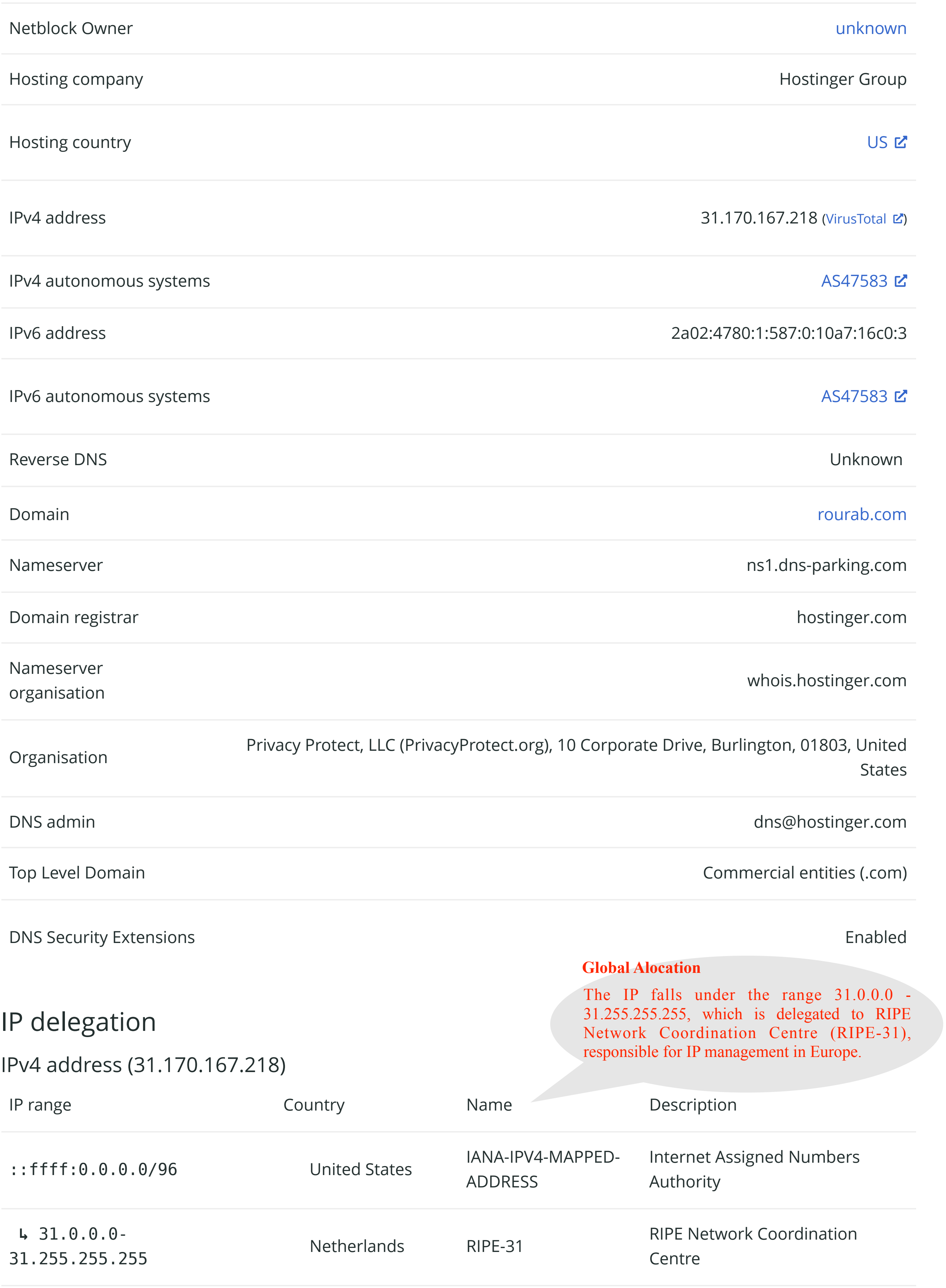
Share: Y

Background

Site title	rourab
Site rank	189649
Description	professor
Date first seen	September 2019
Primary language	English

