UNITED STATES MILITARY ACADEMY
WEST POINT

Identifying HTTPS-Protected NETFLIX Videos in Real-Time

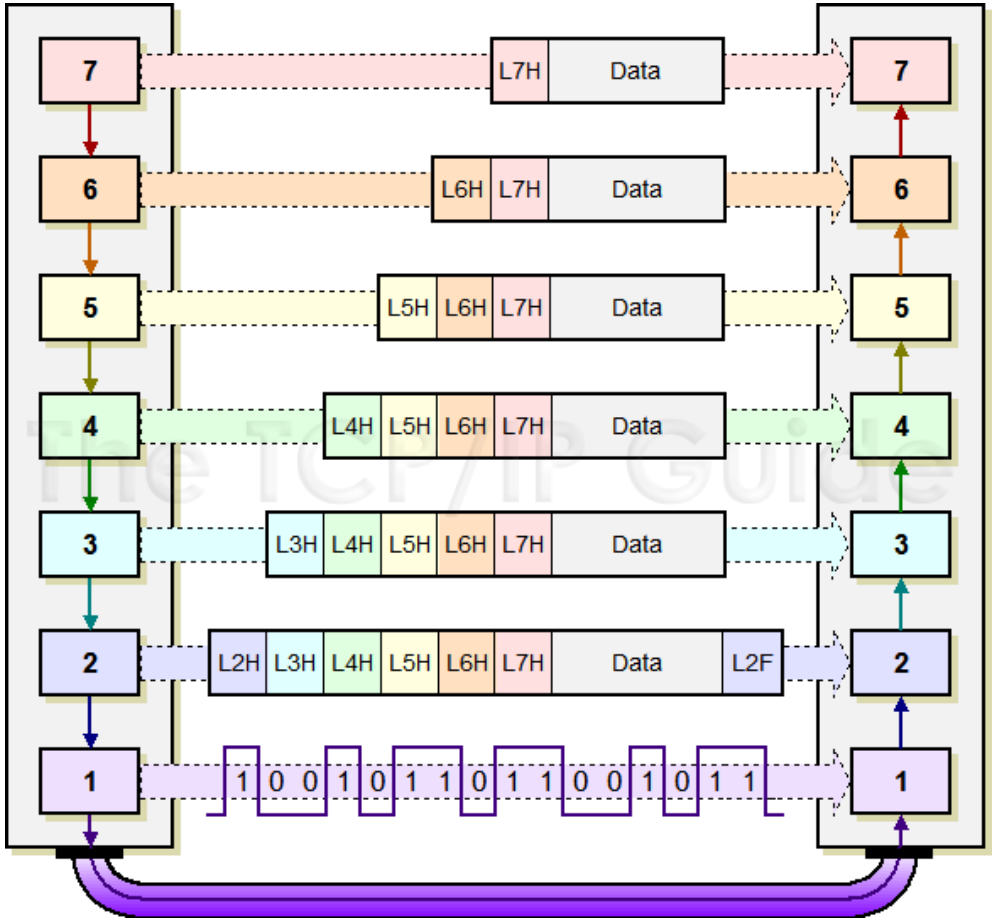
By: Andrew Reed and Michael Kranch

CODASPY 2017 – March 22-24

So What?

- Metadata matters.
- Unique enough data cannot rely on encryption for confidentially.
- Application Data Unit (ADUs) provide an interesting middle ground for Network Traffic analysis.

What is an ADU?



OSI Reference Model Data Encapsulation:
protocol data unit (PDU)

What is an ADU?

Timestamp	Local IP	Dir.	Netflix Server	Size (B)
1471357732.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357736.7	132.240.17.11:31177	<	198.45.63.167:443	2817667
1471357736.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357741.8	132.240.17.11:31177	<	198.45.63.167:443	2816159
1471357741.9	132.240.17.11:31177	>	198.45.63.167:443	756
1471357744.4	132.240.17.11:31177	<	198.45.63.167:443	2822089
1471357744.5	132.240.17.11:31177	>	198.45.63.167:443	756
1471357748.7	132.240.17.11:31177	<	198.45.63.167:443	3117490
1471357748.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357752.7	132.240.17.11:31177	<	198.45.63.167:443	2548098

ADUdump trace of *Home* (3830 kbps encoding).

What is an ADU?

Timestamp	Local IP	Dir.	Netflix Server	Size (B)
1471357732.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357736.7	132.240.17.11:31177	<	198.45.63.167:443	2817667
1471357736.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357741.8	132.240.17.11:31177	<	198.45.63.167:443	2816159
1471357741.9	132.240.17.11:31177	>	198.45.63.167:443	756
1471357744.4	132.240.17.11:31177	<	198.45.63.167:443	2822089
1471357744.5	132.240.17.11:31177	>	198.45.63.167:443	756
1471357748.7	132.240.17.11:31177	<	198.45.63.167:443	3117490
1471357748.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357752.7	132.240.17.11:31177	<	198.45.63.167:443	2548098

ADUdump trace of *Home* (3830 kbps encoding).

What is an ADU?

Timestamp	Local IP	Dir.	Netflix Server	Size (B)
1471357732.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357736.7	132.240.17.11:31177	<	198.45.63.167:443	2817667
1471357736.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357741.8	132.240.17.11:31177	<	198.45.63.167:443	2816159
1471357741.9	132.240.17.11:31177	>	198.45.63.167:443	756
1471357744.4	132.240.17.11:31177	<	198.45.63.167:443	2822089
1471357744.5	132.240.17.11:31177	>	198.45.63.167:443	756
1471357748.7	132.240.17.11:31177	<	198.45.63.167:443	3117490
1471357748.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357752.7	132.240.17.11:31177	<	198.45.63.167:443	2548098

ADUdump trace of *Home* (3830 kbps encoding).

What is an ADU?

