Simulation Modelling Practice and Theory 19 (2011) 2117–2129

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

A metamodeling approach for cross-layer optimization: A framework and application to Voice over WiFi

Ioannis Papapanagiotou [a,*], Fabrizio Granelli [b], Dzmitry Kliazovich [c], Michael Devetsikiotis [a]

[a] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695-7911, USA
[b] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 14, I-38050 Trento, Italy
[c] University of Luxembourg, 6 rue Coudenhove Kalergi, L-1359, Luxembourg

ARTICLE INFO

Article history:
Received 26 March 2011
Received in revised form 22 June 2011
Accepted 23 June 2011
Available online xxxx

Keywords:
Cross-layer design
Metamodeling
Call Admission Control (CAC)
VoWiFi

ABSTRACT

Cross-layer design has been proposed to optimize the performance of networks by exploiting the interrelations among parameters and procedures at different levels of the protocol stack. This paper introduces a quantitative framework for the study of cross-layer interactions, which enables design engineers to analyze and quantify interlayer dependencies and to identify the optimal operating point of the system, by using network economic theory principles. The framework is then used for performance optimization of a single-cell Voice over WiFi (VoWiFi) system. Insights gained from the considered scenario enable us to define a novel cross-layer Call Admission Control (CAC) scheme. The multistage CAC takes into account Quality of Service (QoS) criteria, which provide satisfaction to the end user, as well as revenue criteria that maximize the possible profit of the WiFi provider.

1. Introduction

The layering principle has been long identified as a way to increase the interoperability and to improve the design of telecommunication protocols, where each layer offers services to adjacent upper layers and requires functionalities from adjacent lower ones. However such an approach introduces several limitations. For example in the context of wireless networks, the physical nature of the transmission medium (time varying behavior, interference and propagation environments) severely limits the performance of protocols (e.g., TCP/IP) in other layers. To overcome these limitations, a modification has been proposed, namely, cross-layer design, or "cross-layering". The core idea is to maintain the functionalities associated with the original layers, but to allow coordination, interaction, and joint optimization of protocols crossing different layers Toumpis and Goldsmith [26].

Moreover, Voice over WiFi (VoWiFi) communications represent a challenging scenario, as even in the simplest case of a single IEEE 802.11 cell, performance optimization requires the consideration of several parameters at different levels of the protocol stack. Indeed, codec parameters, link layer and physical parameters (and several others) may have an impact on the overall quality of communication, as it is perceived by the end user Brouzioutis et al. [4].

As limited QoS strategies are typically employed on the wireless link, there is a need for a CAC strategy, in order to limit the number of users in the system and, more generally, to provide possible on-line adjustments to parameters. The CAC decision to accept or deny incoming calls is commonly based on the observed system parameters; and no considerations are made on the possibility to tune these parameters in order to optimize CAC performance.

* Corresponding author.
E-mail addresses: ipapapa@ncsu.edu (I. Papapanagiotou), granelli@disi.unitn.it (F. Granelli), dzmitry.kliazovich@uni.lu (D. Kliazovich), mdevets@ncsu.edu (M. Devetsikiotis).

1569-190X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.simpat.2011.06.005

In view of the above, this paper describes the use of a formal framework to: (a) identify and formalize the interactions crossing the layers of the standardized protocol stack in order systematically to study cross-layer effects in terms of quantitative models; (b) support the design of cross-layering techniques for optimizing network performance and identifying the optimum operating point per configuration. The presented approach, based on techniques well-established in operations research, allows engineers to identify relationships among different design parameters and to estimate the potential advantages (if any) that result from enabling cross-layer interactions. The approach is then instantiated in the framework of a cost-benefit analysis and a Call Admission Control (CAC) strategy is defined, based on both Quality of Service (QoS) and profit maximization. To this aim, (c) one of the contributions presented in this paper is the design principle for CAC schemes, whose decision making process is based on the system model. Differently from what is known, the proposed CAC accepts the maximum possible calls by modifying parameters from different layers.

The structure of this paper is the following. In Section 2 we introduce the cross-layer and cost-benefit terms that are going to be used. In Section 3 a Voice over IP (VoIP) scenario over a WiFi network is presented and analyzed in order to define the maximum number of stations that can be accommodated by a WiFi Access Point (AP) while satisfying the QoS constraints. In Section 3 we also identify the ways that cross-layer parameters affect the performance of the network and the profit of the WiFi provider. Section 4 includes two CAC schemes, one that takes into account the QoS constraints, and another that also incorporates the profit of the provider. In Section 5 we present the previous work, and we conclude the paper with final remarks.

2. Cross-layer design

Cross-layer design allows a large degree of control of the behavior of the system, by enabling a higher level of interaction among the entities at any layer of the protocol stack. Layer K is enabled to control, depending on the specifics, a subset of all the parameters at any of the seven layers of the Open Systems Interconnection (OSI) stack.

