CS489/689 Privacy, Cryptography, Network and Data Security

Confidentiality

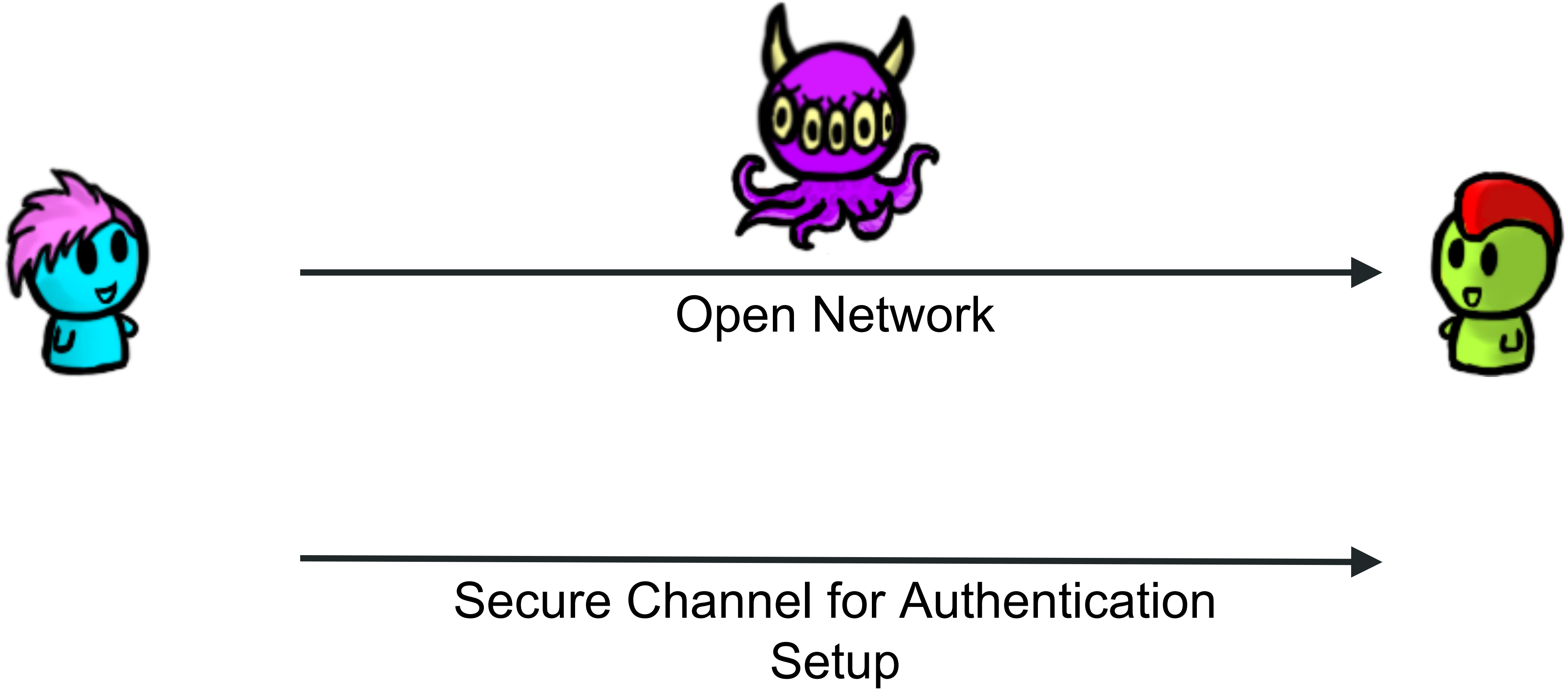
Today's Lecture - Confidentiality

- Traffic snooping
  - Tools and information we can collect
  - Solution is encryption
- What about the data encryption can not protect?
  - Problems and solutions

Confidentiality Goals

- Two parties: Alice and Bob
- They want to communicate securely over a (open) network
- There is one adversary: Eve on the network link
  - Can read traffic on the network
- Alice and Bob want
  - Confidentiality of their messages