Timestamp	Local IP	Dir.	Netflix Server	Size (B)
1471357732.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357736.7	132.240.17.11:31177	<	198.45.63.167:443	2817667
1471357736.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357741.8	132.240.17.11:31177	<	198.45.63.167:443	2816159
1471357741.9	132.240.17.11:31177	>	198.45.63.167:443	756
1471357744.4	132.240.17.11:31177	<	198.45.63.167:443	2822089
1471357744.5	132.240.17.11:31177	>	198.45.63.167:443	756
1471357748.7	132.240.17.11:31177	<	198.45.63.167:443	3117490
1471357748.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357752.7	132.240.17.11:31177	<	198.45.63.167:443	2548098

ADUdump trace of *Home* (3830 kbps encoding).

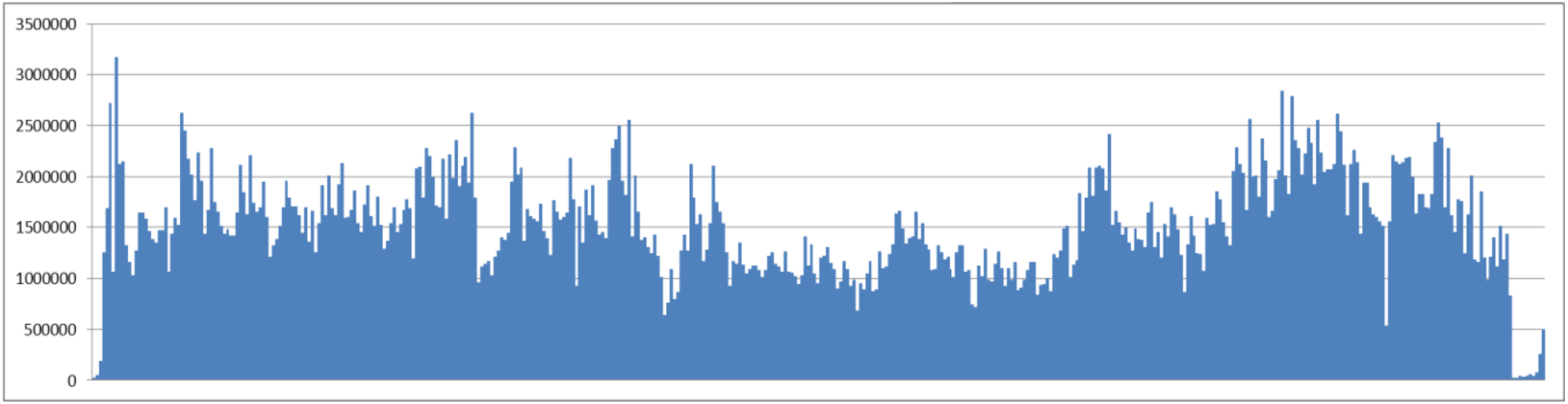
What is an ADU?

Timestamp	Local IP	Dir.	Netflix Server	Size (B)
1471357732.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357736.7	132.240.17.11:31177	<	198.45.63.167:443	2817667
1471357736.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357741.8	132.240.17.11:31177	<	198.45.63.167:443	2816159
1471357741.9	132.240.17.11:31177	>	198.45.63.167:443	756
1471357744.4	132.240.17.11:31177	<	198.45.63.167:443	2822089
1471357744.5	132.240.17.11:31177	>	198.45.63.167:443	756
1471357748.7	132.240.17.11:31177	<	198.45.63.167:443	3117490
1471357748.8	132.240.17.11:31177	>	198.45.63.167:443	756
1471357752.7	132.240.17.11:31177	<	198.45.63.167:443	2548098

ADUdump trace of *Home* (3830 kbps encoding).

VBR / DASH

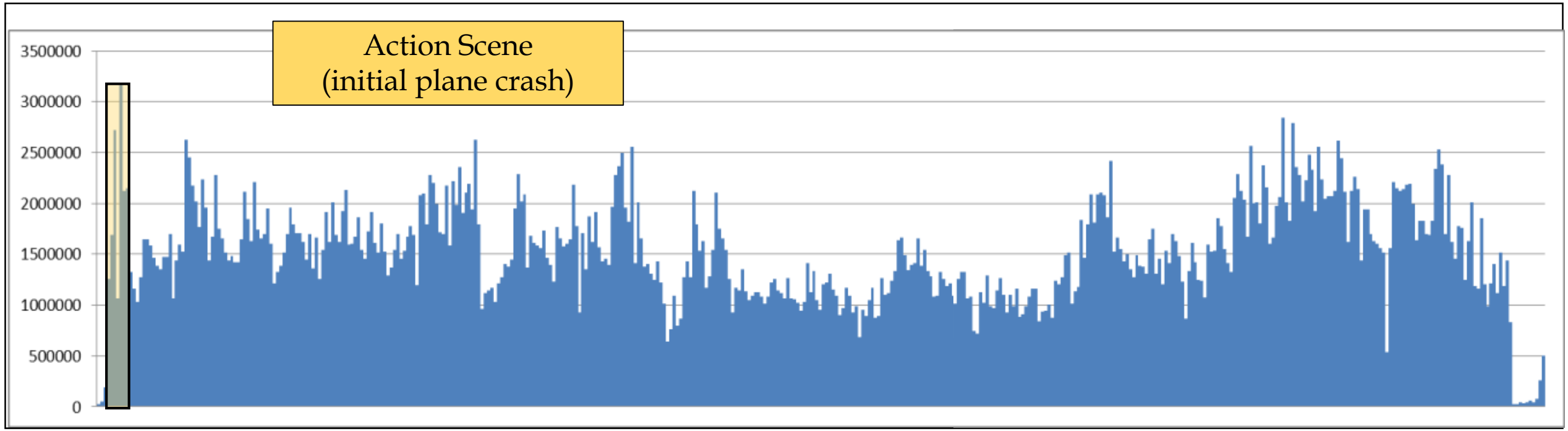
- Variable Bitrate (VBR)
- Dynamic Adaptive Streaming over HTTP (DASH)



