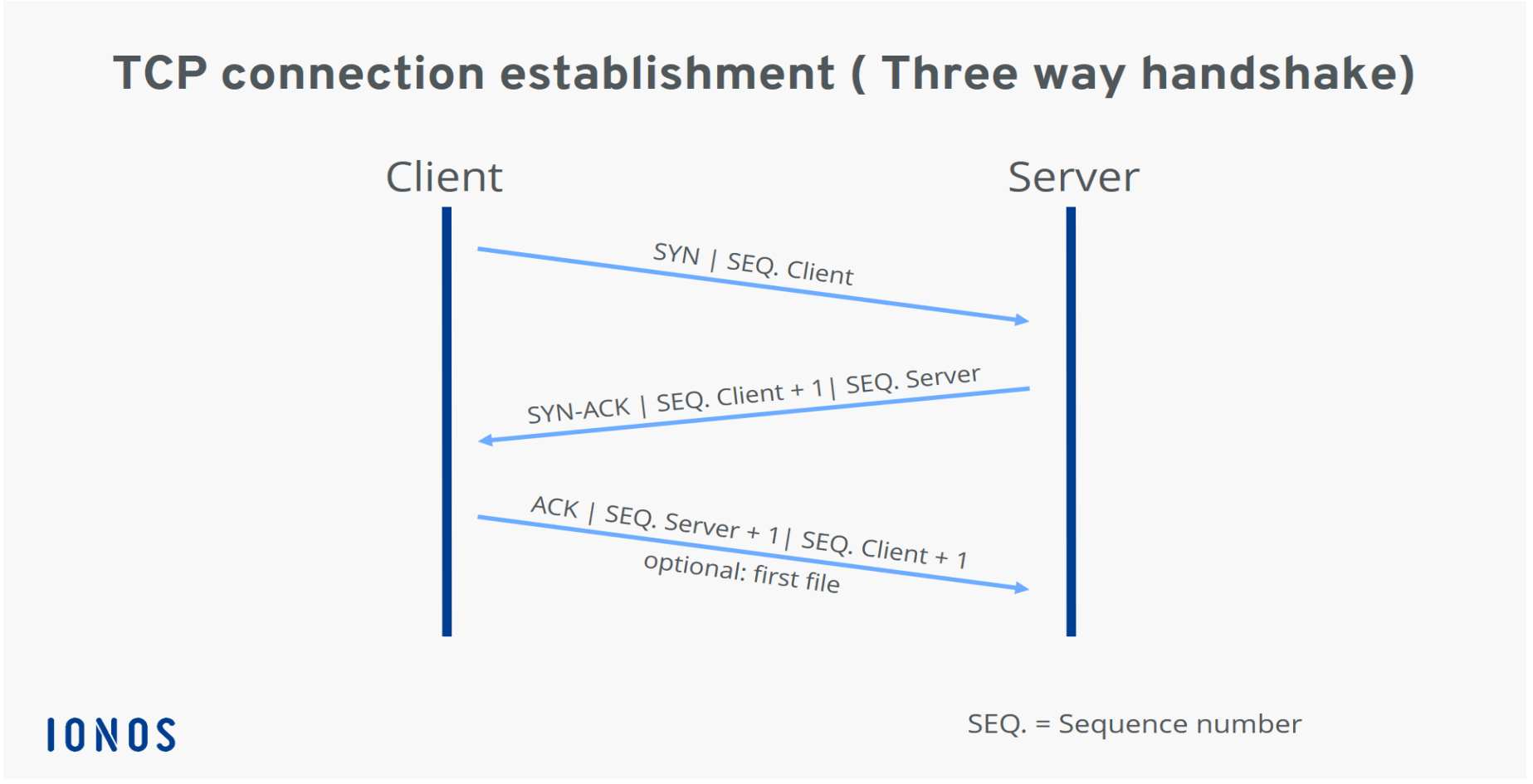
Unit 1:

Introduction to web technology and information Gathering

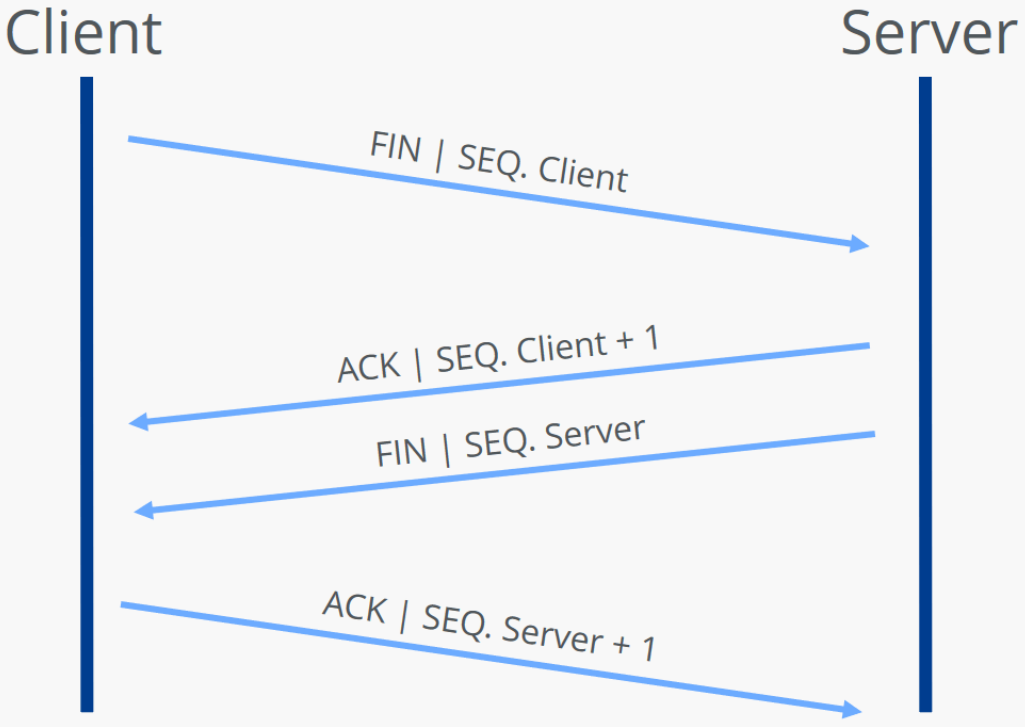
TCP (Transmission Control Protocol)

- The Transmission Control Protocol, or TCP protocol for short, is a standard for exchanging data between different devices in a computer network.
- The current version of the TCP protocol allows two way transmission of data.
- The term TCIP/IP protocol stack is also commonly used to refer to the Internet protocol suite since the TCP protocol is almost always based on the Internet protocol (IP) and this connection is the foundation for the majority of public and local networks and network services.

