

An Optimum Network Selection Solution for Multihomed Hosts using Hopfield Networks

Jorge Espi, Robert Atkinson, David Harle and Ivan Andonovic
Centre for Intelligent Dynamic Communications
University of Strathclyde
Glasgow, United Kingdom
{jorge.espi, r.atkinson, d.harle, i.andonovic}@eee.strath.ac.uk

Abstract—This work introduces a Hopfield Neural Network approach to network selection for multihomed hosts which considers a range of relevant network parameters including available radio access technologies and traffic types (VoIP, video streaming, Web browsing and FTP-based). Also proposed is a novel utility function that further improves network selection. Results show that, in terms of QoS, the allocation obtained using proposed algorithm outperforms other two reference allocation schemes under a range of different scenarios.

Keywords—Network selection algorithms; Hopfield neural networks.

I. INTRODUCTION

Much work is examining the interworking of heterogeneous network technologies to provide improved coverage and QoS differentiation [1]. Within 3GPP much work has focussed on the interworking between 3G and 2G and also Wireless LAN (WLAN) systems [2]. Within the IEEE a Layer 2 triggering mechanism for intersystem handoff, 802.21, is being developed. And within the IETF extensions to the Internet Protocols have been proposed to support mobile devices [3], [4]; the layer 3 nature of these protocols makes them equally applicable to intersystem handoff. Furthermore, at layer 4, there are a number of contributions to improve the performance of TCP and the QoS perceived by either real or non-real time services for mobile devices [5].

In an ideal world, each service should be supported by the most appropriate Radio Access Technology (RAT) taking into account the QoS requirements of the service and the characteristics of the underlying bearers in keeping with Ericsson's "Always Best Connected (ABC)" paradigm [6]. Common Radio Resource Management (CRRM) plays a major role on assessing each network's state and managing the resources in an unified manner across each of the heterogeneous technologies. This enables efficient service delivery to the end user for a range of disparate service types across a range of disparate technologies. Mapping services to technologies in a dynamic manner is the core of this paper.

Much research in this domain has focussed on devices with multiple interfaces that select a RAT on a service per service basis. The underlying assumption in much of this research is that the terminal will be connected to any

one of the many RATs at any particular time. Advances in layer 3 technologies within the IETF provides the potential to be connected to more than one network simultaneously. This approach, known as multihoming, has been the subject of research in a number of IETF working groups [7]. Multihoming can be regarded as a significant enabler towards the ABC paradigm because it supports always-on connectivity to multiple networks and hence negating interruption of service due to layer 2 handoff. This paper focusses on multihomed mobile hosts that are supporting multiple (dissimilar) services concurrently; the challenge is therefore is to map each of the individual services to one of the available RATs.

Network selection (performed at session set-up time) and network reselection (handoff) can be conducted in a network-centric or user-centric fashion. With the former, centralized or hierarchical distributed control can be exercised to optimize resource utilization across the interworked networks to the network operators' satisfaction. Network-centric approaches are limited to scenarios where a single operator owns and controls multiple RATs (3G, WLAN, etc.), or where business agreements exists between partner network operators. User-centric approaches are not bound by this constraint: indeed the competition between network operators could be exploited by users to increase competition in the marketplace and hence enable cost efficient connectivity.

In summary, this paper is focussed on a multihomed host performing user-centric network selection. To the best of the authors' knowledge, this problem has not been tackled in the literature to date. The only work to consider network selection for multihomed hosts to date focusses on a network-centric approach [8]. Specifically, this paper presents a novel multihomed user-centric approach to network selection based on Hopfield Neural Networks (MUC-HNN). HNNs are well suited to solving complex optimization problems within tight time frames in comparison with constrained optimization algorithms [9].

The rest of this paper is structured as follows. Section II defines the problem. Section III formulates the problem from the definition of HNN. Section IV illustrates a user case scenario for numerical evaluation of the MUC-HNN

algorithm. Its performance is compared with other three allocation algorithms. Finally, conclusions are summarized in Section V.

II. PROBLEM DEFINITION

Network selection algorithms must dynamically manage the allocation and de-allocation of traffic to the available networks. Their target is to optimize the available networks resources allocation to the running applications so that every ongoing communication's QoS is maximized.

