A Learning-Based Approach for Network Selection in WLAN/3G Heterogeneous Network

Nadine Abbas, Sireen Taleb, Hazem Hajj and Zaher Dawy

Department of Electrical and Computer Engineering
American University of Beirut, Beirut, Lebanon
Email: {nfa23, sht06, hh63, zd03}@aub.edu.lb

***Abstract*—To meet the huge traffic growth, heterogeneous networks composed of wireless local area networks (WLAN) and 3G cellular networks are used to provide higher capacity and coverage. When the two networks are available, selecting the best network for downloading data with minimum device energy consumption and high quality of service (QoS) becomes a challenging issue especially that mobile devices have limited energy capacity. This paper proposes a learning based approach for performing network selection based on real-network implementations. The main contributions are first, presenting an approach for building training data as a basis for machine learning of network selection and then developing the classification model for network selection. The model considers the features that affect the selection decision known by the user: availability of the networks, signal strength reflecting the channel quality, data size, battery life, speed of the user, location, and type of application. The training data set is based on experimental measurements of WiFi and 3G links using a Samsung Galaxy SII device. The network class annotation chooses the network that provides the user either highest QoS, lowest energy consumption or highest energy efficiency based on its current features status and service requirements. For real-time network selection, the developed model uses decision tree classification. Testing the performance of the classifier using cross validation demonstrated high accuracy for selecting between WiFi and 3G networks.**

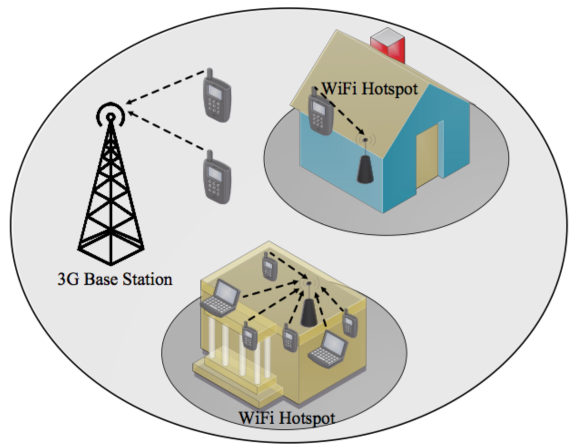

Fig. 1. Heterogeneous network scenarios formed by 3G macro cell and WiFi hotspots

I. Introduction

According to Cisco statistics, the global IP traffic will reach 1.3 zettabytes per year in 2016 [1]. To meet these tremendous traffic demands, 3GPP considered studying the feasibility of interworking between WiFi and cellular systems [2]. The small WiFi cells can complement the cellular systems coverage and capacity and reduce the traffic load on the cellular network. As shown in Figure 1, the network is formed of macro cells which are large cells covered by 3G/4G cellular networks and pico cells which are small cells served by WiFi hotspots. When the two networks co-exist, the user can have the ability to select the best network for downloading data. In addition, the mobile services are now so popular and diverse while battery-powered mobile devices offer limited energy capacity. Accordingly, new energy efficient network selection algorithms need to be proposed to maintain high quality of service while reducing the energy consumption of the mobile device. In this scope, several researches were conducted to propose efficient network selection algorithms based on machine learning. For instance, authors in [3] proposed a handoff algorithm based on fuzzy logic that allows switching between WLAN and 3G depending on signal strength and traffic load. In [4] and [5], authors used analytic hierarchy process or grey relational analysis for network selection. The authors did not consider all the parameters known by the user that can affect the network selection decision. In addition, some results were based on simulations or arbitrarily generated values for signal quality, network load and achievable data rate.

This paper proposes a new learning-based approach for performing WLAN/3G network selection based on real network implementations. The main contributions are: 1. A new approach for building training data as a basis for machine learning of network selection and 2. A new network classification model based on decision tree for network selection. Selecting the network providing minimum device energy consumption and high quality of service (QoS) to the user is a challenging issue due to the diversity of factors affecting the performance of the systems. Some of these parameters cannot be determined by the user such as load on the networks, interference, available real-time resources, and system capacity. Our proposed model considers the parameters that affect the selection decision and that are known by the user. We propose the following features for real-time selection, and develop a model to decide on the network based on the following features: (a) Availability of the networks, (b) Signal strength reflecting the channel quality between the user and the WiFi access point (AP) or the 3G base station, (c) Data size, (d) Battery life of the user device, (e) Velocity of the user, (f) Location of the user and (g) Type of application. In particular, battery life, location and type of service determines the priority of the user in selecting the network that provides the user either the highest QoS, lowest energy consumption or highest energy efficiency based on pre-defined rules. For instance, when the battery life is critical, the device should select the lowest energy consuming network to