How exactly do TCP connections work?



TCP connection termination (TCP Teardown)

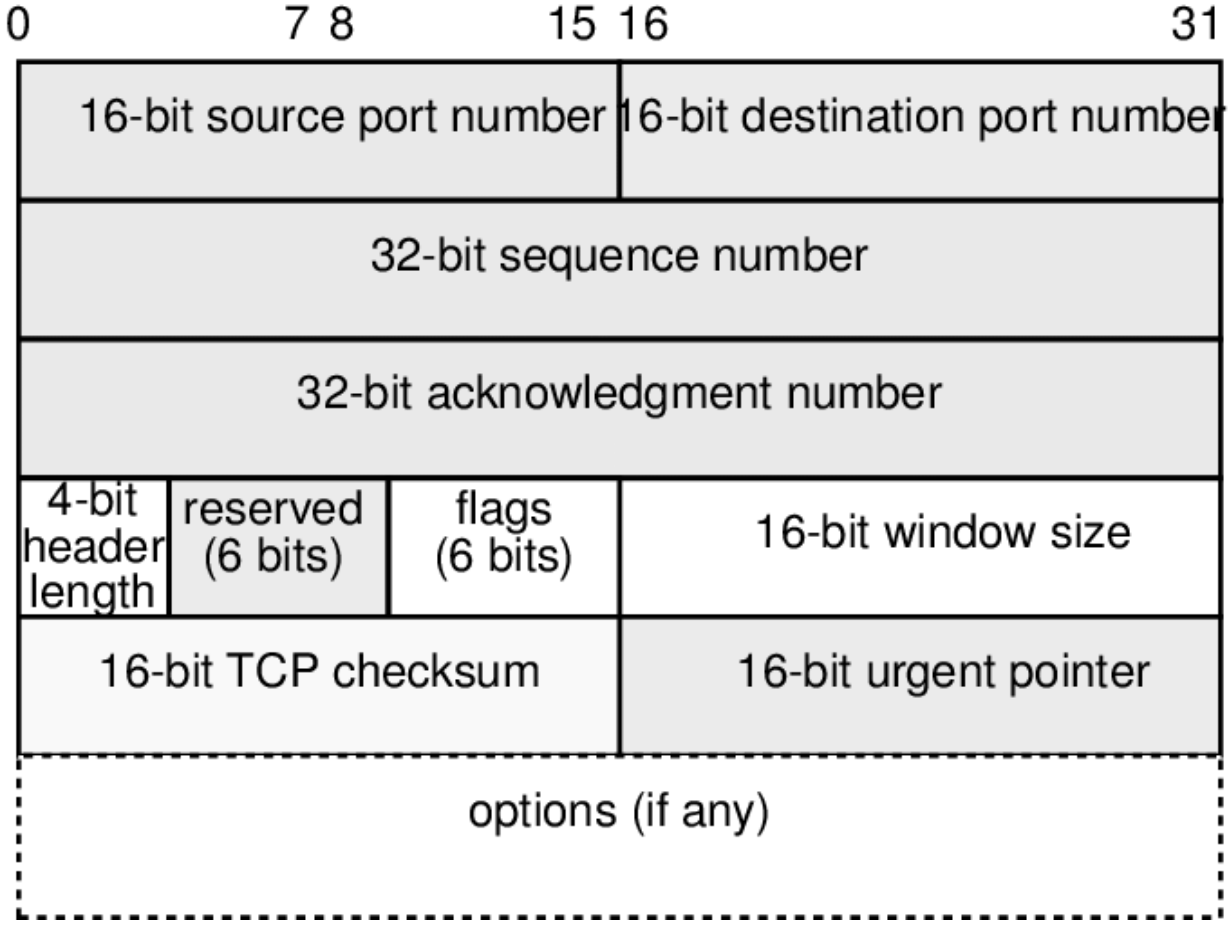


SEQ. = Sequence number

TCP Layers

<table>
<tr><th>OSI Model</th><th>TCP/IP Model</th></tr>
<tr><td>Application Layer</td><td rowspan="3">Application layer</td></tr>
<tr><td>Presentation Layer</td></tr>
<tr><td>Session Layer</td></tr>
<tr><td>Transport Layer</td><td>Transport Layer</td></tr>
<tr><td>Network Layer</td><td>Internet Layer</td></tr>
<tr><td>Data link layer</td><td rowspan="2">Link Layer</td></tr>
<tr><td>Physical layer</td></tr>
</table>

What is the structure of the TCP header?



Source port (16 bits): Identifies the port number of the sender.

Destination port (16 bits): Identifies the port number of receiver.

Sequence number (32 bits): The sequence number specifies the first byte of attached payload data or is sent when the connection is established or terminated. It is also used for validating and sorting the segments after transmission.

Acknowledgment number (32 bits): This field contains the next sequence number that the sender is expecting. An ACK flag (in the offset field) is a precondition for validity.

Offset (4 bits): The “offset”field specifies the length of the TCP header in 32-bit words to highlight the starting point of the payload data. This starting point varies from segment to segment due to the variable ofset field.

Reserved(6 bits): Reserved for future use according to RFC 793 and not yet in use. This field must always be set to 0.

Flags (6 bits): The six possible single bits in the Flags field enable various TCP actions for organizing communication and data processing. The following flags are either set or not set for these actions:

