Connecting the Dots in the Sky

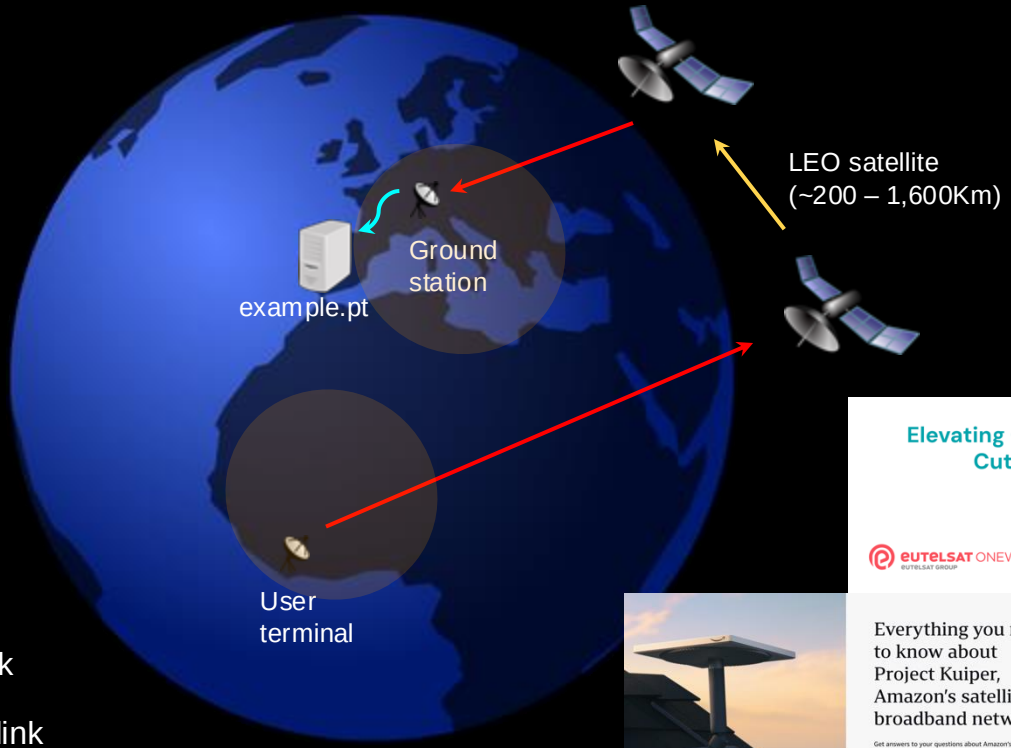
Website Fingerprinting in Low Earth Orbit Satellite Internet

Prabhjot Singh, Diogo Barradas, Tariq Elahi, Noura Limam

SpaceSec Workshop
San Diego, CA, USA
1 March, 2024

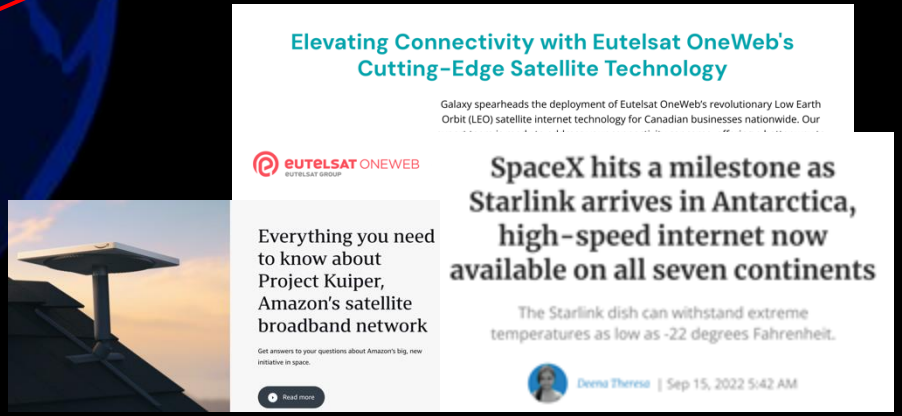
UNIVERSITY OF WATERLOO
THE UNIVERSITY of EDINBURGH