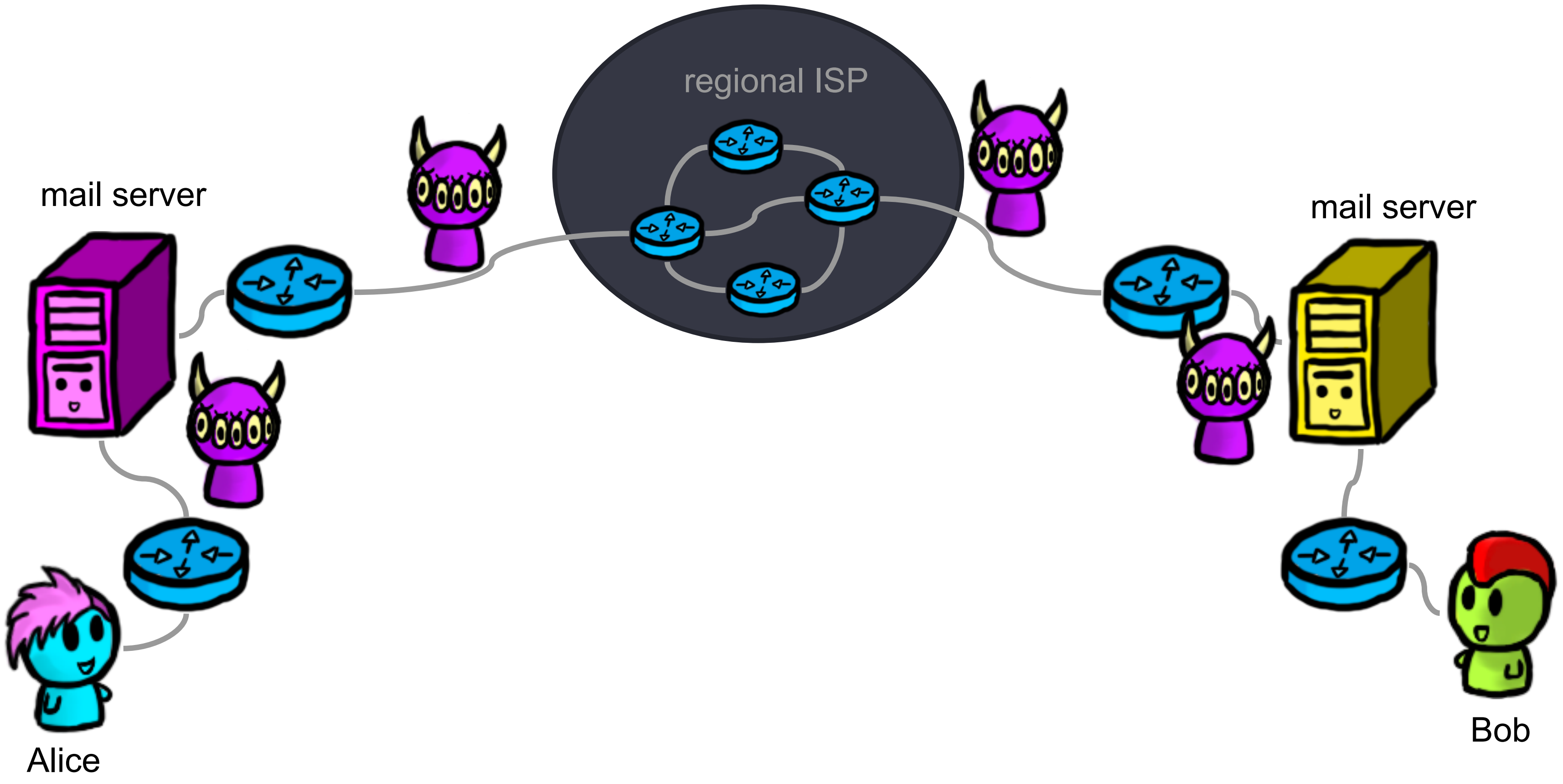
Our Model



Easy case

- Wiretap
  - Broadcast networks
    - Wired/Wireless
  - Router/Host-based
- What does the adversary see?
  - Data Link Layer Information
  - IP Layer Information
  - TCP Layer Information
  - Application Layer Information

Traffic Analysis



Easy attack surface:

- Mallory has access to one of the many hops traffic takes on the internet

Communication media (WiFi)

- WiFi
  - Can be easily intercepted by anyone with a
- WiFi-capable (mobile) device
      - Don’t need additional hardware, which would cause suspicion
- Maybe from kilometers away using a directed antenna
- WiFi also raises other security problems
  - Physical barriers (walls) help against random devices being connected to a wired network, but are (nearly) useless in case of wireless network