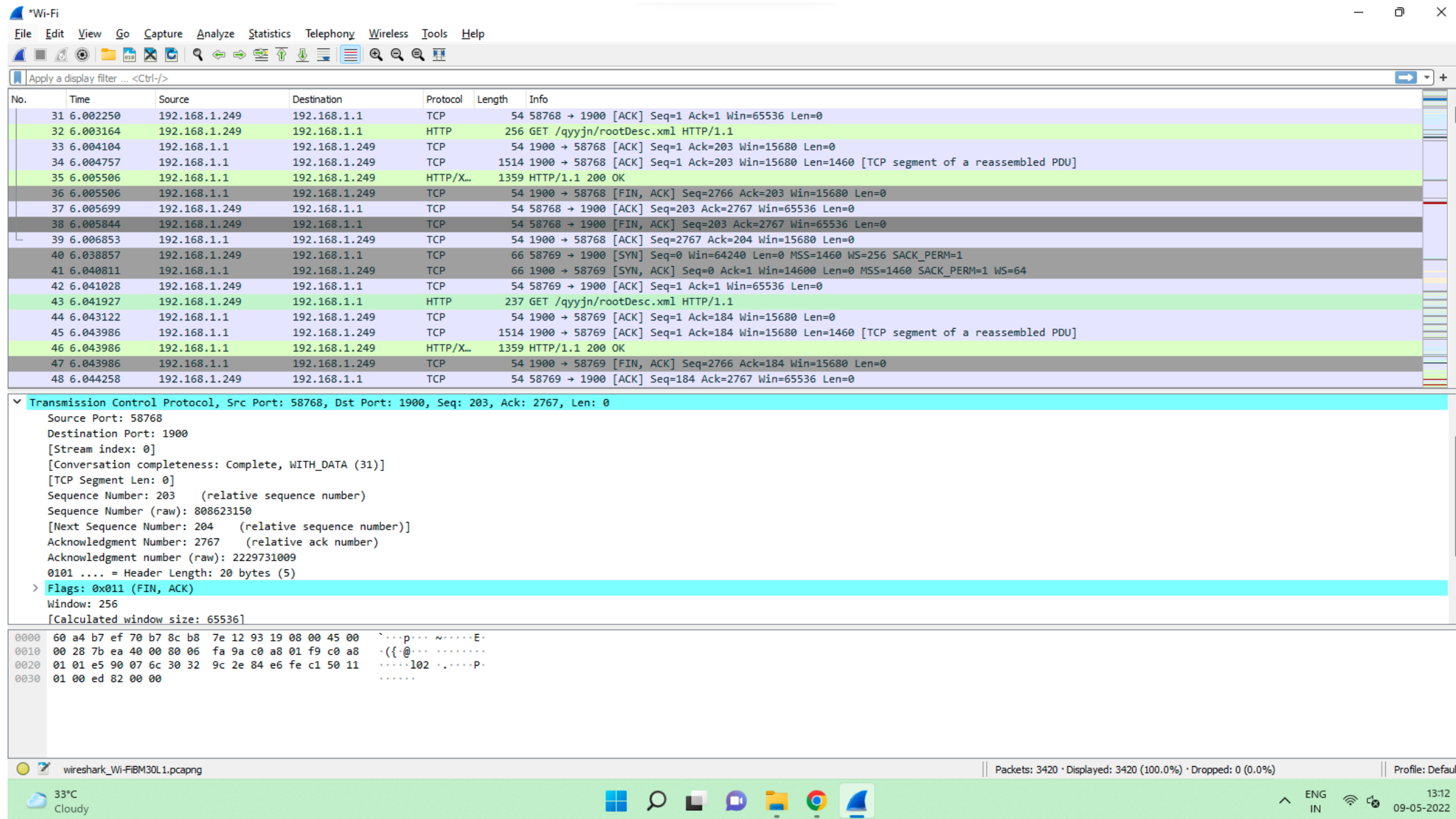
- **URG**: The "Urgent" flag signals to the TCP application that the payload data must be processed immediately up to the set Urgent pointer (see above).
- **ACK**: In combination with the acknowledgment number, the ACK flag acknowledges the receipt of TCP packets. If the flag is not set, the confirmation number is also invalid.
- **PSH**: The "Push" flag ensures that a TCP segment is immediately pushed through without first being sent to the buffer of the sender and receiver.
- **RST**: If there is an error during transmission, a TCP packet with the RST flag set can be used to reset the connection.
- **SYN**: Messages that have SYN flag set represent the first step of the three-way handshake, meaning they initiate the connection.
- **FIN**: The "Finish" flag signals to the other party that a sender is ending the transmission.

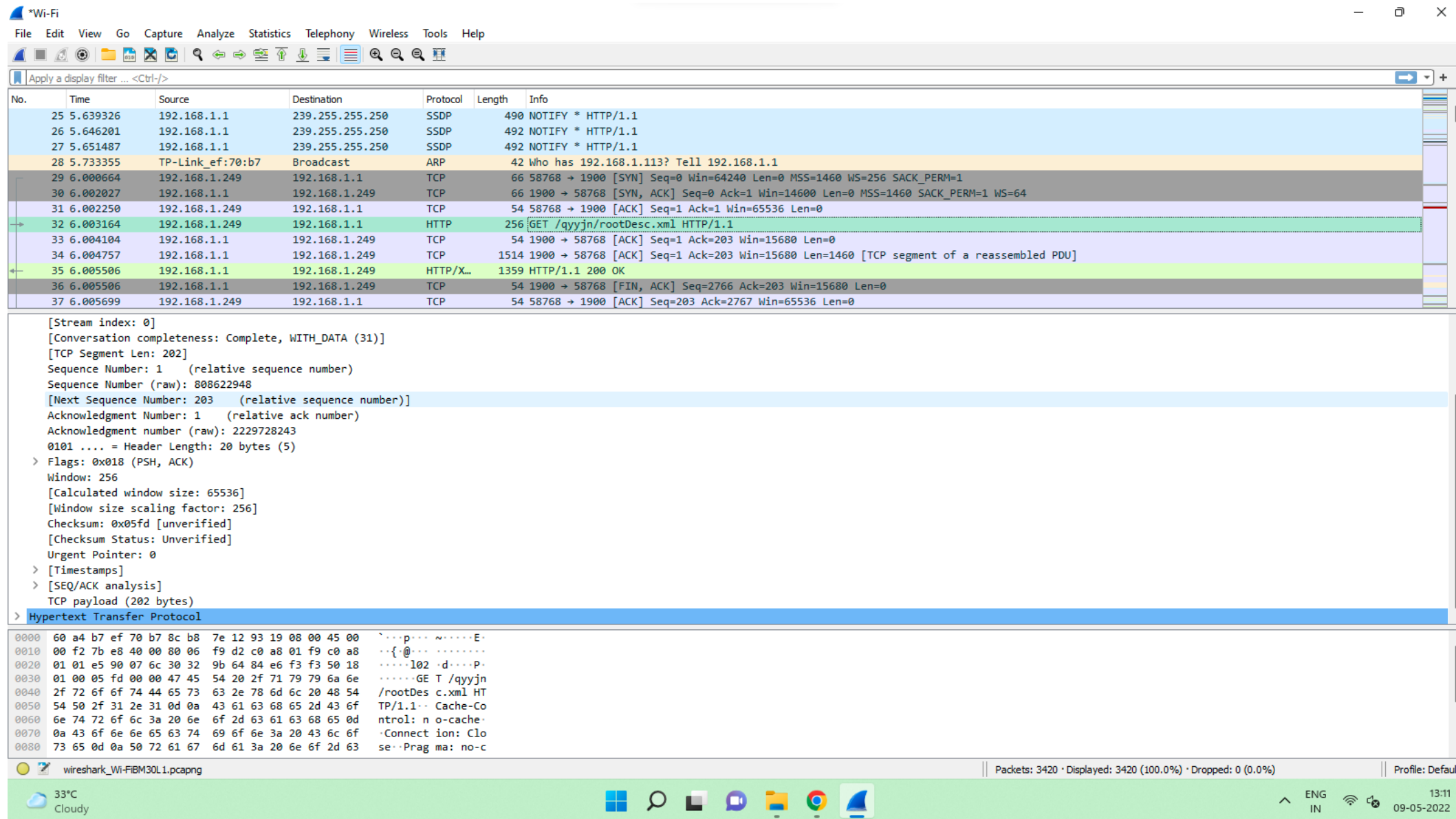
Window size (16 bits): This field specifies the number of bytes that the sender is willing to receive.