Network

Site	https://rourab.com



Netblock Owner	unknown
Hosting company	Hostinger Group
Hosting country	US
IPv4 address	31.170.167.218 (VirusTotal)
IPv4 autonomous systems	AS47583
IPv6 address	2a02:4780:1:587:0:10a7:16c0:3
IPv6 autonomous systems	AS47583
Reverse DNS	Unknown
Domain	rourab.com
Nameserver	ns1.dns-parking.com
Domain registrar	hostinger.com
Nameserver organisation	whois.hostinger.com
Organisation	Privacy Protect, LLC (PrivacyProtect.org), 10 Corporate Drive, Burlington, 01803, United States
DNS admin	dns@hostinger.com
Top Level Domain	Commercial entities (.com)
DNS Security Extensions	Enabled

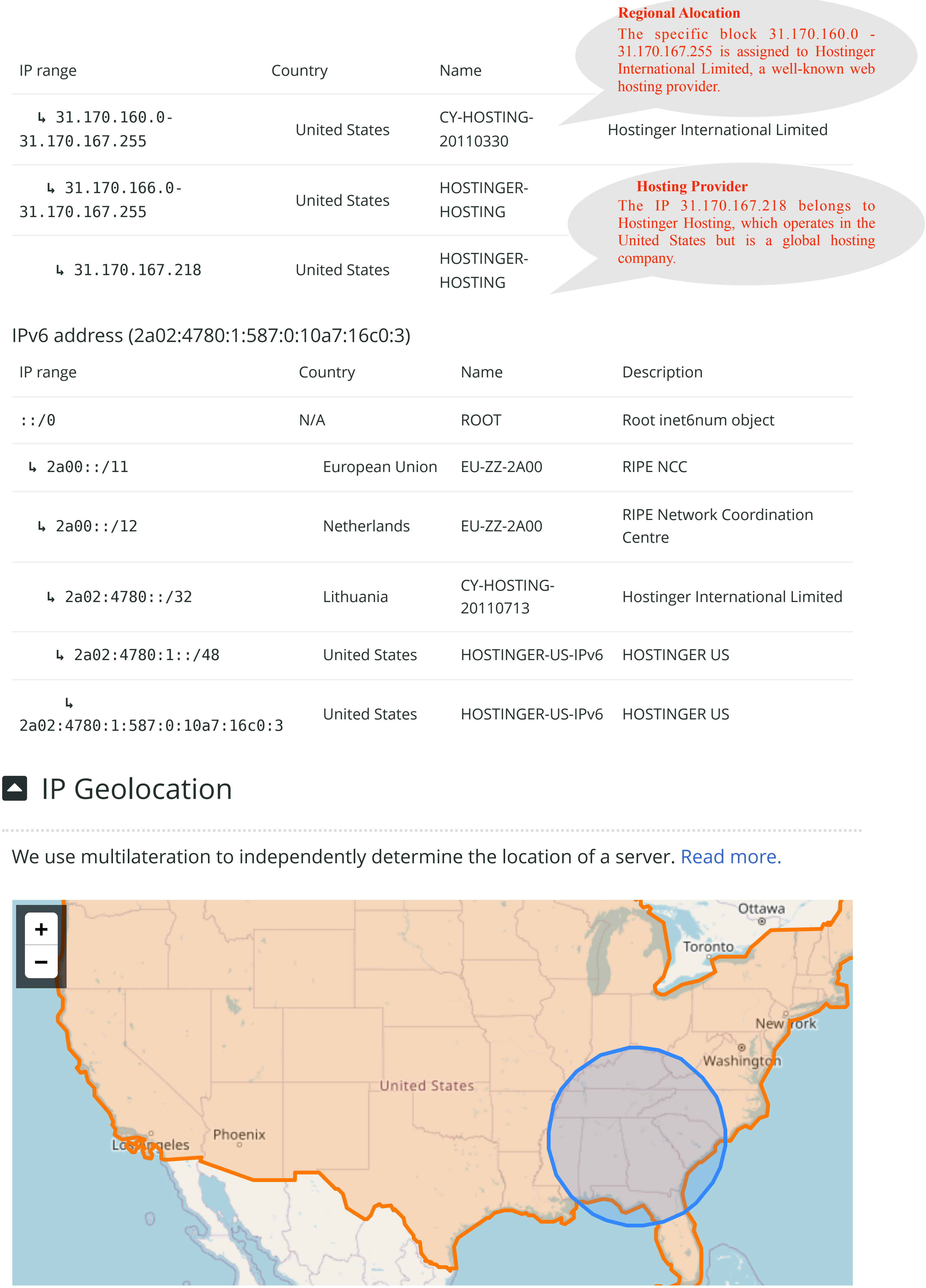
Global Alocation

The IP falls under the range 31.0.0.0 - 31.255.255.255, which is delegated to RIPE Network Coordination Centre (RIPE-31), responsible for IP management in Europe.

IP delegation

IPv4 address (31.170.167.218)

IP range	Country	Name	Description
`::ffff:0.0.0.0/96`	United States	IANA-IPV4-MAPPED-ADDRESS	Internet Assigned Numbers Authority
`↳ 31.0.0.0-31.255.255.255`	Netherlands	RIPE-31	RIPE Network Coordination Centre



Regional Alocation

The specific block 31.170.160.0 - 31.170.167.255 is assigned to Hostinger International Limited, a well-known web hosting provider.

IP range	Country	Name	
