Can my neighbors figure out what fanfic I read online?

(HTTPS Side Channels)

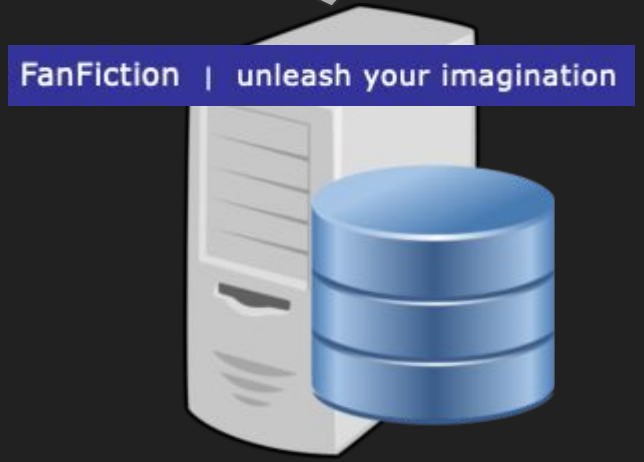
Matthew Pabst
CS 386W

Background

The year is 1995...

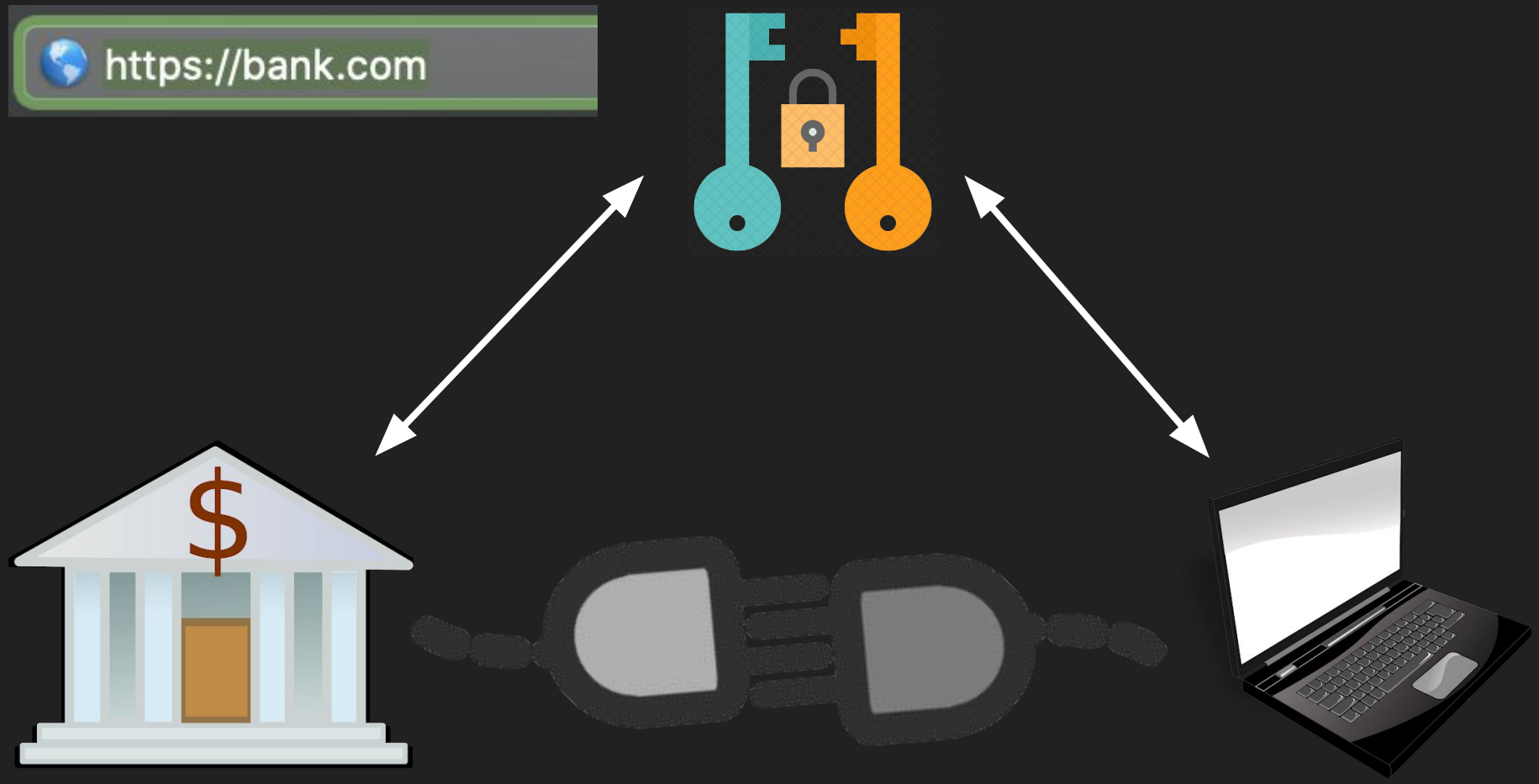
GET /s/13196616/1/The-Magician-an
HTTP/1.1
accept-language: en-US, en
accept: text/html
...