Checksum (16 bits): The Transmission Control Protocol can reliably detect transmission errors. The checksum calculated from the header, the payload data and the pseudo-header is used for this purpose.

Urgent pointer (16 bits): The urgent pointer indicates the position of the first byte after the payload data that is to be processed urgently. As a result, this field is only valid and relevant if the URG flag is set.

Options (0 - 320 bits): Use the Options field if you want to include TCP functions that belong in the general header, for example if you want to define the maximum segment size. The length of the options must always be a multiple of 32, otherwise zero-bit padding is required.


*Wi-Fi
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
No. Time Source Destination Protocol Length Info
31 6.002250 192.168.1.249 192.168.1.1 TCP 54 58768 → 1900 [ACK] Seq=1 Ack=1 Win=65536 Len=0
32 6.003164 192.168.1.249 192.168.1.1 HTTP 256 GET /qyyjn/rootDesc.xml HTTP/1.1
33 6.004104 192.168.1.1 192.168.1.249 TCP 54 1900 → 58768 [ACK] Seq=1 Ack=203 Win=15680 Len=0
34 6.004757 192.168.1.1 192.168.1.249 TCP 1514 1900 → 58768 [ACK] Seq=1 Ack=203 Win=15680 Len=1460 [TCP segment of a reassembled PDU]
35 6.005506 192.168.1.1 192.168.1.249 HTTP/X… 1359 HTTP/1.1 200 OK
36 6.005506 192.168.1.1 192.168.1.249 TCP 54 1900 → 58768 [FIN, ACK] Seq=2766 Ack=203 Win=15680 Len=0
37 6.005699 192.168.1.249 192.168.1.1 TCP 54 58768 → 1900 [ACK] Seq=203 Ack=2767 Win=65536 Len=0
38 6.005844 192.168.1.249 192.168.1.1 TCP 54 58768 → 1900 [FIN, ACK] Seq=203 Ack=2767 Win=65536 Len=0
39 6.006853 192.168.1.1 192.168.1.249 TCP 54 1900 → 58768 [ACK] Seq=2767 Ack=204 Win=15680 Len=0
40 6.038857 192.168.1.249 192.168.1.1 TCP 66 58769 → 1900 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 WS=256 SACK_PERM=1
41 6.040811 192.168.1.1 192.168.1.249 TCP 66 1900 → 58769 [SYN, ACK] Seq=0 Ack=1 Win=14600 Len=0 MSS=1460 SACK_PERM=1 WS=64
42 6.041028 192.168.1.249 192.168.1.1 TCP 54 58769 → 1900 [ACK] Seq=1 Ack=1 Win=65536 Len=0
43 6.041927 192.168.1.249 192.168.1.1 HTTP 237 GET /qyyjn/rootDesc.xml HTTP/1.1
44 6.043122 192.168.1.1 192.168.1.249 TCP 54 1900 → 58769 [ACK] Seq=1 Ack=184 Win=15680 Len=0
45 6.043986 192.168.1.1 192.168.1.249 TCP 1514 1900 → 58769 [ACK] Seq=1 Ack=184 Win=15680 Len=1460 [TCP segment of a reassembled PDU]
46 6.043986 192.168.1.1 192.168.1.249 HTTP/X… 1359 HTTP/1.1 200 OK
47 6.043986 192.168.1.1 192.168.1.249 TCP 54 1900 → 58769 [FIN, ACK] Seq=2766 Ack=184 Win=15680 Len=0
48 6.044258 192.168.1.249 192.168.1.1 TCP 54 58769 → 1900 [ACK] Seq=184 Ack=2767 Win=65536 Len=0
Transmission Control Protocol, Src Port: 58768, Dst Port: 1900, Seq: 203, Ack: 2767, Len: 0
Source Port: 58768
Destination Port: 1900
[Stream index: 0]
[Conversation completeness: Complete, WITH_DATA (31)]
[TCP Segment Len: 0]
Sequence Number: 203 (relative sequence number)
Sequence Number (raw): 808623150
[Next Sequence Number: 204 (relative sequence number)]
Acknowledgment Number: 2767 (relative ack number)
Acknowledgment number (raw): 2229731009
0101 .... = Header Length: 20 bytes (5)
Flags: 0x011 (FIN, ACK)
Window: 256
[Calculated window size: 65536]
0000 60 a4 b7 ef 70 b7 8c b8 7e 12 93 19 08 00 45 00
0010 00 28 7b ea 40 00 80 06 fa 9a c0 a8 01 f9 c0 a8
0020 01 01 e5 90 07 6c 30 32 9c 2e 84 e6 fe c1 50 11
0030 01 00 ed 82 00 00
wireshark_Wi-FiBM30L1.pcapng
Packets: 3420 · Displayed: 3420 (100.0%) · Dropped: 0 (0.0%)
Profile: Defaul
33°C Cloudy
ENG IN
13:12 09-05-2022