↳ 31.170.160.0-31.170.167.255	United States	CY-HOSTING-20110330	Hostinger International Limited
↳ 31.170.166.0-31.170.167.255	United States	HOSTINGER-HOSTING	
↳ 31.170.167.218	United States	HOSTINGER-HOSTING	

Hosting Provider

The IP 31.170.167.218 belongs to Hostinger Hosting, which operates in the United States but is a global hosting company.

IPv6 address (2a02:4780:1:587:0:10a7:16c0:3)

IP range	Country	Name	Description
::/0	N/A	ROOT	Root inet6num object
↳ 2a00::/11	European Union	EU-ZZ-2A00	RIPE NCC
↳ 2a00::/12	Netherlands	EU-ZZ-2A00	RIPE Network Coordination Centre
↳ 2a02:4780::/32	Lithuania	CY-HOSTING-20110713	Hostinger International Limited
↳ 2a02:4780:1::/48	United States	HOSTINGER-US-IPv6	HOSTINGER US
↳ 2a02:4780:1:587:0:10a7:16c0:3	United States	HOSTINGER-US-IPv6	HOSTINGER US

IP Geolocation

We use multilateration to independently determine the location of a server. Read more.

SSL/TLS

Assurance	Domain validation
Common name	rourab.com
Organisation	Not Present
State	Not Present
Country	Not Present
Organisational unit	Not Present
Subject Alternative Name	rourab.com, www.rourab.com
Validity period	From Dec 25 2024 to Mar 25 2025 (3 months)
Matches hostname	Yes
Server	LiteSpeed
Public key algorithm	rsaEncryption
Protocol version	TLSv1.3
Public key length	4096
Certificate check	ok
Signature algorithm	sha384WithRSAEncryption