The algorithm should be triggered whenever: (a) a new session set-up request is made; (b) the user changes his/her preferences or requirements¹; (c) the user's terminal detects a new network; (d) an ongoing service can no longer be supported by a particular radio link, e.g., due to signal degradation.

In the proposed scheme, individual micro-flows are not distributed across RATs because many applications utilize TCP which favors paths that are as symmetrical as possible, i.e. data and acknowledgments will traverse the same RAT. Likewise, the different latency across the different links would disrupt the communication timings in real time applications, which are mostly based on UDP. Therefore, it is considered that same type of traffic must be allocated to only one interface.

The mathematical model developed in this paper is valid for an arbitrary number of interfaces and available networks. However, although the host may be provided with several air interfaces, it is assumed to have only one of each type of RAT.

Therefore, network selection comes down to a constraint optimization problem. There is a plethora of constraint satisfaction algorithms which could be used for network selection: In this paper, the authors explore a HNN-based approach. HNNs can efficiently provide solutions for complex problems: they are more scalable than classical constraint satisfaction approaches, reducing both computation time (processing capability) and spatial complexity (memory required).

III. MUC-HNN ALGORITHM

A. HNN Dynamics

The HNN is the simplest form of neural network that consist of a single layer of neurons fully interconnected via symmetrical weights. Each neuron has a nonlinear output as described by an activation function (and associated activation value) [10]. The neurons update their activation values and weights asynchronously and independently of other neurons. The HNN iteratively converges from an arbitrary input state towards a sub-optimum solution as determined

¹On the forthcoming discussion, the author will differentiate between user preferences and user requirements. Preferential attributes are those to be maximised. Requirements are mandatory values for attributes.

by an energy function. In an N -neuron HNN, the energy can be described as shown in (1).

$$E = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N V_i V_j \omega_{ij} - \sum_{i=1}^N V_i I_i \quad (1)$$

being V_i the i th neuron output, I_i the bias vector and ω_{ij} the associated weight to the $i - j$ neurons interconnection.

The neuron activation function is calculated as shown in (2).

$$V = \frac{1}{2} \left(1 + \tanh \frac{U}{U_0} \right) \quad (2)$$

where U is the neuron input signal, V is the output signal and U_0 is a constant.

For the network to converge, the activation values are updated using the Euler method:

$$U_i^{t+\Delta t} = U_i^t + \Delta t \left(\sum_{j=1}^N V_j \omega_{ij} + I_i - \frac{U_i}{\tau} \right) \quad (3)$$

where τ is the time constant of the network, Δt is the time step and U_i^t is the input of neuron i at time instant t .

The energy function of a HNN has various minima (basins of attraction) that represent the (sub-optimal) solutions. An unknown input pattern represents a particular point in the energy landscape. As the network iterates in its way to a solution, the point moves through the landscape towards one of the hollows, i.e., local solutions [10].

B. Problem Formulation

In this paper, communication network selection is defined as an optimization problem. The HNN will iteratively converge to a solution where the cost associated with mapping a particular service to a particular RAT is minimized. For this purpose, the cost function (4) has been derived.

The network selection problem has been formulated using a 2-D HNN. The network has $N_{net} \times N_{tt}$ neurons, where N_{net} is the number of available networks and N_{tt} is the number of types of traffic that the application level is generating. A neuron ($V_{n,t}$) will be activated when traffic t is allocated to network n .

The proposed energy (cost) function, inspired in [8], consists of six terms. The first term forces the neurons to have a '0' or '1' output signal, or near these values. The second, guarantees that the same traffic type is not shared among several radios or networks. The third ensures that only one network is chosen from a particular type: e.g., a mobile device equipped with only one WLAN interface can only connect to one WLAN network simultaneously. The fourth is intended to enhance the user's network selection. The fifth precludes the user from demanding more than the maximum available bandwidth in each system. Finally, the

sixth maximizes the traffic allocation and hence the total resource utilization. Thus,