LEO Satellite Internet



Low latency:
~10s to 100 milliseconds

High throughput:
~100s Mbit to Gbps speeds



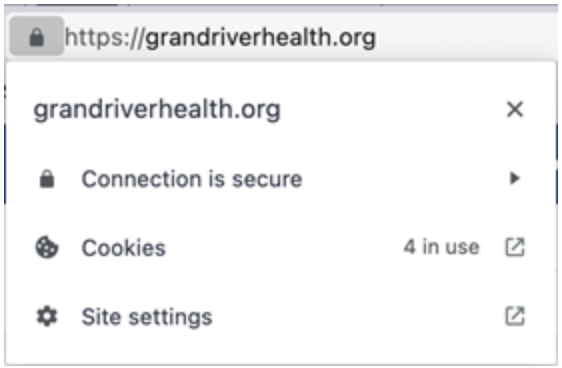
RF link

Fiber link

Laser link

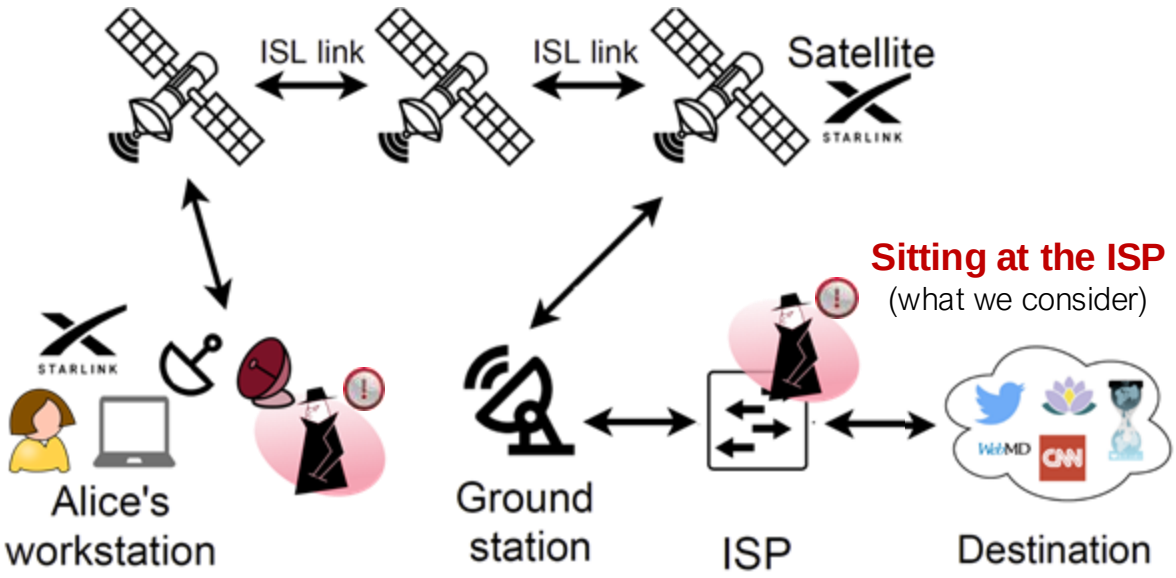
Satellite Internet is still “just” the Internet

- Transport Layer Security (TLS) encrypts the content of communications
- Widely adopted
- Your browser can even do it for you



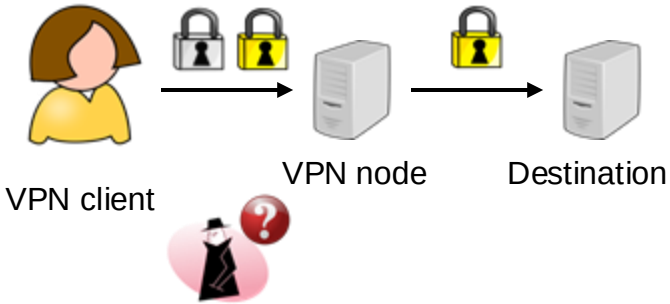
But TLS does not hide everything!
e.g., destination, connection duration

Can I Browse the Internet Privately?