Serial number	0xfa1e3bb87f42b3d2b77bbbd4d63f6fe7
Cipher	TLS_AES_256_GCM_SHA384
Version number	0x02
Perfect Forward Secrecy	Yes
Supported TLS Extensions	RFC8446 key share, RFC8446 supported versions, RFC7301 application-layer protocol negotiation, RFC4366 status request
Application-Layer Protocol Negotiation	h2
Next Protocol Negotiation	Not Present
Issuing organisation	ZeroSSL
Issuer common name	ZeroSSL RSA Domain Secure Site CA
Issuer unit	Not Present
Issuer location	Not Present
Issuer country	AT
Issuer state	Not Present
Certificate Revocation Lists	Not Present
Certificate Hash	sAPyY1ujghFFgbP5N4oVPK8dB7s
Public Key Hash	fa52409150fb6e134d71b42bbe205ff1420ab0fc0123f5fc43c17bd35fa9e374
OCSP servers	http://zerossl.ocsp.sectigo.com
OCSP stapling response	Certificate valid
OCSP data generated	Feb 16 22:18:03 2025 GMT
OCSP data expires	Feb 23 22:18:02 2025 GMT

Certificate Transparency

Signed Certificate Timestamps (SCTs)

Source	Log	Timestamp	Signature Verification
Certificate	*Unknown* zxFW7tUufK/zh1vZaS6b6RpxZ0qwF+ysAdJbd87MOwg=	2024-12-25 09:00:13	*Unknown*
Certificate	*Unknown* zPsPaoVxCWX+lZtTzumyfCLphVwNl422qX5UwP5MDbA=	2024-12-25 09:00:13	*Unknown*

SSLv3/POODLE

This site does not support the SSL version 3 protocol.

More information about SSL version 3 and the POODLE vulnerability.

Heartbleed was a severe security vulnerability in OpenSSL's implementation of TLS's Heartbeat extension

Heartbleed

The site did not offer the Heartbeat TLS extension prior to the Heartbleed disclosure, and so was not exploitable.

This test does not exploit the Heartbleed vulnerability but uses information from conventional HTTPS requests. More information about Heartbleed detection.

SSL Certificate Chain

Sender Policy Framework

A host's Sender Policy Framework (SPF) describes who can send mail on its behalf. This is done by publishing an SPF record containing a series of rules. Each rule consists of a qualifier followed by a specification of which domains to apply this qualifier to. For more information please see open-spf.org.

Qualifier	Mechanism	Argument
+ (Pass)	include	_spf.mail.hostinger.com
~ (SoftFail)	all	

DMARC

DMARC (Domain-based Message Authentication, Reporting and Conformance) is a mechanism for domain owners to indicate how mail purporting to originate from their domain should be authenticated. It builds on SPF and DKIM, providing a method to set policy and to give reporting of failures. For more information please see dmarc.org.

Raw DMARC record:

```
v=DMARC1; p=none
```

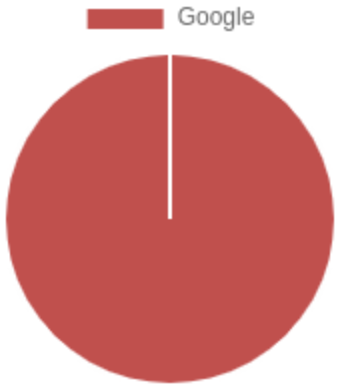
Tag	Field	Value
p=none	Requested handling policy	None: no specific action to be taken regarding delivery of messages.

Web Trackers

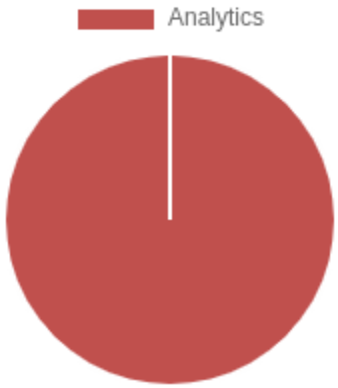
Web Trackers are third-party resources loaded onto a webpage. Trackable resources include social sharing widgets, javascript files, and images. These trackers can be used to monitor individual user behaviour across the web. Data derived from these trackers are primarily used for advertising or analytics purposes.

1 known tracker was identified.

Companies



Categories



Company	Primary Category	Tracker	Popular Sites with this Tracker
Google	Analytics	Googletagmanager	www.virustotal.com, www.avito.ru, www.coingecko.com

Site Technology (fetched today)

Server-Side

Includes all the main technologies that Netcraft detects as running on the server such as PHP.