The system response is modeled as a response $f(\mathbf{p}^1,\ldots,\mathbf{p}^7)$, i.e., as a function of all vectors $\mathbf{p}^j$ of parameters across the layers $j=\{1,\ldots,7\}$. The sensitivity of the system response and the interactions among factors, within and across layers, can then be captured naturally as the partial derivatives $\frac{\partial f}{\partial p_i^j}$ and $\frac{\partial^2 f}{\partial p_i^j \partial p_m^l}$ (ith parameter at layer j, which corresponds to the ith element of vector $\mathbf{p}^j$). Subsequently, one can then strictly or nearly optimize the performance of e_i (performance metric for each layer j) with respect to a subset of $p^{TOT}=\{p_i^j|\forall i,j\}$ under general constraints by using any available method, such as steepest ascent, stochastic approximation, ridge analysis, and stationary points, Box and Draper [3] and Kleijnen [14].

The function $f()$ across the layers can be analytically calculated or empirically estimated. Since closed form mathematical expressions are often unattainable for real systems, in Granelli and Devetsikiotis [10], we outlined a mathematical modeling procedure based on *metamodeling*. In this paper, we continue and extend our work on metamodeling of wireless systems, by (meta)modeling the performance of a multiuser VoWiFi system. On top of that we introduce an admission control scheme for wireless networks.

Our "raw" performance metrics, e_i, are further incorporated into a utility or "benefit" function $U(e_j)$ that expresses how valuable the (net) system performance is to the system owner or user. In general, the exact functional form of the utility function and the resulting objective function are less important than their curvature (often concave, to denote a certain "saturation") and their ability to preserve a relative ordering of the engineering alternatives, to enable ultimate design decisions.

With such an approach, the results achieved during the system optimization phase can then be employed to define guidelines for system design. Indeed, by employing the proposed framework, it is possible to select:

- the sensitivity of the system utility with respect to individual parameters;
- the optimal operating point of the system (direct consequence of the optimization process);
- the proper cross-layer interactions to enable (based on sensitivity of the system); and
- the proper signaling architecture to employ (allowing to identify the set of parameters and measurements to use).

In this paper, we will address the effects of cross-layering from the system and the service provider perpective. However, the same design principles would hold for the end-user perspective (e.g. QoE).

3. A VoIP WiFi scenario

The model is built in a four-dimensional domain defined by a set of parameters considered crucial for the overall system performance, namely, physical bandwidth, link error rate, maximum number of link layer retransmissions, and VoIP frame generation interval. The chosen set of parameters is spread over several layers of the protocol stack, making it difficult to predict the optimal operating point using ad hoc or intuitive methods. Generally many other cross-layer interactions and parameters could be taken into account in the development of the VoIP WiFi scenario.

3.1. System model

3.1.1. Network model

The network model is shown in Fig. 1. The network is an infrastructure WLAN with one Access Point (AP) serving N client nodes. Each client node initiates a bidirectional VoIP call with the AP. For each call, we use the ITU G.711 64 kbps codec,

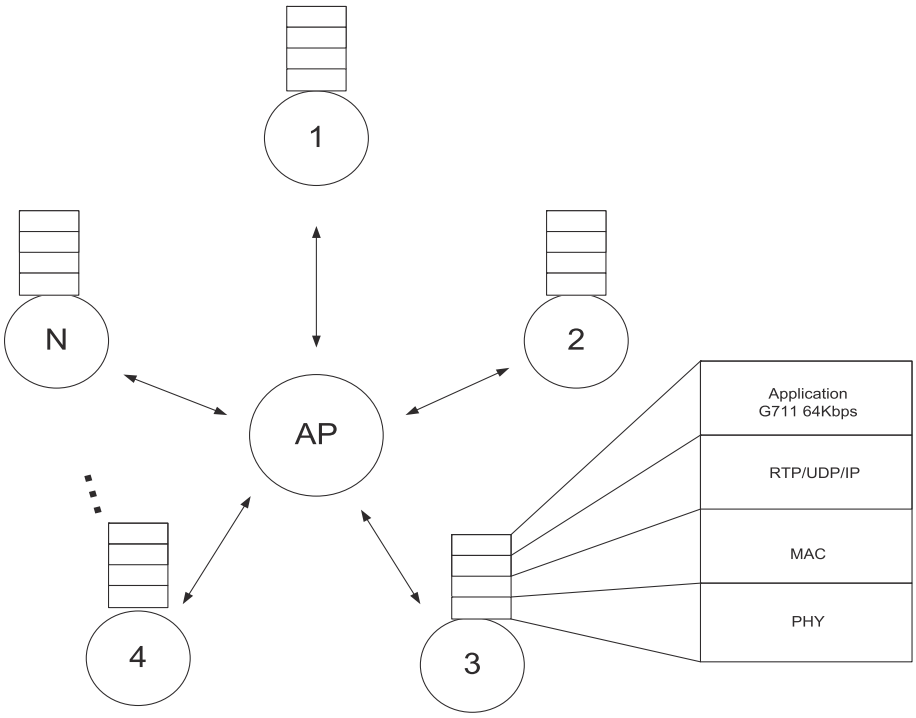

Fig. 1. Simulation scenario of the layered network.

where frames are sent for transmission at regular time intervals. The frames produced by the voice coder are then encapsulated by RTP/UDP/IP layers of the protocol stack adding an overhead of 40 bytes. At the MAC layer, we use IEEE 802.11 DCF basic access mode with no RTS/CTS exchange.