$$\begin{aligned}
E &= \frac{A}{2} \sum_{n=1}^{N_{net}} \sum_{t=1}^{N_{tt}} V_{n,t} (1 - V_{n,t}) + \frac{B}{2} \sum_{t=1}^{N_{tt}} \left(\sum_{n=1}^{N_{net}} V_{n,t} - 1 \right) \\
&+ \frac{C}{2} \sum_{n=1}^{N_{net}} \sum_{\substack{n'=1 \\ n' \neq n}}^{N_{net}} \eta_{n,n'} \left(\sum_{t=1}^{N_{tt}} V_{n,t} \right) \left(\sum_{t=1}^{N_{tt}} V_{n',t} \right) \\
&+ \frac{D}{2} \sum_{n=1}^{N_{net}} \sum_{t=1}^{N_{tt}} V_{n,t} f_{n,t}^u + \frac{E}{2} \sum_{n=1}^{N_{net}} \sum_{t=1}^{N_{tt}} V_{n,t} \xi_{n,t} \\
&+ \frac{F}{2} \sum_{n=1}^{N_{net}} \sum_{t=1}^{N_{tt}} V_{n,t} \frac{f_{n,t}}{f_{min,t}}
\end{aligned} \tag{4}$$

, where

$$\eta_{n,n'} = \begin{cases} 1 & \text{if } n \text{ and } n' \text{ same RAT,} \\ 0 & \text{otherwise.} \end{cases}$$

, and

$$\xi_{n,t} = u \left(\frac{B_a}{B_n} - 1 \right) \tag{5}$$

being

$$B_a = B_t + \sum_{n=1}^{N_{net}} \sum_{\substack{t=1 \\ t \neq t'}}^{N_{tt}} V_{n,t} B_t \tag{6}$$

, B_t standing for bandwidth required for traffic t and B_n available bandwidth at network n , $f_{n,t}$ the cost associated to selecting network n for traffic t and $f_{n,t}^u$ the cost from the user's perspective, as explained in the next section. Thus, by comparison with (1) and (4) the parameters ω and I are

$$\omega_{n,t,n',t'} = A \delta_{n,n'} \delta_{t,t'} - B \delta_{t,t'} - C (1 - \delta_{n,n'}) \eta_{n,n'} \tag{7a}$$

$$I_{n,t} = -\frac{A}{2} + B - \frac{D}{2} f_{n,t}^u - \frac{E}{2} \xi_{n,t} \tag{7b}$$

The HNN will tend to stabilize at the state that entails a minimum energy, i.e., the best network according to the user's perspective cost function $f_{n,t}^u$. However, the definition of this function is nontrivial: This function merges the network attributes and the user requirements and preferences

C. Cost Function Definition

The cost associated to selecting network n for traffic t is computed as:

$$f_{n,t} = \frac{B_t}{B_n} \tag{8}$$

Thus, the network resources are allocated proportionally to the traffic demands. However, this effect is subordinated to the perceived benefit from the ongoing applications.

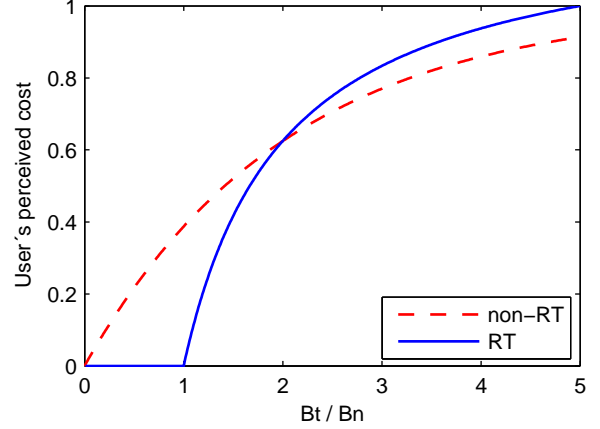


Figure 1. Costs associated with RT and non-RT traffic

In this work, a novel user's perspective cost function, $f_{n,t}^u$, is proposed. The cost value associated with it decreases as the network exceeds the user's QoS requirements (in terms of available bandwidth), while network failure to fulfil the user's requirements is highly penalized. Therefore, the term that refers to the user's perspective of costs (fourth term) of (4) will have a minimum value when only those networks that have the lowest cost values are activated.