Communication media

- Copper cable
  - Inductance allows a physically close attacker to eavesdrop without making physical contact
  - Cutting cable and splicing in secondary cable is another option
- Optical fiber
  - No inductance, and signal loss by splicing is likely detectable
- Microwave/satellite communication
  - Signal path at receiver tends to be wide, so attacker close to receiver can eavesdrop
- All these attacks are feasible in practice, but require physical expenses/effort

Traffic Analysis

- TCP/IP has each packet include unique addresses for the packet's sender and receiver end nodes, which makes traffic analysis easy
- The attacker simply needs to sniff packets to determine what is going where and when.
  - Can be sensitive info such as two CEOs talking or a whistle blower.
- tcpdump is a text-based traffic analysis tool

Tcpdump (1 of 3)

```
14:47:26.566195 IP 192.168.2.2.22 > 192.168.1.1.41916: Flags [P.], seq 196:568, ack 1, win 309, options
[nop,nop,TS val 117964079 ecr 816509256], length 372
```

- 14:47:26.566195 the timestamp of the received packet
- IP is the network layer protocol (IPv4)
- 192.168.2.2.22 is the source IP address and port
- 192.168.1.1 is the destination IP address and port

Tcpdump (2 of 3)

```
14:47:26.566195 IP 192.168.2.2.22 > 192.168.1.1.41916: Flags [P.], seq 196:568, ack 1, win 309, options
[nop,nop,TS val 117964079 ecr 816509256], length 372
```

- TCP Flag (`Flags [P.]`) fields include:

Value	Flag Type	Description
S	SYN	Start Connection
F	FIN	End (Finish) Connection
P	PUSH	Push data
R	RST	Reset connection
.	ACK	Acknowledgement

Tcpdump (3of 3)

```
14:47:26.566195 IP 192.168.2.2.22 > 192.168.1.1.41916: Flags [P.], seq 196:568, ack 1, win 309, options
[nop,nop,TS val 117964079 ecr 816509256], length 372
```

- `seq 196:568` is the sequence number of the data contained in the packet (196 bytes to 568 bytes)
- `ack 1 is the ack number, which is` 1 (sender) or the next expected byte (receiver)
- `win 309` is the number of bytes available in the receiving buffer
- `options [nop,nop,TS val 117964079 ecr 816509256],` are the TCP options
- `length 372` is the length, in bytes, of the payload data (the difference between the first and last byte in the sequence number)

Tcpdump activity – if time at the end

- Launch the Labtainer VM
- In the terminal type: **`labtainer telnet`** and hit enter
- Open the lab manual
- Complete the lab
- Run **stoplab** to finish

