Identifying HTTPS-Protected Netflix Videos in Real-Time

UNITED STATES MILITARY ACADEMY
WEST POINT

Research Problem

Video traffic currently dominates global IP traffic, accounting for an estimated 70% of all traffic in 2015. Dynamic Adaptive Streaming over HTTP (DASH) is one of the most popular video streaming techniques and is used by some of the market's biggest players (e.g. Netflix and Amazon). Previous work showed that DASH with variable bitrate (VBR) is vulnerable to identification but left a few significant questions unanswered.

Question 1: Can we accurately identify DASH videos at scale?

The previous work was only able to identify one video at a time from a pre-defined set of 50 manually cataloged videos. Netflix alone has a library of over 20,000 videos which changes monthly. Can we do this identification in an automated fashion given any Netflix video and simultaneous users?

Sub-Question 1: Can we fingerprint and identify every single Netflix video?

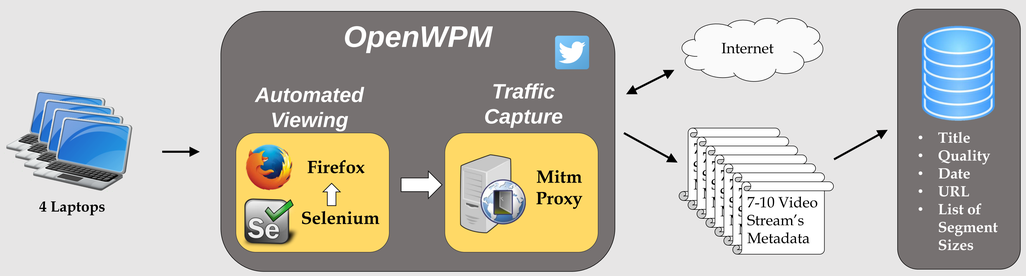
Sub-Question 2: Can our identification algorithm handle ISP equivalent network traffic volume?

Question 2: Can we do this identification with encrypted traffic?

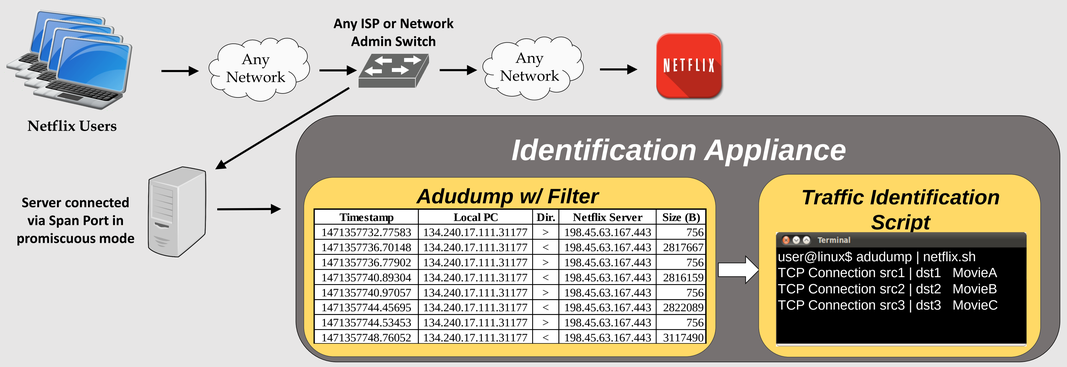
The previous work also only identified HTTP Netflix traffic. Netflix recently switched to using HTTPS to both authenticate and encrypt their video streams in order to improve their privacy. While this switch prevents many previously disclosed deep-packet inspection video identification techniques, does this change prevent us from using Application Data Unit (ADU) sizes to identify DASH videos?

System Setup

Fingerprint Collection



Video Identification



Timestamp	Local PC	Dir.	Netflix Server	Size (B)
1471357732.77583	134.240.17.111.31177	>	198.45.63.167.443	756
1471357736.70148	134.240.17.111.31177	<	198.45.63.167.443	2817667
1471357736.77902	134.240.17.111.31177	>	198.45.63.167.443	756
1471357740.89304	134.240.17.111.31177	<	198.45.63.167.443	2816159
1471357740.97057	134.240.17.111.31177	>	198.45.63.167.443	756
1471357744.45695	134.240.17.111.31177	<	198.45.63.167.443	2822089
1471357744.53453	134.240.17.111.31177	>	198.45.63.167.443	756
1471357748.76052	134.240.17.111.31177	<	198.45.63.167.443	3117490

Results

Feasibility

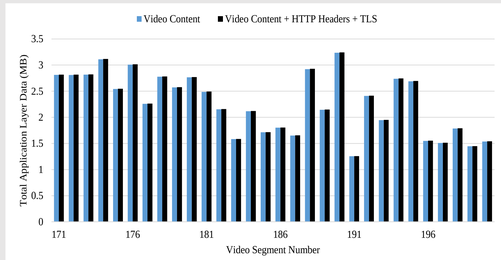


Figure 1: Netflix video overhead due to HTTP headers and TLS (*Home*, 3830 kbps).