For real-time (RT) services, the associated cost of traffic allocation to networks with higher capacity than the required bit rate is zero since then there is no packet dropping as result of congestion. For non-RT services, the function definition assumes a best effort cost characterization: The cost increases exponentially up to a saturation level. The scale parameter of the costs functions has been arbitrarily set considering same perceived cost at 50% of the user requirements ($\frac{B_t}{B_n} = 2$).

$$f_{n,t}^{NRT} = 1 - e^{-0.49 \frac{B_t}{B_n}} \tag{9a}$$

$$f_{n,t}^{RT} = 1.25 \frac{B_t - B_n}{B_t} u \left(\frac{B_t}{B_n} - 1 \right) \tag{9b}$$

Fig. 1 shows the cost characterization for both types of traffic. Thus, the allocation costs at application level are direct consequence of the available bandwidth for each traffic regarding the application demands, and the consistency of the application performance at scarce of resources.

IV. NUMERICAL EVALUATION

A. Simulation Scenario

A simulated network has been implemented to conduct numerical evaluation of the MUC-HNN solution. In this simulation, a multihomed device is provided with three interfaces (UMTS, IEEE 802.11b and Ethernet) that can connect to three access points (and hence RATs) simultaneously. The assumed available bandwidth at the UMTS link

is 60 Kb/s [11]. For the other two links, it is assumed a maximum capacity of 2 Mb/s and 10 Mb/s respectively, and a congestion level uniformly distributed between 0% and 80%. The device supports three concurrent sessions: VoIP, video streaming and file transfer. All three services have the same preference level. The algorithm is able to allocate zero or more services to an interface.

The traffic models for the VoIP and video streaming have been extracted from [12], [13] respectively: the terminal is assumed to generate a constant bit rate of 64 Kb/s for the former, and 5 Mb/s (download) for the latter. FTP file download traffic runs over TCP Reno.

The parameters of the proposed HNN have been calculated in Appendix:

$$\begin{array}{lll}
 A = 10 & B = 20000 & C = 20000 \\
 D = 1000 & E = 15000 & F = 500 \\
 \tau = 1 & U_0 = 0.1 & \Delta t = 10^{-4}
 \end{array}$$

To evaluate MUC-HNN's performance, it is compared with two other algorithms, following the same approach as in [8].

1) *Round Robin (RR)*: This technique allocates the resources from the available networks to each traffic cyclically. The maximum available bandwidth is allocated for each type of traffic considering no traffic sharing between two or more interfaces.

2) *Optimum Bit Rate (OBR)*: This technique allocates to each type of traffic the network capable of supporting it. The algorithm assigns the lowest bit rate greater than the traffic bandwidth requirements.

B. Numerical Results

From the bandwidth allocated to each type of traffic, the following QoS metrics have been calculated: VoIP packet dropping probability; video stream buffering time (as a percentage over the visualization time); and FTP service latency (for a 1MB file download).

Results are based on average outcome for 1000 simulation runs.

Fig. 2 illustrates the cumulative distribution function (CDF) of the packet dropping probability for the VoIP traffic. MUC-HNN offers a 0% blocking probability in any scenario, while OBR achieves it in 90% of the cases and RR only in 25%.

Figs. 3 and 4 show the CDF of the percentage of video stream buffering time and the 1MB-file download latency respectively. The proposed algorithm achieves better performance overall. It is noted, however, that in approximately 7% of the user case scenarios its video stream performance is slightly lower than the RR and OBR allocation. This is a consequence of the algorithm enhanced allocation of resources to both the VoIP and the FTP traffic. MUC-HNN reduces FTP service latency by 50% in 25% of the cases