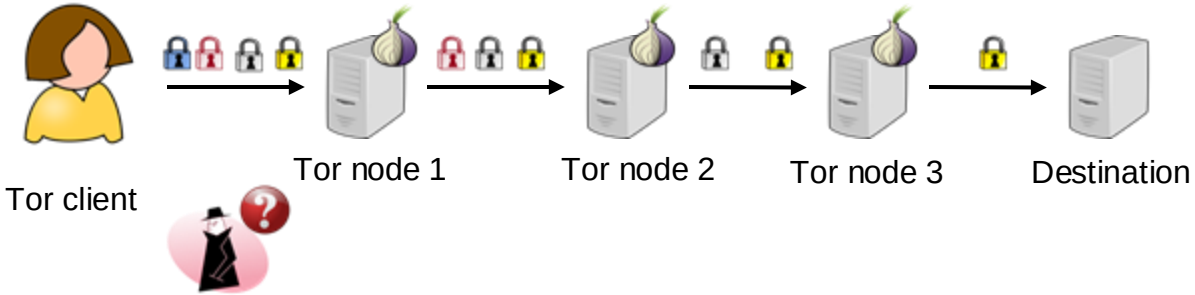


What can we do to protect this information?

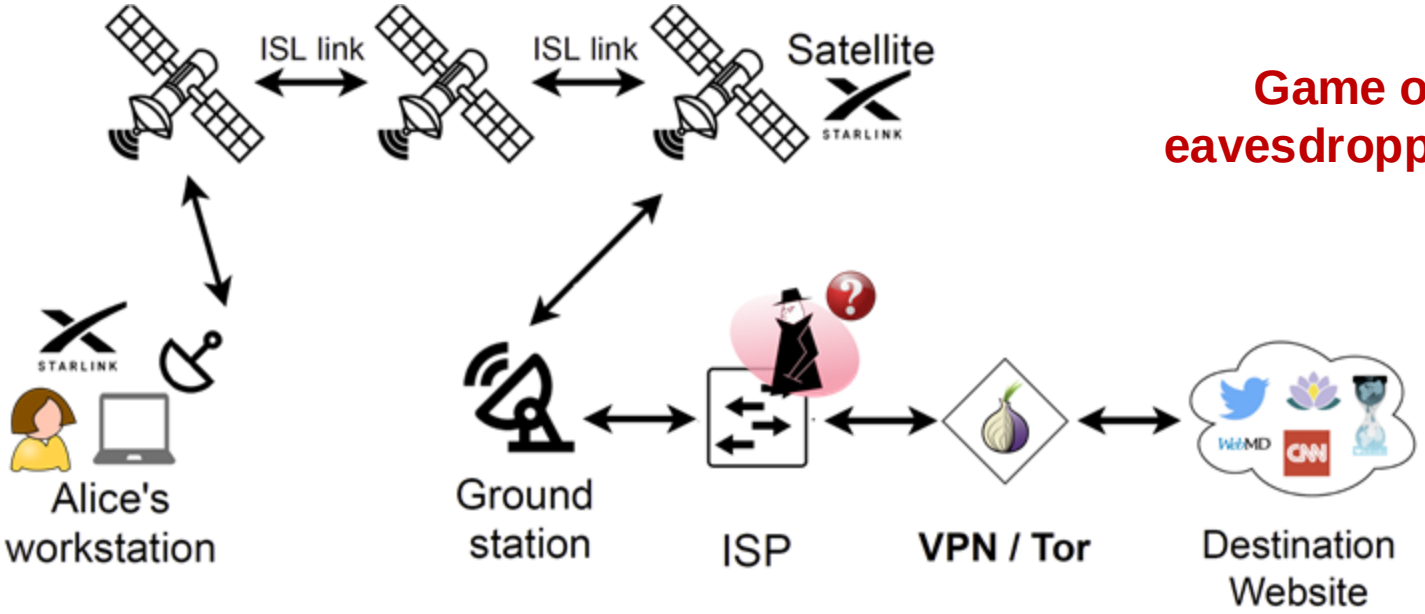
Virtual Private Networks