*Wi-Fi
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
No. Time Source Destination Protocol Length Info
25 5.639326 192.168.1.1 239.255.255.250 SSDP 490 NOTIFY * HTTP/1.1
26 5.646201 192.168.1.1 239.255.255.250 SSDP 492 NOTIFY * HTTP/1.1
27 5.651487 192.168.1.1 239.255.255.250 SSDP 492 NOTIFY * HTTP/1.1
28 5.733355 TP-Link_ef:70:b7 Broadcast ARP 42 Who has 192.168.1.113? Tell 192.168.1.1
29 6.000664 192.168.1.249 192.168.1.1 TCP 66 58768 → 1900 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 WS=256 SACK_PERM=1
30 6.002027 192.168.1.1 192.168.1.249 TCP 66 1900 → 58768 [SYN, ACK] Seq=0 Ack=1 Win=14600 Len=0 MSS=1460 SACK_PERM=1 WS=64
31 6.002250 192.168.1.249 192.168.1.1 TCP 54 58768 → 1900 [ACK] Seq=1 Ack=1 Win=65536 Len=0
32 6.003164 192.168.1.249 192.168.1.1 HTTP 256 GET /qyyjn/rootDesc.xml HTTP/1.1
33 6.004104 192.168.1.1 192.168.1.249 TCP 54 1900 → 58768 [ACK] Seq=1 Ack=203 Win=15680 Len=0
34 6.004757 192.168.1.1 192.168.1.249 TCP 1514 1900 → 58768 [ACK] Seq=1 Ack=203 Win=15680 Len=1460 [TCP segment of a reassembled PDU]
35 6.005506 192.168.1.1 192.168.1.249 HTTP/X… 1359 HTTP/1.1 200 OK
36 6.005506 192.168.1.1 192.168.1.249 TCP 54 1900 → 58768 [FIN, ACK] Seq=2766 Ack=203 Win=15680 Len=0
37 6.005699 192.168.1.249 192.168.1.1 TCP 54 58768 → 1900 [ACK] Seq=203 Ack=2767 Win=65536 Len=0
[Stream index: 0]
[Conversation completeness: Complete, WITH_DATA (31)]
[TCP Segment Len: 202]
Sequence Number: 1 (relative sequence number)
Sequence Number (raw): 808622948
[Next Sequence Number: 203 (relative sequence number)]
Acknowledgment Number: 1 (relative ack number)
Acknowledgment number (raw): 2229728243
0101 .... = Header Length: 20 bytes (5)
Flags: 0x018 (PSH, ACK)
Window: 256
[Calculated window size: 65536]
[Window size scaling factor: 256]
Checksum: 0x05fd [unverified]
[Checksum Status: Unverified]
Urgent Pointer: 0
[Timestamps]
[SEQ/ACK analysis]
TCP payload (202 bytes)
Hypertext Transfer Protocol
0000 60 a4 b7 ef 70 b7 8c b8 7e 12 93 19 08 00 45 00 `···p··· ~·····E·
0010 00 f2 7b e8 40 00 80 06 f9 d2 c0 a8 01 f9 c0 a8 ··{·@··· ····
0020 01 01 e5 90 07 6c 30 32 9b 64 84 e6 f3 f3 50 18 ·····l02 ·d····P·
0030 01 00 05 fd 00 00 47 45 54 20 2f 71 79 79 6a 6e ·····GE T /qyyjn
0040 2f 72 6f 6f 74 44 65 73 63 2e 78 6d 6c 20 48 54 /rootDes c.xml HT
0050 54 50 2f 31 2e 31 0d 0a 43 61 63 68 65 2d 43 6f TP/1.1·· Cache-Co
0060 6e 74 72 6f 6c 3a 20 6e 6f 2d 63 61 63 68 65 0d ntrol: n o-cache·
0070 0a 43 6f 6e 6e 65 63 74 69 6f 6e 3a 20 43 6c 6f ·Connect ion: Clo
0080 73 65 0d 0a 50 72 61 67 6d 61 3a 20 6e 6f 2d 63 se··Prag ma: no-c
wireshark_Wi-FiBM30L1.pcapng
Packets: 3420 · Displayed: 3420 (100.0%) · Dropped: 0 (0.0%)
Profile: Defaul
33°C Cloudy
ENG IN
13:11 09-05-2022

HTTP/S Protocol Basics

URL Example

URL Example With Port

Response Status codes

They are split into 5 groups each group has a meaning and a three digit code.

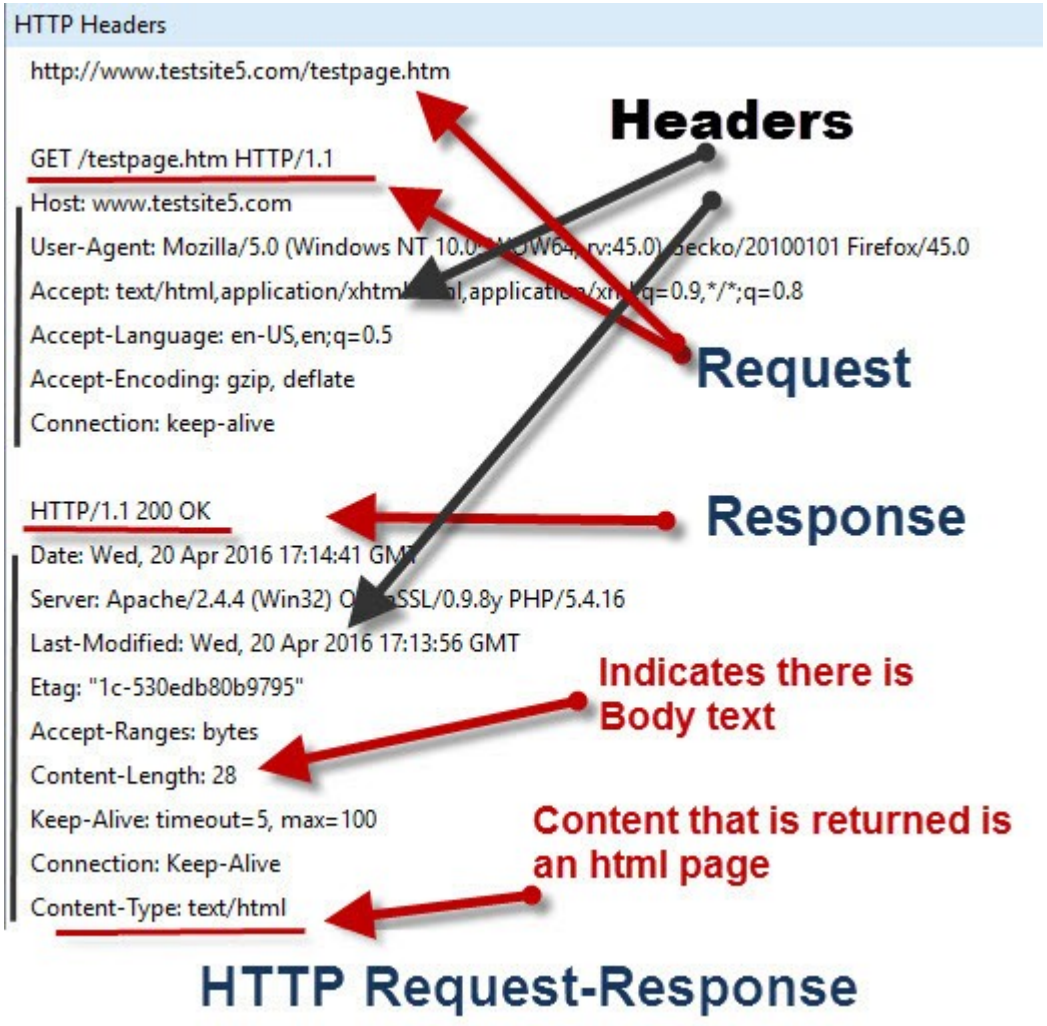
1xx - Informational

2xx - Successful

3xx -Multiple Choice

4xx - Client Error

5xx -Server Error



Request Types, Methods or Verbs

The HTTP protocol now support 8 request types, also called methods or verbs in the documentation,they are:

GET - Requesting resource from server

POST - submitting a resource to a server (e.g. file uploads)

PUT - As POST but replaces a resource

DELETE- Delete a resource from a server

HEAD - As GET but only return headers and not content

OPTIONS - Get the options for the resource

PATCH -Apply modifications to a resource

TRACE -Performs message loop-back

Encoding

- Encoding is the process of converting data from one form to another.
- There are several types of encoding, including image encoding, audio and video encoding, and character encoding.
- Media files are often encoded to save disk space. By encoding digital audio, video, and image files, they can be saved in a more efficient, compressed format.
- eg. .wave ->.mp3 , .MPG -> .DV,
- Character encoding : converting character to bytes (use codec for conversion).

Origin server

The purpose of an origin server is to process and respond to incoming internet requests from internet clients. The concept of an origin server is typically used in conjunction with the concept of an edge server or caching server.

At its core, an origin server is a computer running one or more programs that are designed to listen for and process incoming internet requests. An origin server can take on all the responsibility of serving up the content for an internet property such as a website, provided that the traffic does not extend beyond what the server is capable of processing and latency is not a primary concern.

Cookies

What Are Cookies?

Cookies are text files with small pieces of data — like a username and password — that are used to identify your computer as you use a computer network. Specific cookies known as HTTP cookies are used to identify specific users and improve your web browsing experience.

Data stored in a cookie is created by the server upon your connection. This data is labeled with an ID unique to you and your computer.

When the cookie is exchanged between your computer and the network server, the server reads the ID and knows what information to specifically serve to you.

Cookies

Different types of cookies

"**Magic cookies**" would be used for a login to computer database systems, such as a business internal network. This concept predates the modern "cookie" we use today.

HTTP cookies are a repurposed version of the "magic cookie" built for internet browsing. He recreated this concept for browsers when he helped an online shopping store fix their overloaded servers.

The HTTP cookie is what we currently use to manage our online experiences. It is also what some malicious people can use to spy on your online activity and steal your personal info.

Cookies

What Are Cookies Used For?

Session management: cookies let websites recognize users and recall their individual login information and preferences.

Personalization: Customized advertising is the main way cookies are used to personalize your sessions.

Tracking: Shopping sites use cookies to track items users previously viewed, allowing the sites to suggest other goods they might like.

Why Cookies Can Be Dangerous ?

Sessions

A session is the total time devoted to an activity. In computer systems, a user session begins when a user logs in to or accesses a particular computer, network, or software service. It ends when the user logs out of the service, or shuts down the computer.

A session can temporarily store information related to the activities of the user while connected. A cookie is used in web pages for storing information in case the user leaves the web page or closes down their Internet browser.

Sessions

```
<?php
session_start();
?>


<form action ="session2.php" method="POST">
    <label for="studentName">Student name:
</label>
    <input type="text" id="studentName"
name="studentName">
    <button type="submit">Click here to add a
student</button>
</form>
```

```
<?php
session_start();
$studentName = $_POST['studentName'];
if(!isset($listOfStudents)) {
    $listOfStudents = $_SESSION['listOfStudents'];
}
$listOfStudents[] = $studentName;
$_SESSION['listOfStudents'] = $listOfStudents;
//echo "you are trying to add $studentName to our students";
foreach($listOfStudents as $i) {
    echo "$i is a student in our class <br>"; }
?>
<br> <br>
<a href="session.php">Click here to add another student</a>
```

Finger printing the web server

Web server fingerprinting is one of the critical task for a penetration tester. By knowing the version and type of a running web server allows testers to determine known vulnerabilities and the appropriate exploits to use during testing.

Geting the information about the types and version of the services that uses in the web server helps to find known vuln and exploites for that services during the test.

