De-anonymizing Encrypted Video Streams

Master's Thesis

7 September 2018

Stefano Peverelli
pstefano@student.ethz.ch

supervised by
Melissa Licciardello

Systems Group
Department of Computer Science
ETH Zürich

Abstract

In the last recent years streaming services such as Netflix, Youtube, Amazon Prime Video, Hulu and others, have become the main source for video content delivery to the public. With the effort of private companies and of the AOM consortium, various coding formats and streaming techinques have been refined and have gained popularity. *Adaptive Bitrate Streaming*, between others, enables high quality streaming of media content over HTTP, and represents nowadays the industry's standard.

DASH *Dynamic Adaptive Streaming over HTTP* is an instance of Adaptive Bitrate Streaming originally developed by MPEG. In DASH each media file gets encoded at multiple bitrates, which are then partitioned into smaller segments and delivered to the user over HTTP. Netflix's use of DASH services is no mistery, indeed it is already five years that each title on Netflix sits with its own different bitrate copies on a CDN, waiting to be served to clients in a particular area of the planet. [1]

Despite a recent upgrade in Netflix infrastructure to provide HTTPS encryption of each video stream, research shows that the privacy of the end user is at risk, more precisely there exist techniques to identify the content the client is playing as Reed et Al. [2] have shown. They make use of adudump [3] a command live program that uses TCP sequence and ACKS to infer the sizes of application data unit *ADUs* transferred over each TCP connection. Our approach is mainly based on their work, with few differences highlighted in Section X.

Contents

1

Introduction

According to the latest Cisco's VNI [4], video will account for 82% of all IP traffic in Europe by 2021; moreover the overall IP traffic per person will triplicate from 13*GB* to 35*GB*. These forecast clearly picture the growth of the streaming industry, posing, at the same time an important question on the present and future states of the final user's privacy.

As shown by Reed et Al. [2] anonimity of user's viewing activity is at risk. Not for the use that Netflix or other streaming services do of user's session data, but because of the risk of a man-in-the-middle attack *MITM* by an *evil* party.

In particular, they have shown how the adoption of HTTPS to protect video streams from Netflix *CDN*s to user's end devices, does not hold against passive traffic analysis.

1.1 Motivation

Our intent is to replicate the work of Reed et Al. and to observe if recent changes in Netflix infrastructure can guarantee anonimity of user's viewing activity.

Bibliography

[1] Netflix TechBlog. Per-title encode optimization, 2015. URL https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2.

[2] Andrew Reed and Michael Kranch. Identifying https-protected netflix videos in real-time. In *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*, CODASPY '17, pages 361–368, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4523-1. doi: 10.1145/3029806.3029821. URL http://doi.acm.org/10.1145/3029806.3029821.

[3] Jeff Terrell, Kevin Jeffay, F. Donelson Smith, Jim Gogan, and Joni Keller. Passive, streaming inference of tcp connection structure for network server management. IEEE International Traffic Monitoring and Analysis Workshop, 2009.

[4] Business Insider Intelligence. Video will account for an overwhelming majority of internet traffic by 2021, 2019. URL https://www.businessinsider.com/heres-how-much-ip-traffic-will-be-video-by-2021-2017-6?r=US&IR=T.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

De-anonymizing encrypted video streams

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Stefano	Peverelli

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zurich, 07.09.2019	

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.