Wireshark

Source: Miti Mazmudar

Network layers and encapsulation

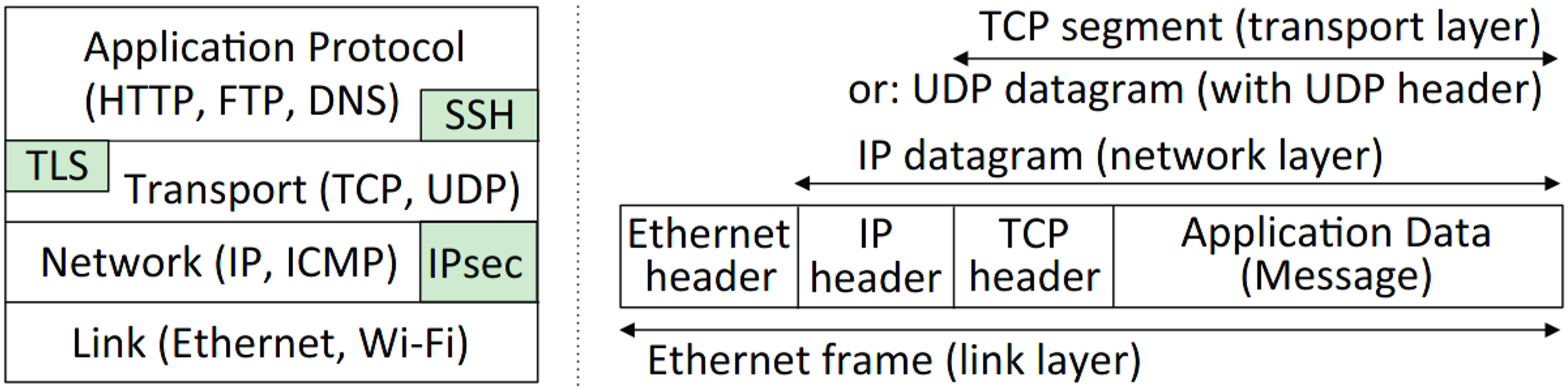
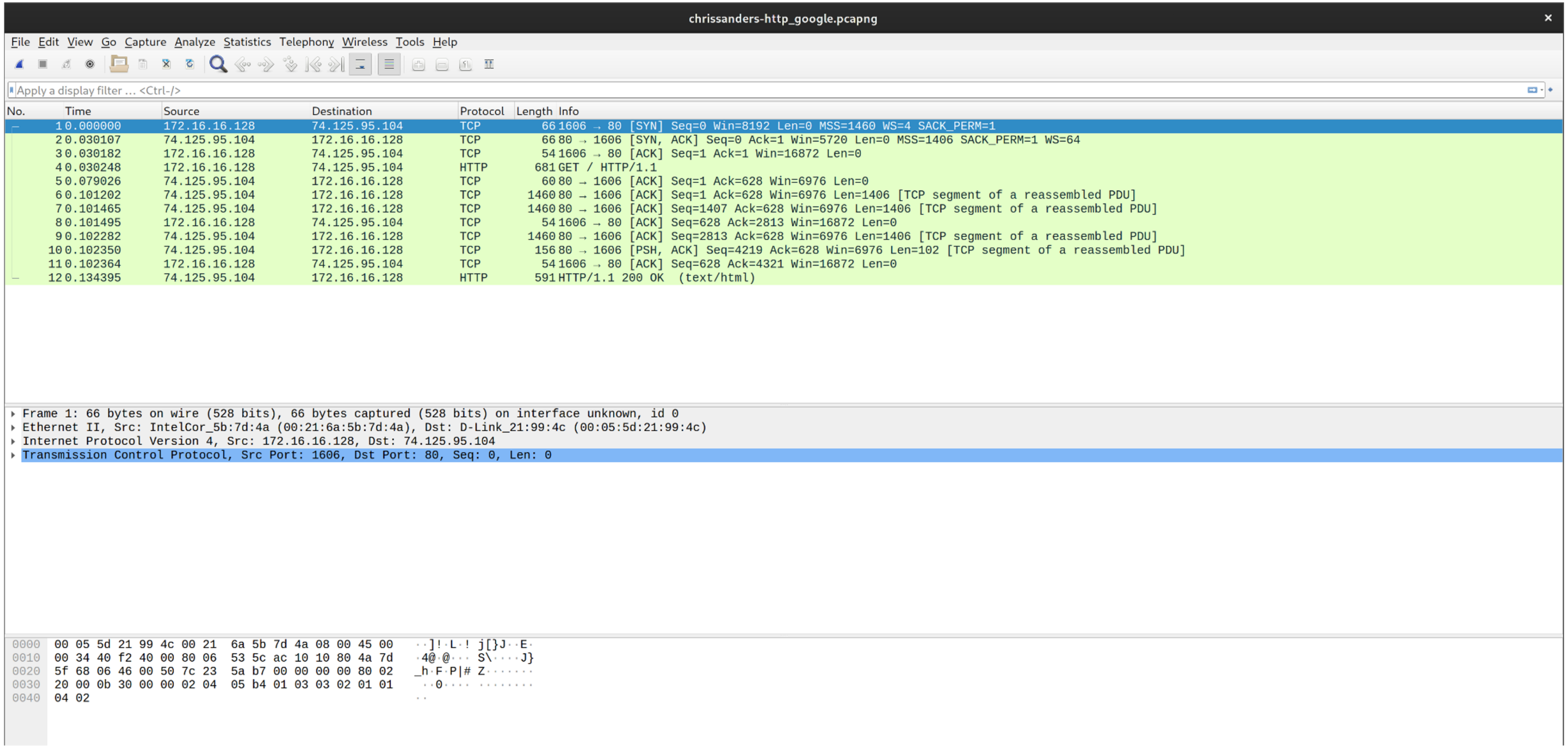


Figure 10.10: Network protocol stack (TCP/IP model) and encapsulation. In the seven-layer OSI model, between Application (7) and Transport are Presentation and Session layers, and Link is Data Link, above Physical (1). Wi-Fi denotes IEEE 802.11 wireless.