Anonymity networks
e.g., Tor

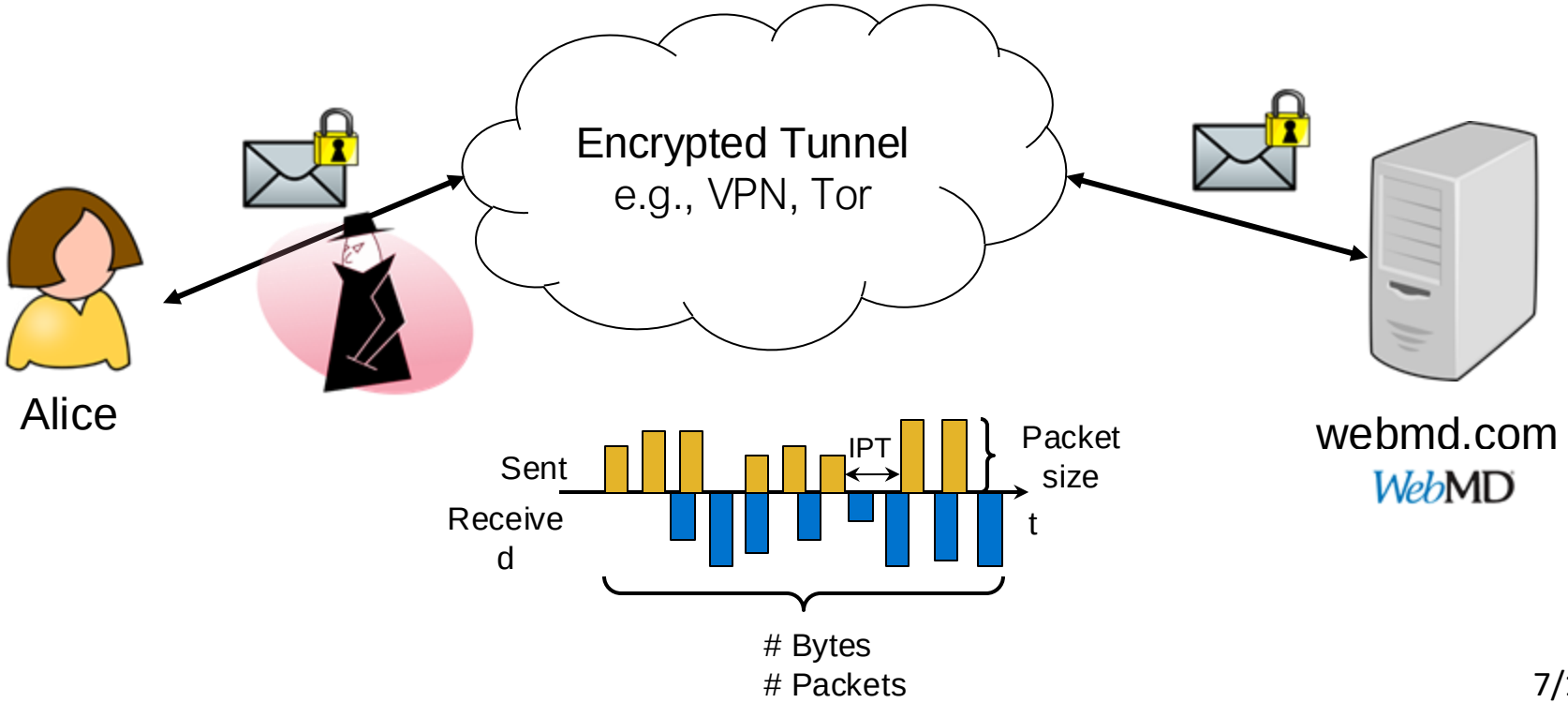


Securing Satellite Networking Traffic



Game over for eavesdroppers, right?