<!DOCTYPE html><html><head>
<meta charset='utf-8'>
<META NAME='ROBOTS'
CONTENT='NOARCHIVE'>
<META http-equiv='X-UA-Compatible' ...



HTTP allows network attackers to read your data in-transit!

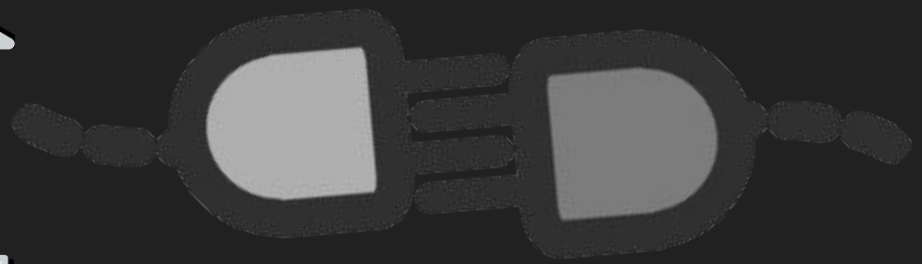
accept-language: en-US, en
...
credit-card-number=123456789012345
security-code=123


https://bank.com

HTTPS uses encryption to prevent traffic sniffing!

2ff9b8a9566b9bfc42e3bd156ce
c1e62071b8c2a4...

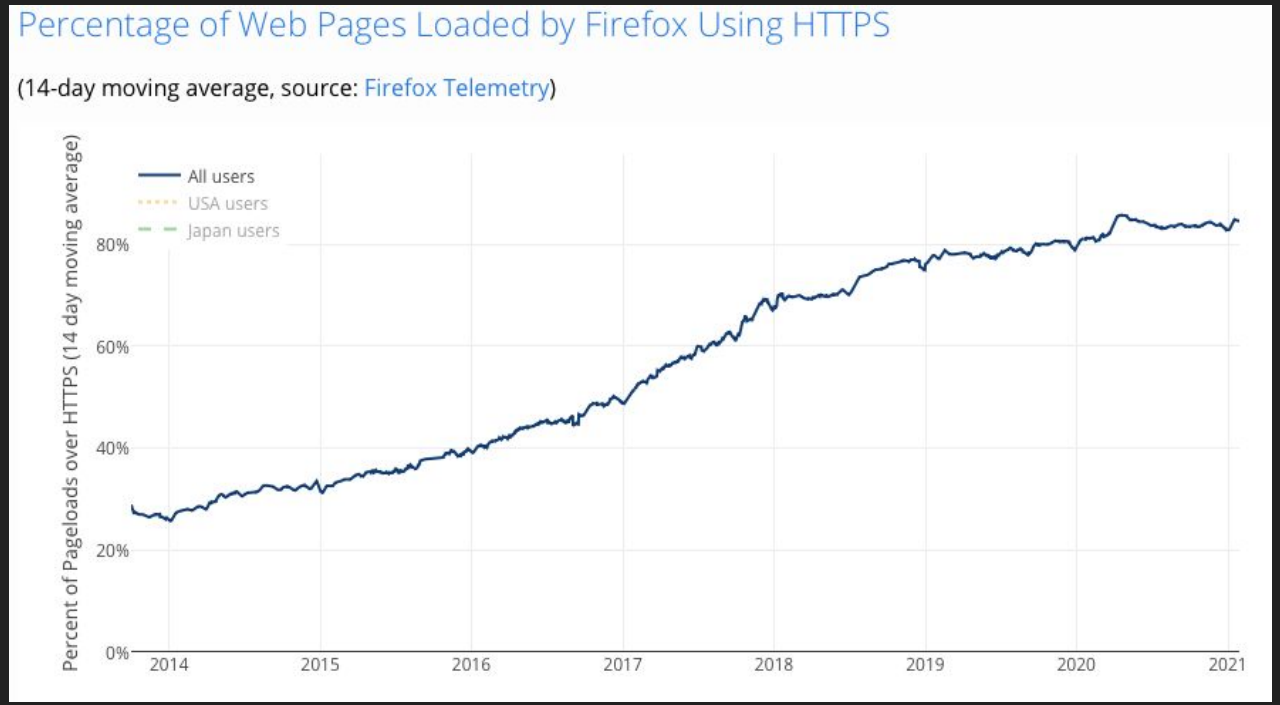
4920db8b4d590db391a9a9c79
b6414eaaa8c590b3...



