Idio Guarino
Traffic analysis of mobile communication and collaboration apps via ML approaches

Tutor: Prof. Antonio Pescapè

Cycle: XXXVI   Year: Second

My background

- **MSc degree**: Computer Engineering, University of Naples Federico II

- **Research group/laboratory**: Traffic Group/ARCLab

- **PhD start date**: 01/11/2020

- **Scholarship type**: no scholarship. Funded by Consortium GARR through the awarded "O.Carlini" research grant.

- **Collaboration**: Huawei R&D center in Paris, France

Mobile traffic growth

Mobile network traffic grows year after year (**+29%** of global bandwidth from **2020** to **2021, Sandvine 2022**)

Starting from the beginning of the Covid-19 pandemic, the increasing use of **communication and collaboration apps (CC Apps)** (e.g., Zoom, Teams, etc.) has caused rapid and sudden growth in traffic volumes.

Global mobile network data traffic (EB per month)
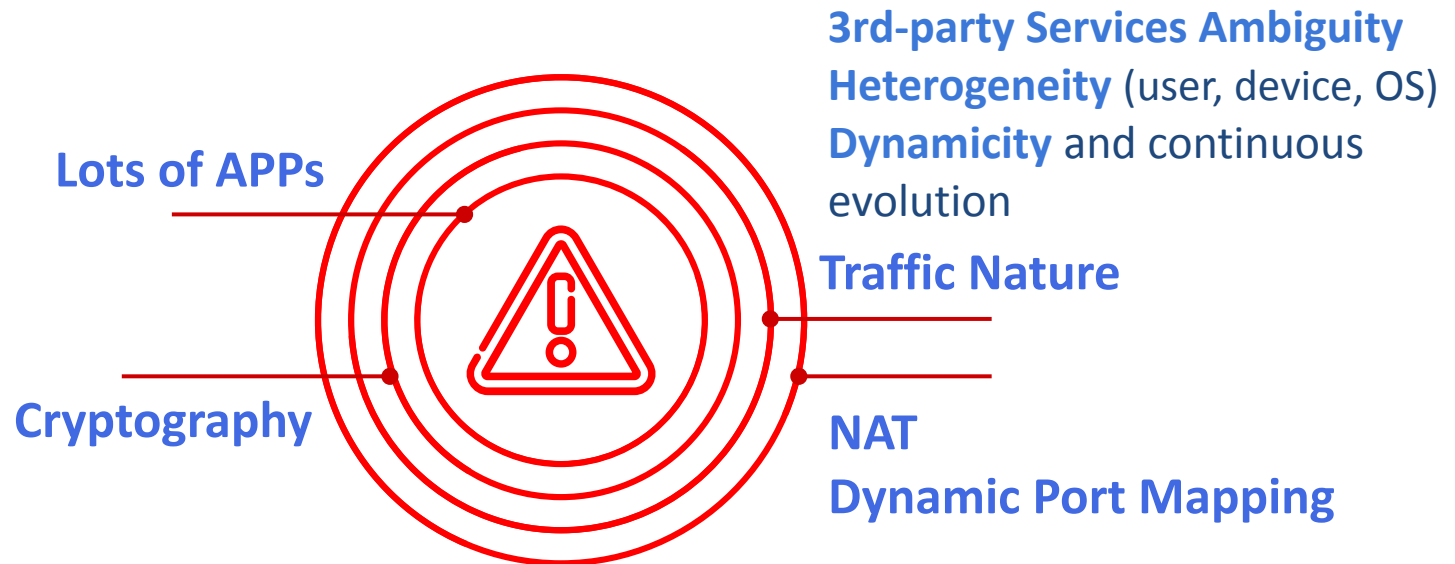Ericsson Mobility Report | June 2022
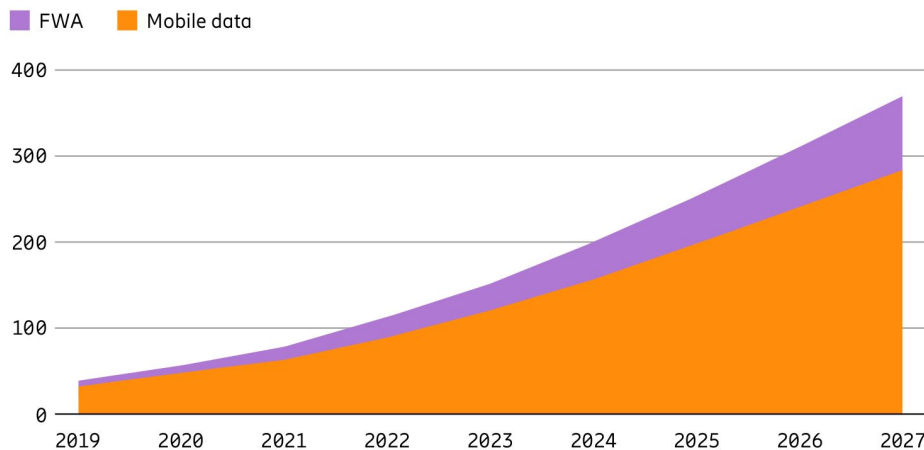
*Sandvine, The Global Internet Phenomena Report, 2022

