Antonio Montieri

Tutor: Prof. Antonio Pescapè

XXXII Cycle - III year presentation

Methodologies for Mobile and Encrypted Traffic Classification via Machine Learning Approaches

Università degli Studi di Napoli Federico II

Background

- ***Graduation:*** M.Sc. Degree in Computer Engineering, cum laude
- ***DIETI Group:*** Computer Networks *COMICS* and *TRAFFIC* research groups
- ***Cooperations:*** HUAWEI simulamet institute imdea networks
- ***Fellowship:*** University Ph.D. grant

Credits Summary

Student: Antonio Montieri **Tutor: Antonio Pescapè** **Cycle XXXII**

antonio.montieri@unina.it antonio.pescape@unina.it

	Credits year 1							Credits year 2							Credits year 3								
	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	**Summary**	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	**Summary**	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	**Summary**	**Total**	**Check**
Modules	4	3	0	6	4	6	**23**	0	3	2	0	1	0	**7**	0	2	0	0	1	1	**3**	32,6	30-70
Seminars	5	2	2	1	0	0	**10**	1	0	4	0	0	0	**6**	0	0	2	0	0	0	**2**	17	10-30
Research	1	5	8	4	6	7	**30**	9	7	4	10	8	10	**48**	10	8	8	9	9	9	**52**	130,4	80-140
	10	10	10	10	10	13	**63**	10	10	10	10	10	10	**60**	10	10	10	9	10	10	**57**	180	180

Experiences Abroad

simulamet

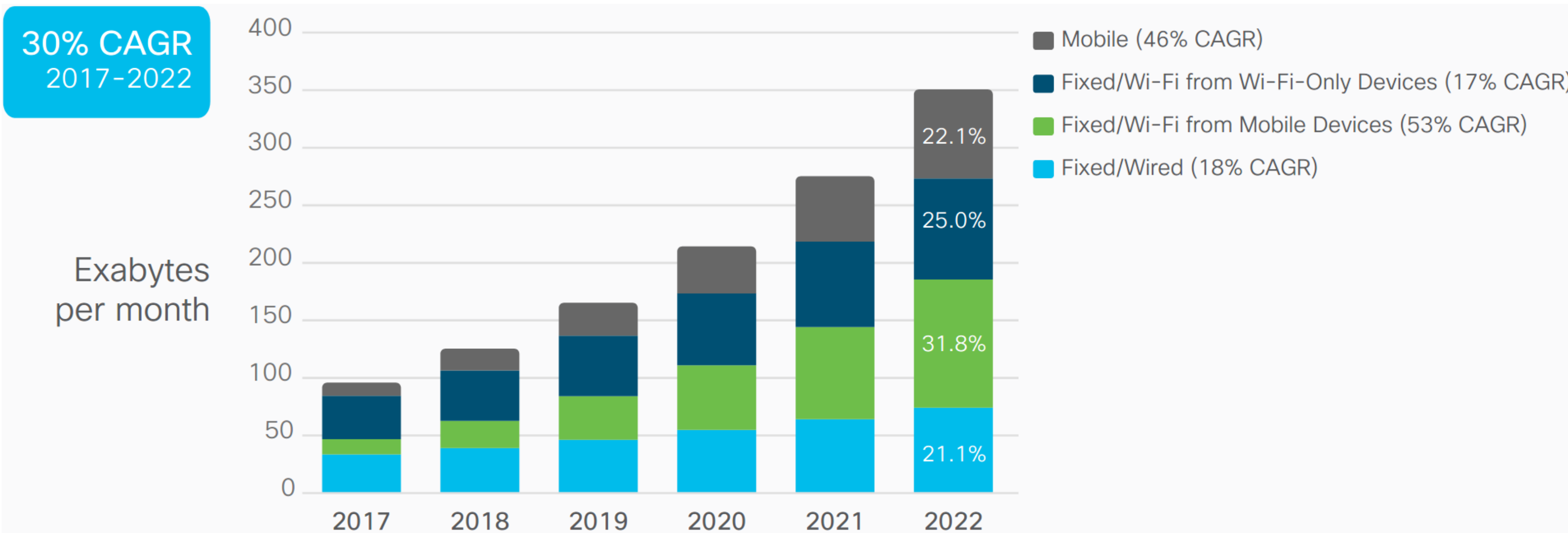
- ***Where:*** Simula Metropolitan Center for Digital Engineering, Oslo, **Norway**
- ***Duration:*** 1.7 months
- ***End:*** 19/07/2018 - 08/09/2018
- ***Advisor:*** Prof. Özgü Alay
- ***Topic:*** Analysis of the Quality of Service and Quality of Experience of communication networks in **high-mobility scenarios**

institute imdea networks

- ***Where:*** IMDEA Networks, Leganes (Madrid), **Spain**
- ***Duration:*** 2.6 months
- ***End:*** 24/01/2019 -10/04/2019
- ***Advisor:*** Prof. Narseo Vallina-Rodriguez
- ***Topic:*** Investigation of mobile-traffic **privacy** and the **3rd-party tracking ecosystem**

Motivations:
Mobile Traffic Growth

Wi-Fi and mobile devices will account for **79% of Internet traffic** by 2022



Source: Cisco 2019 VNI Global IP Traffic Forecast, 2017-20

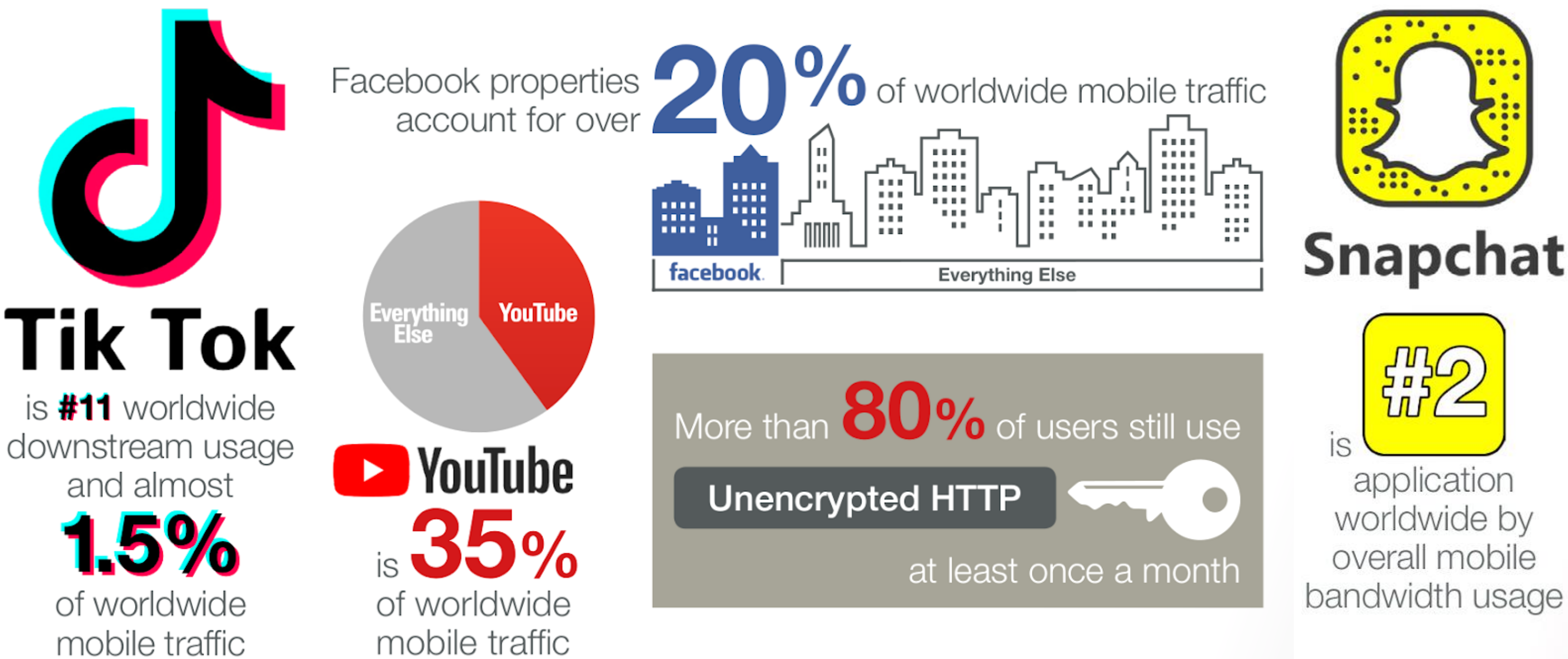
Mobile Traffic Classification

What is flowing through my (mobile) network?

Mobile Traffic Classification

What is flowing through my (mobile) network?

Associating traffic classification objects with the mobile apps that generate them



Source: Sandvine, The Mobile Internet Phenomena Report, 2019

Mobile Traffic Classification: Main Drivers

- Classification of mobile traffic **provides valuable information** for
  - Advertisers
  - Insurance companies
  - Security agencies
  - Infrastructure Operators
  - ...

- But also **raises privacy issues**
  - Indiscriminate surveillance
  - Context-sensitive apps
  - Bring your own device policy
  - ...

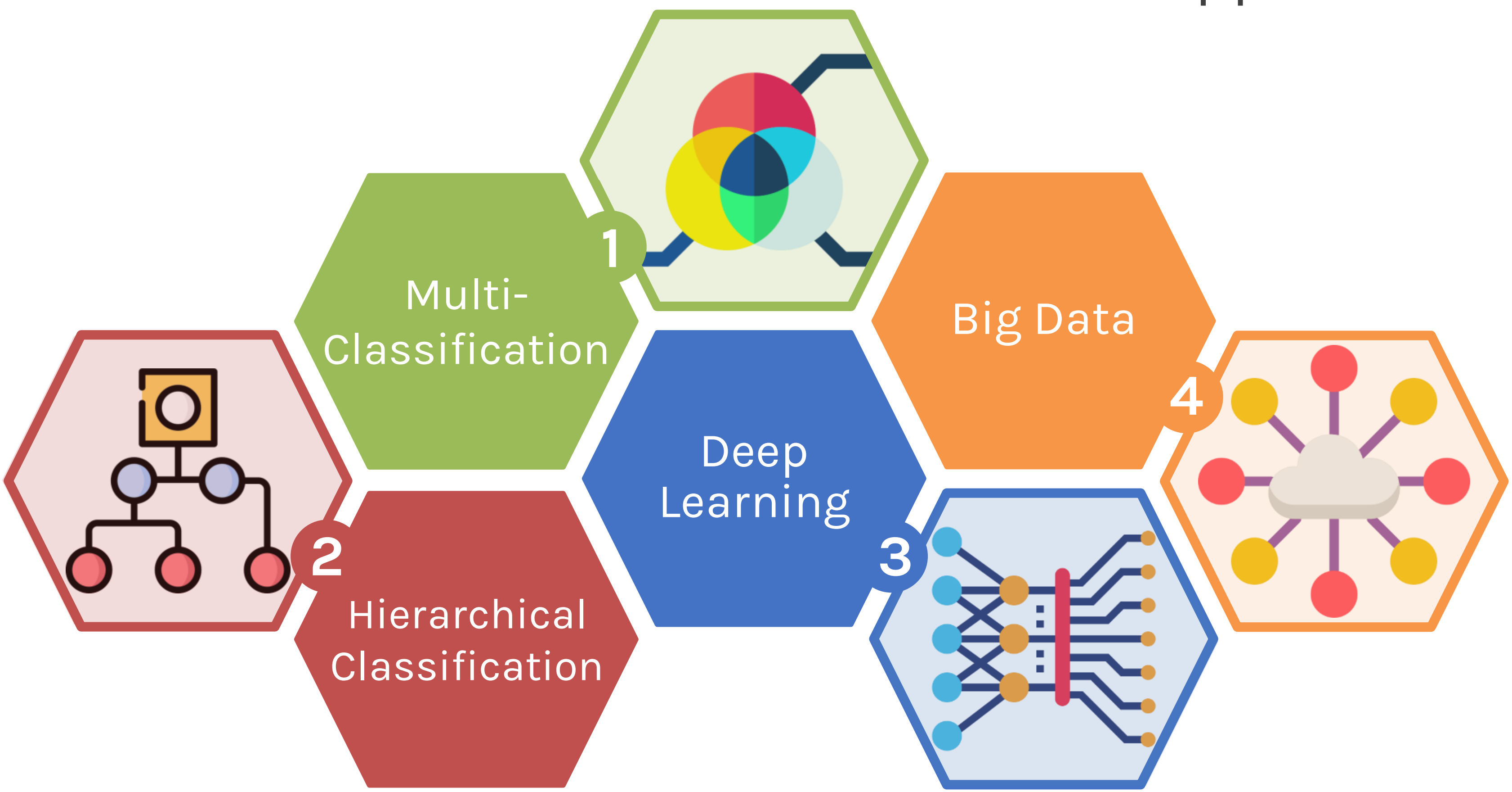
Mobile Traffic Classification: Main Challenges