Lost, Season 1, Ep. 1 (3000 kbps encoding)

VBR / DASH

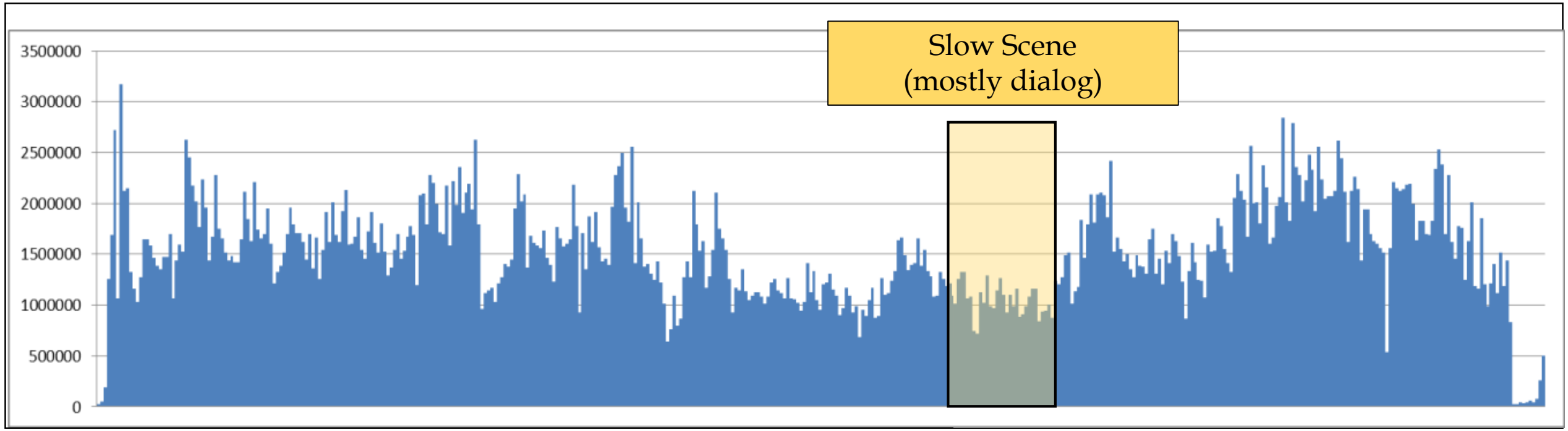
- Variable Bitrate (VBR)
- Dynamic Adaptive Streaming over HTTP (DASH)



Lost, Season 1, Ep. 1 (3000 kbps encoding)

VBR / DASH

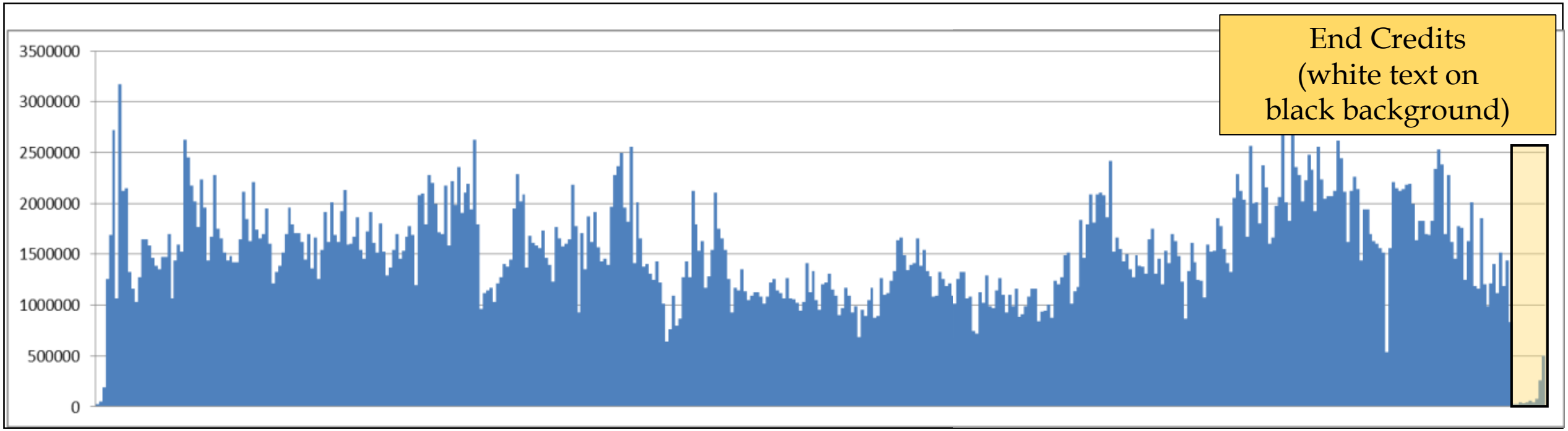
- Variable Bitrate (VBR)
- Dynamic Adaptive Streaming over HTTP (DASH)



Lost, Season 1, Ep. 1 (3000 kbps encoding)

