NMLRG #5 meeting in Seoul, Korea

"Service classification in 5G networks"

P.Demestichas [1], K.Tsagkaris[2,3],
S. Vassaki[2,3], A.Georgakopoulos[2,3]
V. Kosmatos[2,3]

(1) University of Piraeus
(2) WINGS ICT Solutions, www.wings-ict-solutions.eu/
(3) Incelligent, www.incelligent.net

November 2016

Outline

- Motivation/Objectives
- ML approaches in service classification
- Problem description-Service characteristics
- Classification approach
  - Problem setup
  - High-level description
  - Development of classifiers
  - Model evaluation
  - Model comparison
- Conclusions/Next steps

Service Classification in 5G Networks* – Motivation & Objectives

- **Motivation**
  - Existence of diverse vertical/services with different requirements in terms of QoS & capacity:
    - Mobile Broadband (MBB)
    - Massive Machine Type Communications (MTC)
    - Mission Critical Communications (MCC)
    - Broadcast/Multicast Services (BMS)
    - Vehicular to X (V2X)
  - 5G system management → meet the requirements resulting from a large variety of services to be provided simultaneously optimizing the network in order to be resource and energy efficient
  - Prioritization of services and efficient allocation of resources → need for automated service classification schemes

- **Our approach**: use of machine learning (ML) techniques to solve service classification problem

- **Goal**: Accurate identification of services to promote an efficient network tuning (optimal assignment of resources to satisfy the diverse QoS requirements)

*[Investigated under the framework of FANTASTIC-5G project, H2020 G.A.671660, http://fantastic5g.eu/]

ML Approaches in Service Classification

Possible approaches for service classification problem:

- Clustering: Use of Unsupervised Learning techniques
  - Without provision of predefined classes
  - Find patterns in the input data → creation of clusters
  - Possible algorithms: K-means algorithm, Incremental Clustering, Probability based clustering
  - Pros: automatic discovery of classes
  - Cons: labeling of clusters is necessary

    number of clusters> number of application classes
- Classification: Use of Supervised Learning techniques
  - Use of predefined classes of training instances
  - 3 phases: training, cross-validation, application of classifier
  - Goal: from a training dataset, find a function *f(x)* of the input features that best predicts the outcome of the output class *y* for any new unseen values of *x*
  - Possible algorithms : Decision Trees, Naïve Bayes classification algorithms

Problem description - Service Characteristics

- MBB → diverse services (file downloading, streaming) usually larger packets
- MMC → periodic communication (inter-arrival time), small packet size
- MCC → usually small packets (except P2P communications)
- BMS → larger packets, multicast/broadcast communication (not individual destination)
- V2X (V2V or V2I) → high speed of nodes & combination with 4 others services

Service / Parameter	MBB (Video streaming)	MMC	MCC	BMS	V2X
Packet length	Usually large	Small	Small	Medium	Depends on other characteristics
Statistics of Packet length (e.g. std)	Medium	Very small	Very small	Medium-Large	Depends on other characteristics
Packet inter-arrival time	Non periodic	Periodic	Non periodic (burst effect)	Non periodic	Usually non periodic
Statistics of Packet Inter-arrival time	Medium-Large	Very small	Large	Medium	Medium
Flow Direction	DL: large packets UL:small packets	UL	UL	DL	UL/DL
Flow length	Large	Small	Small	Medium	Small-Medium
Statistics of flow length (std)	Small-Medium	Small	Small	Medium	Medium
Mobility	Average	Low	Low	Low-Average	High

Classification approach – Problem Setup

- We consider 3 types of services: MCC, MMC, MBB
- Generation of different types of MCC /MMC traffic following traffic models in 802.16p[1]
- Generation of MBB traffic following traffic models for video streaming (e.g. YouTube)[2]

- **Information of the collected traces:**
  - Simulation time
  - Source
  - Destination
  - Direction (Uplink/Downlink)
  - Packet Size
  - Device Name (e.g. Mobile128)

- **Separation of traces in flows**
  - Same source/destination
  - Interarrival time < threshold