Wireshark



From Chris Sanders' PCAP file collection at https://github.com/chrissanders/packets/ (http_google.pcapng)

Wireshark - Walkthrough

1. Find total number of packets and bytes captured in a .pcap file.

Wireshark - Walkthrough

1. Find total number of packets and bytes captured in a .pcap file.
2. Isolate packets by the upper-most *protocol*.
    a. Sort by protocol field.
    b. Apply a display filter: protocol.type_of_message

Wireshark - Walkthrough

1. Find total number of packets and bytes captured in a .pcap file.
2. Isolate packets by the upper-most *protocol*.
    a. Sort by protocol field.
    b. Apply a display filter
3. Encapsulation in action.
4. How to find what type of payload follows a given header.

Wireshark - Walkthrough

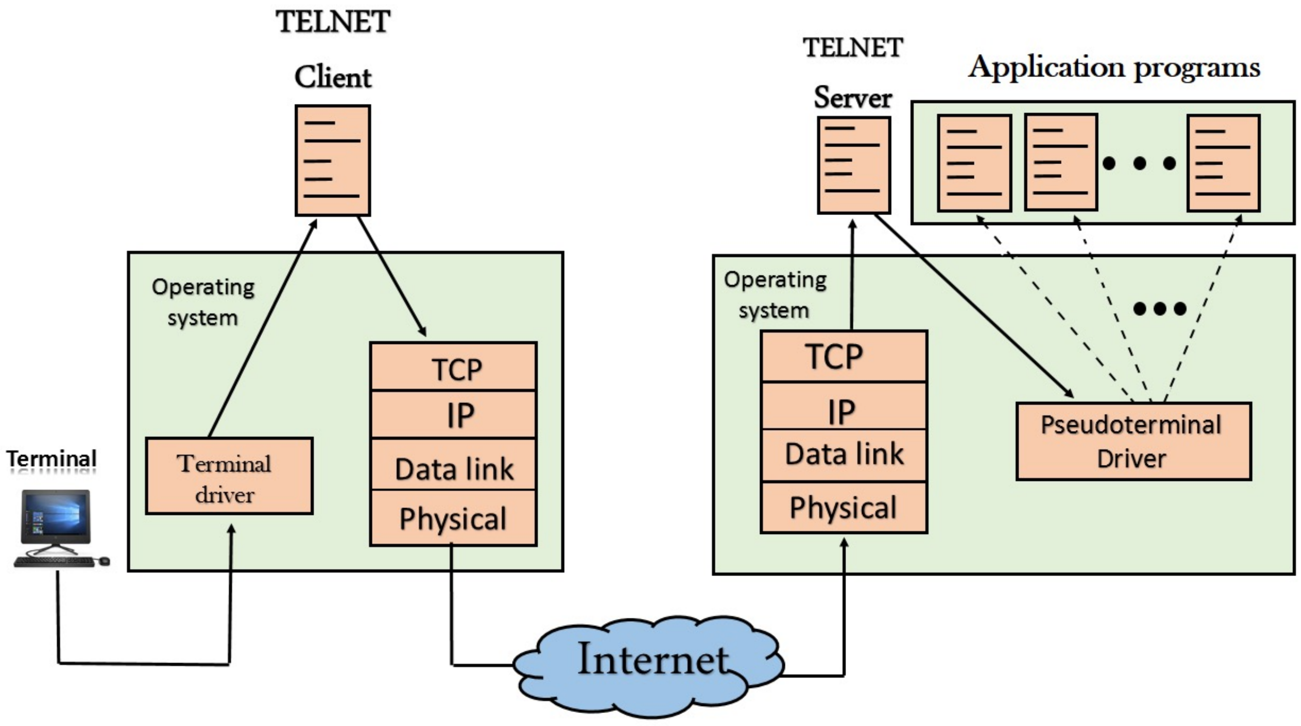
1. Find total number of packets and bytes captured in a .pcap file.
2. Isolate packets by the upper-most *protocol*.
    a. Sort by protocol field.
    b. Apply a display filter:
3. Encapsulation in action.
4. How to find what type of payload follows a given header.
5. Parsing the payload.
6. Addresses --- IP and MAC addresses.
7. TCP flags --- SYN and ACK flags.