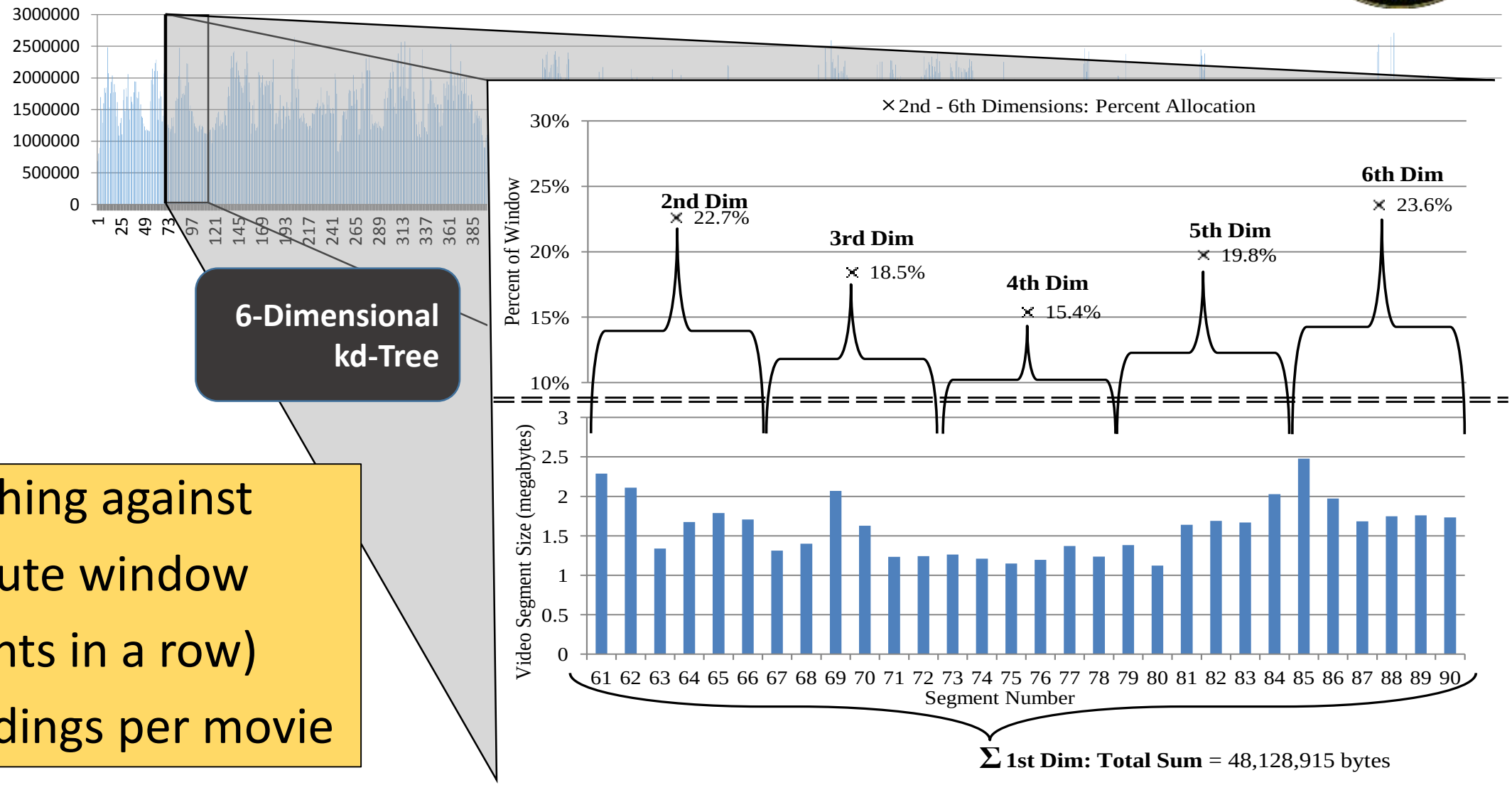
VBR / DASH

- Variable Bitrate (VBR)
- Dynamic Adaptive Streaming over HTTP (DASH)



Lost, Season 1, Ep. 1 (3000 kbps encoding)

This is our **Fingerprint**



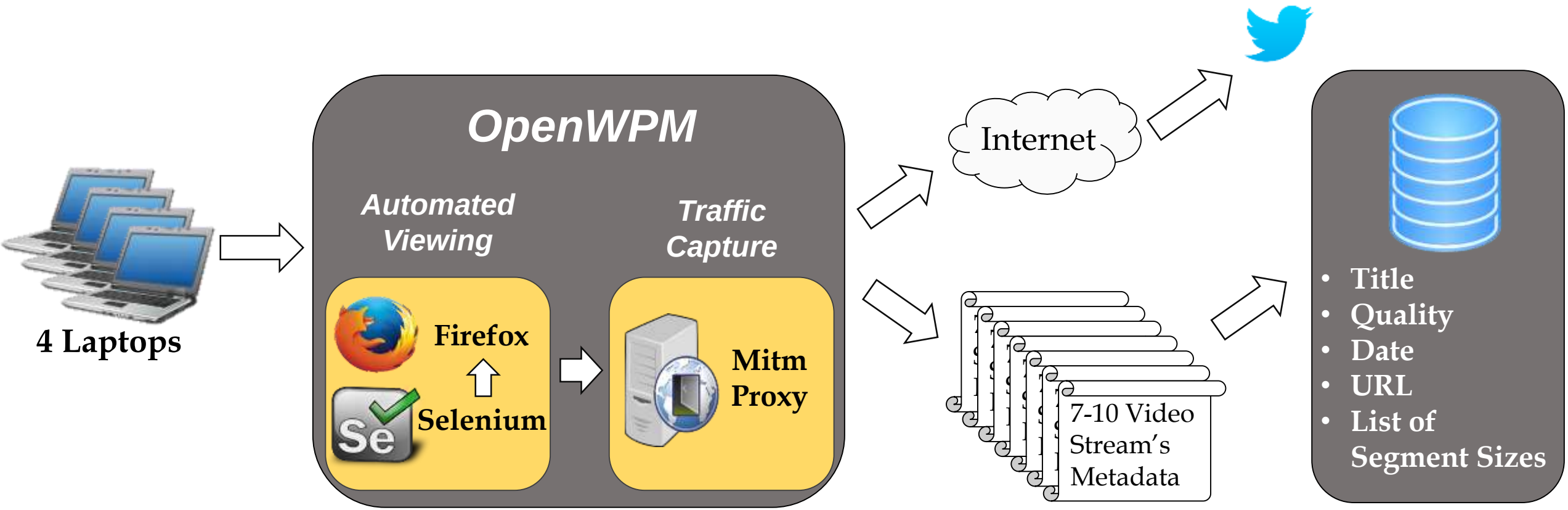
- Only matching against a two minute window (30 segments in a row)
- 7-10 encodings per movie