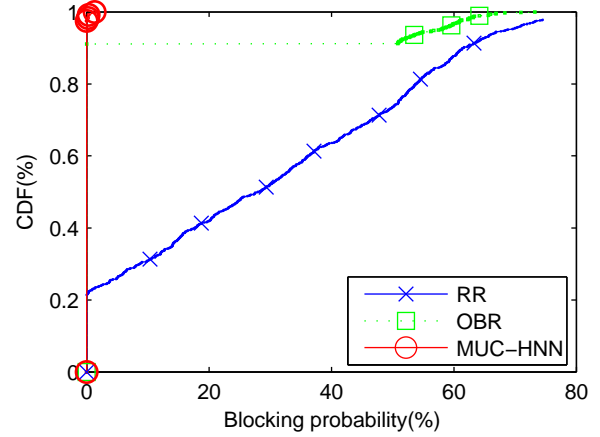


Figure 2. CDF for packet dropping probability of the VoIP service

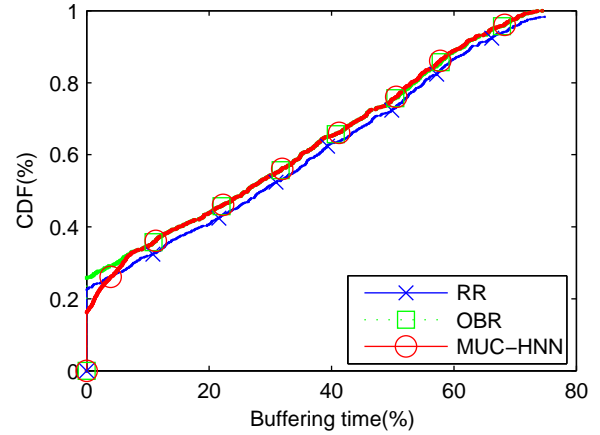


Figure 3. CDF for percentage of buffering time over visualization time (video stream service)

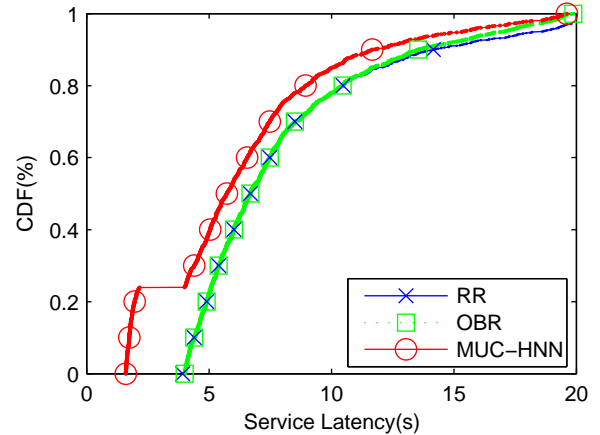


Figure 4. CDF for 1 MB file download latency (FTP service)

considered in comparison with RR and OBR algorithms. In the remaining cases, also attains better performance (approximately 20% latency reduction).

V. CONCLUSION

A novel network selection algorithm for multihomed users is proposed. This algorithm is based on HNNs and a newly defined cost function that describes, from a user's perspective, the cost associated to the user's traffic-network allocation. It is applicable in heterogeneous networks and can deal with any number of available interfaces.

Simulation results show that it has better performance than other reference resource allocation algorithms in a heterogeneous (UMTS, IEEE 802.11b and Ethernet) environment under different congestion levels.

APPENDIX A.

ENERGY FUNCTION WEIGHTING COEFFICIENTS CALCULATION

The weighting coefficients have been calculated as in [8]. The worst case scenarios are considered so that the choice of weights permits increased convergence rates. Choices that imply a benefit from the user perspective, not necessarily reduce the value of (4). Favorable choices are denoted by +, while negative choices are denoted by -.

A. First term

The first term of the energy function permits faster convergence of the HNN.

$$\frac{\partial E}{\partial V_{n,t_+}} = \frac{A}{2} (1 - 2V_{n,t_+}) + \frac{D}{2} f_{n,t_+}^u + \frac{F}{2} \frac{f_{n,t_+}}{f_{min,t}}$$

$$\frac{\partial E}{\partial V_{n,t_-}} = \frac{A}{2} (1 - 2V_{n,t_-}) + \frac{D}{2} f_{n,t_-}^u + \frac{F}{2} \frac{f_{n,t_-}}{f_{min,t}}$$

The condition for converge towards the minimum is:

$$\frac{\partial E}{\partial V_{n,t_+}} < \frac{\partial E}{\partial V_{n,t_-}}$$

In worst case scenario, $V_{n,t_+} = 1$, $V_{n,t_-} = 0$, $f_{n,t_-}^u = f_{n,t_+}^u$.

$$\frac{A}{2} (1 - 2V_{n,t_+}) + \frac{F}{2} \frac{f_{n,t_+}}{f_{min,t}} < \frac{A}{2} (1 - 2V_{n,t_-}) + \frac{F}{2} \frac{f_{n,t_-}}{f_{min,t}}$$

$$A > \frac{\frac{F}{2} \min(f_{n,t_-} - f_{n,t_+})}{f_{min,t}}$$

B. Fifth term

This term avoids network capacity overloading.