Technology	Description	Popular sites using this technology
PHP Enabled	Server supports PHP	www.calculator.net, www.gsmarena.com, www.singpost.com
SSL	A cryptographic protocol providing communication security over the Internet	accounts.google.com, saas-aftral.octime.net, campus-1001.ammon.cloud
PHP	PHP is supported and/or running	www.pixiv.net, www.skillacloud.com, www.whois.com

Client-Side

Includes all the main technologies that run on the browser (such as JavaScript and Adobe Flash).

Technology	Description	Popular sites using this technology
JavaScript	Widely-supported programming language commonly used to power client-side dynamic content on websites	discord.com, chatgpt.com, x.com

Client-Side Scripting Frameworks

Frameworks or libraries allow for easier development of applications by providing an Application Program Interface (API) or a methodology to follow whilst developing.

Technology	Description	Popular sites using this technology
Google Tag Manager	*No description*	www.nexusmods.com, www.avito.ru, www.virustotal.com
jQuery	A JavaScript library used to simplify the client-side scripting of HTML	www.amazon.fr, www.amazon.in, webmail.vinccihoteles.com

Web Stats

Web analytics is the measurement, collection, analysis and reporting of internet data for purposes of understanding and optimizing web usage.

Technology	Description	Popular sites using this technology
Google Webmaster Tools	Set of tools allowing webmasters to check indexing status and optimize visibility of their websites on Google	www.roblox.com, www.ebay.com, app.powerbi.com

Character Encoding

A character encoding system consists of a code that pairs each character from a given repertoire with something else such as a bit pattern, sequence of natural numbers, octets, or electrical pulses in order to facilitate the transmission of data (generally numbers or text) through telecommunication networks or for data storage.

Technology	Description	Popular sites using this technology
UTF8	UCS Transformation Format 8 bit	www.amazon.com, www.netflix.com, www.twitch.tv

HTTP Compression

HTTP compression is a capability that can be built into web servers and web clients to make better use of available bandwidth, and provide greater transmission speeds between both.

Technology	Description	Popular sites using this technology
Gzip Content Encoding	Gzip HTTP Compression protocol	www.amazon.co.uk, www.amazon.co.jp, www.amazon.es

Web Browser Targeting

Web browser targeting enables software applications to make use of specific functions of the browser as well as optimizing the application for specific browser versions.

Technology	Description	Popular sites using this technology
Content Security Policy	Detect and mitigate attacks in the browser	mail.proton.me

Doctype

A Document Type Declaration, or DOCTYPE, is an instruction that associates a particular SGML or XML document (for example, a webpage) with a Document Type Definition (DTD).

Technology	Description	Popular sites using this technology
HTML5	Latest revision of the HTML standard, the main markup language on the web	

