DATX02-22-19

Streaming Video Identification

Target

Are you a D, DV and IT student and interested in **networking** and **network security** and would like to learn about and manipulate:

- **Scapy** for *Packet and Packet Traces Manipulation* and **Tcpdump** for capturing network traffic;
- **KdTrees** (k-dimensional data structure) for building a fingerprint database;
- **Sklearn** for *supervised learning* using **SVM** (Support-Vector Machines);

Then this project is for you!

Project description

Background: There are numerous well-known cases of user and content identification using among others GPS traces, comments on IMDB, medical "anonymized" records and recently **encrypted network packet traces**. *Network traffic analysis*, using information such as packet sizes and their time of arrival, is a quite interesting case as it represents hardly controllable data that can easily be captured or leaked and demonstrates to what extend one can learn about someone's identity and his/her behaviors solely based on encrypted network traffic. You can easily imagine the privacy concerns raised by divulging one's sleep patterns (from sleep-tracking device's traffic) or presence at home (from motion detection camera's traffic).

Project's idea: This bachelor project explores network trafic analysis and focuses on identifying streamed video content based on network traffic. This project suggests 3 mostly independent key components:

1. **Video Recognizer**. From a small dataset of 25 videos, the starting objective is to reproduce and/or adapt the methods that were developed to identify encrypted Netflix videos [1]. You should explore then (and develop/check your own solutions) the possibility to recognize videos that start playing from an arbitrary point. Also if time allows, check if your solution can work on another streaming service (ie youtube or Apple TV+).
2. **Local Network Test**. Reproduce and analyze the following experiment: 2 computers are set within the same network, one streaming a video through an online platform and the other running the recognizer.
3. **Local Traffic Interface**. Develop a basic web interface to display some analysis (eg number of connections, packet per second etc) of traffic from computers on same LAN.

Outcome: Putting all the parts together, the output could be a working real-time video content recognizer but the project can also be driven by students' interests in exploring some aspects more in depth.

Suggested starting material

Some introductory videos on video streaming technologies:

- What is a Content Delivery Network (CDN)? *https://www.youtube.com/watch?v=Bsq5cKkS33I*
- How Does Netflix Work? *https://www.youtube.com/watch?v=YXQpgAAeLM4*
- An Introduction to the Netflix Stack *https://www.youtube.com/watch?v=8iWNUVe346o*
- Apple TV+ Explained | Everything You Need To Know *https://www.youtube.com/watch?v=dtetDGWcS_E*
- Technical details of Netflix *https://en.wikipedia.org/wiki/Technical_details_of_Netflix*

Starting scientific publication:

- **[1]** Andrew Reed and Michael Kranch. Identifying https-protected netflix videos in real-time. In Proceedings of the 7th ACM Conference on Data and Application Security and Privacy, 2017.

Special prerequisites

- a course in Computer Communication (EDA343, EDA344, LEU062) or equivalent knowledge.
- a genuine interest in networking technologies is also recommended.

Proposal author & Supervisor

Romaric Duvignau

Communication language for the project **English**; writing language for the reports English or Swedish.