978-1-4673-5307-6/13/$31.00 ©2013 IEEE

increase its battery life. To decide on the performance of each network in terms of rate and energy efficiency, the following attributes are used: user location, data size, and experimental values for WiFi and 3G signal strengths.

This paper is organized as follows. Section II presents the literature review conducted for the network selection. Section III describes the proposed network selection model including the proposed features for classification, the method for generating training model, and the machine learning model for network selection. The experimental results are described in Section IV. The paper is concluded in Section V.

II. Literature Review

Several researches were conducted for network selection in heterogeneous networks where different networks co-exist such as WLAN and 3G. The authors in [6] defined the best network as the one that best suits the user needs depending on the coverage, cost, bandwidth and application QoS requirements, capacity and security as well as personal preferences. Authors in [7] analyzed the energy efficiency of the always best connected algorithm. Measuring the mobile energy consumption, their results showed that handover is energy consuming but feasible if there is handover to a more efficient link. The authors in [3] first presented the conventional handoff algorithm that allows switching from one interface to another only when the channel quality signal strength exceeds a certain fixed threshold. They enhanced the conventional algorithm by using fuzzy logic based algorithm that uses user's speed and traffic load to decide on the threshold that was fixed in the conventional method. The authors in [8] proposed a user-centric network selection method that allows the user to make decisions according to its context and preferences. The user preferences are based on cost, power gain, rate, load and link quality priorities. The network, which provides higher utility function value, will be selected in addition to energy efficient pre-processing rules. In [4] and [5], the network selection was based on enhancing the quality of service by using analytic hierarchy process (AHP) and grey relational analysis (GRA). The AHP was used to evaluate the user QoS requirements and preferences such as throughput, security and cost. The GRA is then used to process the current network performance and compare it to the required user specifications; the network closer to the required preferences will be selected. The authors in [9] presented a novel radio access technology (RAT) selection policy that takes advantage of congestion information. They proposed a general Markovian framework for the allocation of multiple services into multiple RATs based on load balancing, service-based and congestion-aware RAT selection policies. They also presented analytical expressions for computing congestion probabilities of the RATs. The authors extended their work in [10] to consider the effect of coverage overlapping conditions with the RATs supported by the mobile.

III. The Proposed Network Selection Model

In this section, we present the details of the proposed learning-based approach for performing network selection. The first subsection presents the proposed features that can be available to every user's phone, and that can be used as discriminating attributes for deciding on the network. Subsection III-B covers the proposed approach for developing training data, which can then be used for deriving a classification model for network selection. The model for network selection is presented in Subsection III-C.

A. Feature Selection

Selecting the network providing minimum device energy consumption and high quality of service (QoS) to the user is a challenging issue due to the diversity of factors affecting the performance of the system. Some of these parameters cannot be determined by the user such as load on the networks, interference level, available resources and system capacity. Our proposed model considers all the parameters or attributes that affect the selection decision and that are well known by the user. These proposed features are:

- **Availability of the WiFi and 3G networks (A_{WiFi} and A_{3G}):**
  The two attributes A_{WiFi} and A_{3G} are defined to indicate the availability of the WiFi and 3G networks, respectively. These attributes have binary possible values indicating whether the particular network is available or not.
- **Signal strength (S_{WiFi} and S_{3G}):**
  The two attributes S_{WiFi} and S_{3G} are real values in dBm and they represent the channel quality signal strength between the user and the WiFi access point and 3G base station (BS), respectively.
- **Data size (D):**
  This attribute is the size of the data in Bytes that the user requests to download.
- **Battery life (B):**
  The battery life attribute represents the remaining battery life percentage of the device. In this paper, B is considered critical when the remaining battery life percentage is less than 20%.
- **Speed (V):**
  This attribute is a real value in m/s representing the speed of the user while downloading.
- **Location (L):**
  The location attribute is a binary attribute indicating whether the user is at home or outside home.
- **Type of application (T):**
  This attribute indicates if the type of the application is delay sensitive or not.