3.1.2. Inputs

We have selected the following four parameters, as inputs to the model:

- *Physical data rate (D)* is the data rate available for transmission at the physical layer. In order to comply with IEEE 802.11b, physical data rate values are taken equal to 1 Mbps, 2 Mbps, 5.5 Mbps and 11 Mbps.
- *Packet Error Rate (PER)*: Wireless systems are usually characterized by a high error rate, corrupting data transmitted at the physical layer. In order to evaluate system performance also in most channel conditions, we decided to vary PER between 10^{-9} and 10^{-1}.
- *Maximum number of retransmissions (R)*: The task of link layer Automatic Repeat reQuest (ARQ) is to compensate for high error rates on the wireless channel. The crucial parameter for ARQ scheme performance is the maximum number of retransmission attempts performed before the link layer gives up and drops the frame. Each retransmission consumes the same physical resources as the original frame transmission, thus reducing the overall capacity of the cell. On the other hand, retransmissions increase packet delivery delay. In our network model, the value of parameter *R* varies between 0 and 5, where 0 corresponds to the case when no retransmissions are performed at the link layer.
- *Voice packet interval (I)*: defines the time interval between successive frames generated by the voice codec. Voice packets are then encapsulated using RTP over UDP/IP protocols. Voice frames produced by the codec are relatively small (usually smaller than 100 bytes). As a result, a significant fraction of the nominal network capacity is wasted due to protocol overhead (40 bytes per packet). The parameter I varies from 10 to 90 ms in the considered scenario.

While, there can be unlimited number of input parameters, for modeling purposes we chose only those four. In fact, analytical studies of the IEEE 802.11 have shown the importance of those parameters on the performance of wireless networks [20,17]. In fact the selection of inputs could not only be limited to those that are modifiable, but at least measurable (e.g. PER on the wireless connection) in order to enable optimization. For example, in case PER is known and/or variable, the proposed system can use such information to understand the potential impact of the other parameters and have a complete picture of their impact, therefore enabling optimization.

3.1.3. Outputs

The output response of interest, $N = f(D, \text{PER}, R, I)$, is the maximum number of Voice over IP (VoIP) calls that can be supported by the Wireless Local Area Network (WLAN) cell with a satisfactory quality, which is defined by the following constraints.

3.1.4. Constraints

Several factors affecting VoIP performance can be mainly divided into human factors and network factors. Human factors define the perception of the voice quality by the end-user. The most widely accepted metric, called the Mean Opinion Score (MOS) P.800 [19], provides the arithmetic mean of all the individual scores, and can range from 1 (worst) to 5 (best).

The factors affecting the MOS ranking are related to network dynamics and include end-to-end propagation delay and frame loss, P.800 [19] and Schulzrinne et al. [24]. The delay includes the encoder's processing and packetization delay, queuing delay, channel access, and propagation delay. For this reason, in order to ensure an acceptable VoIP quality, we limit the delay parameter to 100 ms measured between the unpacketized voice data signal at codecs located at the sender and the receiver nodes. The second factor, frame loss rate, affects the VoIP quality due to non-ideal channel conditions. The chosen ITU G.711 64 kbps codec, P.800 [19], shows acceptable MOS rating (MOS = 3) for frame loss rate up to 5%, Ding and Goubran [6].

3.2. Cross-layer model of VoIP

As a closed form analytical model across the layers is clearly intractable, we define a quantitative model for the VoIP capacity as $N = f(D, \text{PER}, R, I)$ estimated via response surface (meta)modeling.

3.2.1. Implementation of cross-layer signaling

The network model described above is implemented in the NS2 network simulator (version 2.33) NS-2 [18]. The simulation parameters are summarized in Table 1. The ITU G.711 64kbps codec is emulated using a Constant Bit Rate (CBR) generator source, producing blocks of data at regular intervals specified by the voice interval I input parameter. In addition to the voice codec, the Cross-Layer Control (CLC) module is added at the application layer of the protocol stack (see Fig. 2). CLC is able to read the externally measured values of D and PER from the physical and link layers (cross-layer). By external measurements, we mean those that do not belong to the same layer. Moreover CLC is able to read the internal values of I from the application layer (intra-layer). Finally, it can set *R*, *I*, or *D* to the desired value.

3.2.2. Model definition

For each combination of input parameters, that is, *D*, *PER*, *R*, and *I*, we run a series of simulations with the number of VoIP flows incrementally set from 1 to 25 (fractional factorial design). Then we find the maximum number of VoIP flows *N* accepted by the system as the output for which the quality of the voice signal remains above a satisfactory level (as defined in Table 1, with end-to-end delay less than 100 ms and frame error rate less than 5%), by checking every voice frame.