Remember all that Snowden stuff?

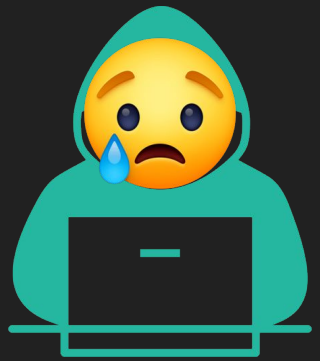
Most sites only used HTTPS for login/payment pages until ***recently***!

Network attackers (your ISP, governments, etc) knew exactly what you were doing online for ***20 years***!



Intuition

So if I’m using HTTPS, I’m completely safe browsing fanfic on the free Starbucks WiFi?



Not exactly ...

Lower layers in the networking stack leak information!

src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...

src: 192.168.1.5
dst: 143.182.253.12
size: 258
...

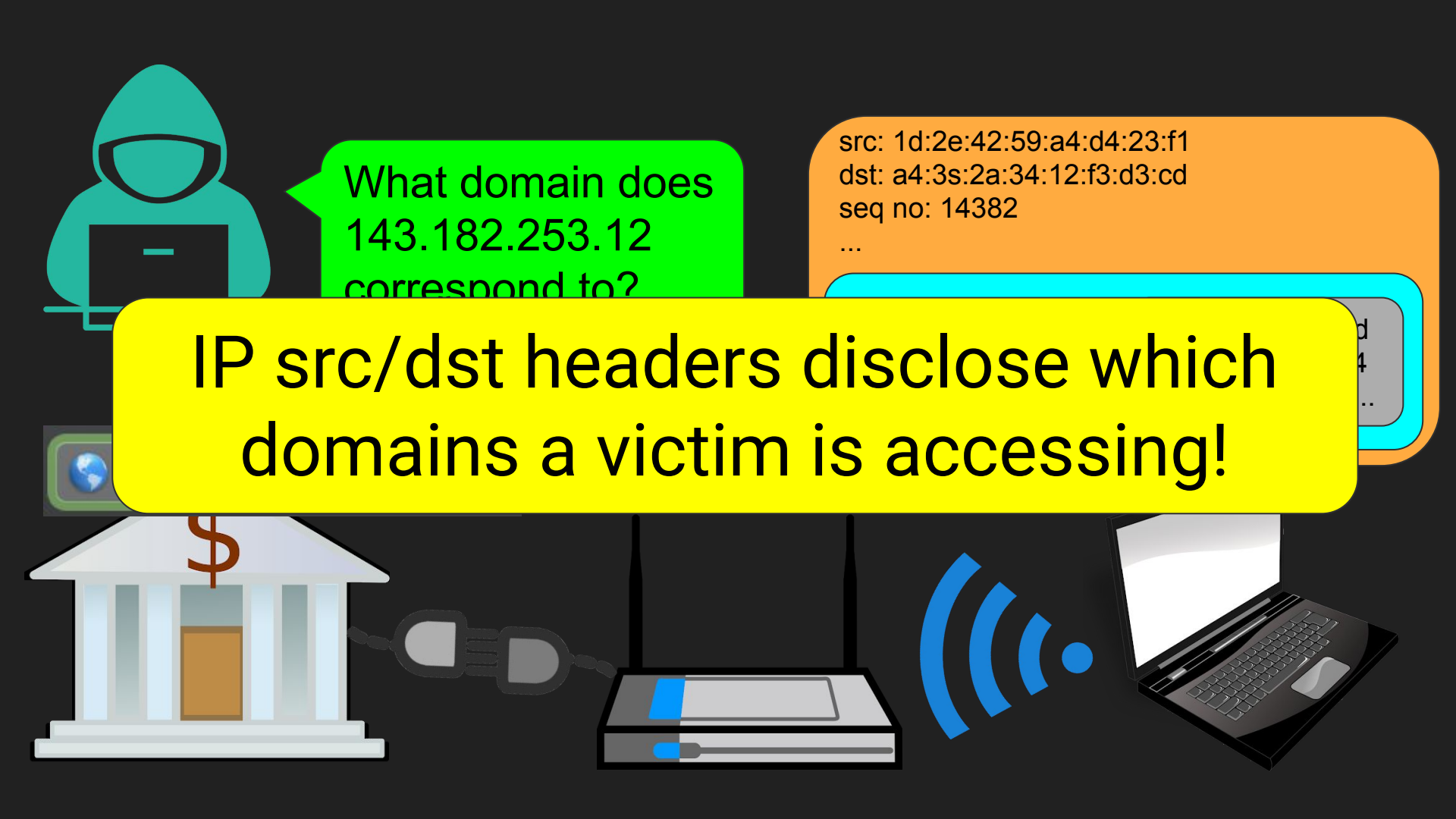
4920db8b4d590d
b391a9a9c79b64
14eaaa8c590b3...