B. Training Model Development

To develop training data that can be useful for the network classification model, supervised annotation of network class label is needed in association with every set of the features' values. We propose the rules shown in Table I for the annotation of network choice. Table I describes the rule choices under different scenarios when both WiFi and 3G networks are available. The first four columns of the table represent some of the features available to the mobile: the location L, the speed V of the user, the battery life B, and the service type T. The network selection rules are:

TABLE I. NETWORK SELECTION RULES

Features				Decision
V	L	B	T	Rule
$\geq$5m/s	-	-	-	3G
<5m/s	home	-	-	highest rate
<5m/s	not home	critical	not delay sensitive	lowest energy
<5m/s	not home	critical	delay sensitive	highest energy eff.
<5m/s	not home	not critical	-	highest energy eff.

- If the user is moving faster than 5m/s, the user uses 3G cellular network since the WLAN coverage is limited and cannot manage high mobility [8].
- If the user is at home, battery life is neglected since the phone can be charged anytime. The link that offers the highest rate will be selected.
- If the user is not at home and the battery life is less than 20%, it is considered to be critical thus the network selected will be the network that offers lowest energy consumption if the service type is not delay sensitive.
- If the user is not at home and the battery life is critical, the network that offers highest energy efficiency is selected if the service is delay sensitive to provide a balance between rate and energy.
- If the battery life is not critical and the user is not at home, the user needs a good quality of service while increasing the battery life; therefore, the network with highest energy efficiency will be selected.

To derive annotation from the proposed rules, additional measures are needed to fire the different rules. These measures are: rate R, energy consumption E, and energy efficiency η. Figure 2 summarizes the proposed approach for building the training data set based on the selected features, rules and measurements. The figure shows the basic components of the network selection model. The following features: A_{WiFi}, $A_{3\text{G}}$, V, L, B and T are used for rule generation. While, the combination of other features: D, L, S_{WiFi} and $S_{3\text{G}}$ are used to define the rate, energy, and energy efficiency accordingly. Combining the rules with the measurement results, the annotation for network selection is derived for each set of features' values, and ultimately leading to the complete training data.

The additional features are measured experimentally for generating training data only, and they are not available as features in actual classification. The measurements are collected under different scenarios of the following features: the location L, the signal strength S_{WiFi}, $S_{3\text{G}}$ and the size D of the data that needs to be downloaded. The rate R and power consumed P are measured directly from the application and from a data acquisition device when experiments are conducted. Section IV presents more details and sample results. Then, energy E and energy efficiency η are derived indirectly from measurements as follows:

$$E(\text{Joules}) = P(\text{W}) \cdot Time(\text{s}) = P(\text{W}) \cdot \frac{D(\text{bits})}{R(\text{bps})} \quad (1)$$

$$\eta(\text{bits/Joule}) = \frac{D(\text{bits})}{E(\text{Joules})} \quad (2)$$

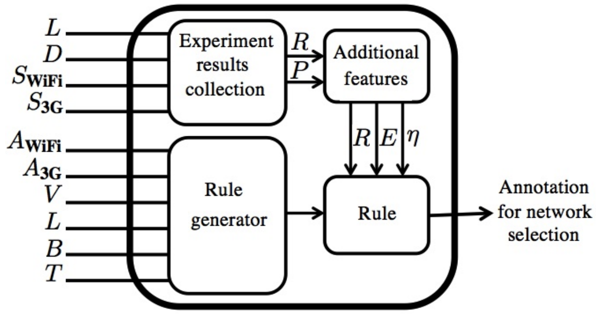

Fig. 2. Basic components of building the training data for the proposed network selection model

To illustrate a sample of the data, Table II presents measurements that are collected experimentally under a specific set of conditions for L, S_{WiFi}, $S_{3\text{G}}$ and D. In this scenario, the user is at home, close to the WiFi hotspot, having a bad 3G signal and needs to download a 1MB file. When the file is downloaded, the rate and power consumption are measured to determine the energy consumption and energy efficiency of each link.