Tools for web server fingerprinting

https://linuxsecurity.expert/security-tools/web-application-fingerprinting-tools

https://infosecland.com/tools/fingerprint

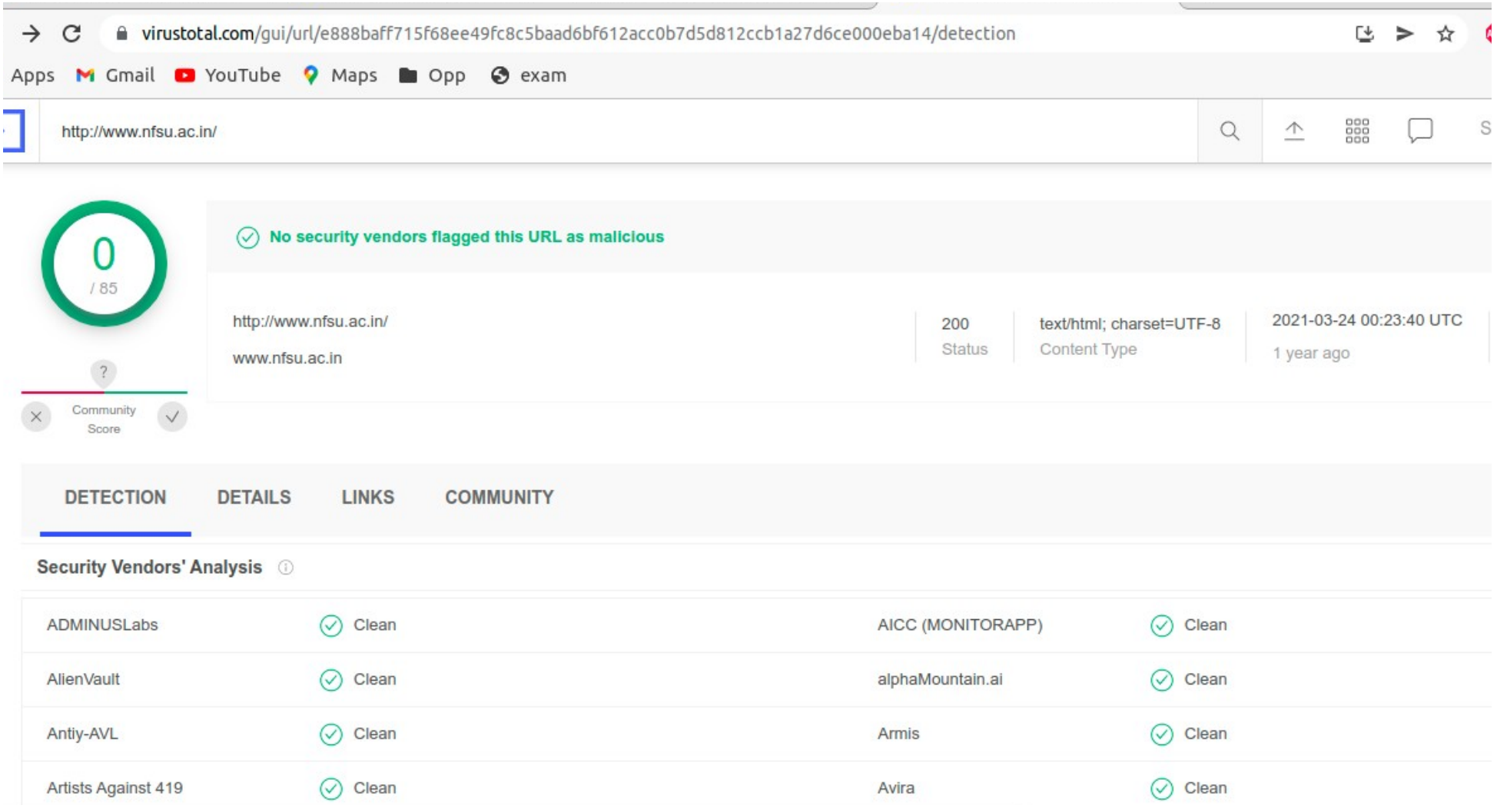
Subdomain’s enumeration

Subdomain enumeration is an information gathering technique. It can be used to define the all sites opened to the internet in a company. In large organizations, it is very common to have some forgotten websites that having vulnerabilities or some sensitive data.

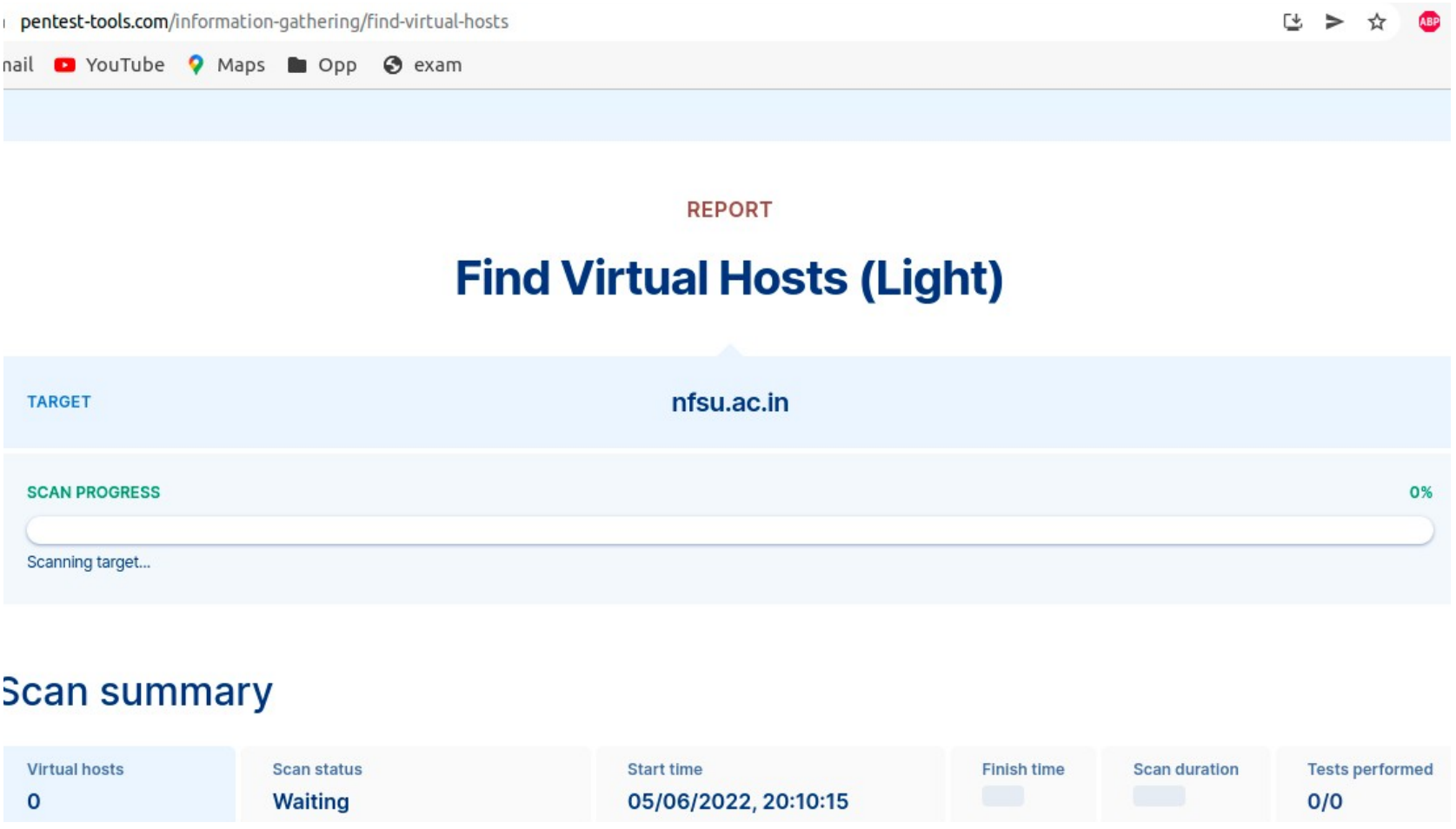
"v=spf1 include:_spf.google.com include:spf.mandrillapp.com ~all"