- **Huge volume** of mobile traffic evolving at an unprecedented pace
  - **One-click** installation
  - Quick-paced automatic **updates**
  - **Different versions** of apps and/or operating systems running on **different devices**

- Increasing adoption of **encrypted protocols**
  - hinders methods based on Deep Packet Inspection
  - requires approaches based on **Machine Learning (ML)**

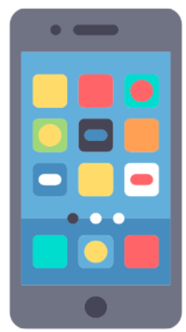
My Contribution

Proposing **novel methodologies** for encrypted and mobile Traffic Classification (TC) via ML approaches



Data quality is critical

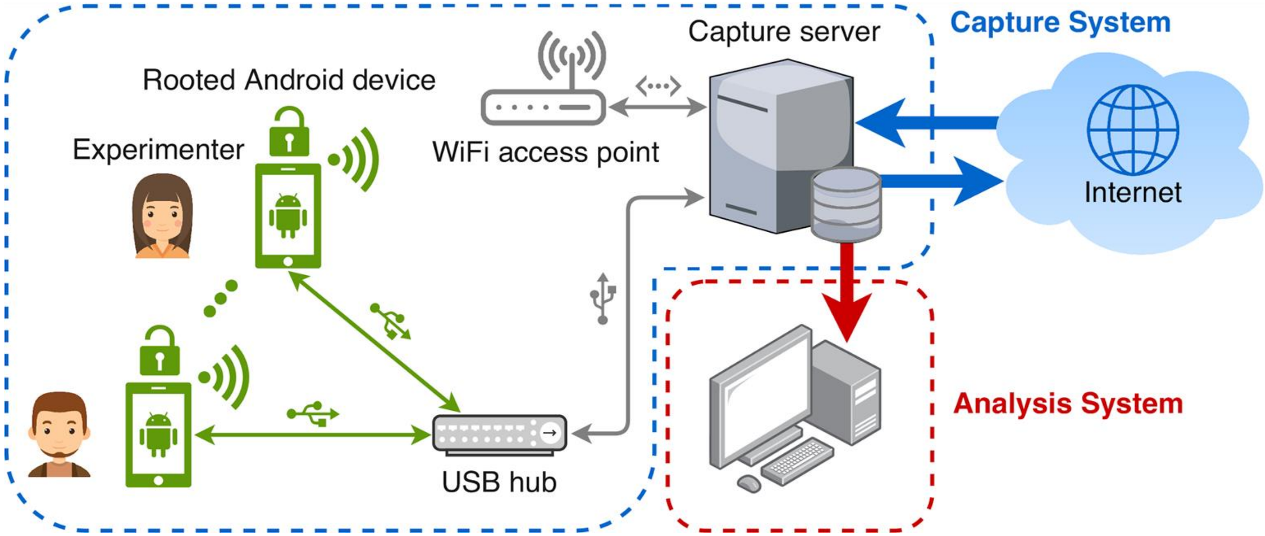
Data-driven TC methodologies require **reliably labeled datasets** to ensure proper design, realization, and validation



MIRAGE

Reproducible architecture for generating **mobile-app traffic** and automatically creating the related high accurate **ground-truth**

MIRAGE Architecture



- Provides **connectivity** to mobile devices
- Collects **network traffic** and system-call **log-files**
- Can handle **multiple devices** at the same time

- Performs the **Ground-Truth** building
- Constructs the final **mobile-app traffic** dataset
- Extracts the **MIRAGE-2019** public version

Benchmarking TC

Human-generated **mobile and encrypted traffic datasets** are used to assess the set of TC methodologies proposed

1 MIRAGE

FB/FBM

2

Android

3

iOS

Mobile-app traffic datasets

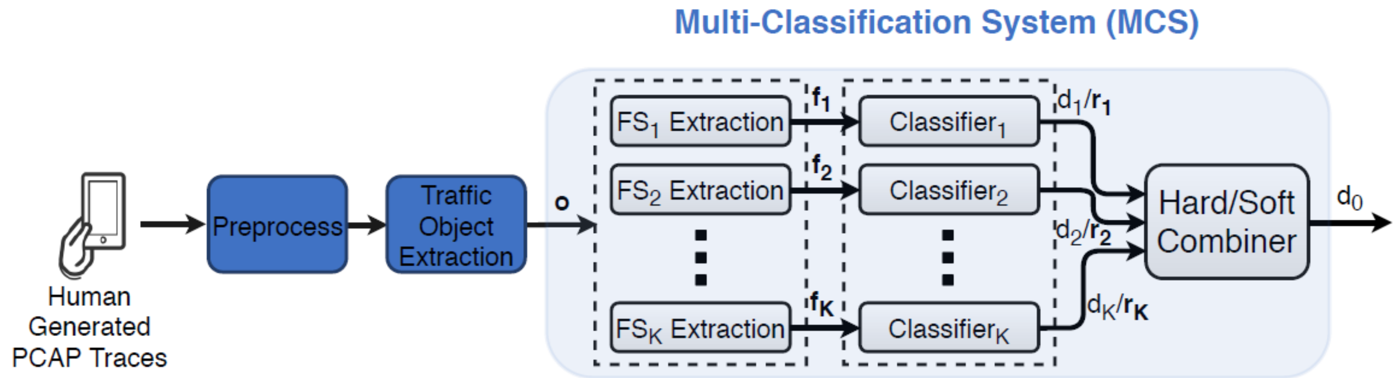
Anon17 public dataset [1]

[1] K. Shahbar and A. N. Zincir-Heywood, "Packet momentum for identification of anonymity networks"

Multi-Classification

Outputs from ML classifiers can be combined to perform **Multi-Classification (MC)** tasks

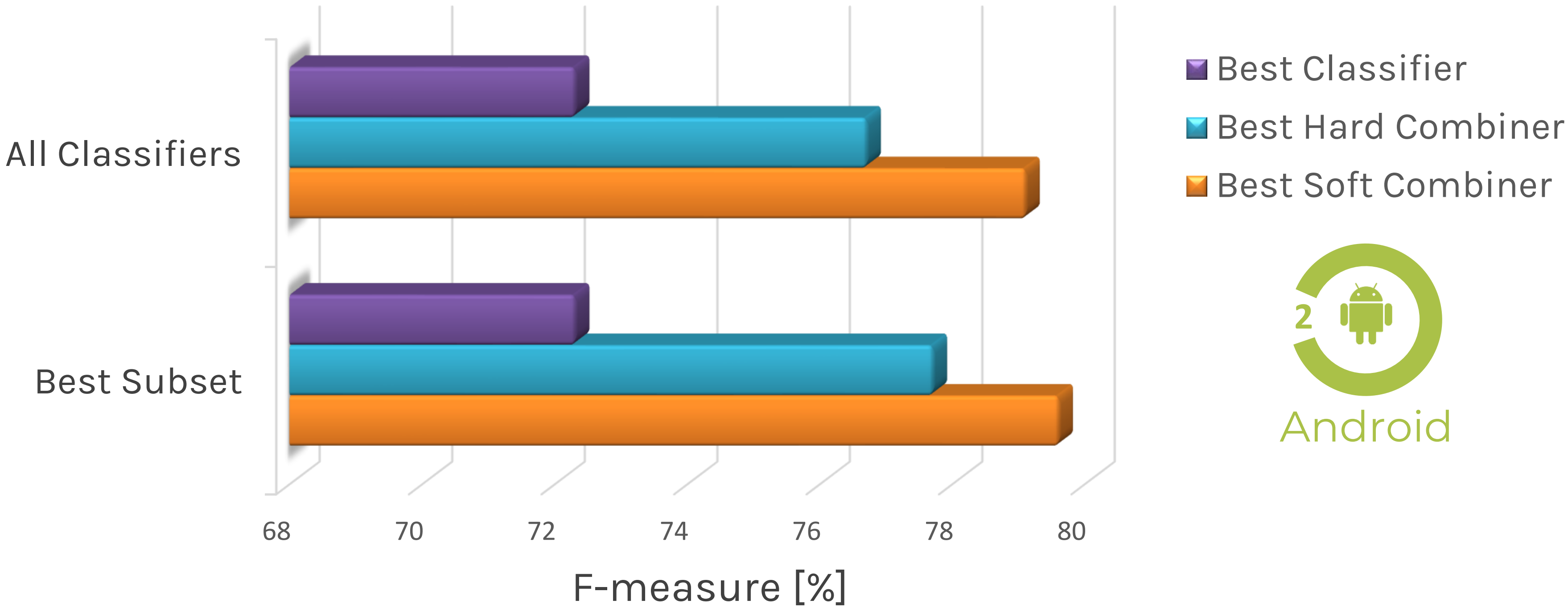
Various **classifier fusion rules** have been proposed in the literature [2, 3] based on both **hard** and **soft** approaches



The aim is **to improve classification performance** on encrypted mobile apps' traffic

[2] A. Dainotti, A. Pescapé, and C. Sansone, "Early classification of network traffic through multi-classification"
[3] L. I. Kuncheva, "Combining pattern classifiers: methods and algorithms"

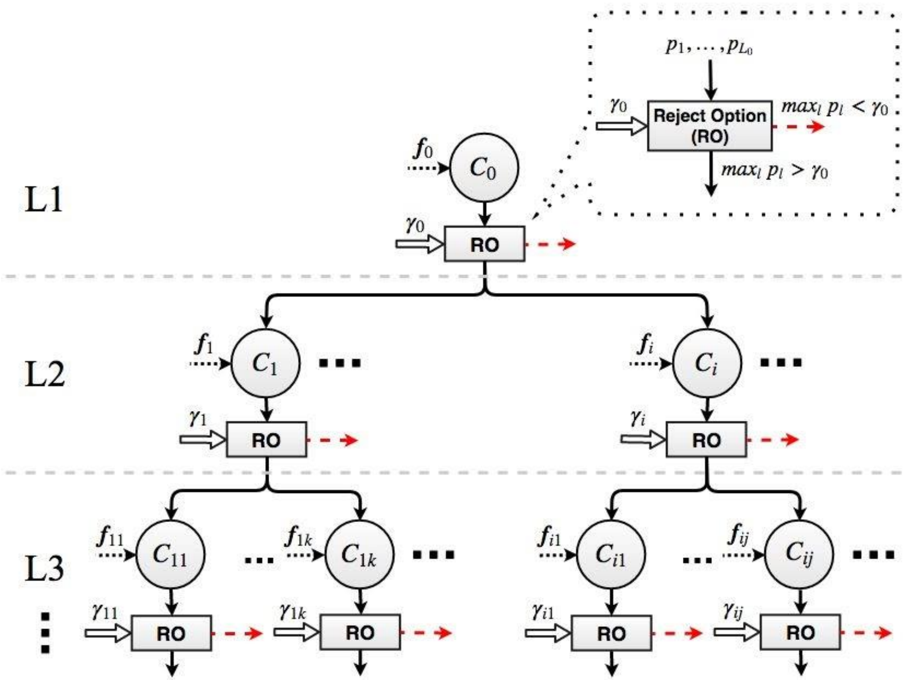
Performance of base classifiers are improved



Android

Careful selection of combination rule and classifiers subset allows to obtain up to **+7.3% F-measure** increment