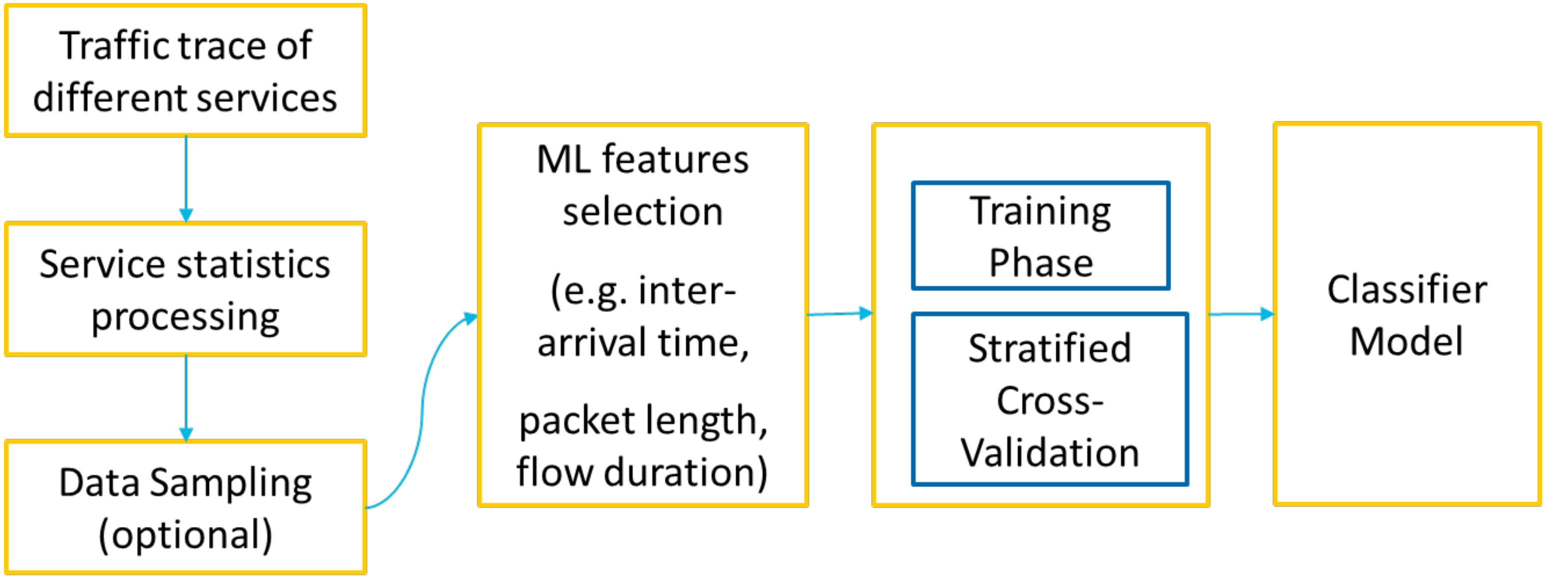
Features generation for each flow:

- Interarrival time (Mean value/ Std)
- Packet Size (total/Mean/Std/Min/Max)
- Number of packets
- Source
- Destination
- Direction

[1] IEEE 802.16p Machine to Machine (M2M) Evaluation Methodology Document (EMD)

[2] Ameigeiras, Pablo, et al. "Analysis and modelling of YouTube traffic." *Transactions on Emerging Telecommunications Technologies* 23.4 (2012): 360-377.

Classification approach – High-level description



Classification

- Use of predefined classes of training instances
- 3 phases: training, cross-validation, application of classifier
- Goal: from the training dataset , find a function *f(x)* of the input features that best predicts the outcome of the output class *y* for any new unseen values of *x*

Classification approach – Development of classifiers

- Algorithms for investigation :
  - Naïve Bayes
  - Support Vector Machine (SVM)
  - Decision Tree
  - k-Nearest Neighbors
  - Logistic Regression
  - Random Forest

Classification Mechanism	Accuracy
Naive Bayes	0.808
Support Vector Machine	0.662
Decision Tree	0.976
K Nearest Neighbor Classifier	0.952
Logistic Regression	0.685
Random Forest Classifier	0.988