Research Questions

- This technique documented in previous work
  - Leaky Streams by Andrew Reed and Ben Klimkowski (CCNC '16)

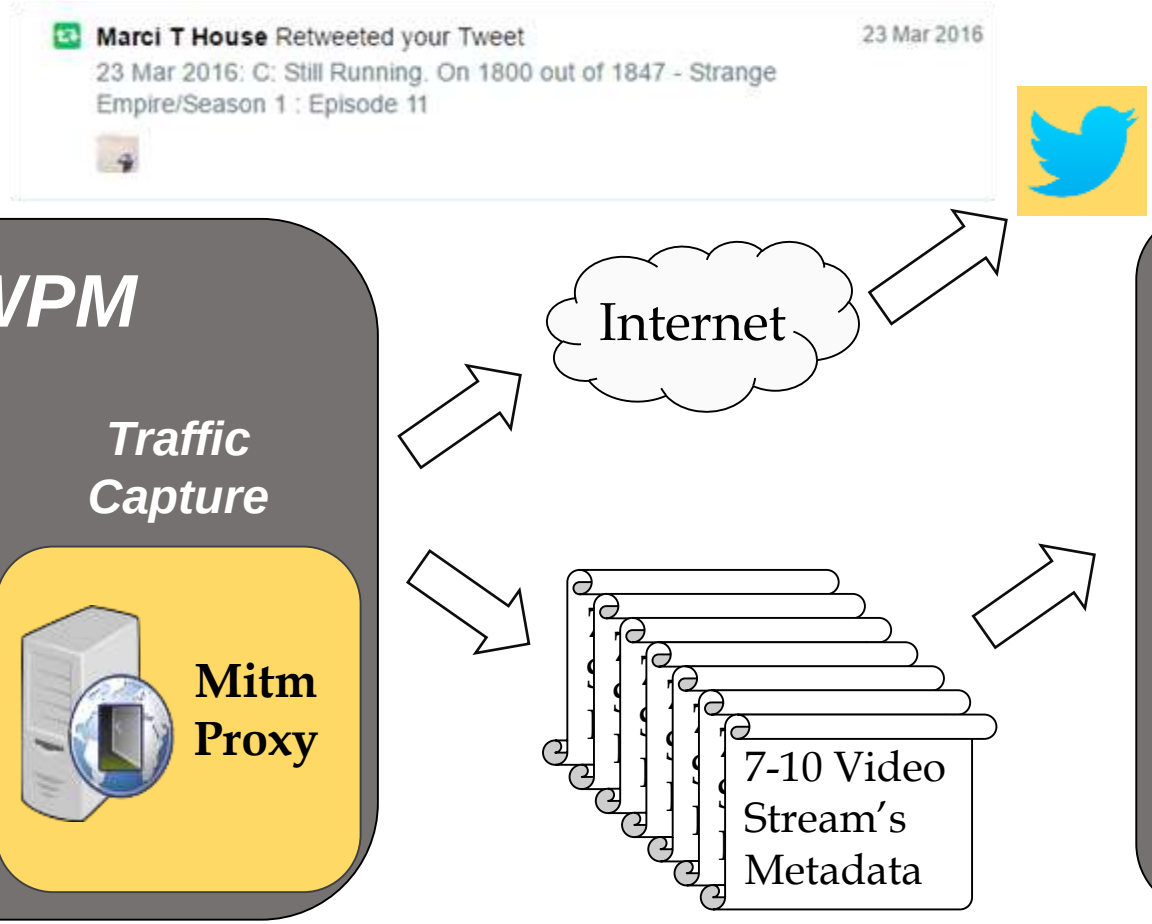
- Previous work had several problems:
  - From a predefined set of 50 videos
    - **Does these windows stay unique over all videos?**
  - Only handled one user at a time
    - **Is this analysis fast enough to handle multiple users and ISP level traffic?**