What can I do with this information?

src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...

src: 192.168.1.5
dst: 143.182.253.12
size: 258
...

4920db8b4d590d
b391a9a9c79b64
14eaaa8c590b3...

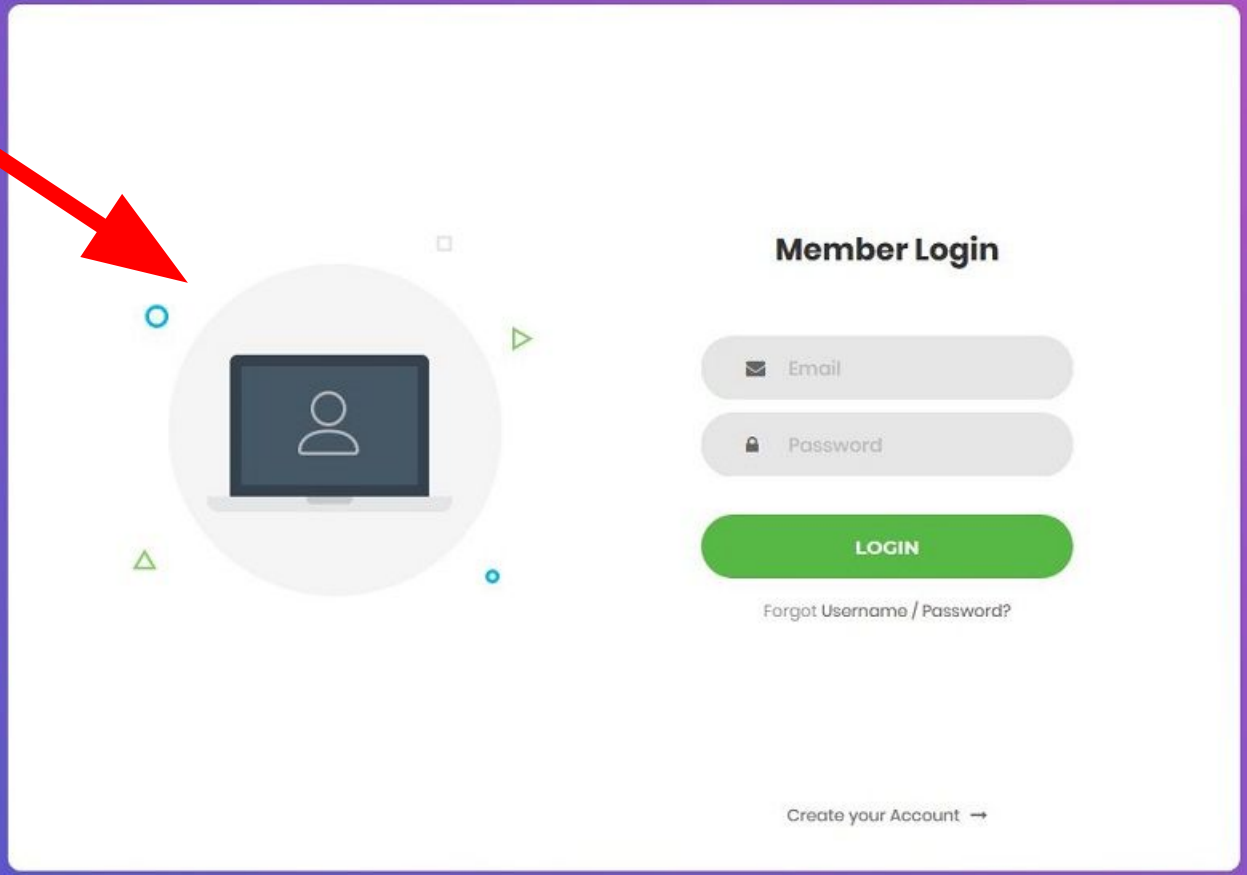