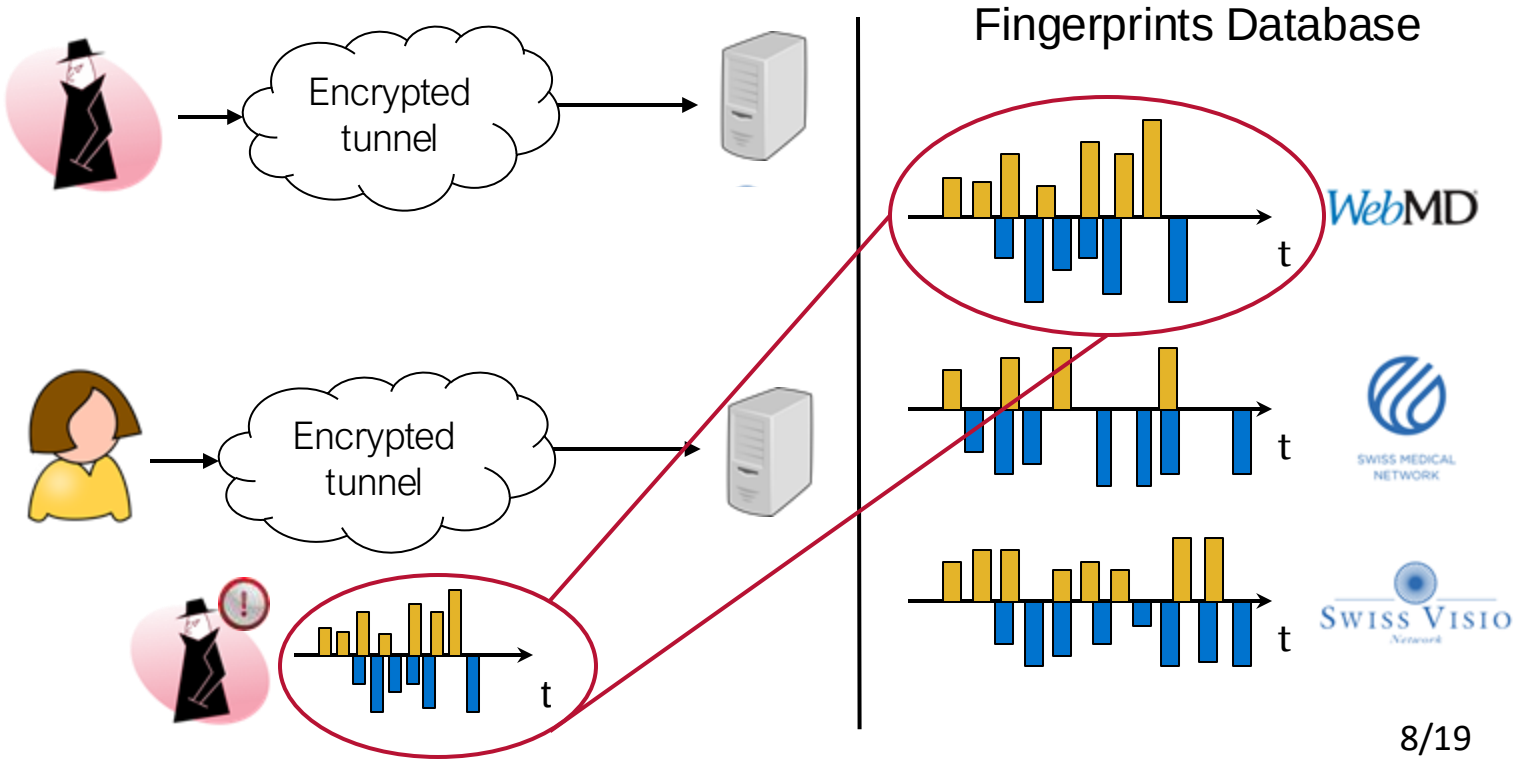
Encrypted Connections Leak Metadata



Website Fingerprinting (WF)

Step 1:
Build fingerprint database

Step 2:
Match Alice's traffic



WF across Network Environments

Fiber networks

- Traditional/de facto target

Cellular networks

- Different link properties than fiber links
  - Added perturbations: latency, jitter, packet drops...
- Different transport protocol behaviour/encapsulation

What about LEO networks?