Collection Set-up



System to fingerprint *ALL* Netflix Videos

Collection Set-up

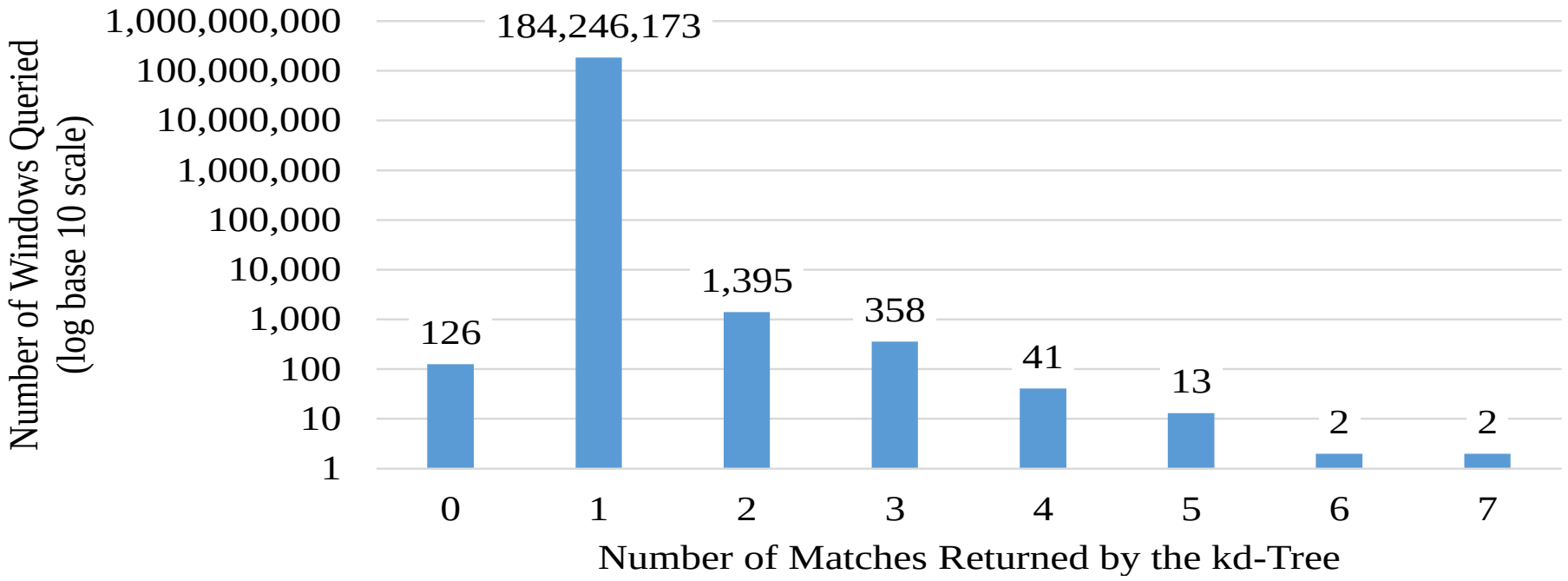


System to fingerprint *ALL* Netflix Videos

Collection Results

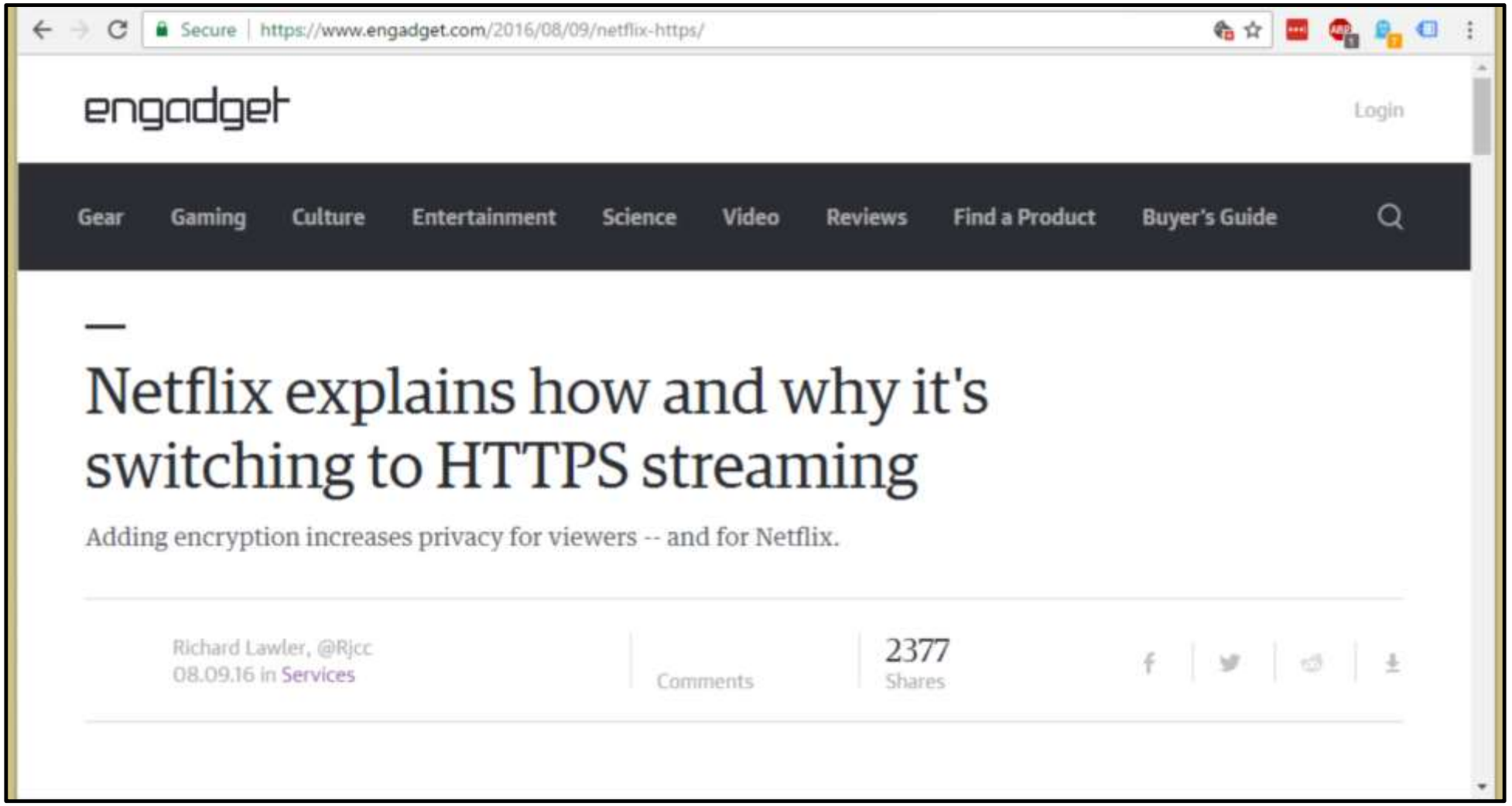
- 42,027 unique videos
  - 38,780 shows (92%)
  - 3,247 movies (8%)
- 330,264 total fingerprints (7.86 per video)
- 184,248,110 total 2-minute video windows
- Average length 38:54
  - Average movie length 1:33:30
  - Average show length 0:34:17
- 1.37GB of storage space

Collection Uniqueness



- 99.9989% of two-minute windows unique
- Non-unique windows all from abnormal video windows