The goal is to define how a change of the input parameters affects the VoIP capacity of the networks. Table 2 shows the values of the input parameters used in the experiment. For homogeneity, the same input values were used per simulation run and for all stations in the network. In order to fit the simulation results with a model, we used the JMP [13] and a second order polynomial RSM model, in order to identify the interactions and to define the corresponding coefficients. Although a polynomial regression technically is a special case of multiple linear regression, it can capture the effect that the underlying monomials can be highly correlated (e.g. PER and *D*). Those are presented in the following equation (note that the interaction between I and R is not significant, therefore it is excluded from the model). Results show that the squared coefficient of multiple correlations of the fitted model is equal to 0.81.

$$\begin{aligned} N = max\{0, &-5.1027 + 1.5575D + 292.8806I + 1.3677R - 157.3738\text{PER} + 5.9569D * I \\ &+ 0.1980D * R - 5.1210D * \text{PER} - 891.6851I * \text{PER} + 3.7706R * \text{PER} - 0.1186D^2 - 2710.813I^2 - 0.2935R^2 \\ &+ 1644.7405\text{PER}^2\} \end{aligned} \quad (1)$$

Table 1
Simulation parameters based on IEEE 802.11.

Parameter name	Value
Slot	20 μs
SIFS	10 μs
DIFS	50 μs
PLCP preamble and header	192 μs
Data rate	1, 2, 5.5 or 11 Mbps
Basic data rate	1 Mbps
Propagation model	Two-ray ground
RTS/CTS	OFF

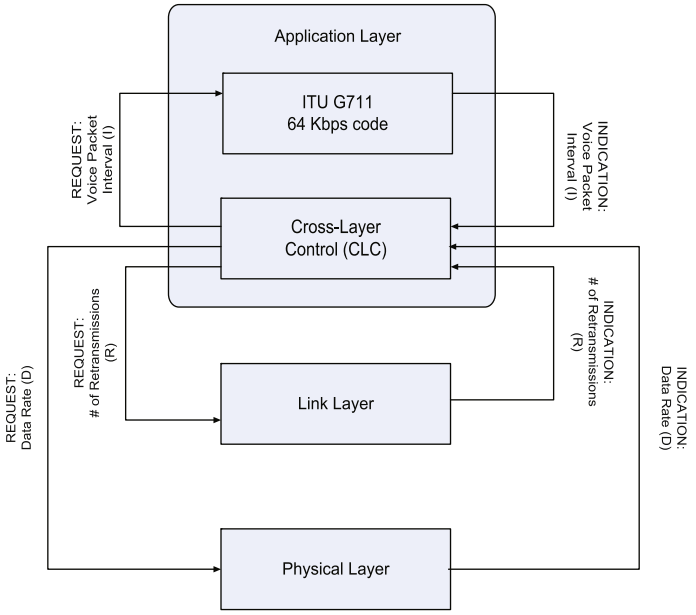

Fig. 2. Cross-Layer Control (CLC) module and cross-layer interactions.

Table 2
Experiment design parameters.

	Parameter name	Abbreviation	Levels	Values
Inputs	Data rate	D (Mbps)	4	1, 2, 5.5, 11
	packet error rate	PER	9	10^{-9}, 10^{-8} 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}
	Retransmissions	R	6	0, 1, 2, 3, 4, 5
	voice packet interval	I(ms)	9	10, 20, 30, 40, 50, 60 70, 80, 90
Constraints	Voice E2E delay	(ms)	–	<100
	Frame loss rate	FLR	–	<5%

Figs. 3–5 illustrate the obtained metamodel function N in all four dimensions of D, I, R, and PER. The maximum of N with respect to I is located between 0.05 and 0.07 seconds as it is evident in Fig. 3. Obviously, with the increase of I, client nodes generate fewer packets, thus increasing network capacity. However, the voice packet interval was chosen to take such values such that it is lower than the maximum Voice E2E delay. Consequently, after a certain threshold, an additional increase of I becomes unfavorable, leading to an overall network capacity decrease. A similar observation can be made for the maximum number of retransmissions configured at the link layer. With a higher R, the system can sustain a higher error rate at the wireless link. However, each retransmission consumes bandwidth resources from the shared channel. For high data rate scenarios ($D = 11$ Mb/s), retransmissions take just a small fraction of the entire bandwidth while for low data rate scenarios ($D = 1$ or 2 Mb/s), the portion of bandwidth used for retransmissions becomes considerable (see Fig. 4). As a result, the N is maximized at R equal to 3 for low data rates. From the comparison of Figs. 3 and 4 it can be observed that N is not sensitive to changes of PER. However in Fig. 3c and for PER = 10^{-1} the number of stations, that the network can hold, is decreased to almost a half for high throughput cases.

In order to determine the optimal solution, we solve a nonlinear constrained relaxed model

$$\begin{aligned} &\max N \\ &\text{subject to } [0,0,0,0] \leqslant [D,I,R,\text{PER}] \leqslant [11,0.09,5,10^{-1}] \end{aligned} \tag{2}$$

Since the inequality constraints form a closed convex set in $\mathbb{R}^4$, the active set method converges to the optimal solution, Gill et al. [9]. The optimal solution is determined to be $D = 11$ Mb/s, $I = 0.066$ s, $R = 5$, PER = 10^{-9}, and achieves a value $N = 19.92 \approx 20$, something that can be also verified by Figs. 3–5. The reader should note that this maximum corresponds