$$\frac{\partial E}{\partial V_{n,t_+}} = \frac{A}{2} (1 - 2V_{n,t_+}) + \frac{D}{2} f_{n,t_+}^u + \frac{F}{2} \frac{f_{n,t_+}}{f_{min,t}}$$

$$\frac{\partial E}{\partial V_{n,t_-}} = \frac{A}{2} (1 - 2V_{n,t_-}) + \frac{D}{2} f_{n,t_-}^u + \frac{F}{2} \frac{f_{n,t_-}}{f_{min,t}}$$

In worst case scenario, $V_{n,t_+} = 1$, $V_{n,t_-} = 0$, $f_{n,t_-}^u = f_{n,t_+}^u$, $f_{n,t_-} = f_{n,t_+}$

$$E > A$$

C. Second term

Users must not allocate same type of traffic to different interfaces simultaneously. The rationale behind this is that users are not enabled to handle connections at transport level seamlessly across several interfaces.

Terms forth and sixth decrease neuron outputs. Considering δ the maximum desired distance to the desired sum value, equilibrium is achieved when $|\sum_{n=1}^{N_{net}} V_{n,t} - 1| < \delta$. Considering the worst case scenario:

$$\left| \frac{D}{2} + \frac{F}{2} \right| < \delta B, B > \frac{D + F}{2\delta}$$

D. Third term

This term must decrease neuron output if $\eta_{n,n'} = 1$.

$$\frac{\partial E}{\partial V_{n,t}} = \frac{A}{2} (1 - 2V_{n,t}) + B \left(\sum_{n'=1}^{N_{net}} V_{n',t} - 1 \right)$$

$$+ \frac{C}{2} \eta_{n,n'} \left(\sum_{t'=1}^{N_{tt}} V_{n',t'} \right) (1 - \delta_{n,n'})$$

$$+ \frac{D}{2} f_{n,t}^u + \frac{E}{2} + \frac{F}{2} \frac{f_{n,t}}{f_{min,t}}$$

$$\frac{\partial E}{\partial V_{n,t}} > 0$$

Since $B > A$, worst case scenario is $V_{n,t} = 0$, $V_{n',t} = 0$.

$$\left(\sum_{t'=1}^{N_{tt}} V_{n',t'} \right) C > A + 2B - D f_{n,t}^u - E - F \frac{f_{n,t}}{f_{min,t}}$$

Therefore $C \gg 0$.

REFERENCES

- [1] S. L. Tompros and S. Denazis, "Interworking of heterogeneous access networks and QoS provisioning via IP multimedia core networks," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 52, no. 1, pp. 215-227, Jan. 2008.
- [2] 3GPP. (2009, Nov.) TSG Core Network and Terminals WG4. [Online]. Available: <http://www.3gpp.org>
- [3] D. Johnson *et al.*, *Mobility Support in IPv6*, IETF Std. RFC 3775, 2004.
- [4] R. Koodli *et al.*, *Fast Handovers for Mobile IPv6*, IETF Std. RFC 4068, 2004.
- [5] W. Liao *et al.*, "Improving TCP Performance in Mobile Networks," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 569-571, Apr. 2005.
- [6] E. Gustaffson and A. Johnson, "Always best connected," *IEEE Wireless Commun. Mag.*, vol. 10, no. 1, pp. 49-55, Feb. 2003.
- [7] WGs charts. (2009, Nov.) The Internet Engineering Task Force (IETF). [Online]. Available: <http://www.ietf.org>

- [8] D. Calabuig *et al.*, "A Delay-Centric Dynamic Resource Allocation Algorithm for Wireless Communication Systems Based on HNN," *IEEE Trans. Veh. Commun.*, vol. 57, no. 6, pp. 3653–3665, Nov. 2008.
- [9] N. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. The MIT Press, 1998.
- [10] B. Krose and P. Smagt, *An introduction to Neural Networks*. The Amsterdam University Press, 1996.
- [11] J. S. Blogh and L. Hanzo, *Third-Generation Systems and Intelligent Wireless Networking*. Wiley-IEEE Press, 2004.
- [12] ITU-T. (2009, Nov.) Recommendation G.711. [Online]. Available: <http://www.itu.int>
- [13] ——. (2009, Nov.) Annex A-Profiles and levels, Recommendation H.264. [Online]. Available: <http://www.itu.int>