Cheatsheet

- Wireshark cheatsheets online such as:
- https://www.comparitech.com/net-admin/wireshark-cheat-sheet/

TELNET protocol

- Old school, insecure SSH.
  - Remote terminal access to other machines
- All commands are sent in plaintext.
- Not really used these days.
- To connect to a machine
  - telnet <ip>



Note on A2

- Wireshark displays the relative TCP sequence number, by default
    - which equals to the actual sequence number minus the initial sequence number

To see the actual sequence number in a packet:

right click the TCP section of the Wireshark output, and select "Protocol Preference". In the popup window, uncheck the "Relative Sequence Number" option.

Wireshark activity

Act.

- Launch the Labtainer VM
- In the terminal type: `labtainer wireshark-intro` and hit enter
- Open the lab manual
- Complete the lab and upload the .lab file to learn!

- Remember, you can use **checkwork** to see if you completed the task
- **stoplab** to exit the lab

- Student manual is on learn

Easy case remedy

- Encrypt the message
  - Previous/Upcoming lectures
- Remaining issues
  - Information that cannot be (easily) encrypted
    - Timing of a message
    - Length of a message
    - Meta data

Timing of messages

- Consider two network links: L_1 and L_2
  - E.g. after source and before destination
  - E.g. before and after router
- There is a message m_1 on L_1 sent at t_1 to L_2
- There is a message m_2 on L_2 at t_2
- If Alice and Bob are communicating over L_1 and L_2
  - And $t_2 > t_1$ and $t_2 < t_1 + \varepsilon$
  - Then $Pr[m_1 = m_2] > Pr[m_1 = m_i]$ for "most" m_i

Length of a message

- The length of a message cannot be encrypted, only padded
- Padding can be wasteful
- Message Length distribution differs:
  - Data Link Layer
  - IP Layer
  - Application Layer

A very old statistic

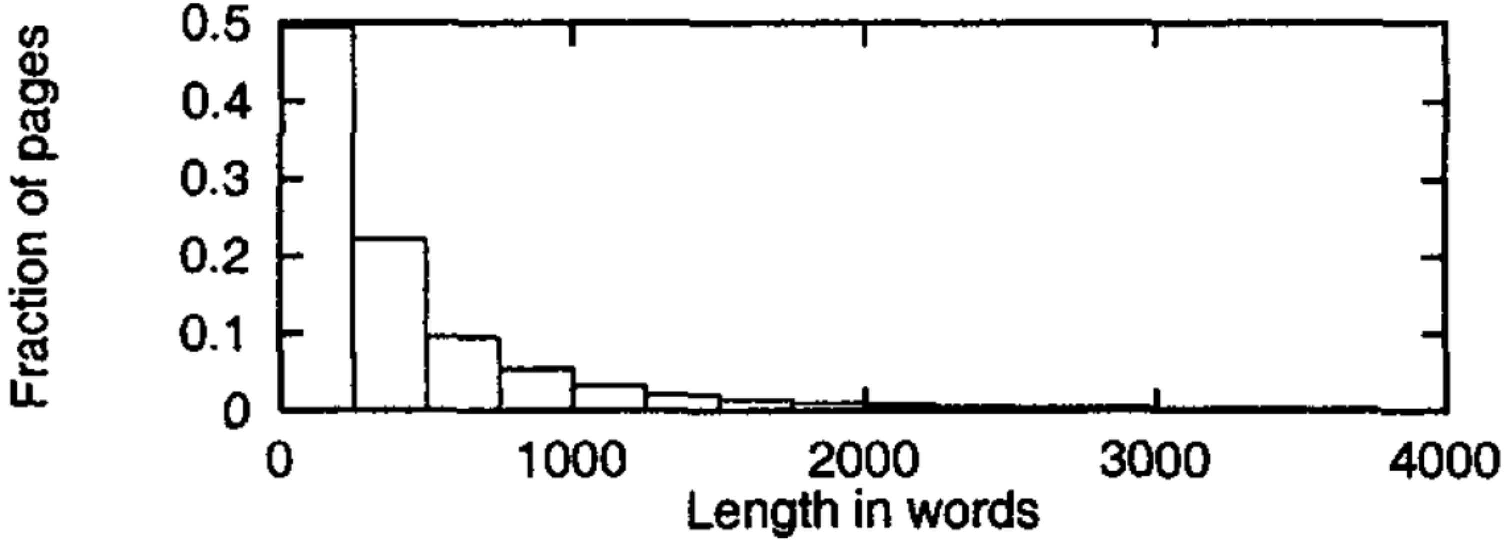


Figure 2: Distribution of content lengths of web pages.