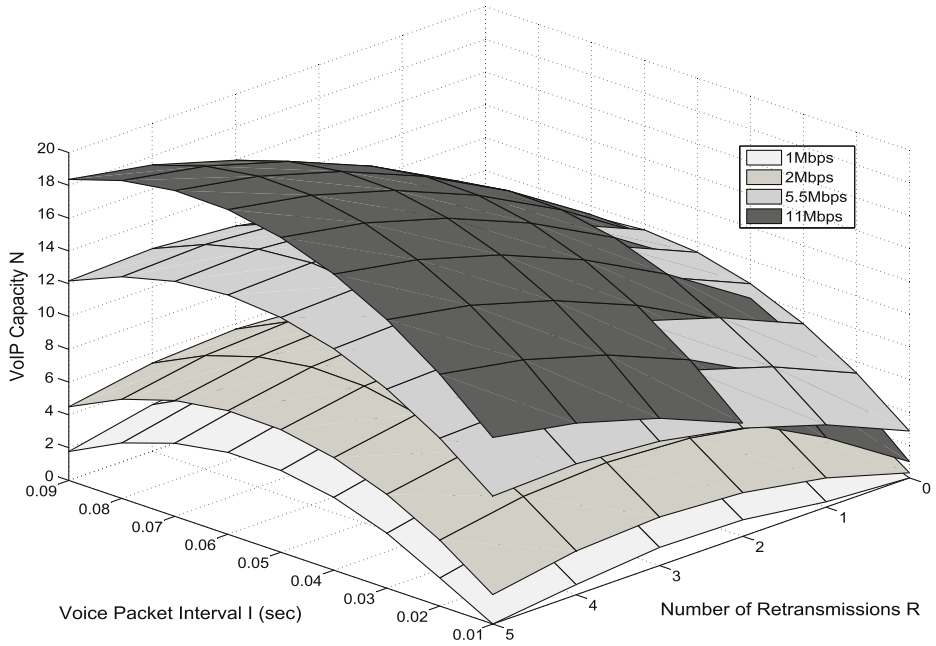

Fig. 3. VoIP call capacity (N) for PER = 10^{-9}.

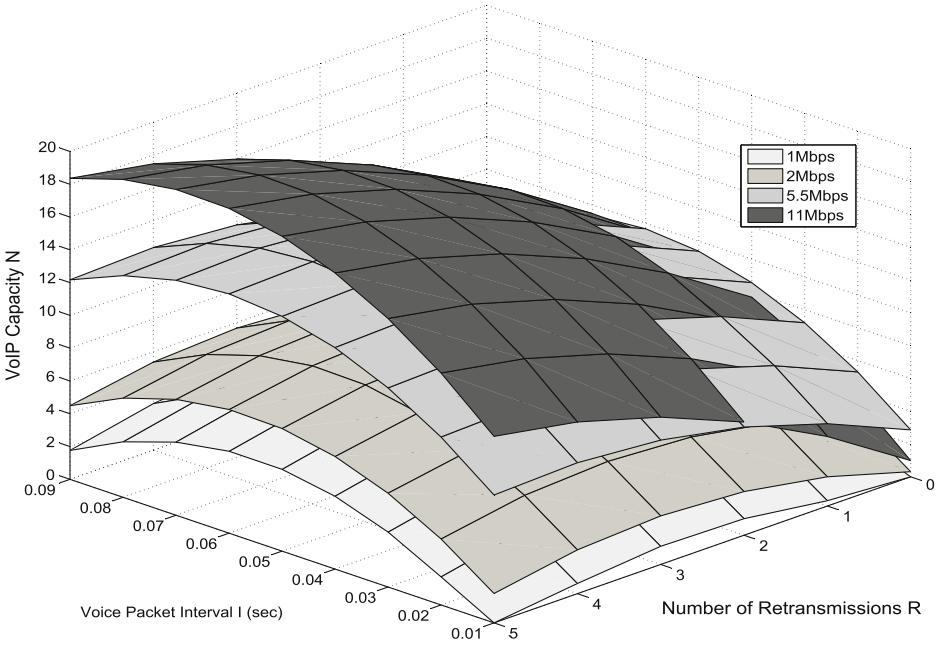

Fig. 4. VoIP call capacity (N) for PER = 10^{-5}.

to approximately 36% utilization of D provided at the physical layer. The remaining 64% of the transmission capacity is wasted on physical and link layer overheads, which become especially large for small (like VoIP) packets. This overhead, which is fixed for packets of any size, can be several times larger than the portion of the packet carrying data payload for small packets. Based on the above model, we proceed to quantify the sensitivity of the response on the four cross-layer variables D, I, R, and PER by a second order analysis, for simplicity. To this end, the Jacobian matrix is:

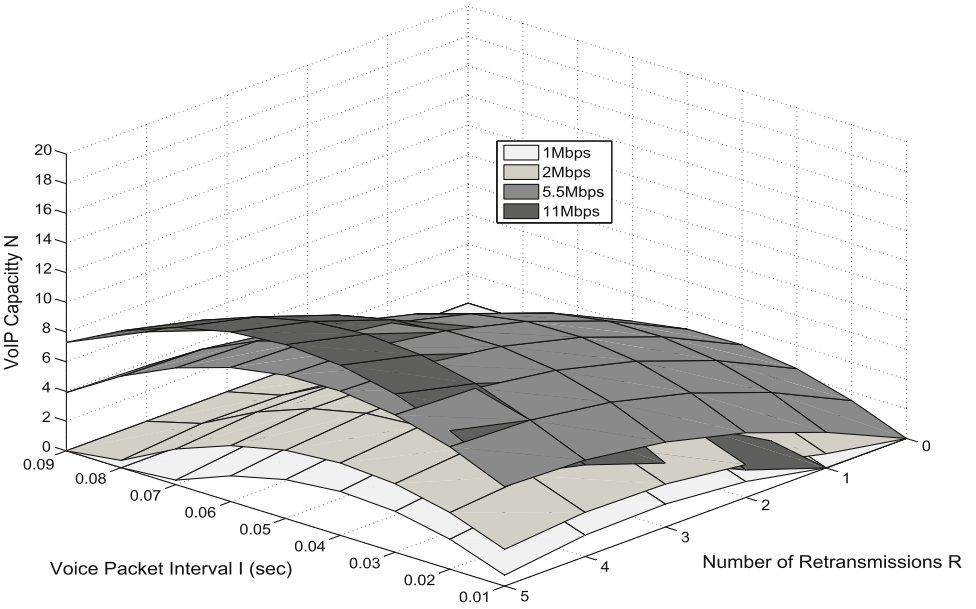

Fig. 5. VoIP call capacity (N) for PER = 10^{-1}.

$$\begin{bmatrix} \frac{\partial N}{\partial D} \\ \frac{\partial N}{\partial I} \\ \frac{\partial N}{\partial R} \\ \frac{\partial N}{\partial \mathrm{PER}} \end{bmatrix} = \begin{bmatrix} -0.2372D + 5.9569I + 0.198R - 5.1209\mathrm{PER} + 1.5575 \\ 5.9569D - 5421.626I - 891.6851\mathrm{PER} + 292.8815 \\ 0.198D - 0.587R + 3.77\mathrm{PER} + 1.3677 \\ -5.12D - 891.6851I + 3.77R + 33289.48\mathrm{PER} - 157.3734 \end{bmatrix} \tag{3}$$

The knowledge of the behavior of the first-order derivatives of N allows the estimation of the impact of each of the parameters. Since N has interactions of more variables, the Hessian matrix is shown below.

$$N = \begin{bmatrix} -0.2372 & 5.9569 & 0.198 & -5.1209 \\ 5.9569 & -5421.626 & 0 & -891.6851 \\ 0.198 & 0 & -0.587 & 3.77 \\ -5.12 & -89.6851 & 3.77 & 3289 \end{bmatrix} \tag{4}$$

The two zero points of the Hessian matrix shows that the interactions between I and R are very small, and have been removed in the regression analysis. In order to show more interactions between cross-layer parameters, a higher order polynomial could have been produced. However the length of the polynomial would also increase.

The absolute maximum values of the derivatives are presented in Table 3. In our case, the voice packet interval I and the Packet Error Rate PER at the physical layer have a higher impact on the maximum number of calls that can be supported by the system.

3.2.3. System optimization

Once the metamodel is established, it is possible to exploit the information it contains to build a utility function and, thus, enable a cost-benefit analysis of the problem. In the following we identify the utility and the profit function from the service provider's perspective and related it to an additional Call Admission Control module that is based on profit values as well as technical constraints.

Table 3
Absolute maximum values of the derivatives of N.

Derivative	Maximum	D (Mb/s)	I (s)	R	PER
$max\left\|\frac{\partial N}{\partial D}\right\|$	2.84	1	0.09	5	0
$max\left\|\frac{\partial N}{\partial I}\right\|$	278.27	1	0.09	$\geqslant 0$	0.1
$max\left\|\frac{\partial N}{\partial R}\right\|$	3.92	11	$\geqslant 0.02$	0	0.1
$max\left\|\frac{\partial N}{\partial \mathrm{PER}}\right\|$	293.95	11	0.09	0	0

3.3. Service provider perspective

From the point of view of the service provider, the assumed main concern is associated with maximization of the profit obtained from the operating network. The profit is directly proportional to the number of calls that can be supported by the system simultaneously, while satisfying the QoS constraints. In other words, it represents the consumer's preference over the system conditions. For example, if the provider charges P_{call} (marginal income from a single call), then more calls would result in higher profit, taking into account that the Service Level Agreements (SLAs) are satisfied. Moreover, the profit function must also have components related to capital and operating costs.

Definition 1. We define an indirect utility function (borrowing the term from microeconomics) for the VoWiFi system, as the function that shows the profit from the service provider perspective given specific price and system conditions.