Research field of interest

Network operators need to know the characteristics of the traffic flowing through their network, to perform a better management and optimization

Many Challenges

3rd-party Services Ambiguity
Heterogeneity (user, device, OS)
Dynamicity and continuous evolution

Lots of APPs

Traffic Nature

Cryptography

NAT
Dynamic Port Mapping

How to manage these changes?

Traffic Classification

What's going through my network?

Traffic Prediction

How will traffic characteristics evolve?

Application Scenarios

QoS **Quality of Service**

Resource Management

Security

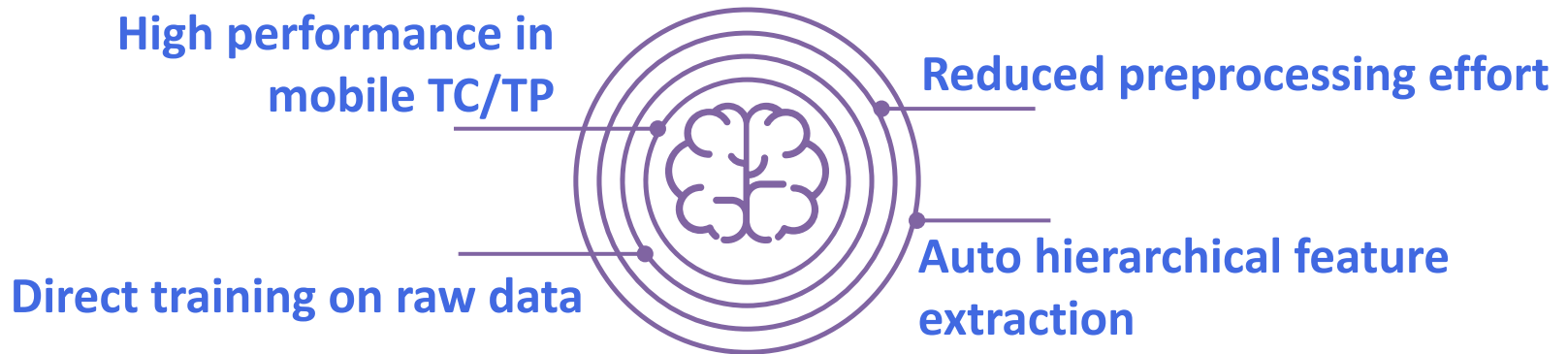**Traffic Engineering**

My Research

Design and evaluation of **Deep-Learning approaches** for Traffic Classification and Traffic Prediction, mainly focused on traffic generated by **CC Apps**

High performance in mobile TC/TP

Reduced preprocessing effort

Direct training on raw data

Auto hierarchical feature extraction

Our Dataset: MIRAGE COVID-CCMA-2022

Skype, Teams, Zoom, etc.
9 Android APPs

Human Generated Traffic Data

3 User ACTIVITIES
Audio-call, Chat, and Video-call

**Reliable labeling**
bi(directional)-flow level

Traffic Classification of CC Apps

Focus on **CC apps** and the different **activities** the user can performs with the same each of them (e.g. Chat, Audio/Video-Call, etc.)

State-of-the-art Traffic Classifiers

● 😁 Excellent performance when used for *app classification* (**App-TC**)