What domain does 143.182.253.12 correspond to?
src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...
IP src/dst headers disclose which domains a victim is accessing!

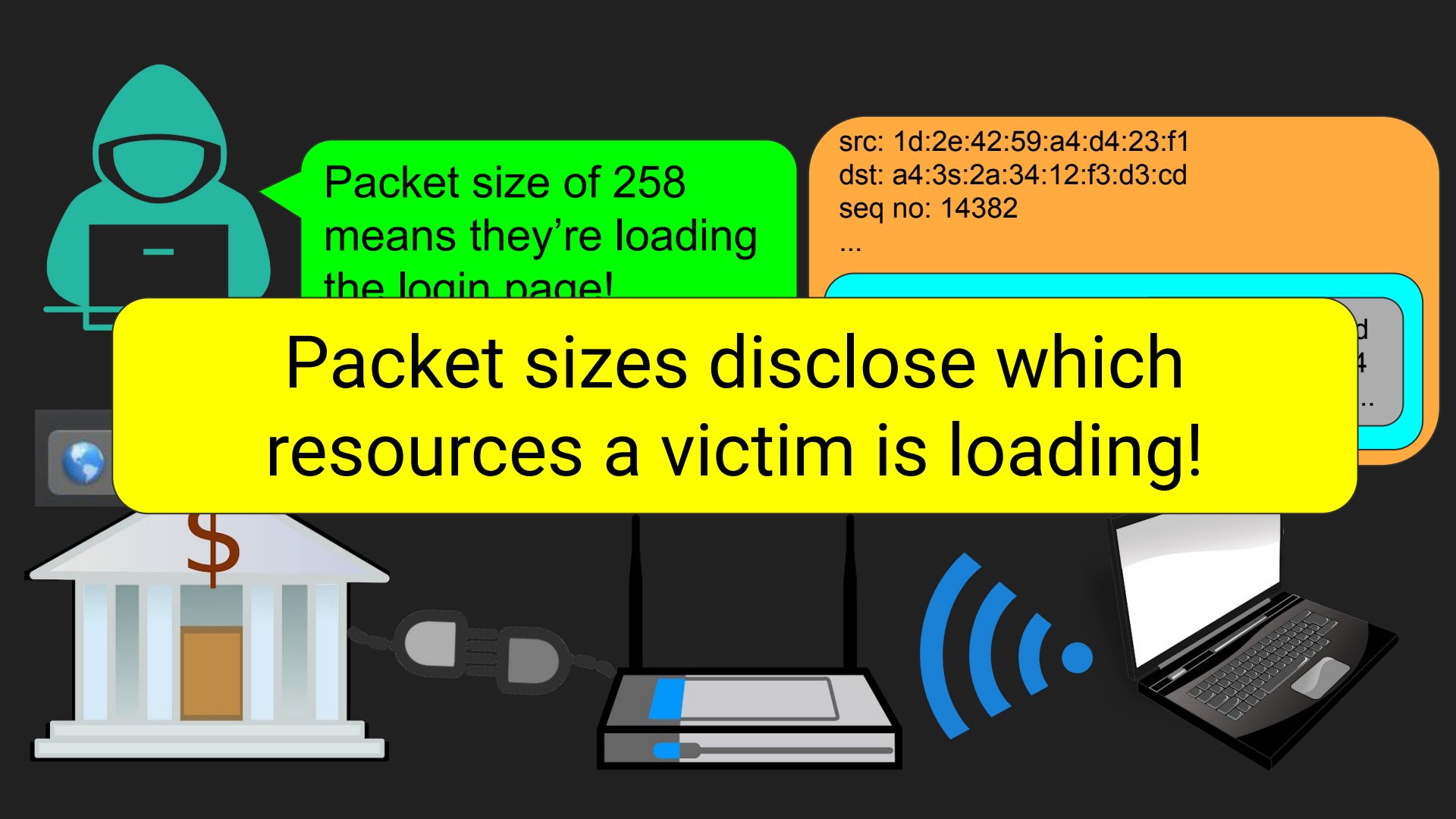
What does a packet size of 258 correspond to?