- Confusion Matrix- Evaluation Metrics for classification

Service \ Classification Result	MMC	Other services
MMC	TP	FN
Other services	FP	TN

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

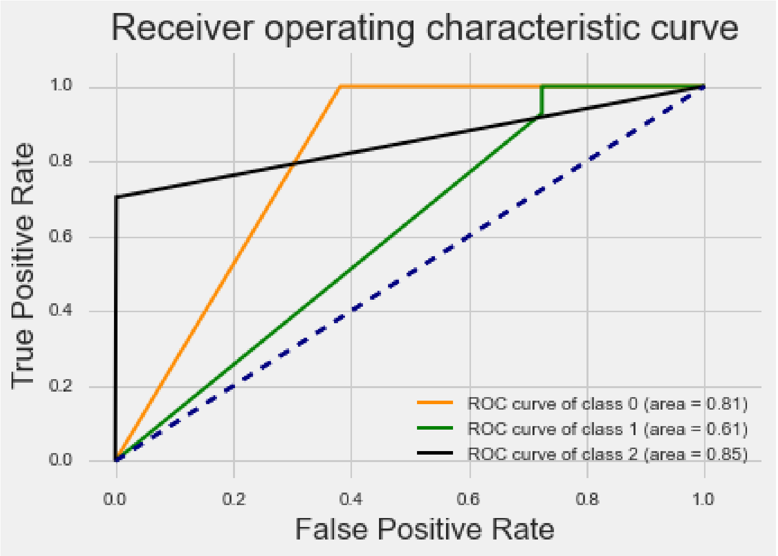
Classification approach – Model Evaluation

- Consider a set of different evaluation metrics
- Analysis of the tradeoff between metrics (ROC curve)
- Optimization of metrics depending on the service classification objectives:
  - ✓ MCC services→ high sensitivity (recall) values
  - ✓ MBB/MMC services → high accuracy values [eliminate classification error]
- Definition of new evaluation metric [$Perf_{ind}$] based on weighted values of Recall metric [for MCC class] and accuracy metrics [for MBB/MMC class].

Classification approach – Model Comparison

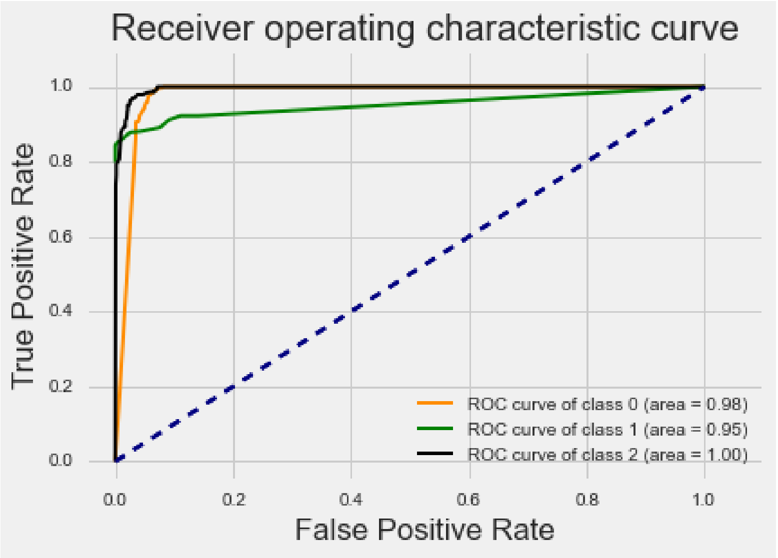
Naïve Bayes

Class\ Metrics	Precision	Recall	F1-score	$Perf_{ind}$
MMC	0.72	1	0.84	0.77
MCC	1.00	0.53	0.69	
MBB	1.00	0.72	0.84	
Avg/total	0.86	0.81	0.80	