Source: Davison, 2000.

Timing and Length

- Timing and Length can be combined
- Message length rarely changes across the network
  - Exception: Fragmentation
- Hence when matching messages based on time, one can consider only messages of the same length

Metadata

- Who is communicating with whom
- Difficult to hide, not easy to encrypt
  - When encrypting end-to-end, how does a router forward?
  - When encrypting to the next hop, what if the adversary controls the router?
- Default
  - At its network layer, addresses are unencrypted
  - E.g. TLS does not encrypt IP addresses or TCP port numbers

Side-Channels

- Timing side-channels
  - The runtime of code depends on sensitive data
- Storage side-channels
  - Storage is not properly cleaned and then shared
    - Cache side-channels, Cold boot attack
- Hardware side-channels
  - Power, electro-magnetic waves, sound
- Fault injection
  - Leakage from error behavior

Private Database Systems Example

Timing Attack

```
WHERE ssn = '123-45-6789'
    AND disease = 'condition'
    AND col1 IN (SELECT col2 FROM table)
```

Error Messages

```
WHERE ssn = '123-45-6789'
    AND disease = 'condition'
    AND SQRT(age-1000) = 1
```

Source: https://people.mpi-sws.org/~francis/side-channel.pdf

Classic Timing Side-Channel Example

- Square-and-Multiply Algorithm for Modular Exponentiation
- Input: x, e, p
- Output: x^e mod p
- Note* for decryption e is often a secret key

```
s = x, r = x
for each bit b of the binary representation of e (skip first bit)
        s = s^2 (mod p)
        if b = 1
                s= s * r  (mod p)    The execution of this line depends on e
return s
```

How to exploit this side channel

- Assume (mod p) is implemented as

1: If r > p-1
2: r = r − p

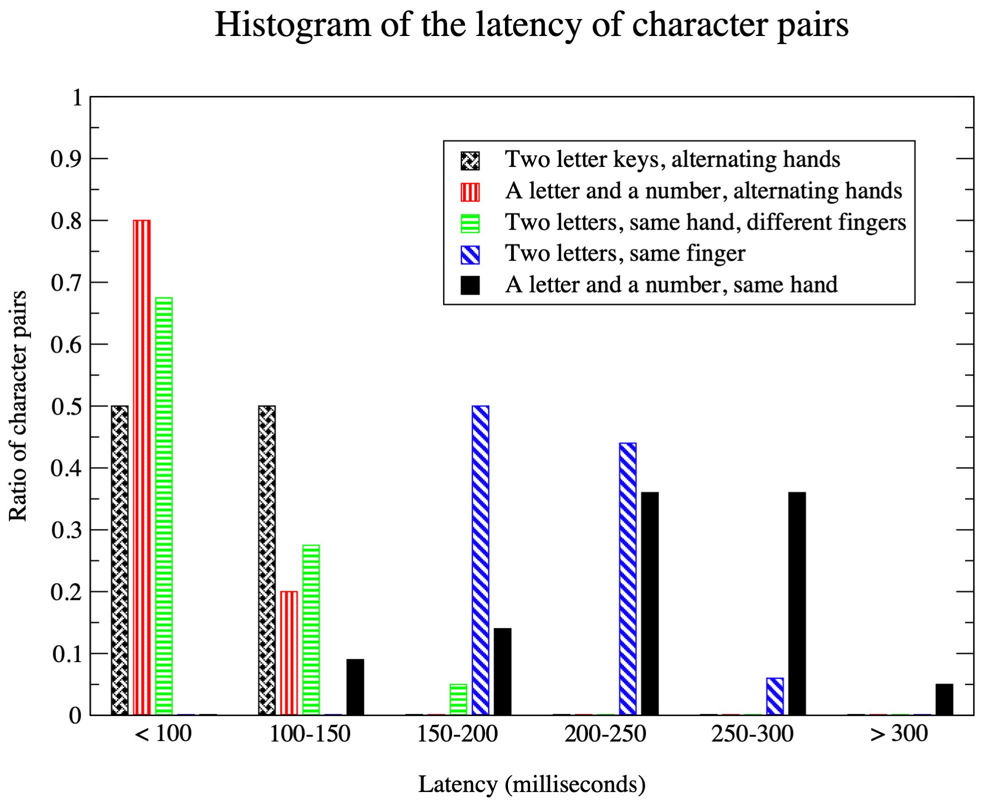
- We give some example inputs x for decryption
- For each bit of e, we choose many values for x
    - Some such that we expect the mod to require computation, some not.
- If the values take the same time likely the bit of e was 0
- Once we know the first bit, repeat.
- For more details: https://www.cs.sjsu.edu/faculty/stamp/students/article.html