src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...

src: 192.168.1.5
dst: 143.182.253.12
size: 258
...

4920db8b4d590d
b391a9a9c79b64
14eaaa8c590b3...


Member Login
Email
Password
LOGIN
Forgot Username / Password?
Create your Account →


Packet size of 258 means they're loading the login page!
src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...
Packet sizes disclose which resources a victim is loading!

Adversaries sniffing public WiFi can guess your browsing history!

This is an old idea

The first paper[1] to introduce this idea was from 2002!

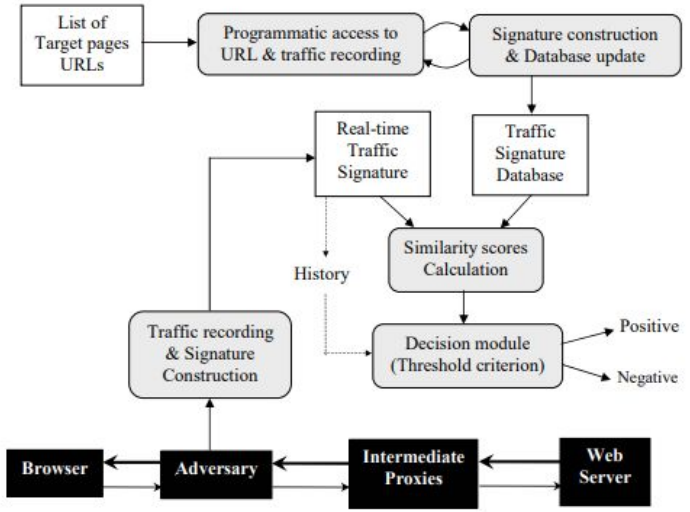


Figure 1. The adversary architecture.

Idea

Disclaimer:

Sniffing people's network traffic is ***illegal***, unless you own the network and have their consent!

Design:

How to guess a victim’s browsing history by sniffing their WiFi traffic

First, access pages and see which packets are sent and received

32 bytes to upload.wikimedia.org	US
36 bytes to en.wikipedia.org	Facebook
48 bytes to en.wikipedia.org	Wikipedia, US
8,532 bytes from upload.wikimedia.org	US
12,435 bytes from upload.wikimedia.org	Facebook
16,400 bytes from upload.wikimedia.org	Wikipedia

143.182.253.12 -> upload.wikimedia.org

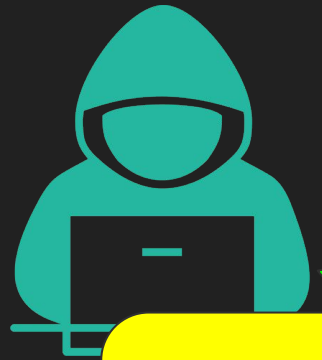
8,532 bytes from upload.wikimedia.org means ...

src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...

src: 143.182.253.12
dst: 192.168.1.5
size: 8,532
...

4920db8b4d590d
b391a9a9c79b64
14eaaa8c590b3...

32 bytes to upload.wikimedia.org	US
36 bytes to en.wikipedia.org	Facebook
48 bytes to en.wikipedia.org	Wikipedia, US
64 bytes to en.wikipedia.org	Wikipedia
72 bytes to meta.wikimedia.org	Facebook, US
8,532 bytes from upload.wikimedia.org	US
12,435 bytes from upload.wikimedia.org	Facebook
16,400 bytes from upload.wikimedia.org	Wikipedia



143.182.253.12 -> upload.wikimedia.org

8,532 bytes from upload.wikimedia.or

src: 1d:2e:42:59:a4:d4:23:f1
dst: a4:3s:2a:34:12:f3:d3:cd
seq no: 14382
...