STARLINK OneWeb iridium

Main Contributions

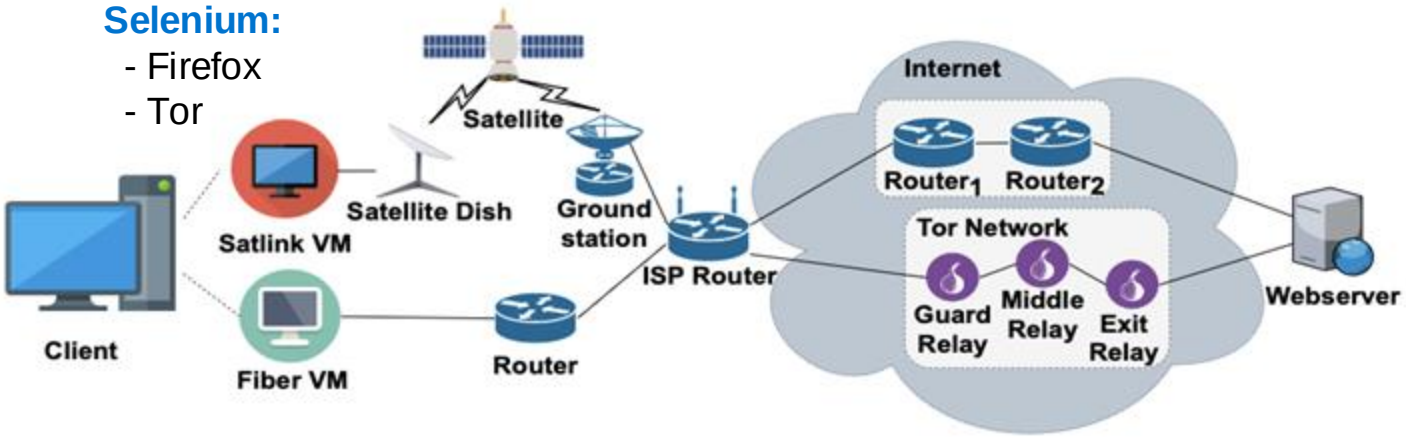
1. A novel dataset of website access traces:
   - via Starlink and traditional fiber
     - over Tor and plain Firefox

2. A comparison of the traffic characteristics in connections established:
   - via Starlink and traditional fiber
     - over Tor and plain Firefox

3. A study on the success of website fingerprinting attacks on satellite links
   - as compared to traditional fiber links

Experimental Testbed & Dataset

Selenium:
- Firefox
- Tor

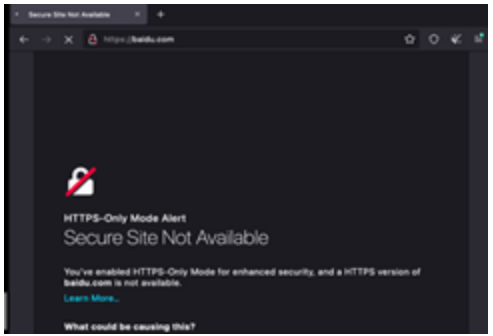


Collected 125x traces from the top-125 websites on the Tranco list

(But had to get rid of a few)

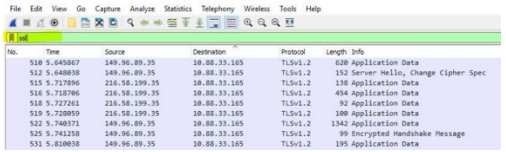
Data Pre-Processing

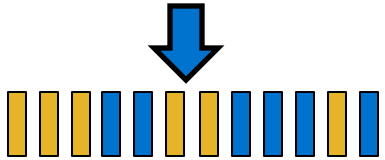
- Identify and remove errors
  - Remove timeout pages, blank pages, pages with captchas
  - Remove pages unaccessible via Tor (server-side blocking?)



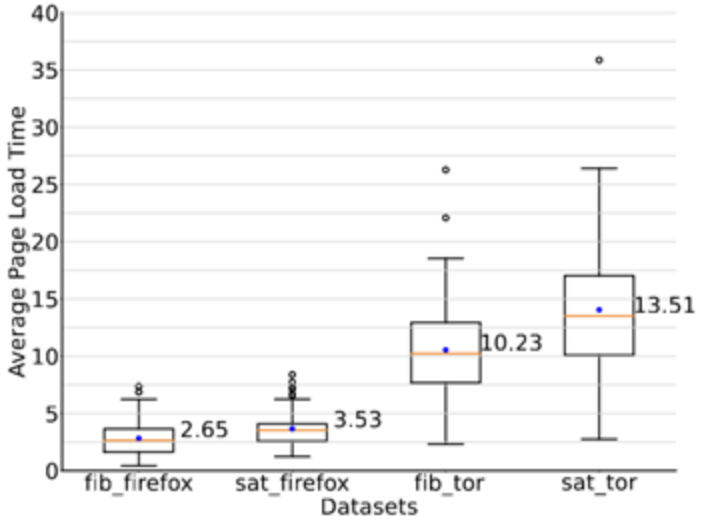
80 traces for 75 websites in each config
== 24 000 traces