TABLE II. FEATURES AND MEASUREMENTS: SAMPLE VALUES

Features				Measurements R(Mbps) and P(W)			
L	D	S_{WiFi}	$S_{3\text{G}}$	R_{WiFi}	P_{WiFi}	$R_{3\text{G}}$	$P_{3\text{G}}$
home	1MB	-38dBm	-101dBm	1.8	1.044	0.6	1.45

In this scenario, the user has a speed lower than 5m/s and is at home. As a result, based on the rules in Table I and the collected measurements, the network with highest rate is selected. Combining the rule with the collected measurements, the network that provides highest rate is WiFi. Thus, the annotation corresponding to this set of features is set to WiFi.

C. Network Selection Model

For the development of the network selection model, we propose supervised learning with the developed training data in Subsection III-B. This is a standard step in machine learning, with several options for classification algorithms such as the use of decision trees, Naive Bayes, and Support Vector Machine (SVM) [11]. Once the model is developed, a new set of features available to the user can be fed to the model, and the real time decision is obtained on the network choice for downloading data based on the user status and network conditions. In this paper, we propose the use of decision trees since the model gives a set of rules that can be logically evaluated for their relevance. The performance accuracy and evaluation of the classifier can be tested using standard methods such as cross validation. Details of the derived decision tree with our experimental measurements are presented in Section IV.

IV. EXPERIMENTAL RESULTS

In this section, we present first the details of the experimental setup and scenarios conducted to collect the needed measurements. In Subsection IV-B, the sample measurements results are first presented and analyzed to demonstrate the validity of the measurements collected. Second, the method for

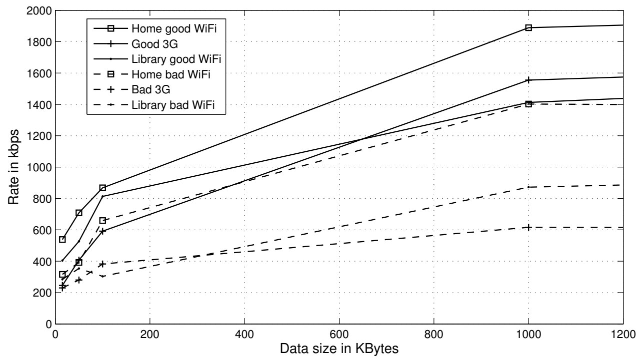

Fig. 3. Data rate variation with 3G and WiFi signal strength and data size

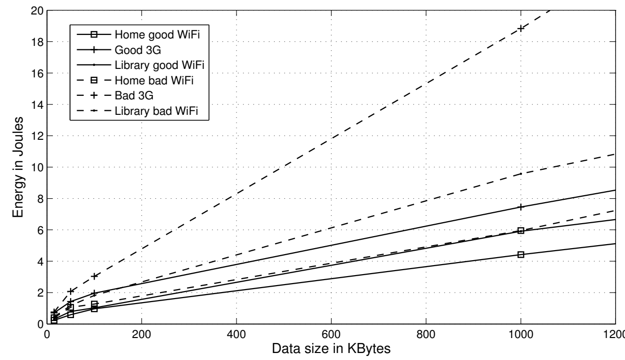

Fig. 4. Energy consumption variation with 3G and WiFi signal strength and data size

building the training data based on the rules and measurements is presented and illustrated by several scenarios. The results for the network selection classification model based on decision tree classification are presented in Subsection IV-C.

A. Experimental Setup

To collect the data needed to build the training dataset, the following setup was used. First, an Android application was developed on a Samsung Galaxy SII device. The purpose of the application was to download different data sizes while varying the device location, thus, changing the signal strength of WiFi and 3G. To measure performance under different conditions, six different locations were chosen: at home near the WiFi hotspot (home with good WiFi signal strength), at home far from the WiFi AP (home with bad WiFi signal strength), at the library where the network is loaded near the WiFi AP (Library with good WiFi signal) and far from the WiFi AP (Library with bad WiFi signal), indoor with bad 3G signal and outdoor with good 3G signal. In each scenario, the download data rate was obtained from the application while power consumption was measured using a data acquisition device (DAQ) monitored by LabView. The energy consumption and energy efficiency were derived offline using (1) and (2), respectively.