K Nearest Neighbors

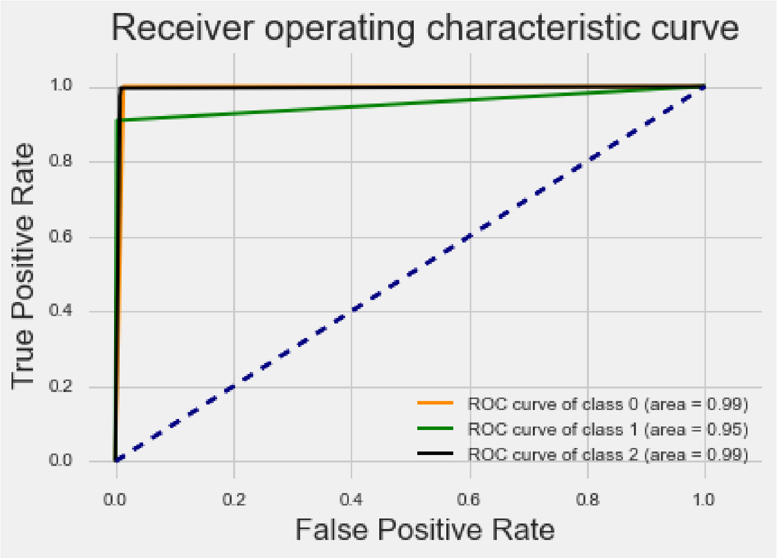
Class\ Metrics	Precision	Recall	F1-score	$Perf_{ind}$
MMC	0.93	1.00	0.96	0.94
MCC	1.00	0.86	0.93	
MBB	0.96	0.95	0.96	
Avg/total	0.95	0.95	0.95	



Classification approach – Model Comparison

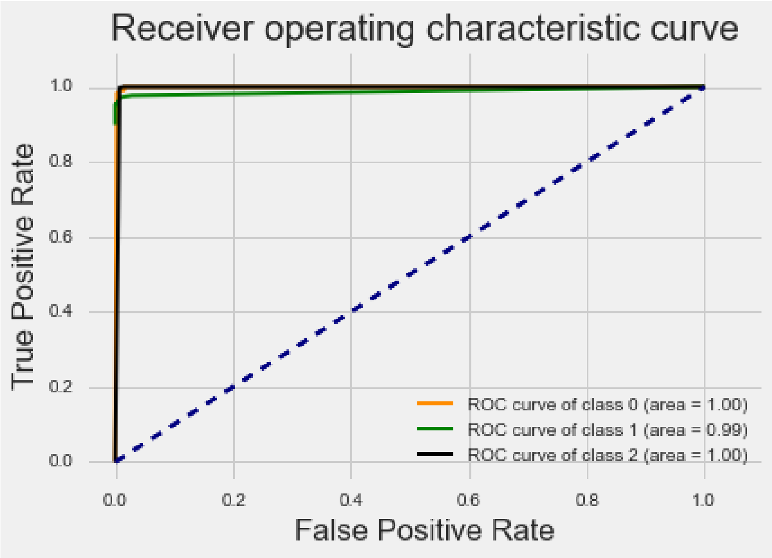
Decision Tree

Class\ Metrics	Precision	Recall	F1-score	$Perf_{ind}$
MMC	0.99	0.97	0.98	0.98
MCC	0.94	0.97	0.96	
MBB	0.98	0.99	0.99	
Avg/total	0.98	0.98	0.98	



Random Forest

Class\ Metrics	Precision	Recall	F1-score	$Perf_{ind}$
MMC	0.99	1.00	0.99	0.98
MCC	0.99	0.97	0.98	
MBB	0.99	0.99	0.99	
Avg/total	0.99	0.99	0.99	



Conclusion - Next steps

- Development & evaluation of automation mechanisms based on machine learning for Service Classification in 5G networks
  - Application of ML models
  - Comparison & selection of the most appropriate model

- Evaluation of predictive mechanisms for real-world mobile network scenario
  - Time series prediction using ML techniques
  - Selection of adequate evaluation metrics

Thank You!

For details you can visit:
http://tns.ds.unipi.gr
http://incelligent.net
http://wings-ict-solutions.eu