- Convert packet traces
  - Raw IP packets converted into simpler representation
  - Tor traffic converted into “cell traces”

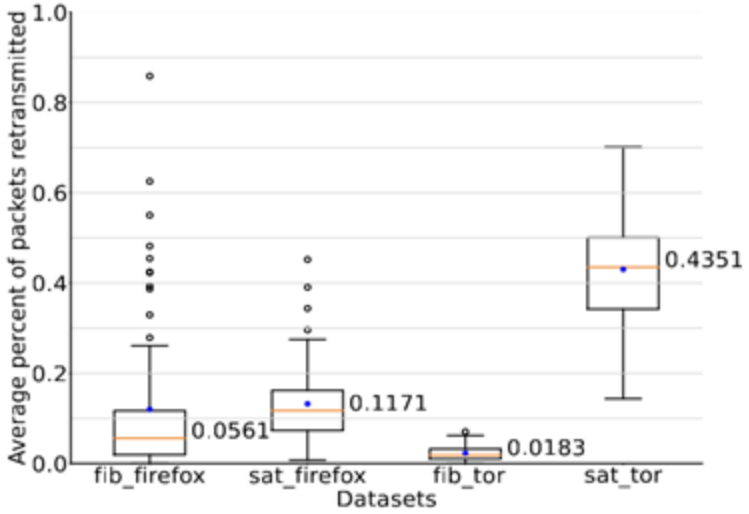




Characterization of Starlink and Fiber Traces (I)

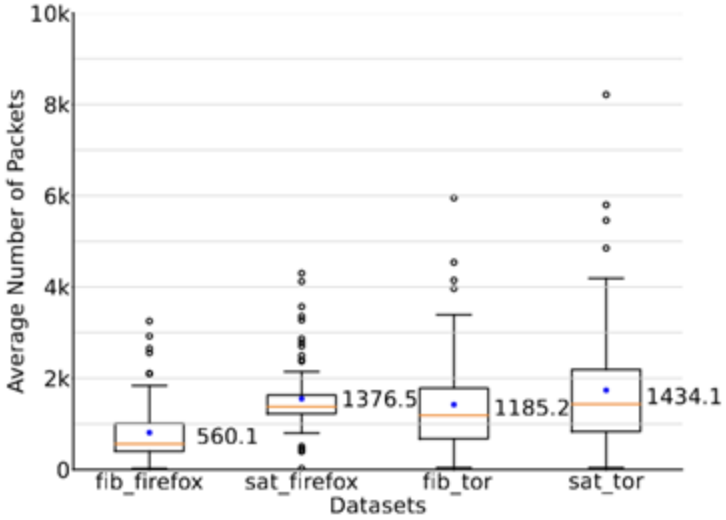


Starlink loads webpages 33% slower than fiber

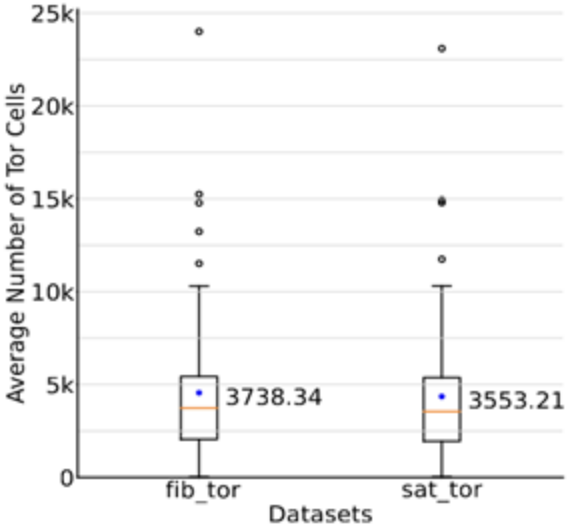


TCP retransmissions are up to 23x more common in Starlink

Characterization of Starlink and Fiber Traces (II)



Starlink exchanges 20% more packets than fiber



Tor exchanges a similar # of cells for Starlink and fiber

Starlink and fiber connections have different characteristics, which may impact fingerprinting results