Over networks

- Timing side channels can be exploited over networks
- Recovery of private key in SSL over LAN [Brumley and Boneh, 2003]
  - http://crypto.stanford.edu/~dabo/papers/ssl-timing.pdf
- Could factor RSA by taking into consideration optimizations in the decryption function!
  - Guess bit by bit in a binary search, took a long time if they were close, shorter if not.

SSH Timing Attack

- Ssh would send each keystroke immediately in a different packet
- Based on time between packets can reconstruct data
- https://people.eecs.berkeley.edu/~daw/papers/ssh-use01.pdf



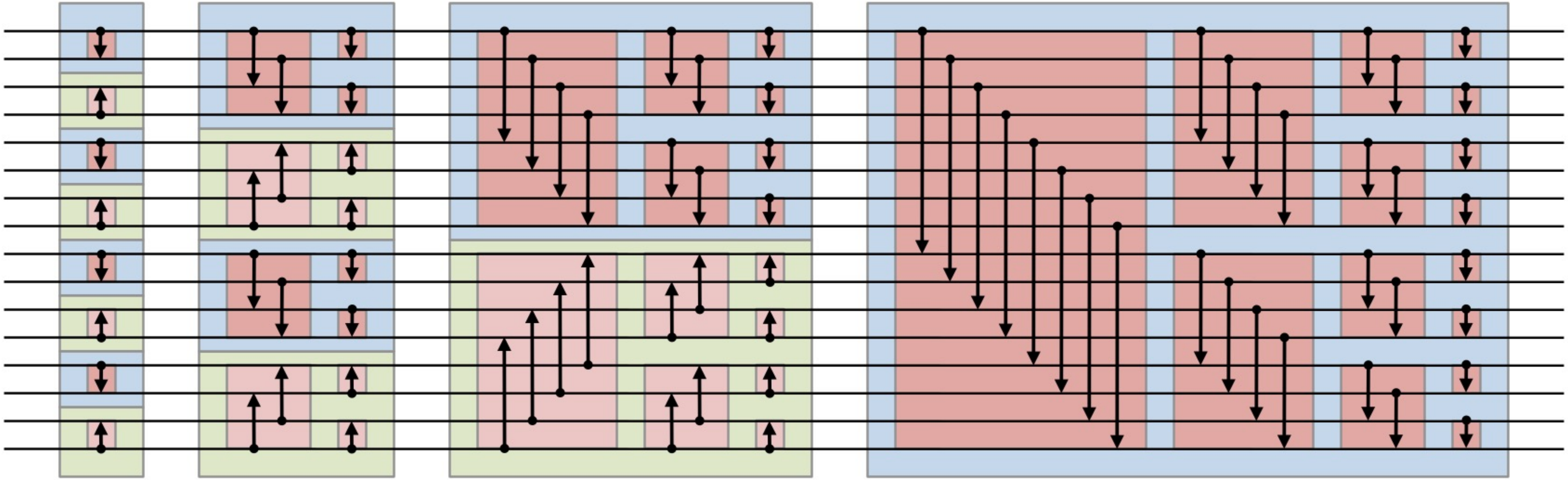
To summarize

- Many, many examples of side channels in practice
- Very hard to foresee them all
- Important to keep up to date with latest attacks
- Better, collaborate!

Mitigating Timing Side Channels

- Constant-time code
  - Can be difficult
  - No branches on sensitive data
  - Called oblivious algorithms
- Randomization
  - Choose random number R
  - Compute S = ModExp(Rx, e, p)
  - Compute T = ModExp(R, e, p)
  - Compute S/T (mod p)

Example: Oblivious Sorting Algorithm



- Conducts the exact same comparisons regardless of the data.

Next class, Applications!