Hierarchical Classification



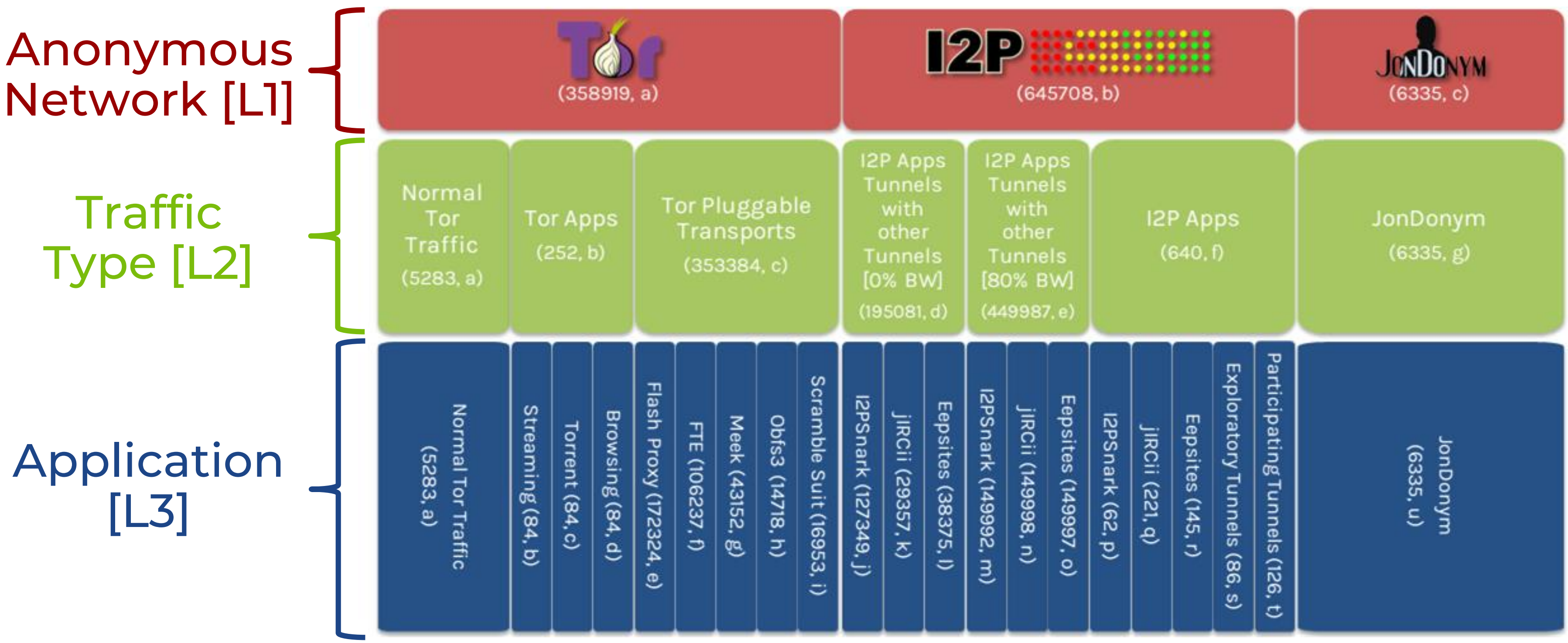
Hierarchical-TC Framework

- **ML-based** classifiers arranged in a **tree** fashion
- **"Divide-et-impera"** approach
  - **Scalability** enhancement
  - **Per-node** tuning and performance
- "Practical" benefits **by design**

Hierarchical Classification (HC) represents a perfect match for encrypted TC at various granularity levels

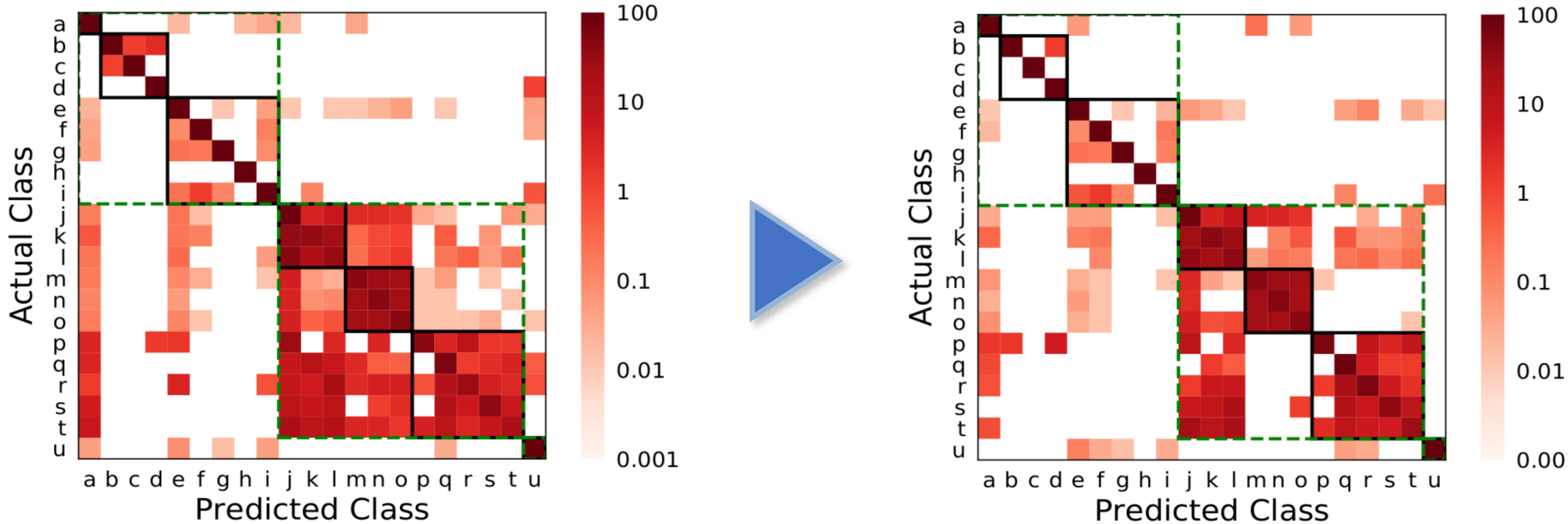
Proposed HC Framework

HC of **Anonymity Tools'** traffic at
three granularity levels according to Anon17



Fine-Grained Performance Improvement

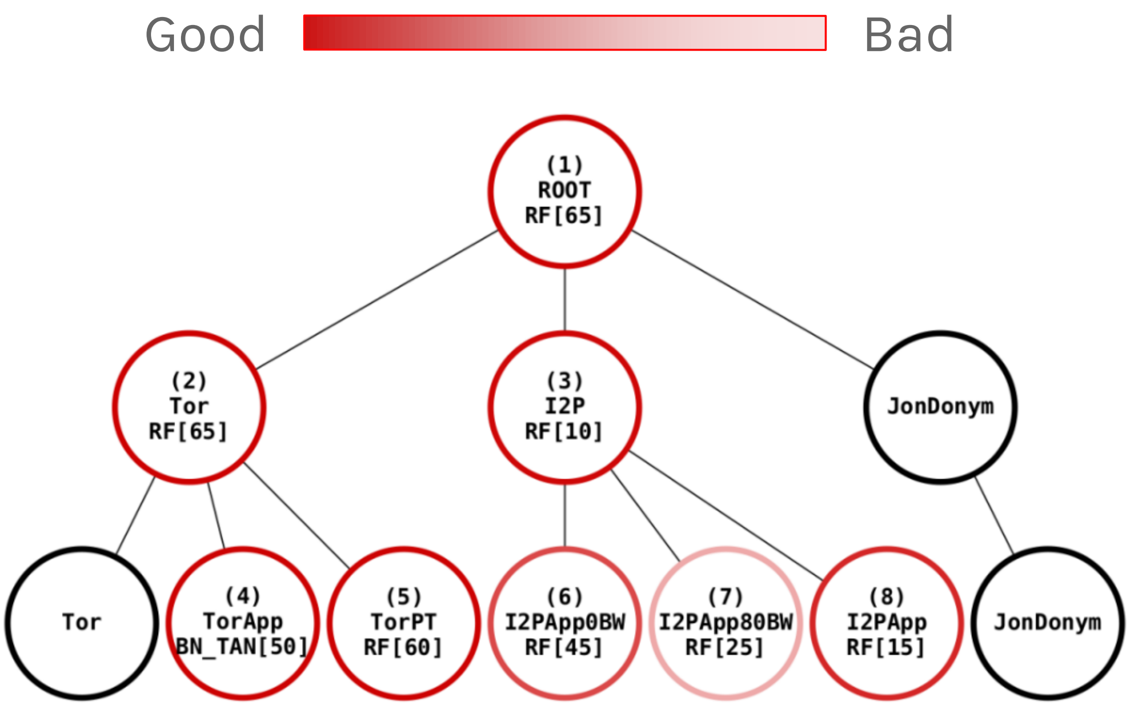
From Flat to Hierarchical Classification



Errors more confined within the
same **Anonymous Network** or **Traffic Type**

Per-node Performance Breakdown

Per-node performance figures allow to accurately evaluate **per-node behaviors**



Random Forest is the best classifier for each node except **Bayesian Networks** for TorApp node

Significant degradation at L3 for **I2PApp80BW**

- Accuracy → 48.94%
- F-measure → 48.90%

Beyond ML-Based TC

Machine Learning (ML) Flow

ML classifiers

rely on domain-expert handcrafted features

- Time-consuming process
- Unsuited to automation
- Rapidly outdated

Difficulty to design accurate and up-to-date mobile and encrypted traffic classifiers

Beyond ML-Based TC

3

Machine Learning (ML) Flow

In → Feature Extractor → Features → ML Classifier → Out

ML classifiers
rely on domain-expert handcrafted features

- Time-consuming process
- Unsuited to automation
- Rapidly outdated

Difficulty to design accurate and up-to-date mobile and encrypted traffic classifiers