More specifically, (P_{call}*D) accounts for the price that the user has to pay for the used bandwidth at time t. Assuming infinite capacity of the backbone link (compared with the wireless capacity), the service provider benefits from sending more packets to the users, with the least retransmission attempts. Thus, the profit function is:

$$\pi(N, D, R) = N(t) * D * \left\{ P_{\text{call}} - P_{\text{power}} \frac{p}{2} \right\} \tag{5}$$

where P_{power} is the marginal cost of a unit of transmitted power measured in mWatts. The second term accounts for the cost that the provider has to pay for the energy spent to connect the nodes in the network. We consider the simplest case of a Gaussian channel, where the power spent is linearly dependent on the number of WiFi clients. The WiFi clients are uniformly distributed in the WiFi cell (therefore $N/2$). The quantity p is a constant expressing the required Watts per mean number of stations. We assume that the maximum output energy is in the interval [−20 dBm, 20 dBm] based on the distance of the station from the AP. For simplicity we set p = 0 dBm = 1 mW. Implementation of other client or channel distributions would require minor changes to the cost function.

Fig. 6 presents the behavior of π. The ratio $P_{\text{call}}/P_{\text{power}}$ is chosen to be equal to 100 in this example. This corresponds to the policy of the service provider charging a transmission rate based pricing scheme of 1$/MByte. We assumed a much bigger cost for P_{call} compared to P_{power}, since the resources reserved for a call usually cost much more for the service provider. Fig. 6a shows reduced profit for the WiFi provider for low values of voice packet interval, which leads to high bandwidth consumption. For high values the profit is also reduced because the QoS constraints are not satisfied. Since this parameter is at the application layer it can easily be controlled by the service provider. For all transmission rates, the maximum profit is achieved with I = 0.06, apart from D = 11 Mbps, where the optimal I is equal to 0.07.

In Fig. 6b, we can see that with higher PER, the profit is less, because N from the metamodel is smaller, thus leading to less profit for the provider. It is useful to note that there are some cases that the profit function is almost zero, corresponding to the case that the ongoing stations are not charged enough for the provider to make a profit. This is specifically shown on

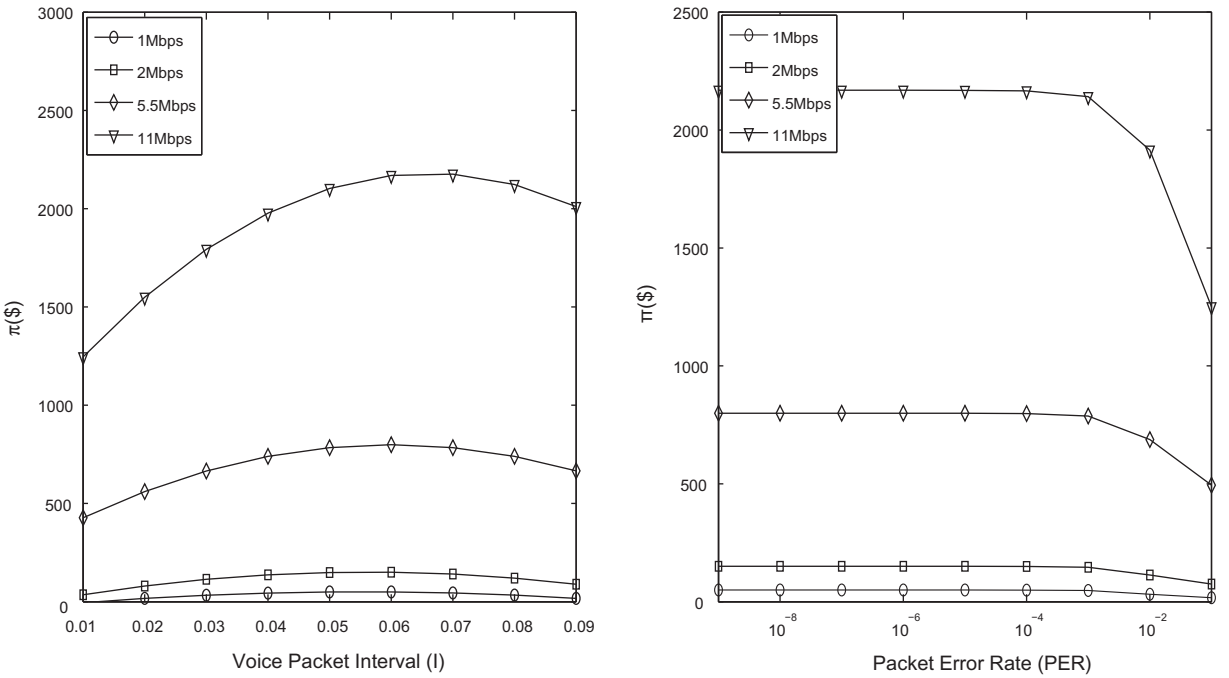

Fig. 6. VoIP service provider profit function π as a function of Voice Packet Interval in seconds and Packet Error Rate.

Fig. 6a, where for 1 Mbps and $I = 0.01$, the provider does not receive enough bytes in order to have high profit. This proves the inefficiency of the flat rate pricing scheme and the need for more adaptive schemes based on the link data rate.

4. Design principles for VoWIFI optimal capacity allocator and Call Admission Control

A Call Admission Control (CAC) algorithm can be used in order to provide a satisfactory performance to VoIP communications. The performance of the overall system significantly depends on several parameters, which can be recognized (and quantified) at different layers of the protocol stack. The proposed centralized Call Admission Control monitors the status of the overall VoIP system and exploits the metamodel information to provide the proper cross-layer parameter settings to perform run-time optimization of the system. Such a CAC is supported by the knowledge of the utility function (see Section 3) and is implemented at the AP as the central point of the cell where all traffic converges. Therefore, differently from what is known in the CAC literature, the proposed methodology incorporates system parameters to optimally the resources.