Host Records (A) ** this data may not be current as it uses a static database (updated monthly)

result.nfsu.ac.in HTTP: Apache/2.4.48 (Unix) OpenSSL/1.0.2k-fips HTTPS: Apache/2.4.48 (Unix) OpenSSL/1.0.2k-fips HTTP TECH: Apache,2.4.48 WordPress HTTPS TECH: Apache,2.4.48 WordPress	3.109.218.170 ec2-3-109-218-170.ap-south-1.compute.amazonaws.com	AMAZON-02 India
admission.nfsu.ac.in	192.46.215.110 admission.nfsu.ac.in	LINODE-AP Linode, LLC India
online.nfsu.ac.in HTTP: Apache HTTPS: Apache	196.12.45.114	SOFTNET-AS-AP Software Technology Parks of India - Bangalore India
law.nfsu.ac.in	13.232.53.47 ec2-13-232-53-47.ap-south-1.compute.amazonaws.com	AMAZON-02 India
career.nfsu.ac.in	192.46.215.110 admission.nfsu.ac.in	LINODE-AP Linode, LLC India
recruitment.nfsu.ac.in HTTP: Apache/2.4.48 (Unix) OpenSSL/1.0.2k-fips	3.109.218.170 ec2-3-109-218-170.ap-south-1.compute.amazonaws.com	AMAZON-02 India


virustotal.com/gui/url/e888baff715f68ee49fc8c5baad6bf612acc0b7d5d812ccb1a27d6ce000eba14/detection
Apps
Gmail
YouTube
Maps
Opp
exam
http://www.nfsu.ac.in/
0
/ 85
No security vendors flagged this URL as malicious
http://www.nfsu.ac.in/
www.nfsu.ac.in
200
Status
text/html; charset=UTF-8
Content Type
2021-03-24 00:23:40 UTC
1 year ago
Community
Score
DETECTION
DETAILS
LINKS
COMMUNITY
Security Vendors' Analysis
ADMINUSLabs
Clean
AICC (MONITORAPP)
Clean
AlienVault
Clean
alphaMountain.ai
Clean
Antiy-AVL
Clean
Armis
Clean
Artists Against 419
Clean
Avira
Clean

finding virtual hosts



finding virtual hosts (what to scan for ?)

Light vs Full Scan

Using the full version of the Find Subdomains scanner allows you to discover more subdomains with additional subdomain discovery techniques.

TESTING AREAS

- DNS records (NS, MX, TXT, AXFR)
- DNS Enumeration
- Common configuration issues
- Certificate Transparency Logs
- HTML links
- SSL certificates
- Google and Bing search
- External APIs
- Reverse DNS enumeration
- Smart DNS search

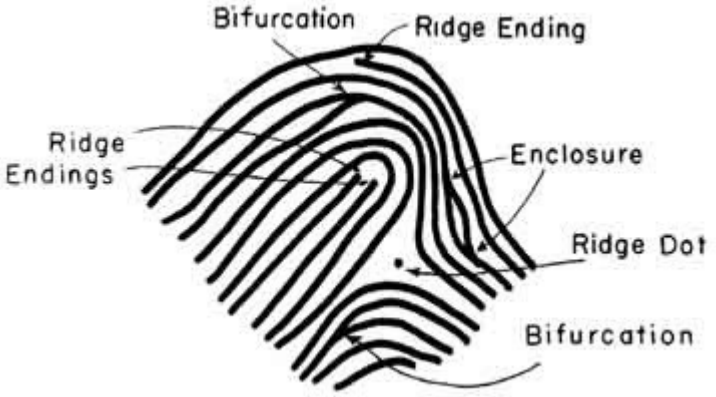
Finger printing custom applications

When performing a Web Application Security Assessment, an important step is Fingerprinting which allows for further exploitation by an attacker. So as a security researcher/pentester, we should do well at fingerprinting the web server, which gives lot of information like application name, software version, web server info, OS, and more. This helps for known vulnerabilities, researching vulnerabilities and exploiting.

So here I will discuss some techniques which are required for this task:

https://resources.infosecinstitute.com/topic/finger-printing-print-the-finger-of-an-application/

Finger Print



Basic and composite ridge characteristics (minutiae)

Minutiae	Example	Minutiae	Example
ridge ending		bridge	
bifurcation		double bifurcation	
dot		trifurcation	
island (short ridge)		opposed bifurcations	
lake (enclosure)		ridge crossing	
hook (spur)		opposed bifurcation/ridge ending	

Server Fingerprinting

The "fingerprint" is a **hash of the server's public key**. It is not the public key itself; however, hash functions are so that it is not feasible to actually compute two public keys with the same fingerprint.

Therefore, if you want to know whether you are talking to the right server (and not some impersonator), then you "just" need to compute the server's key fingerprint (from the public key that the server just sent to you) and compare it with a "reference fingerprint".

How to calculate Hash value ?

https://cyberops.in/blog/how-to-calculate-hash-value/

```
import java.io.*;

class GFG {
	public static void main(String[] args)
	{
		String str = "GFG";
		System.out.println(str);

		int hashCode = str.hashCode();
		System.out.println(hashCode);
	}
}
```

Enumerating resources

Enumeration is defined as a process which establishes an active connection to the target hosts to discover potential attack vectors in the system, and the same can be used for further exploitation of the system. Enumeration is used to gather the following:

- Usernames, group names
- Hostnames
- Network shares and services
- IP tables and routing tables
- Service settings and audit configurations
- Application and banners
- SNMP and DNS details

Enumerating resources

Number	Name of the tool	Web links
01	OpUtils	https://www.manageengine.com/products/oputils/
02	SolarWinds	http://www.solarwinds.com/
03	SNScan	http://www.mcafee.com/us/downloads/free-tools/snscan.aspx
04	SNMP Scanner	http://www.secure-bytes.com/snmp-scanner.php
05	NS Auditor	http://www.nsauditor.com/

Enumerating resources

Number	Name of the tool	Web links
01	Softerra LDAP Administrator	http://www.ldapadministrator.com/
02	Jxplorer	http://jxplorer.org/
03	active directory domain services management pack for system center	https://www.microsoft.com/en-in/download/details.aspx?id=21357
04	LDAP Admin Tool	http://www.ldapadmin.org/
05	LDAP Administrator tool	https://sourceforge.net/projects/ldapadmin/

Relevant information through misconfigurations

Misconfigurations are often seen as an easy target, as it can be easy to detect on misconfigured web servers, cloud and applications and then becomes exploitable, causing significant harm and leading to catastrophic data leakage issues for enterprises.

What is Security Misconfiguration?

Security misconfiguration is the implementation of improper security controls, such as for servers or application configurations, network devices, etc. that may lead to security vulnerabilities.

For example, insecure configuration of web applications could lead to numerous security flaws including:

Incorrect folder permissions

Default passwords or username

Setup/Configuration pages enabled

Debugging enabled

Google hacking

Google hacking, sometimes, referred to as Google dorking, is **an information gathering technique** used by an attacker leveraging advanced Google searching techniques. Google hacking search queries can be used to identify security vulnerabilities in web applications, gather information for arbitrary or individual targets, discover error messages disclosing sensitive information, discover files containing credentials and other sensitive data.

Resources :

https://www.exploit-db.com/google-hacking-database

https://pentest-tools.com/information-gathering/google-hacking

https://www.ma-no.org/en/security/google-hacking-secrets-the-hidden-codes-of-google

https://www.oakton.edu/user/2/rjtaylor/CIS101/Google%20Hacking%20101.pdf