Deep Learning (DL) Flow

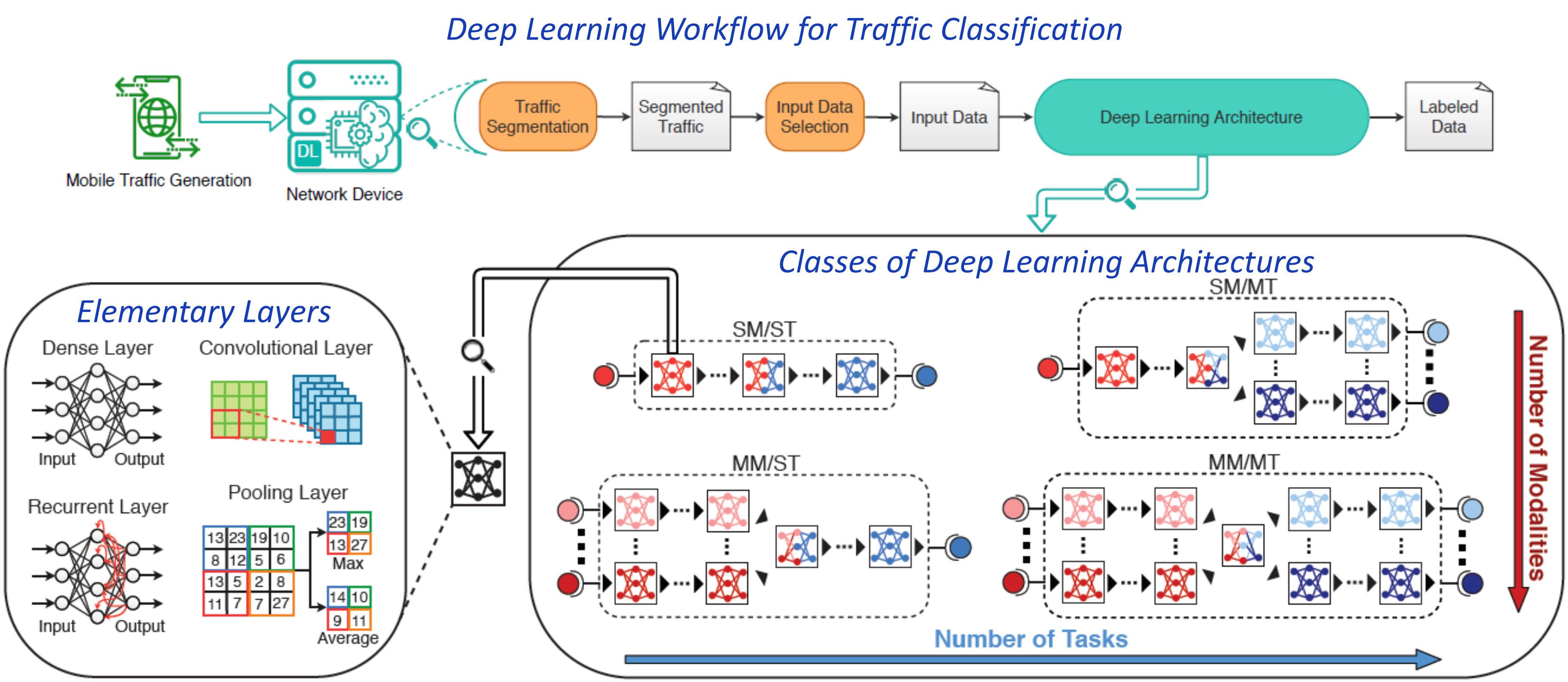
In → DL Classifier → Out

DL classifiers
are trained directly from input data

- Automatic hierarchical feature extraction
- Reduced preprocessing effort

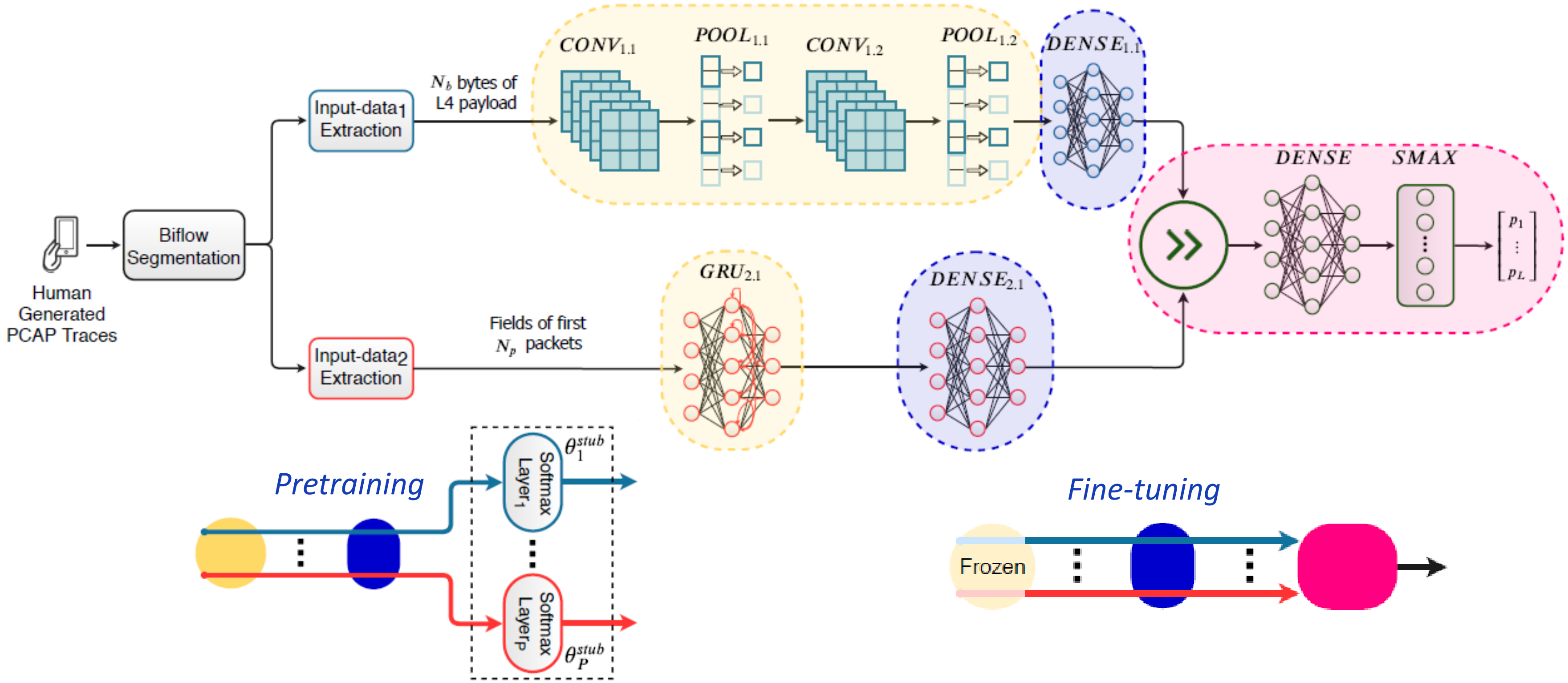
Stepping stone toward high performance in encrypted and mobile TC

DL-Based TC Framework



The MIMETIC Architecture

The proposed framework is employed to design a novel **multi-modal DL-based mobile traffic classification** architecture to exploit the different views of a TC object

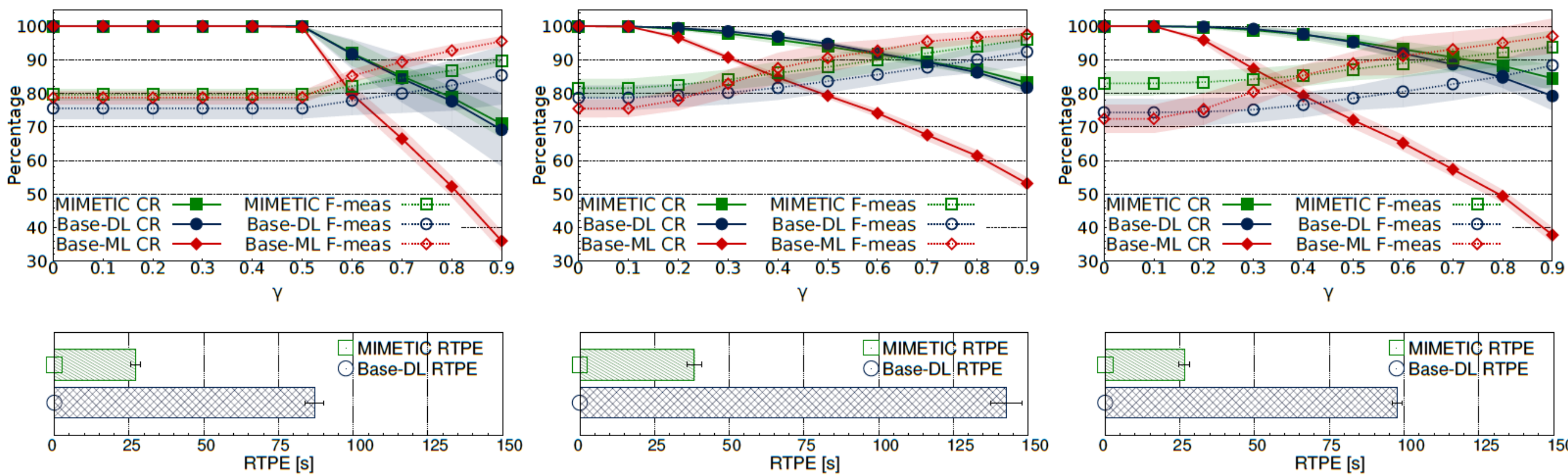


MIMETIC outperforms ML and DL baselines

FB/FBM

Android

iOS



- Up to **+8.6% F-measure** improvement
- Run Time Per Epoch (RTPE) > **3.5× lower**

Big Data-Enabled TC

Training of **DL networks** may result in **completion times** orders-of-magnitude **higher** than those acceptable

Big Data (BD) parallelization **perfectly suits** the repetition of demanding tasks as in DL-based TC

Cloud services provides **practical** and **convenient** tools to address these goals

However…

Big Data-Enabled TC

Training of **DL networks** may result in **completion times** orders-of-magnitude **higher** than those acceptable

Big Data (BD) parallelization **perfectly suits** the repetition of demanding tasks as in DL-based TC

Cloud services provides **practical** and **convenient** tools to address these goals

However…

…DL training is **non-naturally-parallelizable** preventing the transparent application of the BD framework

Big Data-Enabled TC

4

Investigating and experimentally evaluating the adoption of **DL networks** for classifying encrypted mobile traffic via the **BD framework**

Three **intertwined dimensions**

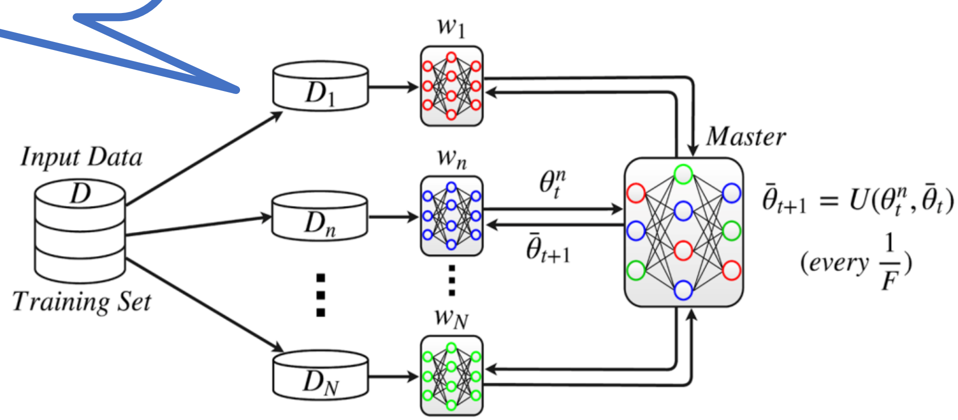
- Classification performance
- Training completion time
- Cloud deployment cost

Cloud-based setup

Training Process on BD

N cooperating workers coordinated by a single central master realizing the data parallelism

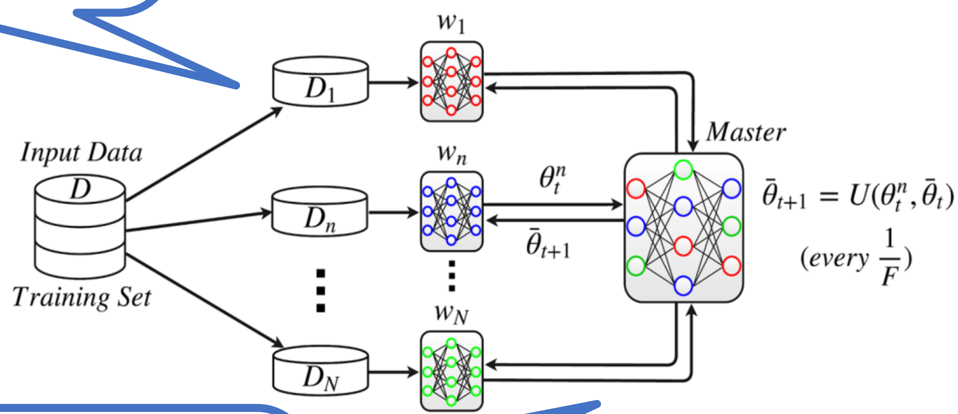
N = {2, 4, 8, 16}



Training Process on BD

N cooperating workers coordinated by a single central master realizing the data parallelism

N = {2, 4, 8, 16}



Communication protocol governing the exchange of commits & pulls between the workers and master

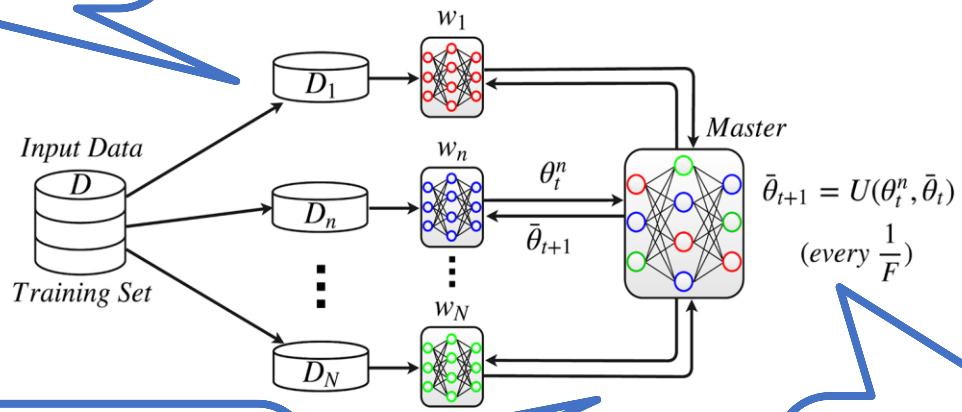
Asynchronous

Training Process on BD

4

N cooperating workers coordinated by a single central master realizing the data parallelism

N = {2, 4, 8, 16}



Communication protocol governing the exchange of commits & pulls between the workers and master

Asynchronous

Update frequency F at which the workers execute a commit

From one per mini-batch to one per worker

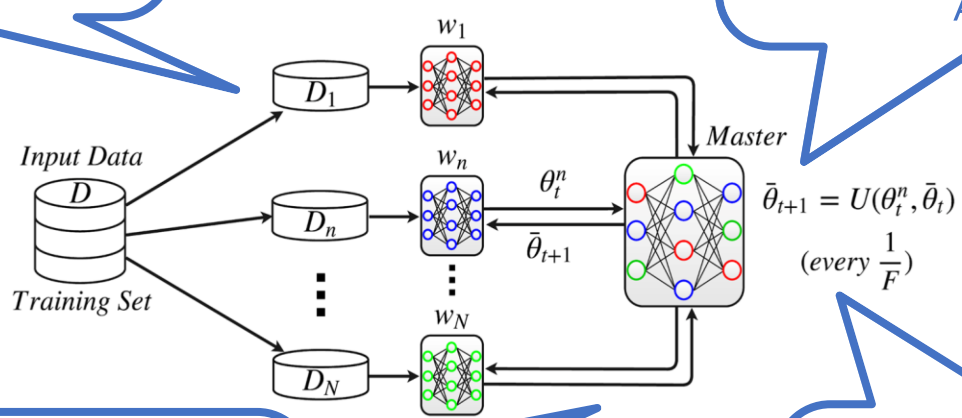
Training Process on BD

N cooperating workers coordinated by a single central master realizing the data parallelism

N = {2, 4, 8, 16}

Federated-optimization algorithm defined by both local workers computation and master update policy

AEASGD



Communication protocol governing the exchange of commits & pulls between the workers and master