● 😡 Poor performance when used for *activity classification* (**Activity-TC**)
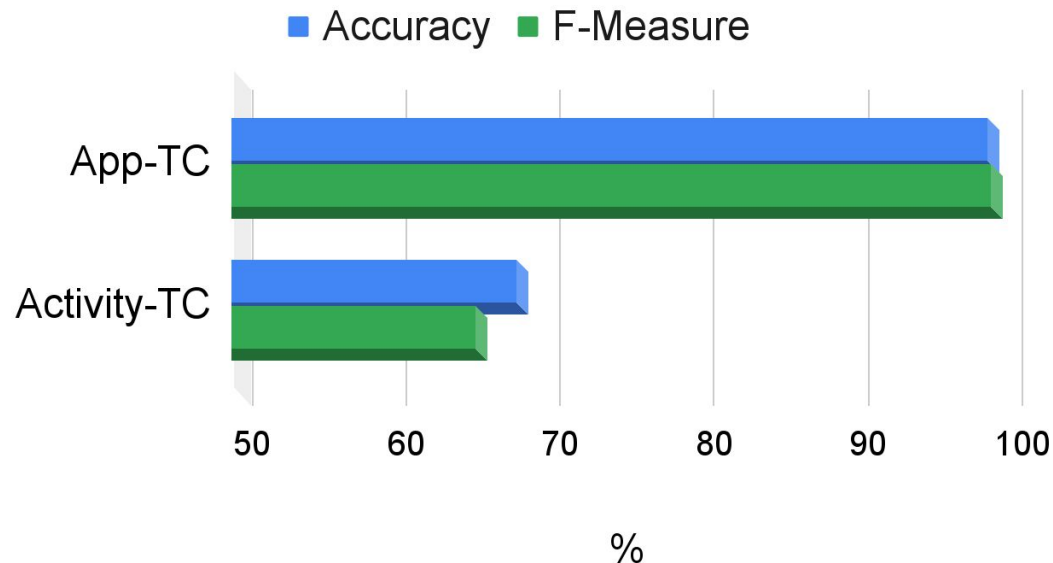
My Proposal

- An **innovative approach** to solve **Activity-TC** based on a novel set of features (**Context Input**) which take into account the context of the biflow to be classified (i.e. the set of co-existing biflows which run in parallel) [J1]
- Design of a multi-modal architecture (**MIMETIC-ALL**) to capitalize on heterogeneous information extracted from network traffic (viz. modalities)
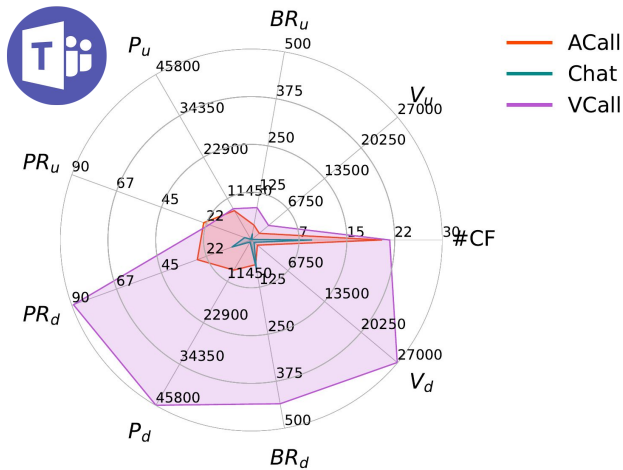

BR=Byte-Rate, **PR**=Packet-Rate,
V=Volume, **P**=Amount of packets,
#CF=Number of Co-existing biflows.
u=upstream, **d**=downstream
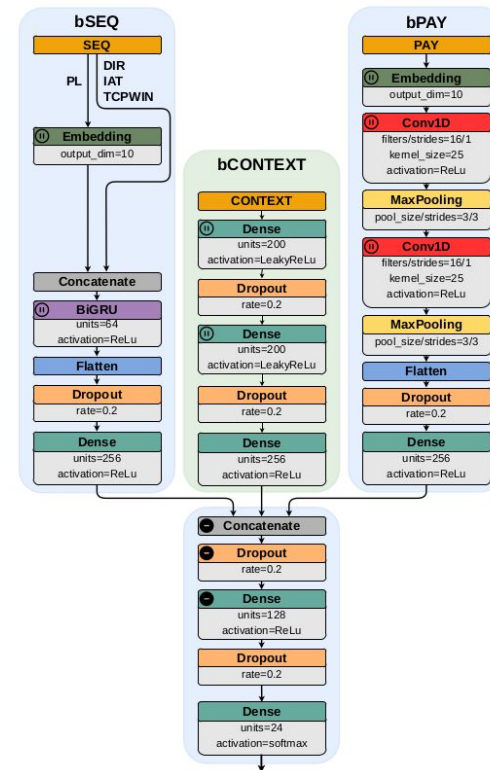
MIMETIC-ALL