B. Training Model Development Results

First, to show the validity of our experimental results, the measurements collected were analyzed. Then, the approach for building the training data set is presented. Figures 3, 4 and 5 capture the variation of rate, energy consumption and energy efficiency with respect to data size (Kbytes), location (home and library) and signal strength (near, far from the WiFi hotspot, indoor and outdoor). Figure 3 shows that rate increases when the user has better signal quality when closer to WiFi AP. These results were as expected since the transmission quality is affected by the channel between the transmitter and the receiver. The data rate increased with data size since the calculations took into consideration the connection setup time; however, data rate saturates when data size is larger than 1MB in our measured scenarios. In addition, the data rate depends on the load on the WiFi network. In our model, we assumed that when the user is outside home, the WiFi network is considered to be loaded such as in a library environment. As expected, the WiFi rate showed lower values in a loaded environment.

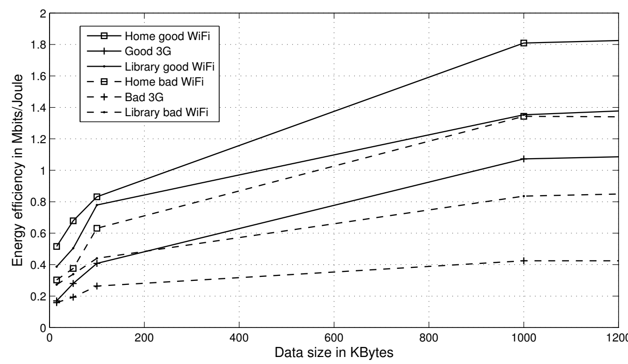

Fig. 5. Energy efficiency variation with 3G and WiFi signal strength and data size

The results for power consumption showed that the receiving power was on average 1.044Watts for WiFi and 1.45Watts for 3G. These values represent the power consumed by the mobile when the application is open and receiving and the brightness of the mobile is medium. The mobile power consumption is approximately constant since when receiving the mobile will be processing the data received. In general, the power consumed when 3G is used is more than the power consumed while receiving via WiFi due due to the additional processing requirements at the device level; these results are similar to the results provided in [7]. The energy consumption is then computed using (1). As shown in Figure 4, the energy consumption increased when the data size increased since it needs more time to download the data. The results also showed lower energy consumption when having good signal strength of WiFi and 3G since the rate is higher and, thus, the time needed to download data will be lower. Figure 5 shows the energy efficiency variations with respect to location and signal strength. The energy efficiency is computed using (2) and is a balance between data rate and power consumption. In our measurements, downloading data from WiFi with good signal strength was always more energy efficient than downloading data via 3G.

The network selection depends on the rules discussed in Table I. These rules needed additional measurements on rate, energy consumption and energy efficiency. The measurements presented above were used to get these additional information. To illustrate the method used for developing the training data set based on the rules and measurements, real examples are presented as follows. We assumed first that a user needs to download a 1MB file from home having a speed lower than 5m/s. Based on the rules, if the user is at home, the link that provides the best rate is selected. Based on measurements, the rates provided by WiFi and 3G are compared. Table III presents sample measured data rate, energy consumption and energy efficiency for different combinations of L, S_{WiFi}, S_{3G} and D. Assuming the user is close to the WiFi AP, and have a bad 3G signal, the rate of WiFi is greater than rate of 3G therefore the annotation is WiFi. Considering another scenario where the user is outside home having a good 3G signal and connected to a far loaded WiFi AP. Assuming the user has a critical battery life and the application is not delay sensitive and following the rules, the least energy consuming link is selected which is in this case the 3G network as provided in Table III.

TABLE III. DATA RATE, ENERGY CONSUMPTION AND ENERGY EFFICIENCY WHEN DOWNLOADING A 1MB DATA FILE FROM DIFFERENT LOCATIONS WITH DIFFERENT SIGNAL STRENGTHS

L	S	R[Mbps]	E[J]	η[Mbits/Joule]
Home	good Wifi -39dBm	1.88	4.44	1.80
Home	bad WiFi -80dBm	1.40	5.96	1.34
Library	good WiFi -39dBm	1.41	5.92	1.35
Library	bad WiFi -80dBm	0.87	9.60	0.83
Outdoor	good 3G -61dBm	1.55	7.48	1.07
Indoor	bad 3G -101dBm	0.61	19.01	0.42

Based on the rules and measurements previously presented, the training data set is developed. It is composed of the following nine attributes: A_{WiFi}, A_{3G}, B, V, T, L, D, S_{WiFi} and S_{3G}. The training data set was formed by 7920 tuples representing the number of scenarios considered for different combinations of the attributes.