Asynchronous

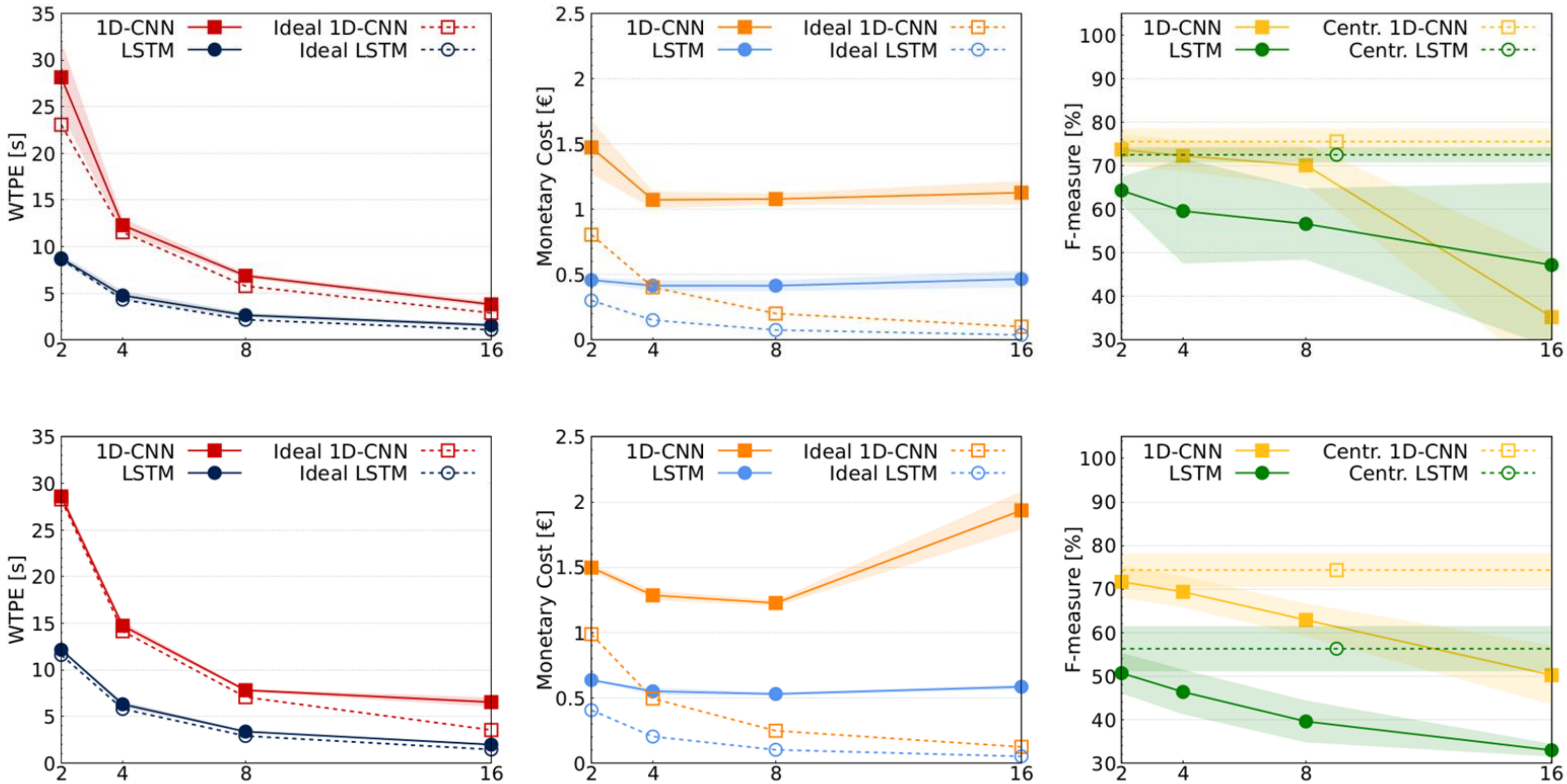
Update frequency F at which the workers execute a commit

From one per mini-batch to one per worker

How does N impact BD-Enabled TC?

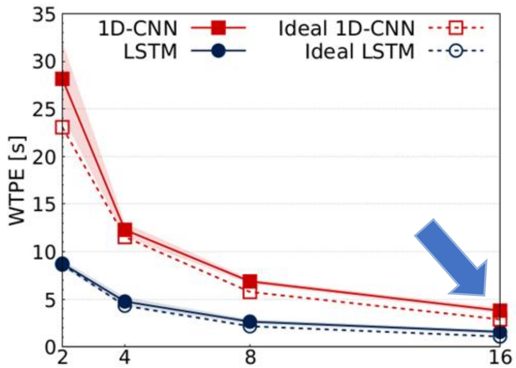
1 FB/FBM

3 iOS

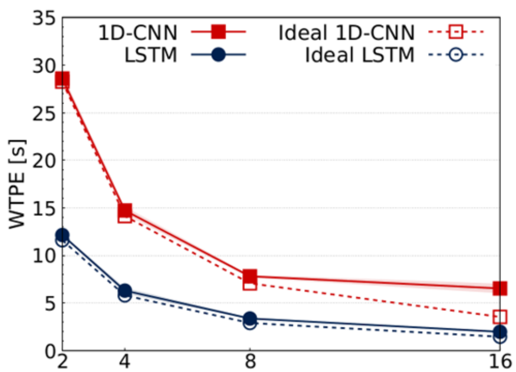


How does N impact BD-Enabled TC?

FB/FBM



iOS



Training time vs. N

- *Decreasing trend* with N for both TC tasks
- 1D-CNN (FB/FBM) shows up to *-91.8% training time* w.r.t. centralized deployment

How does N impact BD-Enabled TC?

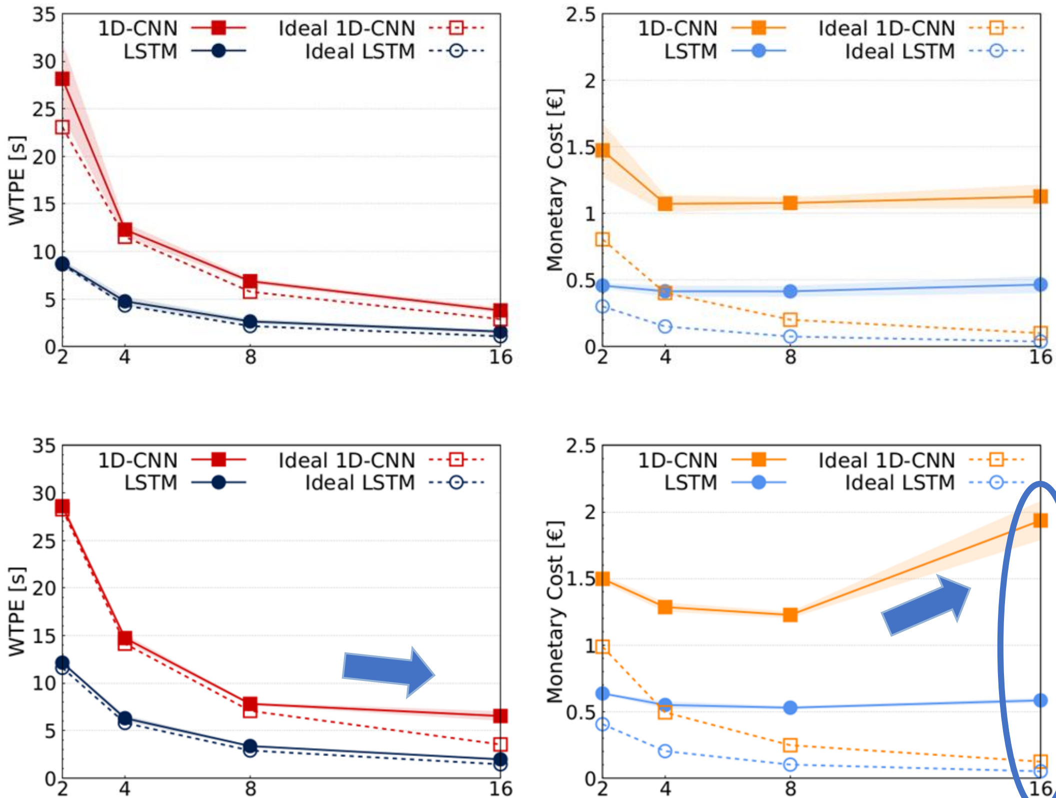
4

1

FB/FBM

3

iOS



Cost vs. N

- Proportional to the *training time*
- 1D-CNN (iOS) shows *+54.2% cost* and only *-2.53% training time* when passing from *N = 8* to *N = 16*

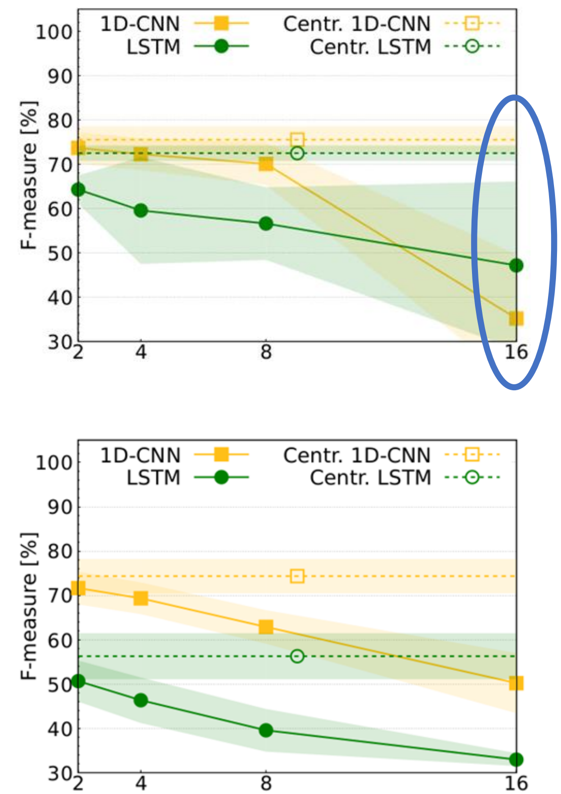
How does N impact BD-Enabled TC?

FB/FBM

iOS

F-measure vs. N

- *Decreasing F-measure* with N for both TC tasks
- 1D-CNN (FB/FBM) shows *-53.4% F-measure* when *N = 16* w.r.t. centralized deployment

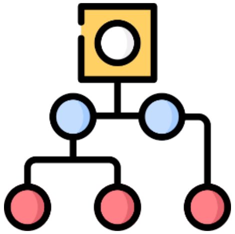


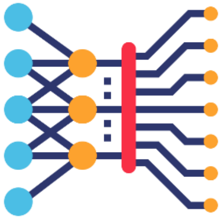
Conclusions

Mobile and encrypted TC plays a role of paramount importance in network management

The effectiveness of traditional methods **is hampered** in these dynamic and challenging scenarios

Novel methodologies for encrypted and mobile TC based on advanced ML and DL approaches are proposed, implemented, and evaluated





Products (1/4)

[J1] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, "MIMETIC: Mobile Encrypted Traffic Classification using Multimodal Deep Learning," Elsevier Computer Networks, Volume 165, 24 December 2019.

[J2] **Antonio Montieri**, Domenico Ciuonzo, Giampaolo Bovenzi, Valerio Persico, Antonio Pescapè, "A Dive into the Dark Web: Hierarchical Traffic Classification of Anonymity Tools," IEEE Transactions on Network Science and Engineering, Early Access, 2019.

[J3] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, "Mobile Encrypted Traffic Classification Using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges," IEEE Transactions on Network and Service Management, Volume 16 , Issue 2 , June 2019.

Products (2/4)

[J4] **Antonio Montieri**, Domenico Ciuonzo, Giuseppe Aceto, Antonio Pescapè, “Anonymity Services Tor, I2P, JonDonym: Classifying in the Dark (Web),” IEEE Transactions on Dependable and Secure Computing, Early Access, 2018.

[J5] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, “Multi-Classification Approaches for Classifying Mobile App Traffic,” Elsevier Journal of Network and Computer Applications Volume 103, 1 February 2018, Pages 131-145.

[J6] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, “Toward Effective Mobile Encrypted Traffic Classification through Deep Learning,” Elsevier Future Generation Computer Systems, under the first round of review, 2020.

Products (3/4)

[C1] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Valerio Persico, Antonio Pescapè, “MIRAGE: Mobile-app Traffic Capture and Ground-truth Creation,” 4th IEEE International Conference on Computing, Communications and Security (ICCCS 2019), October 10-12, 2019, Rome, Italy.
Best Paper Award ICCCS 2019.

[C2] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Valerio Persico, Antonio Pescapè, “Know your Big Data Trade-offs when Classifying Encrypted Mobile Traffic with Deep Learning,” 3rd Network Traffic Measurement and Analysis Conference (TMA 2019), June 17-21, 2019, Paris, France.