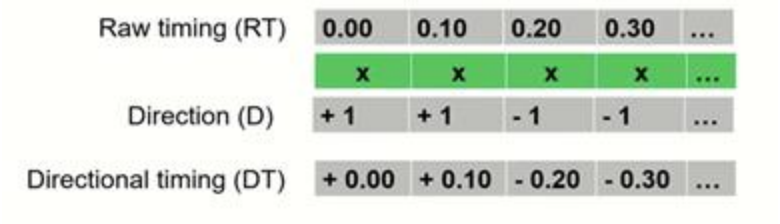
Website Fingerprinting Attacks

- Machine learning-based attacks
  - Manual feature extraction

№	Feature Description
1.	Number of incoming packets.
2.	Number of outgoing packets as a fraction of the total number of packets.
3.	Number of incoming packets as a fraction of the total number of packets.
4.	Standard deviation of the outgoing packet ordering list.
5.	Number of outgoing packets.
6.	Sum of all items in the alternative concentration feature list.
7.	Average of the outgoing packet ordering list.
8.	Sum of incoming, outgoing and total number of packets.
9.	Sum of alternative number packets per second.
10.	Total number of packets.
11-18.	Packet concentration and ordering features list.
19.	The total number of incoming packets stats in first 30 packets.
20.	The total number of outgoing packets stats in first 30 packets.

K-Fingerprinting

- Deep learning-based attacks
  - Automatic feature extraction



DF and Tik-Tok

Closed-world setting:

- Which amongst one the 75 websites was visited?

Attack accuracy on undefended Tor traffic

Dataset (acc. %)	K-FP	DF	Tik-Tok
Tor - Fiber	0.73	0.87	**0.89**
Tor - Starlink	0.64	0.85	**0.87**

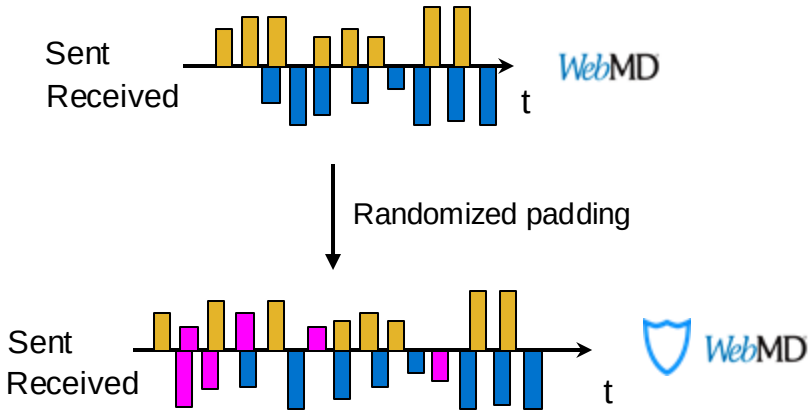
1. WF attacks are more accurate on fiber links

2. The best attacks obtain a similar accuracy for fiber and Starlink (2% diff.)

3. Attacks relying on manual features face a larger accuracy degradation (and a 9% difference between fiber and Starlink)

Defenses against Website Fingerprinting

- Obfuscate the real characteristics of a trace
  - Adaptive (WTF-PAD) and randomized (FRONT) padding
  - Constant-rate padding (Tamaraw and CS-BuFLO)
  - Many more…



Attack accuracy on defended Tor traffic

Defense (acc. %)	Fiber Traces	Starlink Traces
Undefended	0.89	0.87
WTF-PAD	0.84	0.79
FRONT_1	**0.59**	**0.47**
FRONT_2	**0.55**	**0.44**
CS-BuFLO	0.17	0.15
Tamaraw	0.11	0.10

1. Overall, defenses still make traffic hard to fingerprint over satellite links

2. FRONT variants seem particularly susceptible to the change of connection type

Takeaways

- Website fingerprinting (WF) can reveal browsing habits over encrypted traffic
- WF had not yet been explored within the context of LEO satellite internet
- We show WF may be as concerning in satellite networking as in traditional fiber

- **Future work:**
  - How do weather conditions impact fingerprinting?
  - What happens once inter-satellite links are active?
  - Does the success of attacks hold in the open-world setting?

→ Scan to check our pre-print (feedback is welcome!)

Diogo Barradas
diogo.barradas@uwaterloo.ca