HTML 5

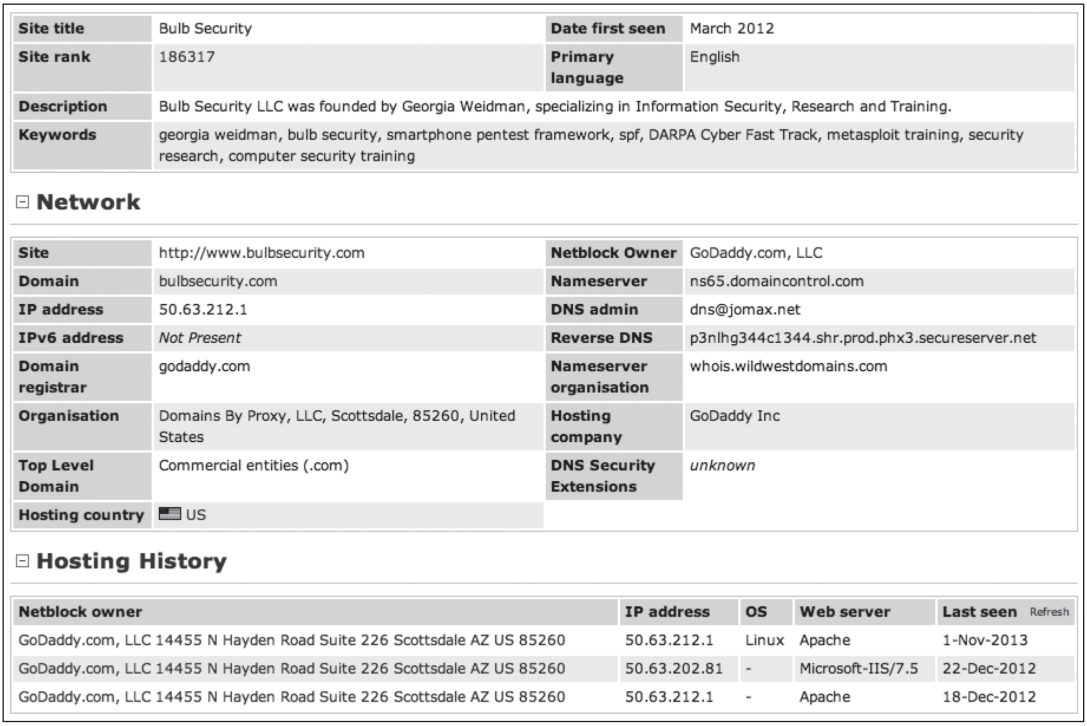
HTML5 is a markup language for structuring and presenting content for the World Wide Web and a core technology of the Internet. It is the fifth revision of the HTML standard.

Technology	Description	Popular sites using this technology
Viewport meta tag	HTML5 tag usually used for mobile optimization	www.tiktok.com, www.canva.com, erp.fxpro.com

CSS Usage

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation semantics (the look and formatting) of a document written in a markup language (such as XHTML).

Technology	Description	Popular sites using this technology
External	Styles defined within an external CSS file	www.instagram.com, www.amazon.de, www.deepl.com
CSS Media Query	*No description*	www.bilibili.com, www.office.com, www.paypal.com



Site title	Bulb Security	**Date first seen**	March 2012
Site rank	186317	**Primary language**	English
Description	Bulb Security LLC was founded by Georgia Weidman, specializing in Information Security, Research and Training.		
Keywords	georgia weidman, bulb security, smartphone pentest framework, spf, DARPA Cyber Fast Track, metasploit training, security research, computer security training		

⊟ Network

Site	http://www.bulbsecurity.com	**Netblock Owner**	GoDaddy.com, LLC
Domain	bulbsecurity.com	**Nameserver**	ns65.domaincontrol.com
IP address	50.63.212.1	**DNS admin**	dns@jomax.net
IPv6 address	*Not Present*	**Reverse DNS**	p3nlhg344c1344.shr.prod.phx3.secureserver.net
Domain registrar	godaddy.com	**Nameserver organisation**	whois.wildwestdomains.com
Organisation	Domains By Proxy, LLC, Scottsdale, 85260, United States	**Hosting company**	GoDaddy Inc
Top Level Domain	Commercial entities (.com)	**DNS Security Extensions**	*unknown*
Hosting country	US		

⊟ Hosting History

Netblock owner	IP address	OS	Web server	Last seen Refresh
GoDaddy.com, LLC 14455 N Hayden Road Suite 226 Scottsdale AZ US 85260	50.63.212.1	Linux	Apache	1-Nov-2013
GoDaddy.com, LLC 14455 N Hayden Road Suite 226 Scottsdale AZ US 85260	50.63.202.81	-	Microsoft-IIS/7.5	22-Dec-2012
GoDaddy.com, LLC 14455 N Hayden Road Suite 226 Scottsdale AZ US 85260	50.63.212.1	-	Apache	18-Dec-2012