[C3] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, “Mobile Encrypted Traffic Classification Using Deep Learning,” 2nd Network Traffic Measurement and Analysis Conference (TMA 2018), June 26-29, 2018, Vienna, Austria.

Products (4/4)

[C4] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, “Traffic Classification of Mobile Apps through Multi-classification,” 2017 IEEE Global Communications Conference (IEEE GLOBECOM 2017); Communication QoS, Reliability and Modeling (CQRM) Symposium, December 4-8, 2017, Singapore.

[C5] **Antonio Montieri**, Giuseppe Aceto, Domenico Ciuonzo, Antonio Pescapè, “Anonymity Services Tor, I2P, JonDonym: Classifying in the Dark,” 29th International Teletraffic Congress (ITC 29), September 4-8, 2017, Genova, Italy.

[C6] Giuseppe Aceto, **Antonio Montieri**, Antonio Pescapè, “Internet Censorship in Italy: an Analysis of 3G/4G Networks,” 2017 IEEE International Conference on Communications (ICC 2017); Communication QoS, Reliability and Modeling (CQRM) Symposium, May 21-25, 2017, Paris, France.

Thank you!

Questions?

antonio.montieri@unina.it

wpage.unina.it/antonio.montieri

traffic.comics.unina.it

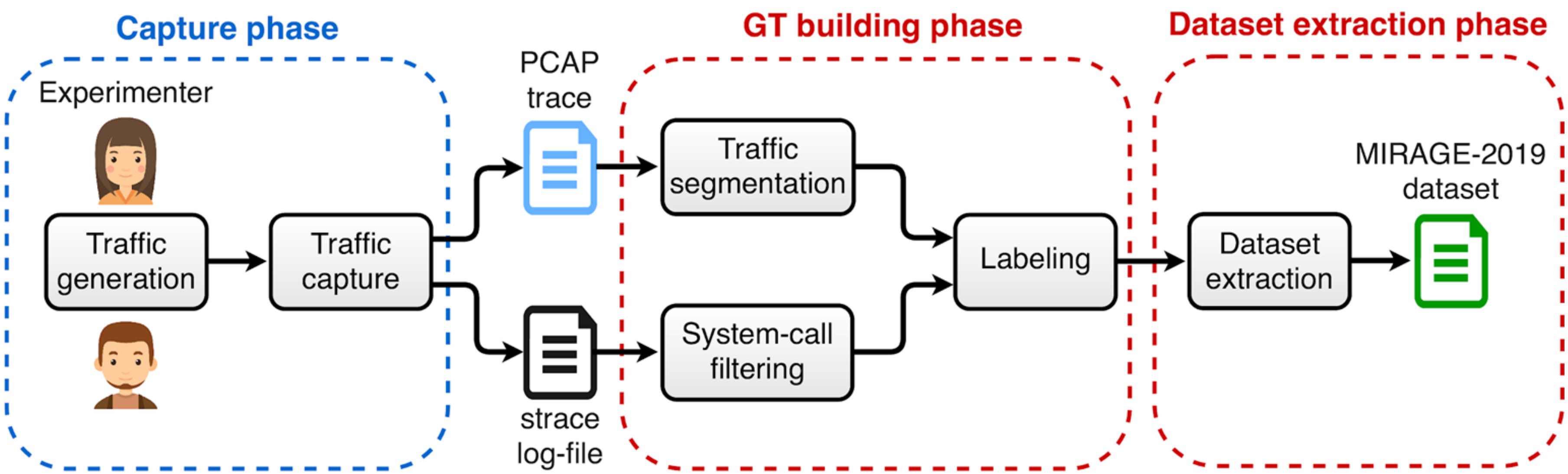
Extra Slides

Paper	Year	Traffic		Technique					Dataset	
		MT	ET	MC	HC	DL	MM	BD	HD	OD
Szabo *et al.* [86]	2007	□	■	♦	◊	◊	◊	◊	●	○
He *et al.* [87]	2008	□	■	♦	◊	◊	◊	◊	●	○
Callado *et al.* [88]	2010	□	■	♦	◊	◊	◊	◊	●	○
Yu *et al.* [89]	2010	□	■	◊	♦	◊	◊	◊	●	○
Dainotti *et al.* [90]	2011	□	■	♦	◊	◊	◊	◊	●	○
Mellia *et al.* [91]	2012	□	■	◊	♦	◊	◊	◊	●	○
Dai *et al.* [72]	2013	■	□	◊	◊	◊	◊	◊	○	○
Stöber *et al.* [41]	2013	■	■	◊	◊	◊	◊	◊	●	○
De Donato *et al.* [40]	2014	□	■	♦	◊	◊	◊	◊	●	○
D'Alessandro *et al.* [92]	2015	□	■	◊	◊	◊	◊	♦	○	◐
Wang *et al.* [74]	2015	■	■	◊	◊	◊	◊	◊	◐	○
Wang [78]	2015	□	□	◊	◊	♦	◊	◊	●	○
Yoon *et al.* [79, 80]	2015	□	□	♦	♦	◊	◊	◊	◐	○
Alan and Kaur [75]	2016	■	■	◊	◊	◊	◊	◊	○	○
Shbair *et al.* [93]	2016	□	■	◊	♦	◊	◊	◊	●	○
Conti *et al.* [76]	2016	■	■	◊	◊	◊	◊	◊	○	○
Saltaformaggio *et al.* [60]	2016	■	■	◊	◊	◊	◊	◊	●	○
Yuan and Wang [94]	2016	□	■	◊	◊	◊	◊	♦	●	◐
Dong *et al.* [95]	2017	□	■	◊	♦	◊	◊	◊	●	○
Li *et al.* [73]	2017	■	□	◊	◊	♦	◊	◊	●	○
Chen *et al.* [83]	2017	□	■	◊	◊	♦	◊	◊	◐	○
Lopez-Martin *et al.* [96]	2017	□	■	◊	◊	♦	◊	◊	●	○
Lotfollahi *et al.* [97]	2017	□	■	◊	◊	♦	◊	◊	●	◐
Vu *et al.* [98]	2017	□	■	◊	◊	♦	◊	◊	●	◐
Wang *et al.* [21]	2017	□	■	◊	◊	♦	◊	◊	●	◐
Wang *et al.* [99]	2017	□	■	◊	◊	♦	◊	◊	●	●
Ke *et al.* [100]	2017	□	■	◊	◊	◊	◊	♦	●	◐
Taylor *et al.* [42, 43]	2018	■	■	◊	◊	◊	◊	◊	○	○
Chen *et al.* [81]	2018	□	■	◊	♦	♦	◊	◊	●	◐
Le *et al.* [77]	2018	■	■	◊	◊	◊	◊	♦	●	○
Huang *et al.* [23]	2018	□	■	◊	◊	♦	◊	◊	●	◐
Shi *et al.* [101]	2018	□	■	◊	◊	♦	◊	◊	●	◐
Zhang *et al.* [102]	2018	□	■	◊	◊	♦	◊	◊	●	◐
Wang *et al.* [103]	2018	□	■	◊	◊	♦	◊	◊	●	◐
Li *et al.* [104]	2018	□	■	◊	◊	♦	◊	◊	●	○
Liu *et al.* [105]	2019	□	■	◊	◊	♦	◊	◊	●	○
Sun *et al.* [106]	2019	□	■	◊	◊	♦	◊	◊	●	◐
Zeng *et al.* [107]	2019	□	■	◊	◊	♦	◊	◊	●	◐
This thesis	*2020*	■	■	♦	♦	♦	♦	♦	●	●

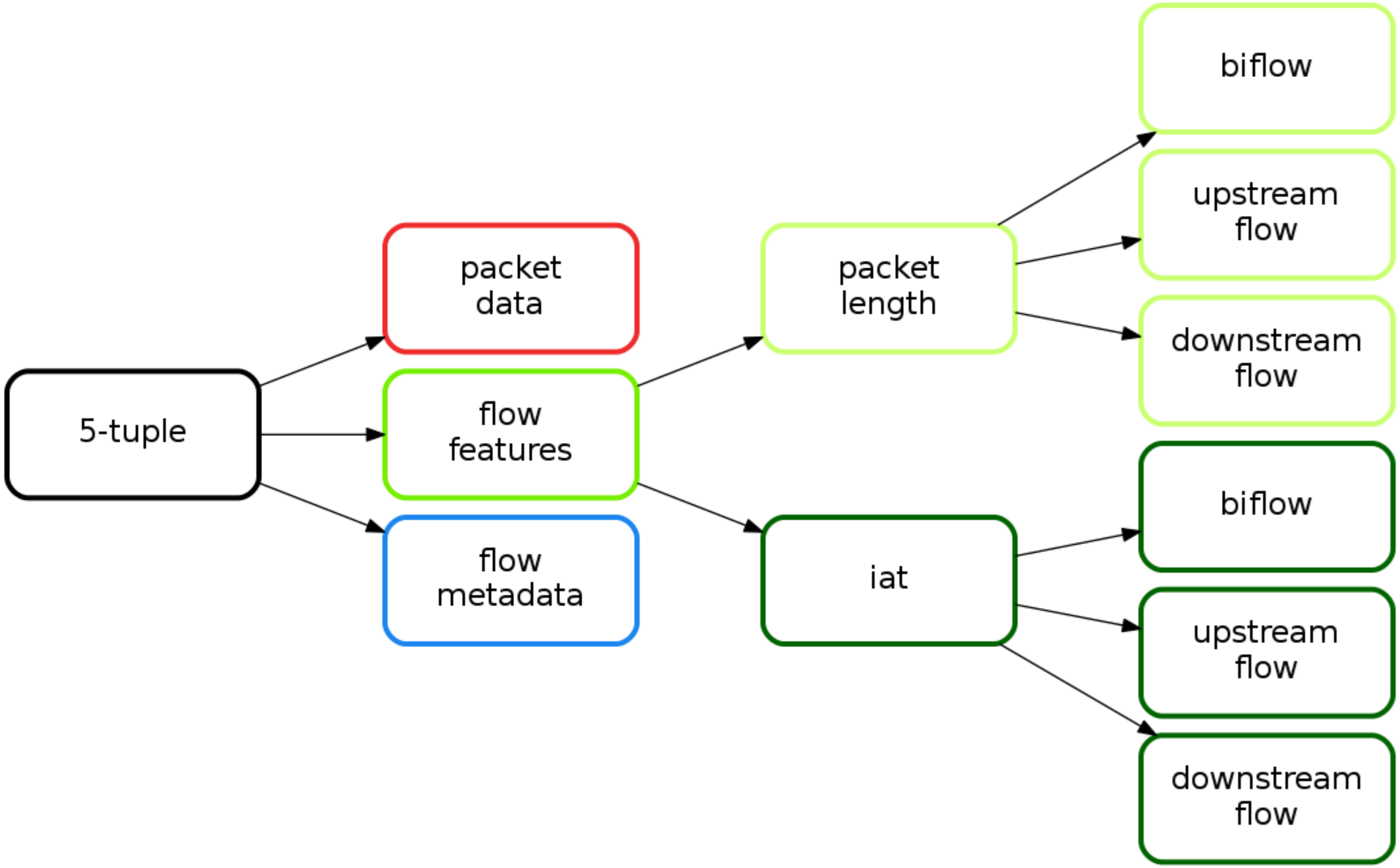
Mobile and encrypted TC works and comparison with the set of proposed methodologies

- Traffic
  - Mobile Traffic (MT)
  - Encrypted Traffic (ET)
- Technique
  - Multi-Classification (MC)
  - Hierarchical Classification (HC)
  - Deep Learning (DL)
  - Multi-modal (MM)
  - Big Data (BD)
- Dataset
  - Human Dataset (HD)
  - Open Dataset (OD)

MIRAGE-2019 Dataset Building



MIRAGE-2019 Dataset Structure



Hard Combiners

<table>
<tr><th>Label</th><th>Technique</th><th>Category</th><th>Training</th></tr>
<tr><td>MV</td><td>Majority Voting</td><td rowspan="2">Vote (Bayesian)</td><td>None (Confusion Matrix)</td></tr>
<tr><td>WMV</td><td>Weighted Majority Voting</td><td>Confidence vector</td></tr>
<tr><td>NB</td><td>Naïve Bayes</td><td>Bayesian</td><td>Confusion Matrix</td></tr>
<tr><td>BKS</td><td>Behavior Knowledge Space</td><td rowspan="2">Behavior Knowledge Space (Bayesian)</td><td rowspan="2">BKS & Confusion Matrix</td></tr>
<tr><td>WER</td><td>Wernecke</td></tr>
<tr><td>ORA</td><td>Oracle</td><td>Oracle</td><td>N/A</td></tr>
</table>