Netflix Upgrades to HTTPS

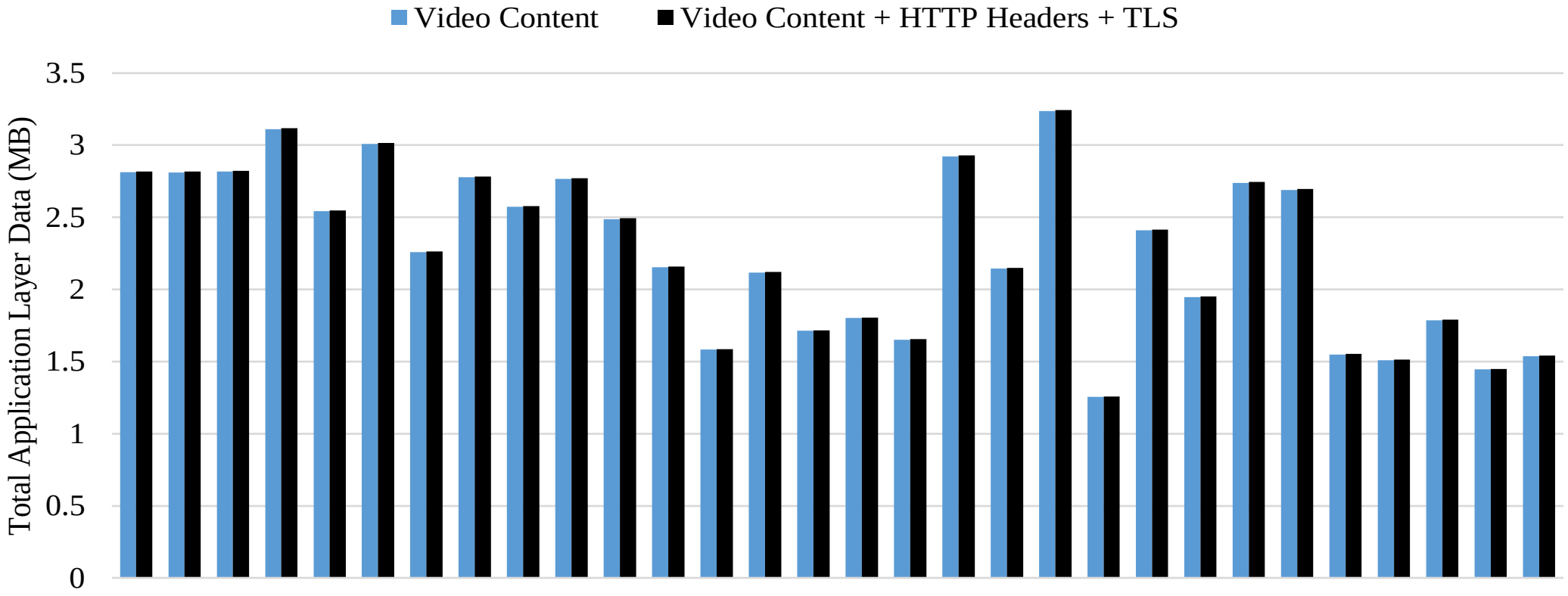


HTTPS Connection Overhead

- Will this work on HTTPS Connections?

HTTPS Connection Overhead

- Will this work on HTTPS Connections?



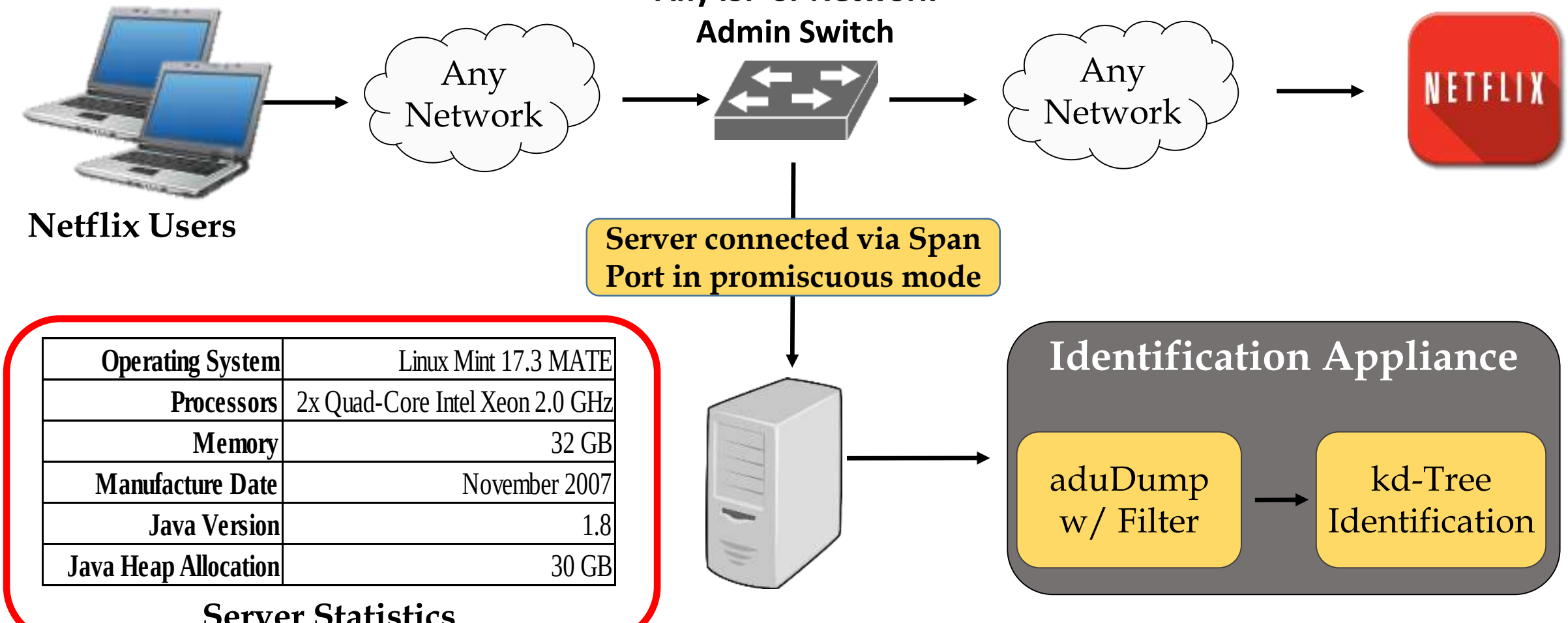
Video overhead due to HTTP headers and TLS (*Home*, 3830 kbps encoding).