Then, sniff victim packets, and match with prior knowledge

Article Ta

Unit

From Wikipedia, the free encyclopedia

"America", "US", "USA", and "United States of America" redirect here. For the landmass comprising North, Central, South America, and the Caribbean, see Americas. For other uses, see America (disambiguation), US (disambiguation), USA (disambiguation), and United States (disambiguation).

The **United States of America** (**USA**), commonly known as the **United States** (**U.S.** or **US**) or **America**, is a country located primarily in North America,

United States of America

Methodology

- The victim randomly selects a Wikipedia page from one of its top visited pages

- The adversary predicts which page the victim was browsing based on sniffed traffic

Results: Wired (packets captured locally)

Data Set	Accuracy	Accuracy (top 5)
Top 10 sites		
Top 50 sites		
Top 100 sites		

Results: Wired (packets captured locally)

Data Set	Accuracy	Accuracy (top 5)
Top 10 sites	**86%**	**100%**
Top 50 sites		
Top 100 sites		

Results: Wired (packets captured locally)

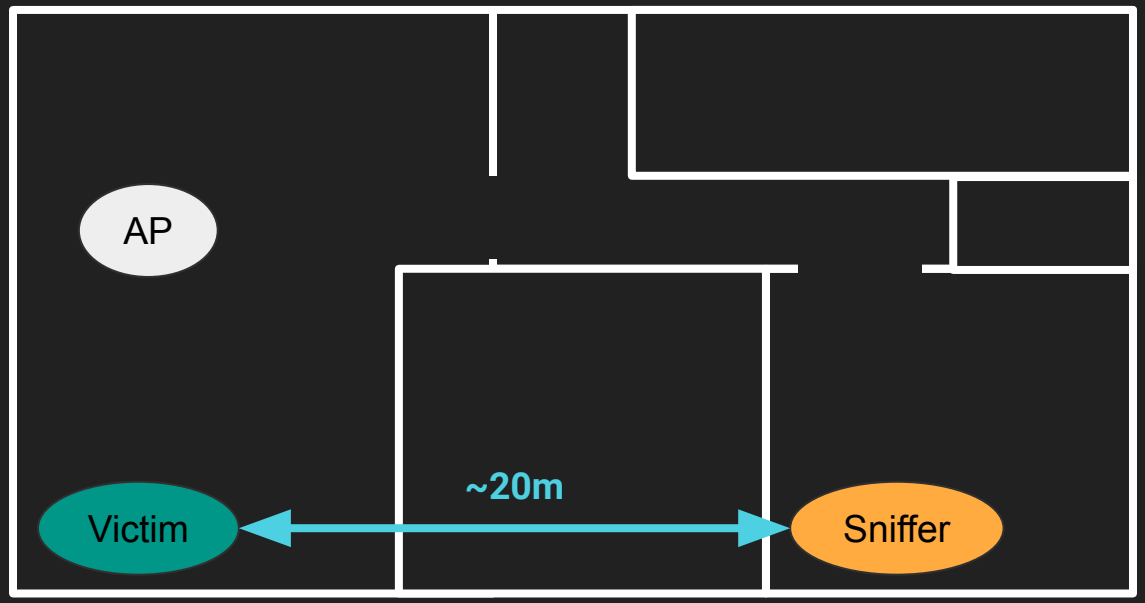
Data Set	Accuracy	Accuracy (top 5)
Top 10 sites	86%	100%
Top 50 sites	**47%**	**87%**
Top 100 sites		

Results: Wired (packets captured locally)

Data Set	Accuracy	Accuracy (top 5)
Top 10 sites	86%	100%
Top 50 sites	47%	87%
Top 100 sites	**41%**	**75%**

Wireless Configuration: Close

Wireless Configuration: Far



Results: Wireless (top 10 sites)

Configuration	Accuracy	Accuracy (top 5)
Close		
Far		

Results: Wireless (top 10 sites)