Figure 5-1: Netcraft's results for bulbsecurity.com

Whois Lookups

All domain registrars keep records of the domains they host. These records contain information about the owner, including contact information. For example, if we run the Whois command line tool on our Kali machine to query for information about *bulbsecurity.com*, as shown in Listing 5-1, we see that I used private registration, so we won't learn much.

```
root@kali:~# whois bulbsecurity.com
  Registered through: GoDaddy.com, LLC (http://www.godaddy.com)
   Domain Name: BULBSECURITY.COM
      Created on: 21-Dec-11
      Expires on: 21-Dec-12
      Last Updated on: 21-Dec-11

   Registrant: ❶
   Domains By Proxy, LLC
   DomainsByProxy.com
   14747 N Northsight Blvd Suite 111, PMB 309
   Scottsdale, Arizona 85260
   United States
   Technical Contact: ❷
      Private, Registration  BULBSECURITY.COM@domainsbyproxy.com
      Domains By Proxy, LLC
      DomainsByProxy.com
      14747 N Northsight Blvd Suite 111, PMB 309
      Scottsdale, Arizona 85260
      United States
      (480) 624-2599      Fax -- (480) 624-2598

   Domain servers in listed order:
      NS65.DOMAINCONTROL.COM ❸
      NS66.DOMAINCONTROL.COM
```

Listing 5-1: Whois information for bulbsecurity.com

This site has private registration, so both the registrant ❶ and technical contact ❷ are domains by proxy. Domains by proxy offer private registration, hiding your personal details in the Whois information for the domains you own. However, we do see the domain servers ❸ for *bulbsecurity.com*.

Running Whois queries against other domains will show more interesting results. For example, if you do a Whois lookup on *georgiaweidman.com*, you might get an interesting blast from the past, including my college phone number.

DNS Reconnaissance

We can also use Domain Name System (DNS) servers to learn more about a domain. DNS servers translate the human-readable URL *www.bulbsecurity.com* into an IP address.

Nslookup

For example, we could use a command line tool such as Nslookup, as shown in Listing 5-2.

```
root@Kali:~# nslookup www.bulbsecurity.com
Server:     75.75.75.75
Address:    75.75.75.75#53

Non-authoritative answer:
www.bulbsecurity.com    canonical name = bulbsecurity.com.
Name:    bulbsecurity.com
Address: 50.63.212.1 ❶
```

Listing 5-2: Nslookup information for www.bulbsecurity.com

Nslookup returned the IP address of *www.bulbsecurity.com*, as you can see at ❶.

We can also tell Nslookup to find the mail servers for the same website by looking for MX records (DNS speak for email), as shown in Listing 5-3.

```
root@kali:~# nslookup
> set type=mx
> bulbsecurity.com
Server:     75.75.75.75
Address:    75.75.75.75#53

Non-authoritative answer:
bulbsecurity.com    mail exchanger = 40 ASPMX2.GOOGLEMAIL.com.
bulbsecurity.com    mail exchanger = 20 ALT1.ASPMX.L.GOOGLE.com.
bulbsecurity.com    mail exchanger = 50 ASPMX3.GOOGLEMAIL.com.
bulbsecurity.com    mail exchanger = 30 ALT2.ASPMX.L.GOOGLE.com.
bulbsecurity.com    mail exchanger = 10 ASPMX.L.GOOGLE.com.
```

Listing 5-3: Nslookup information for bulbsecurity.com*'s mail servers*

Nslookup says *bulbsecurity.com* is using Google Mail for its email servers, which is correct because I use Google Apps.

Host

Another utility for DNS queries is Host. We can ask Host for the name servers for a domain with the command `host -t ns` *`domain`*. A good example for domain queries is *zoneedit.com,* a domain set up to demonstrate zone transfer vulnerabilities, as shown here.

```
root@kali:~# host -t ns zoneedit.com
zoneedit.com name server ns4.zoneedit.com.
zoneedit.com name server ns3.zoneedit.com.
--snip--
```

This output shows us all the DNS servers for *zoneedit.com.* Naturally, because I mentioned that this domain was set up to demonstrate zone transfers, that's what we are going to do next.