Research Questions

- This technique documented in previous work
  - Leaky Streams by Andrew Reed and Ben Klimkowski (CCNC '16)

- Previous work had several problems:
  - From a predefined set of 50 videos
    - ~~Does this metadata stay unique over all videos?~~
  - Worked on HTTP Netflix connections
    - ~~What about HTTPS Netflix traffic?~~
  - Only handled one user at a time
    - Is this analysis fast enough to handle multiple users and ISP level traffic?

Video Identification Setup



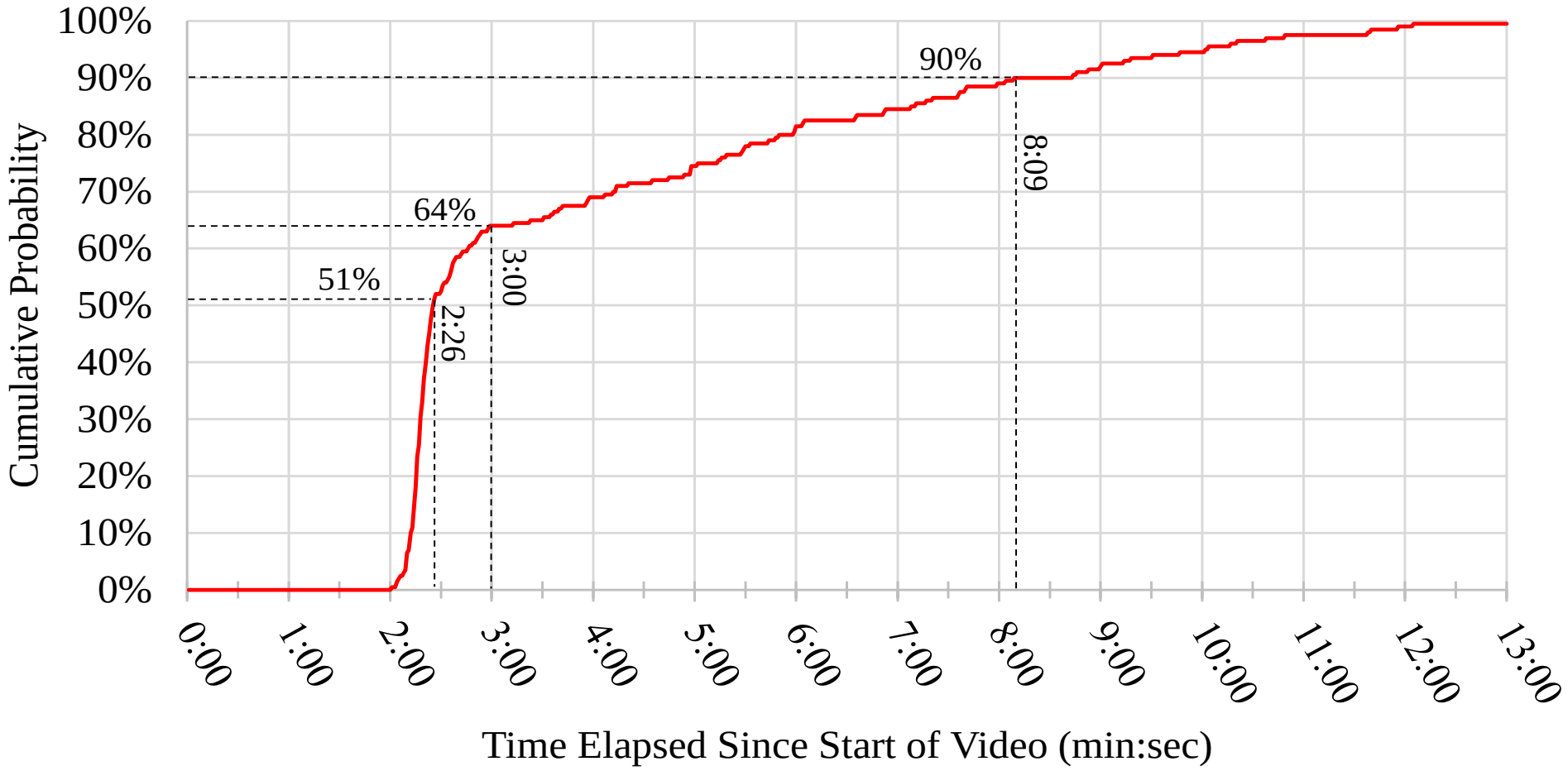
Operating System	Linux Mint 17.3 MATE
Processors	2x Quad-Core Intel Xeon 2.0 GHz
Memory	32 GB
Manufacture Date	November 2007
Java Version	1.8
Java Heap Allocation	30 GB

Server Statistics

Video Identification Results

- Identified 199/200 (99.5%) of videos watched
- Identified an average of 19:19 of the 20:00 minutes watched
- Able to conduct 8, 792 window searches per second
  - A new windows is generated every 4 seconds per video
  - Can identify approx. 35,000 simultaneous streams

Video Identification Result



Cumulative probability of identifying a video over time

This Attack Is Preventable

- Vary segment lengths
- Request multiple segments at once
- Do not request segments in order

Make the data less unique

So What?

- Metadata matters.
- Unique enough data cannot rely on encryption for confidentially.
- Application Data Unit (ADUs) provide an interesting middle ground for Network Traffic analysis.

Thank you!

Questions?

www.mjkranch.com