Configuration	Accuracy	Accuracy (top 5)
Close	**54%**	**90%**
Far		

Results: Wireless (top 10 sites)

Configuration	Accuracy	Accuracy (top 5)
Close	54%	90%
Far	**50%**	**94%**

Interesting Mispredictions

- UK: predicted World War I
- Presidents of the US: predicted Abraham Lincoln
- Elon Musk: predicted Illuminati
- Illuminati: predicted Justin Bieber
- Malware: predicted France and Germany

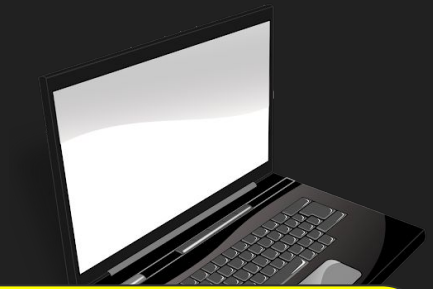
Implications

No one would go through this trouble to spy on me!

- This whole process could be automated, so it wouldn’t be expensive
- Your ISPs could sell your browsing history to advertisers
- Same thing goes for network infrastructure providers (e.g. Google)
- Under laws like the Patriot Act, the US government could legally get this kind of information from companies for bulk surveillance

src: 143.182.253.12
dst: 192.168.1.5
size: 8,532
...

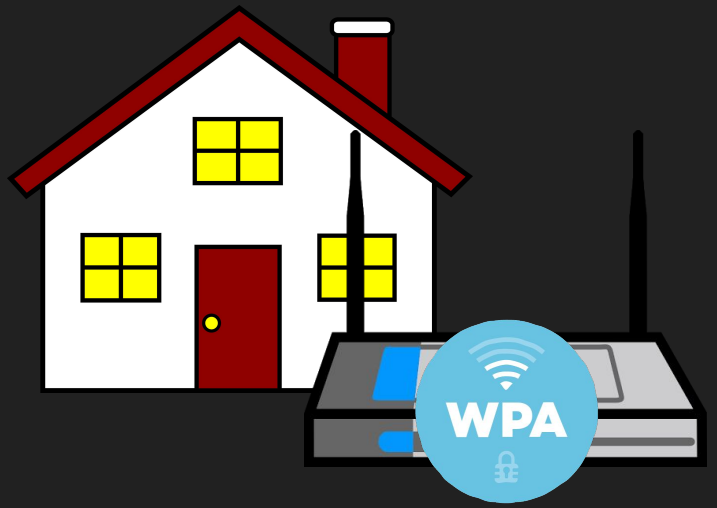
4920db8b4d590d
b391a9a9c79b64
14eaaa8c590b3...

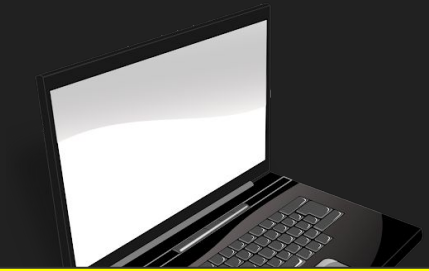


Public networks allow attackers to learn this information!

Defenses:

How can we prevent this attack?





Encrypting wireless traffic hides packet destinations (but not sizes)

Approaches

1. Try to obfuscate lower layer networking details:

a. Request random resources to confuse attackers

b. Add random number of bytes to every packet

2. Encrypt lower layer networking details:

a. Even if our network is unencrypted, we can make an encrypted connection to a proxy server

b. Then, we can send all our real traffic through this encrypted connection

References

1. Sun et al. "Statistical Identification of Encrypted Browsing Traffic". http://infolab.stanford.edu/~qsun/research/identification.pdf

2. Chen et al. ”Side-Channel Leaks in Web Applications: a Reality Today, a Challenge Tomorrow”. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5504714&casa_token=mxvyFaBwUW4AAAAA:H1JIVwwcLfWsQ4VcEL_bTa_IiOBS9f1O5sHXICZpaUMnGAmDbsyAa_9I0qUfGq4kgGL0p6kauw&tag=1

3. My code: https://github.com/PabstMatthew/https-side-channels

Challenges

- Packet loss when sniffing
- Requests to the same resource are not always a consistent size
  - Cookies, different user-agents, compression
- Browser caches
- Dynamic web pages
- Personalized web pages