C. Network Selection Classification Results

The supervised training data set was imported into the RapidMiner data mining tool to generate the decision tree based on gain ratio [12]. As shown in Figure 6, the root attribute was A_{WiFi} and then A_{3G}. The tree showed that if WiFi is not available, the model selects 3G network. As expected from the rules used in the training data, the third highest level attribute is the speed of the user. If the two networks are available and the user's speed is more than 5m/s, 3G is selected. Based on the remaining attributes, the model makes the network selection decision by following the decision tree. The results of the decision tree were consistent with our observations and experimental results analysis. For instance, one path of the decision tree showed that WiFi was always selected when the 3G signal is less than -96dBm and WiFi signal is less than -70dBm. This classification is consistent with the measurements results that showed that in the case of having bad WiFi and 3G signals, WiFi provided better gains in term of rate, energy consumption and efficiency. The performance evaluation of the classifier was tested using cross validation, with 66% of the tuples in the training data set used as training data to build the decision tree and the remaining tuples used to test the decision tree. The classifier for the considered scenarios led to 99.77% average accuracy.

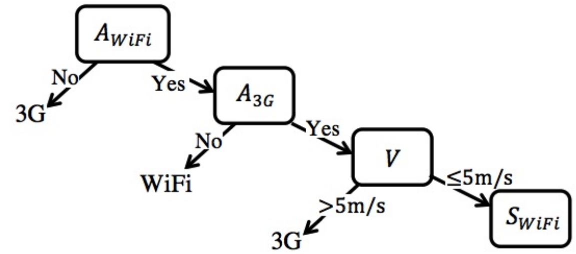

Fig. 6. First four levels of the decision tree

V. CONCLUSION

This paper proposed a learning-based approach for 3G/WLAN network selection based on real network measurements. The paper presented first a novel approach for building training data by applying selection rules based on experimental measurements. We also propose the use of decision tree as the classification model since it provides the decision in forms of rules that can be intuitively interpreted. Our proposed model considered the parameters that affect the selection decision and that are known by the user. The classifier was tested using cross validation showing high average accuracy for the considered scenarios.

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2011-2016," *White Paper*, May 2012.

[2] 3GPP TR 22.934 version 11.0.0 Release 11, "Feasibility Study on 3GPP System to Wireless Local Area Network (WLAN) Interworking," 2012.

[3] A. Majlesi and B.H. Khalaj, "An Adaptive Fuzzy Logic Based Handoff Algorithm for Hybrid Networks," *in Proceedings of the 6th International Conference on Signal Processing*, August 2002.

[4] Q. Song and A. Jamalipour, "Network Selection in an Integrated Wireless LAN and UMTS Environment Using Mathematical Modeling and Computer Techniques," *IEEE Wireless Communications*, vol. 12, no. 3, pp. 42–48, June 2005.

[5] ——, "Quality of Service Provisioning in Wireless LAN/UMTS Integrated Systems Using Analytic Hierarchy Process and Grey Relational Analysis," *in Proceedings of the Global Telecommunications Conference Workshops*, December 2004.

[6] E. Gustafsson and A. Jonsson, "Always Best Connected," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 49–55, February 2003.

[7] J. Kellokoski, J. Koskinen and T. Hamalainen, "Power Consumption Analysis of the Always-Best-Connected User Equipment," *in Proceedings of the 5th International Conference on New Technologies, Mobility and Security*, May 2012.

[8] Q. Nguyen-Vuong, N. Agoulmine and Y. Ghamri-Doudane, "A User-Centric and Context-Aware Solution to Interface Management and Access Network Selection in Heterogeneous Wireless Environments," *Computer Networks*, vol. 52, no. 18, pp. 3358–3372, December 2008.

[9] X. Gelabert, O. Sallent, J. Perez-Romero and R. Agusti, "Radio Access Congestion in Multiaccess/Multiservice Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4462–4475, October 2009.

[10] ——, "Performance Evaluation of Radio Access Selection Strategies in Constrained Multi-Access/Multi-Service Wireless Networks," *Computer Networks*, vol. 55, no. 1, pp. 173–192, January 2011.

[11] J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2012.

[12] RapidMiner Manual, "RapidMiner 5.0 Manual," *Rapid-I GmbH*, 2010.