- **Simple combiners** $\longrightarrow$ No need for training
- **Trainable combiners** $\longrightarrow$ Some free/tunable parameters
  - 2nd level training or validation phase required
  - Random training-validation-test set splitting → 50% - 25% - 25%

Soft Combiners

Technique	Category
Mean	Class Conscious Non-trainable
Maximum	
Minimum	
Median	
Trimmed Mean	
Harmonic Mean	
Geometric Mean	
Generalized Mean	
Probabilistic Product	

Technique	Category
Fuzzy Integral	Class Conscious Trainable
K Weights	
KL Weights	
DT-SE	Class Indifferent
DT-L1	
DT-FSD	
Dempster-Shafer	
Oracle	Oracle

HC Methodology (1/2)

Traffic in Anon17 is split into **directional flows**

Four sets of classification features can be extracted

1. 74 **statistics**
   - Flow direction & duration
   - Packet Length (PL) statistics
   - Inter-Arrival Time (IAT) statistics
   - TCP and IP header-related features
   - Number of connections

   Histograms of

2. PL
3. PL & IAT
4. <PL, IAT> of the **first K packets**

HC Methodology (2/2)

Offline TC → 1 2 3

Early-based TC → 4

Five ML-based classification algorithms

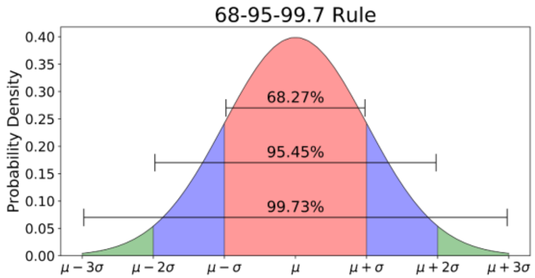
- 1 4 Naïve Bayes **(NB_SD)**
- 2 3 Multinomial Naïve Bayes **(MNB)**
- 1 4 Bayesian Networks **(BN_TAN)**
- 1 4 **C4.5**
- 1 4 Random Forest **(RF)**

Bayesian Approach

Decision Tree

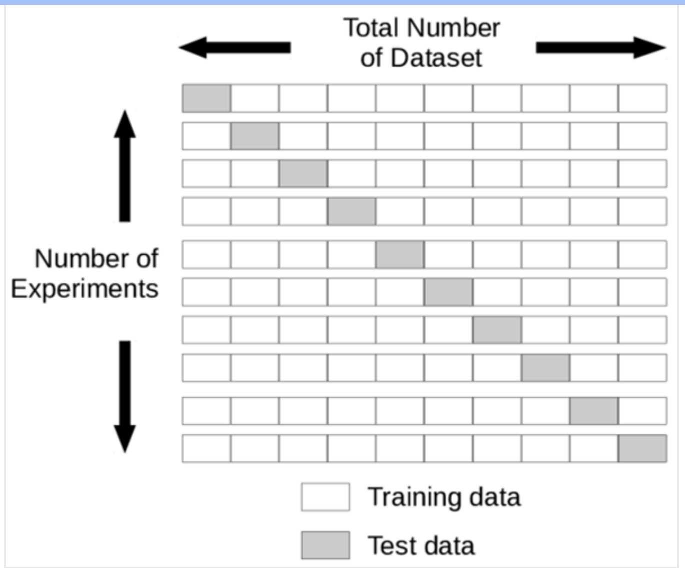
Performance Evaluation

- Accuracy
- **Per-class** measures
  - F-measure & G-mean (macro)
- **Fine-grained** performance
  - Confusion matrix
- Performance measures as **$\mu \pm 3\sigma$**



99.73% under
Gaussian assumption

(Stratified) 10-fold validation



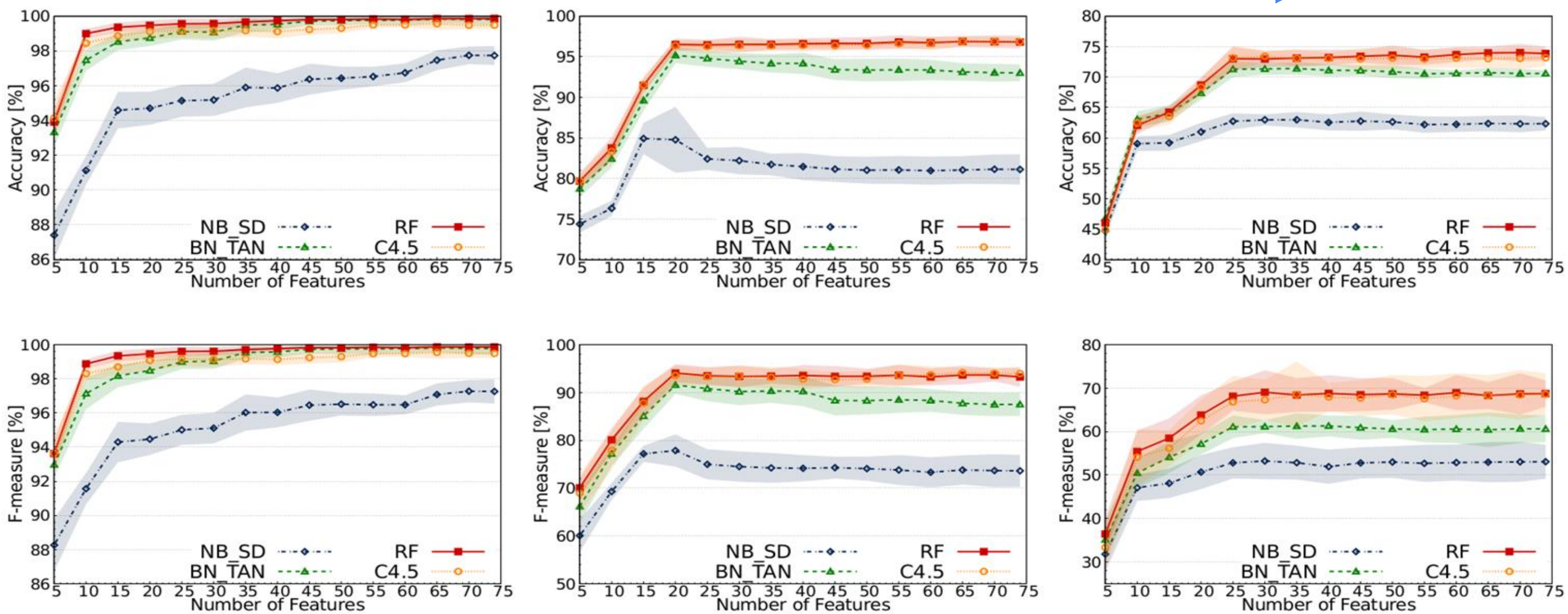
Depth of Accurate TC

Feature set 1

TC-task Complexity → Performance Drop

Accuracy

F-measure

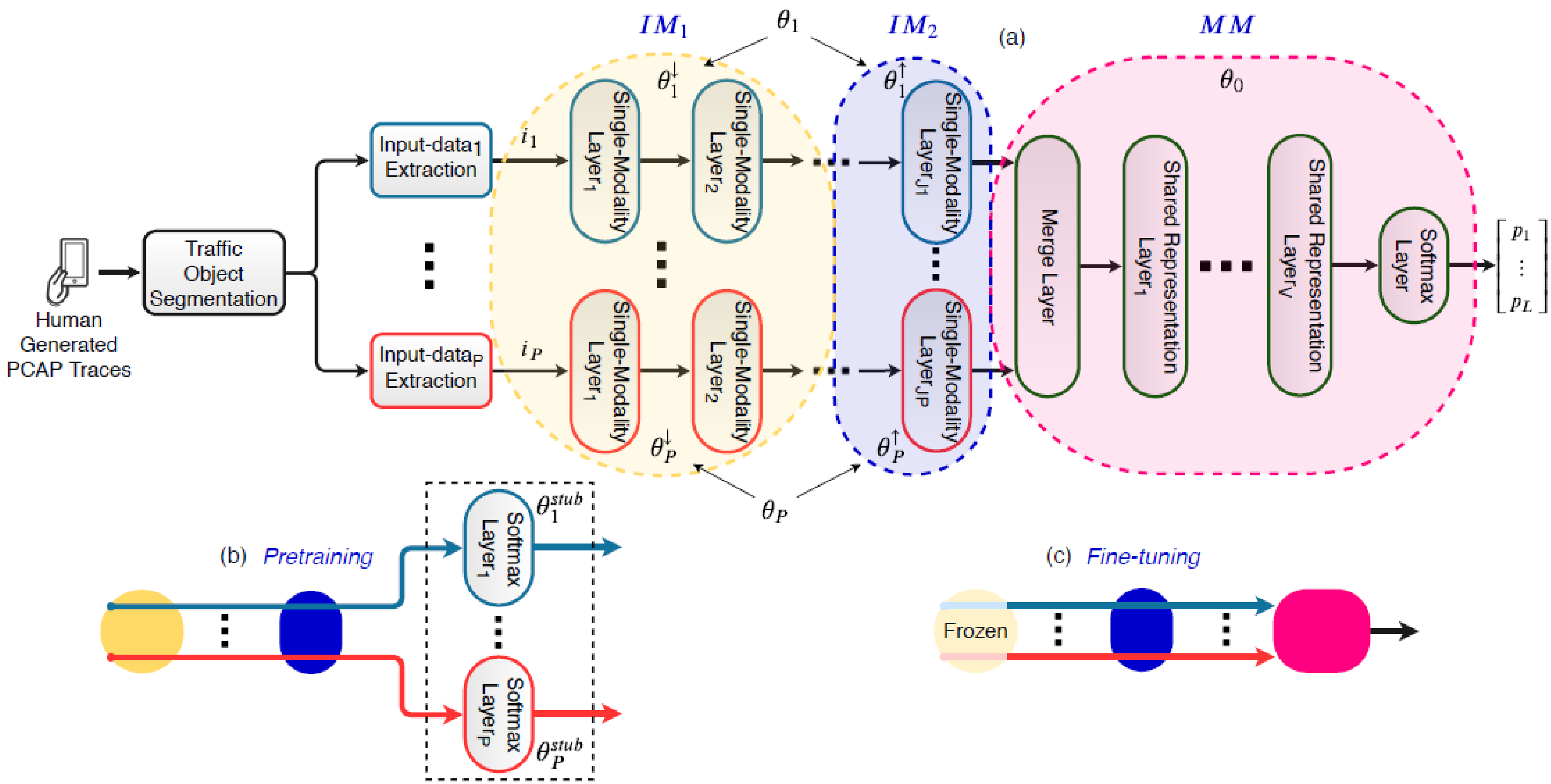


Anonymous Network [L1]

Traffic Type [L2]

Application [L3]

MIMETIC General Framework



MIMETIC Performance

FB/FBM

Table 5.6: Accuracy, F-measure, and G-mean [%] comparison of MIMETIC with the four groups of baselines: (I) best single-modal DL classifiers, (II) shallow neural networks, (III) state-of-the-art ML-based mobile-traffic classifier, (IV) classifier fusion techniques. Results refer to the FB/FBM dataset and are in the format *avg.* *(± std.)* obtained over 10-folds. The last group reports the *Maximum Improvement Over Best - Classifier* (MIOB-C) and the *Maximum Improvement Over Best - Fusion Technique* (MIOB-FT) [%] of MIMETIC architecture. Highlighted values: **overall best classifier**, best baseline classifier (♦), and best baseline fusion technique (‡) for each dataset and performance measure.

	Architecture	*Accuracy*	*F-measure*	*G-mean*
	MIMETIC	**79.98 (± 0.49)**	**79.63 (± 0.51)**	**79.53 (± 0.60)**
I	1D-CNN [99] (L7-784)	76.37 (± 0.73)	75.56 (± 1.01)	74.79 (± 1.76)
I	HYBRID [96] (MAT-20)	74.26 (± 0.98)	73.23 (± 0.95)	72.18 (± 1.05)
II	MLP-1 (L7-784)	74.46 (± 0.88)	73.89 (± 0.86)	73.55 (± 0.89)
II	MLP-1 (MAT-20)	68.93 (± 1.32)	67.86 (± 0.94)	66.98 (± 0.75)
III	Tay_RF [42] (flow-based)	79.56 (± 0.62) ♦	78.73 (± 0.62) ♦	78.37 (± 0.76) ♦
IV	MV	75.13 (± 0.92)	74.48 (± 1.14)	74.02 (± 1.65)
IV	SOA	78.86 (± 0.79) ‡	78.37 (± 1.00) ‡	78.06 (± 1.61) ‡
IV	TLF	74.61 (± 1.57)	73.60 (± 1.80)	72.59 (± 2.14)
	MIOB-C	+ 0.42 (± 0.65)	+ 0.90 (± 0.68)	+ 1.16 (± 0.99)
	MIOB-FT	+ 1.12 (± 0.89)	+ 1.26 (± 1.14)	+ 1.47 (± 1.84)

"HYBRID" refers to an hybrid DL architecture combining 2D convolutional and LSTM layers (viz. LSTM + 2D-CNN) proposed in [96].