Fingerprint Collection

Collection Statistics

- 4 laptops over 7 days (live tweeting)
- **1.37GB of storage space**
- 42,027 unique videos
  - 38,780 shows (92%)
  - 3,247 movies (8%)
- 330,264 total fingerprints
  - Average of 7.86 per video
- 20 seconds per video
- Average length 38:54
  - Average movie length 1:33:30
  - Average show length 0:34:17

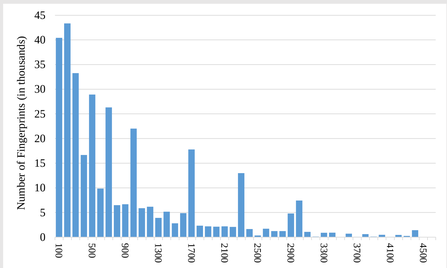


Figure 2: Number of fingerprints by average bitrate in 100 kbps bins.

Video Identification

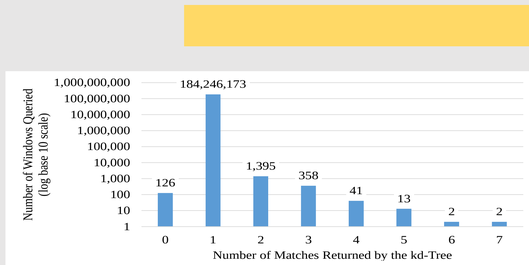


Figure 3: Number of matches returned by the kd-Tree when searching for each window.

Identification Statistics

- **99.9989% of two-minute video windows are unique**
- Two sets of 100 random videos for 20 minutes each
- Identified 99.50% of random test traffic video streams
- Identified 96.12% of total viewing activity
- Minimal computing hardware required :
  - Processor: 2x Quad-Core 2.0 GHz
  - Memory: 32 GB Memory
- **92,000 concurrent streams (Efficient mode)**
- 35,000 concurrent streams (Exhaustive mode)

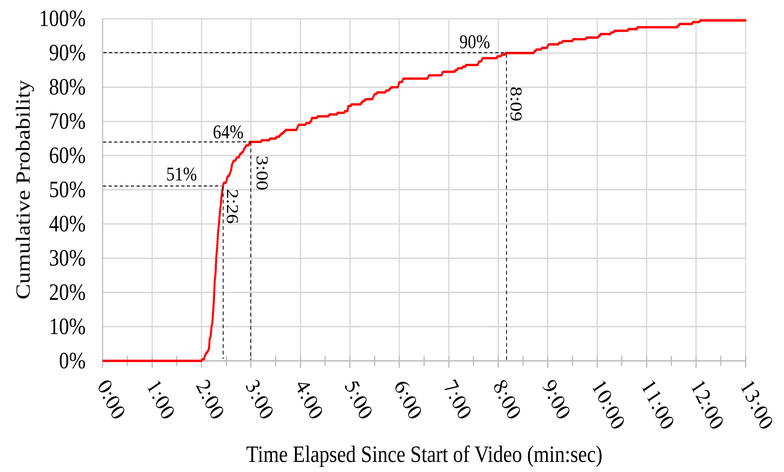


Figure 4: Cumulative probability of identifying a video over time

Conclusions

- Even with encryption, variable bitrate encoding still leaks details of the underlying content.
- Application Data Units provide an interesting vantage point to track data streams without doing packet level analysis.
- An ISP or network administrator could easily do this type of analysis with minimal hardware requirements.
- To prevent this type of attack, Netflix could modify the data requested at the application layer by making non-sequential segment requests or requesting multiple segments worth of data at once.
- ***Application developers need to consider the patterns in the data that they pass to the transport layer instead of relying entirely on encryption to provide confidentiality.***

ACM Conference on Data and Applications Security and Privacy (CODASPY 2017)
March 22 – 24, 2017
Scottsdale, Arizona, USA

Michael Kranch, Andrew Reed
michael.kranch@usma.edu, andrew.reed@usma.edu
Electrical Engineering and Computer Science Department
United States Military Academy, West Point, NY 10996