Zone Transfers

DNS zone transfers allow name servers to replicate all the entries about a domain. When setting up DNS servers, you typically have a primary name server and a backup server. What better way to populate all the entries in the secondary DNS server than to query the primary server for all of its entries?

Unfortunately, many system administrators set up DNS zone transfers insecurely, so that anyone can transfer the DNS records for a domain. *zoneedit.com* is an example of such a domain, and we can use the `host` command to download all of its DNS records. Use the `-l` option to specify the domain to transfer, and choose one of the name servers from the previous command, as shown in Listing 5-4.

```
root@kali:~# host -l zoneedit.com ns2.zoneedit.com
Using domain server:
Name: ns2.zoneedit.com
Address: 69.72.158.226#53
Aliases:

zoneedit.com name server ns4.zoneedit.com.
zoneedit.com name server ns3.zoneedit.com.
zoneedit.com name server ns15.zoneedit.com.
zoneedit.com name server ns8.zoneedit.com.
zoneedit.com name server ns2.zoneedit.com.
zoneedit.com has address 64.85.73.107
www1.zoneedit.com has address 64.85.73.41
dynamic.zoneedit.com has address 64.85.73.112
bounce.zoneedit.com has address 64.85.73.100
--snip--
mail2.zoneedit.com has address 67.15.232.182
--snip--
```

Listing 5-4: Zone transfer of zoneedit.com

There are pages and pages of DNS entries for *zoneedit.com*, which gives us a good idea of where to start in looking for vulnerabilities for our pentest. For example, *mail2.zoneedit.com* is probably a mail server, so we should look for potentially vulnerable software running on typical email ports such as 25 (Simple Mail Transfer Protocol) and 110 (POP3). If we can find a webmail server, any usernames we find may lead us in the right direction so that we can guess passwords and gain access to sensitive company emails.

Searching for Email Addresses

External penetration tests often find fewer services exposed than internal ones do. A good security practice is to expose only those services that must be accessed remotely, like web servers, mail servers, VPN servers, and maybe SSH or FTP, and only those services that are mission critical. Services like these are common attack surfaces, and unless employees use two-factor authentication, accessing company webmail can be simple if an attacker can guess valid credentials.

One excellent way to find usernames is by looking for email addresses on the Internet. You might be surprised to find corporate email addresses publicly listed on parent-teacher association contact info, sports team rosters, and, of course, social media.

You can use a Python tool called theHarvester to quickly scour thousands of search engine results for possible email addresses. theHarvester can automate searching Google, Bing, PGP, LinkedIn, and others for email addresses. For example, in Listing 5-5, we'll look at the first 500 results in all search engines for *bulbsecurity.com*.

root@kali:~# theharvester -h

```
root@kali:~# theharvester -d bulbsecurity.com -l 500 -b all

*******************************************************************
*                                                                 *
* | |_| |__   ___    /\  /\__ _ _ ____   _____  ___| |_ ___ _ __  *
* | __| '_ \ / _ \  / /_/ / _` | '__\ \ / / _ \/ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* TheHarvester Ver. 2.2a                                          *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*******************************************************************


Full harvest..
[-] Searching in Google..
    Searching 0 results...
    Searching 100 results...
    Searching 200 results...
    Searching 300 results...
--snip--

 [+] Emails found:
------------------
georgia@bulbsecurity.com

[+] Hosts found in search engines:
------------------------------------
50.63.212.1:www.bulbsecurity.com

--snip--
```

Listing 5-5: Running theHarvester against bulbsecurity.com

There's not too much to be found for *bulbsecurity.com*, but theHarvester does find my email address, *georgia@bulbsecurity.com*, and the website, *www.bulbsecurity.com*, as well as other websites I share virtual hosting with. You may find more results if you run theHarvester against your organization.