MIMETIC Performance

2 Android

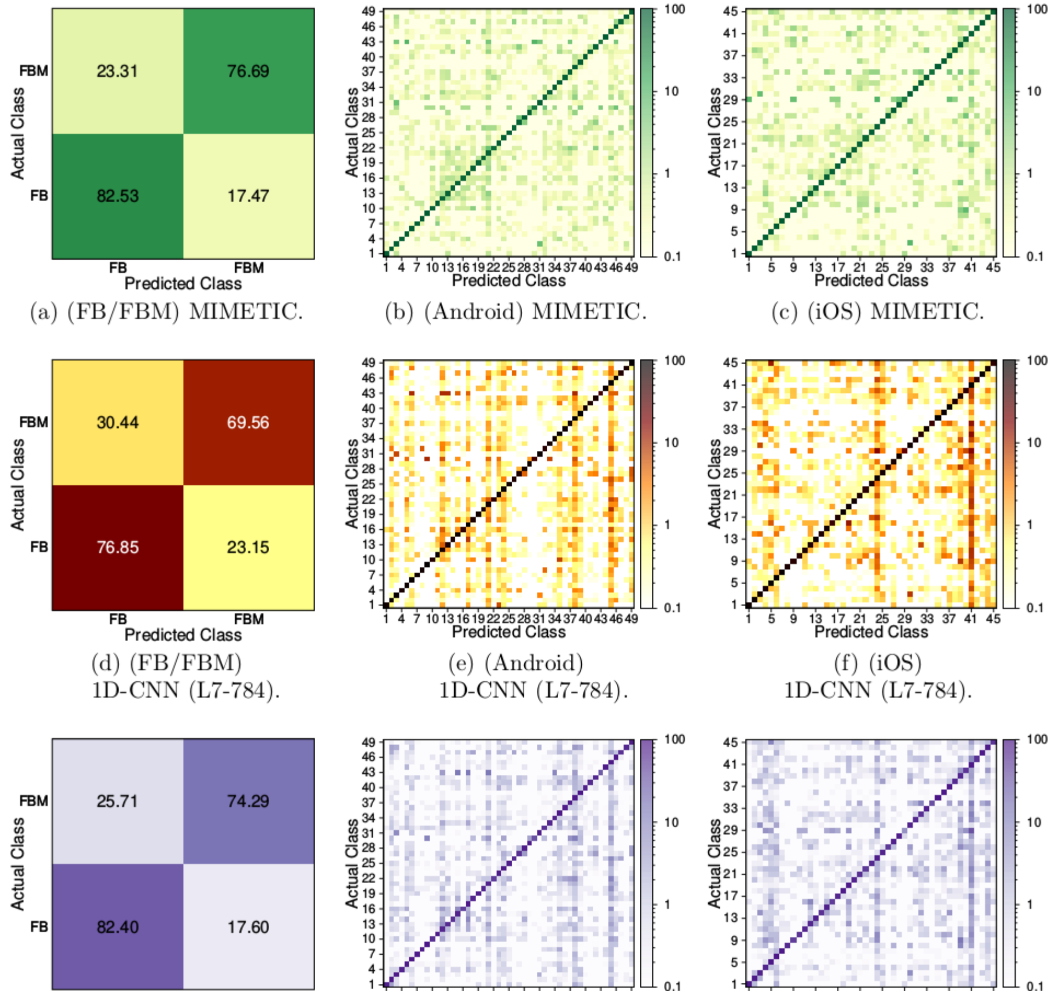
3 iOS

Table 5.7: Accuracy, F-measure, and G-mean [%] comparison of MIMETIC with the four groups of baselines: (I) best single-modal DL classifiers, (II) shallow neural networks, (III) state-of-the-art ML-based mobile-traffic classifier, (IV) classifier fusion techniques. Results refer to the the multi-class datasets are in the format *avg. (± std.)* obtained over 10-folds. The last group reports the *Maximum Improvement Over Best - Classifier* (MIOB-C) and the *Maximum Improvement Over Best - Fusion Technique* (MIOB-FT) [%] of MIMETIC architecture. Highlighted values: **overall best classifier**, best baseline classifier (♦), and best baseline fusion technique (‡) for each dataset and performance measure.

	Architecture	*Android*			*iOS*		
		Accuracy	*F-measure*	*G-Mean*	*Accuracy*	*F-measure*	*G-Mean*
	MIMETIC	**89.49 (± 0.32)**	**81.51 (± 0.93)**	**91.96 (± 0.95)**	**89.14 (± 0.82)**	**82.99 (± 1.14)**	**92.25 (± 0.84)**
I	1D-CNN [99] (L7-784)	85.70 (± 0.45) ♦	78.68 (± 1.20) ♦	86.82 (± 0.87) ♦	82.64 (± 1.63) ♦	74.34 (± 1.29) ♦	84.00 (± 1.31) ♦
I	HYBRID [96] (MAT-20)	77.95 (± 0.41)	64.52 (± 1.17)	76.35 (± 1.45)	69.17 (± 0.64)	58.75 (± 0.76)	72.17 (± 0.75)
II	MLP-1 (L7-784)	78.71 (± 0.65)	69.79 (± 1.17)	81.52 (± 1.38)	77.16 (± 0.63)	67.61 (± 1.07)	80.11 (± 0.99)
II	MLP-1 (MAT-20)	64.94 (± 0.47)	48.26 (± 0.96)	63.10 (± 1.07)	54.42 (± 0.63)	40.86 (± 1.04)	57.56 (± 1.03)
III	Tay_RF [42] (flow-based)	84.78 (± 0.30)	75.49 (± 0.89)	83.86 (± 0.58)	80.77 (± 0.84)	72.39 (± 1.39)	81.88 (± 1.27)
IV	MV	80.41 (± 0.40)	71.28 (± 0.85)	81.74 (± 0.77)	77.24 (± 0.62)	66.49 (± 0.97)	78.92 (± 0.97)
IV	SOA	87.08 (± 0.29) ‡	80.07 (± 0.81) ‡	87.00 (± 0.80) ‡	84.68 (± 0.55) ‡	75.94 (± 1.10) ‡	84.15 (± 0.96) ‡
IV	TLF	68.87 (± 1.05)	48.82 (± 1.92)	62.55 (± 1.86)	62.01 (± 0.97)	39.07 (± 1.52)	54.07 (± 1.94)
	MIOB-C	+ 3.79 (± 0.59)	+ 2.83 (± 1.66)	+ 5.14 (± 1.06)	+ 6.50 (± 2.12)	+ 8.66 (± 1.77)	+ 8.25 (± 1.72)
	MIOB-FT	+ 2.40 (± 0.48)	+ 1.44 (± 1.56)	+ 4.96 (± 1.46)	+ 4.46 (± 1.01)	+ 7.05 (± 1.43)	+ 8.10 (± 1.27)

"HYBRID" refers to an hybrid DL architecture combining 2D convolutional and LSTM layers (viz. LSTM + 2D-CNN) proposed in [96].

MIMETIC Performance



(a) (FB/FBM) MIMETIC. (b) (Android) MIMETIC. (c) (iOS) MIMETIC.

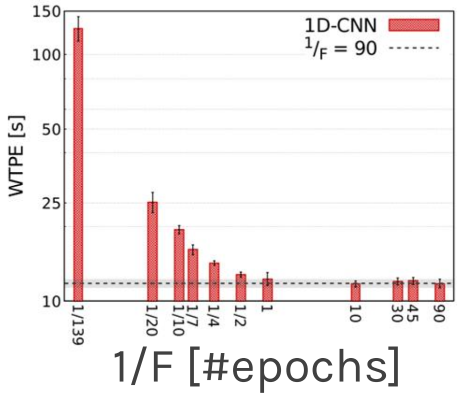
(d) (FB/FBM) 1D-CNN (L7-784). (e) (Android) 1D-CNN (L7-784). (f) (iOS) 1D-CNN (L7-784).

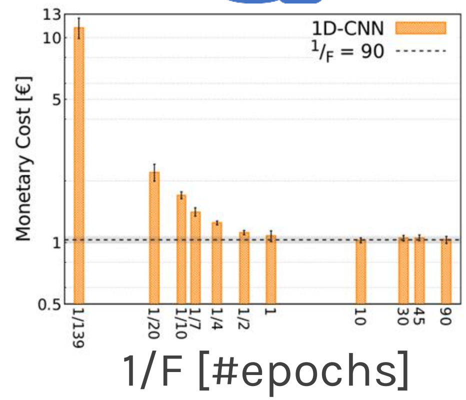
(g) (FB/FBM) SOA. (h) (Android) SOA. (i) (iOS) SOA.

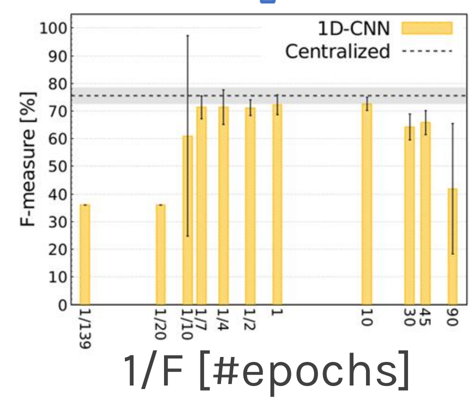
How does F impact BD-Enabled TC?

1D-CNN
N = 4

FB/FBM



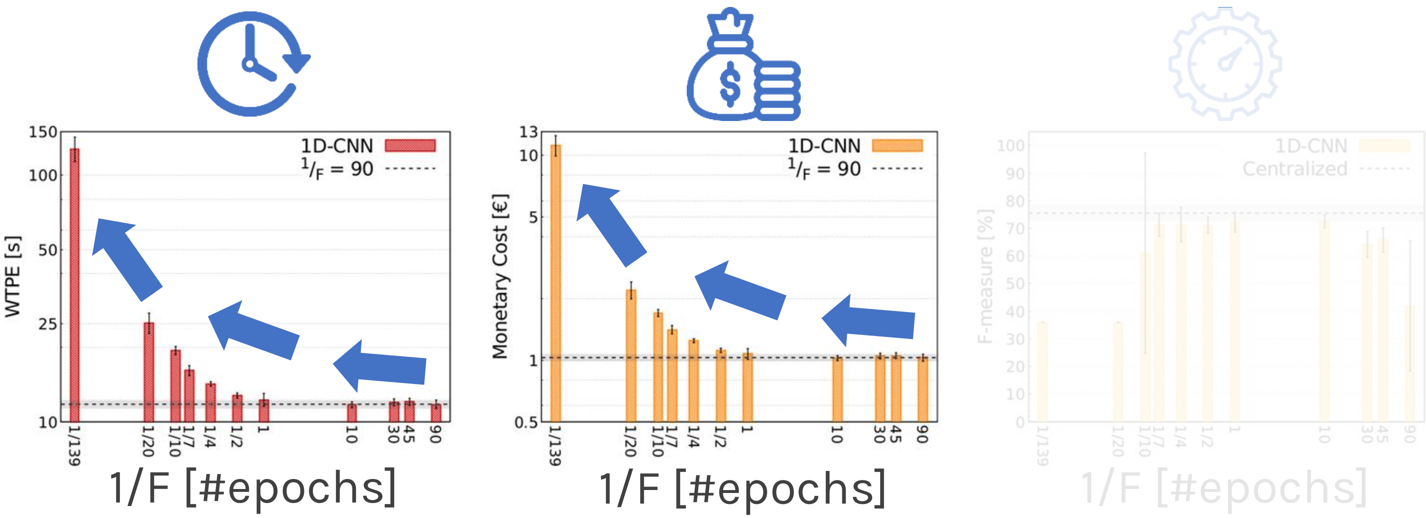




How does F impact BD-Enabled TC?

1D-CNN
N = 4

FB/FBM



- Both **training time** and **cost** increase with frequency update F
  - *+53.2%* passing from *1/F = 90* to *1/F = 1/20*
  - *+80.3%* passing from *1/F = 1/20* to *1/F = 1/139*

How does F impact BD-Enabled TC?

1D-CNN
N = 4

FB/FBM

- **F-measure** varies for different frequency update F intervals
  - Best performance for $1/4 \leq 1/F \leq 10$
  - Significant degradation for $1/F \leq 1/10$ and $1/F \geq 30$
  - *Computational bottleneck* at the master for $1/F \leq 1/10$ hindering the correct collection of the updates from the workers