My Results

MIMETIC-ALL: **Context-inputs** combined with inputs commonly used in the literature (e.g., payload bytes and features extracted from packet headers) leads to a **significant increase** in performance when facing the user activity classification (**+17%** of F-measure) [J1]
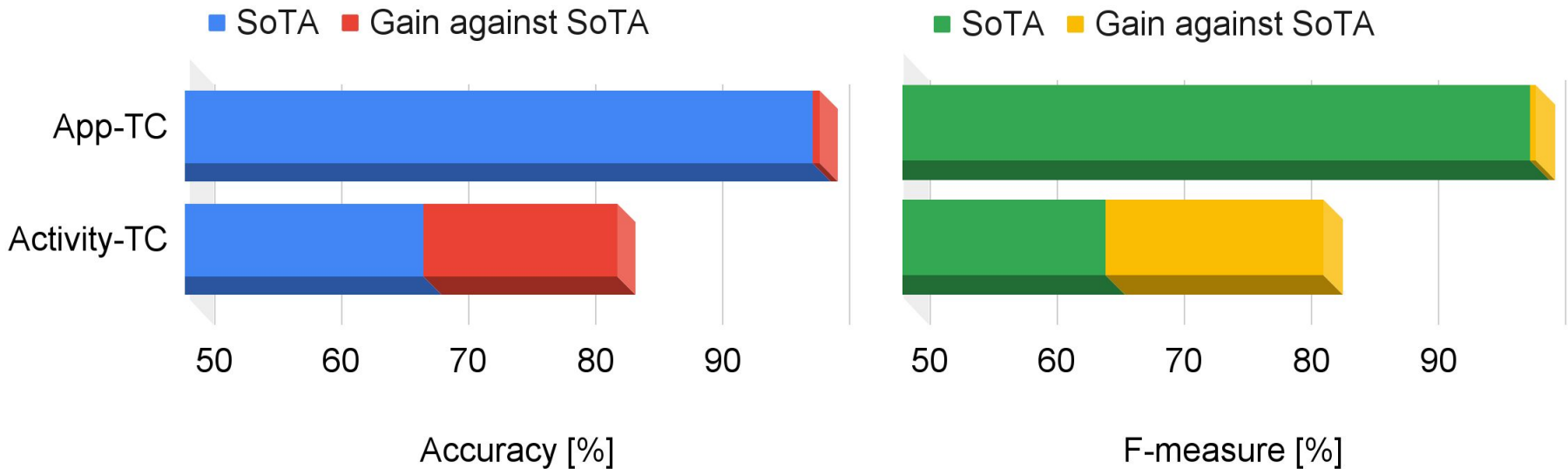
Additionally, **Context Input** makes models **more reliable** and **robust**

My Period Abroad

Where: Huawei R&D Center, Paris, France

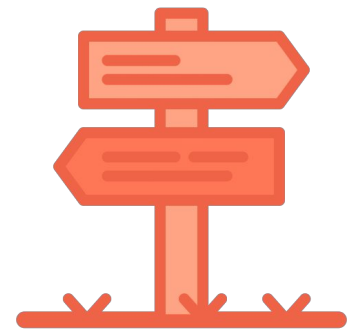**Duration:** 25/07/2022 - ongoing (estimated end 25/01/2023)

Research Topic: apply Few-Shot learning approaches in the context of Internet traffic classification:

- **Few-Shot Learning:** making classification (or regression) based on a very small number of samples
- The **limited number of samples** makes a classification (regression) problem **harder** than a standard supervised problem (with huge samples)
- **Application Scenario:** build a classifier (Intrusion Detection System) able to detect a new app (attack) for which we do not yet have much data available.

Next Year

- Continuation of my research on the application of **Few-Shot Learning** techniques in the field of **internet traffic classification**

- Design and evaluation of a *novel* approach able to **solve Traffic Classification** and **Traffic Prediction simultaneously** by exploiting the *multi-modal* and *multi-task learning*

My Products

[C1]	*"On the use of Machine Learning Approaches for the Early Classification in Network Intrusion Detection"*, I.Guarino, G. Bovenzi, D. Di Monda, G. Aceto, D. Ciuonzo and A. Pescapè, 2022 IEEE International Symposium on Measurements and Networking (M&N), 2022. Published
[J1]	*"Contextual Counters and Multimodal Deep Learning for Activity-Level Traffic Classification of Mobile Communication Apps during COVID-19 Pandemic"*, I. Guarino, G. Aceto, D. Ciuonzo, A. Montieri, V. Persico and A. Pescapé, Elsevier Computer Networks, Special issue on Machine Learning empowered Computer Networks, 2022. Accepted.
[C2]	*"Fine-Grained Traffic Prediction of Communication-and-Collaboration Apps via Deep-Learning: a First Look at Explainability"*, I. Guarino, G. Aceto, D. Ciuonzo, A. Montieri, V. Persico and A. Pescapé, 2023 IEEE International Conference on Communications (ICC). Submitted.

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	6.0	1.5	2.5	0	10.0
Bimonth 2	0	1.5	8.5	0	10.0
Bimonth 3	0	0.7	9.3	0	10.0
Bimonth 4	0	3.4	7.8	0	11.2
Bimonth 5	0	0	8.8	0	8.8
Bimonth 6	0	0	10.0	0	10.0
Total	6.0	7.1	46.9	0	60
Expected	30 - 70	10 - 30	80 - 140	0 – 4.8	

Thanks for the attention!