Before a new VoIP call is initiated by the mobile node, an ADD Traffic Specification (ADDTS) request is sent to reserve network resources (it is assumed that CAC requests and CAC replies are signaled according to the IEEE 802.11e specification, Chen et al. [5]). Nominal MAC Service Data Unit (MSDU) size, minimum and maximum service intervals, data rate, delay bound, and other service specific parameters specified by the TSPEC field of the ADDTS request are obtained from the application based on VoIP codec parameters.

For a new VoIP call request to be accepted, the CAC module replies with an ADDTS response; and after its reception, the mobile node can start the VoIP data flow. Otherwise, a negative response is sent and the VoIP call must be dropped at the mobile node. Whenever the CAC module needs to specify any of the mobile nodes to change its transmission parameters, such as the maximum number of retransmissions configured at the link layer, it encapsulates these requests into beacon frames periodically broadcast by the Access Point.

The proposed CAC is composed of two stages, as shown on Fig. 7. In the first stage the metamodel generates the maximum number of stations $N(t)$ (assuming that each station generates a single VoIP call each time t), taking into account QoS parameters such as that the end-to-end delay is lower than 100 ms and FLR <5%. Note that in this part we add a discrete time dimension t to the output of the metamodel (see Eq. (1)), in order to study the CAC performance over time. If the number of current calls in the system plus the incoming call is smaller than $N(t)$, then the incoming call is accepted. However if the opposite happens the CAC sets new parameters (either increase R or I) and calculates a new $N(t)$. In stage 2, the Phase II of the CAC can be used as a separate module (as shown on Fig. 7) or with the combination of a Revenue Based Admission Control (RBAC). This acts as a first check of whether the new call will result in better profit for the provider. If the new call gives more profit to the provider, it is moved to the second phase and into the Call Admission Control module. The CAC can be implemented independently, if the provider does not wish to implement a profit based allocation mechanism. We use the diode notation to showcase that a module can be enabled/disabled.

Initially $R = R_{min}$ and $I = I_{min}$ are chosen, in order to achieve the lowest possible end-to-end delay. With such parameters, $N(t)$ is calculated in the Stage 1 metamodel. All incoming VoIP calls are accepted until $N(t)$ is reached. If more VoIP calls need

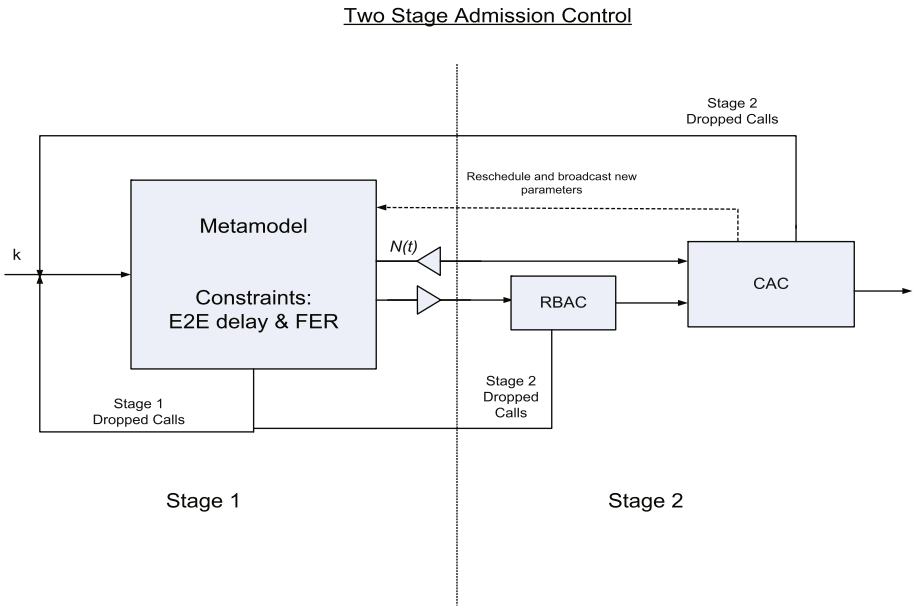

Fig. 7. Two stage admission control modular design.

to access the network, the CAC chooses either to increase the retransmission limit or the voice packet interval. A change in those parameters will result in an increase in the maximum allowed calls in the system $N(t)$. However, before changing the parameters, the RBAC will determine if it is worth to accept the call (increase the profit) or reject it. After it passes that criteria, the choice between changing R or I, is based on which one will allow more VoIP calls ($N(t)$ with respect to both R and I is concave) and whether they have reached the upper bound. We specify R to be the first choice for optimization, since its modification requires less overhead in the system and will lead to faster optimization. The $N(t+1)$ is then calculated with new parameters. If after that calculation the number of calls is smaller than $N(t+1)$ then it is accepted. Otherwise the CAC runs

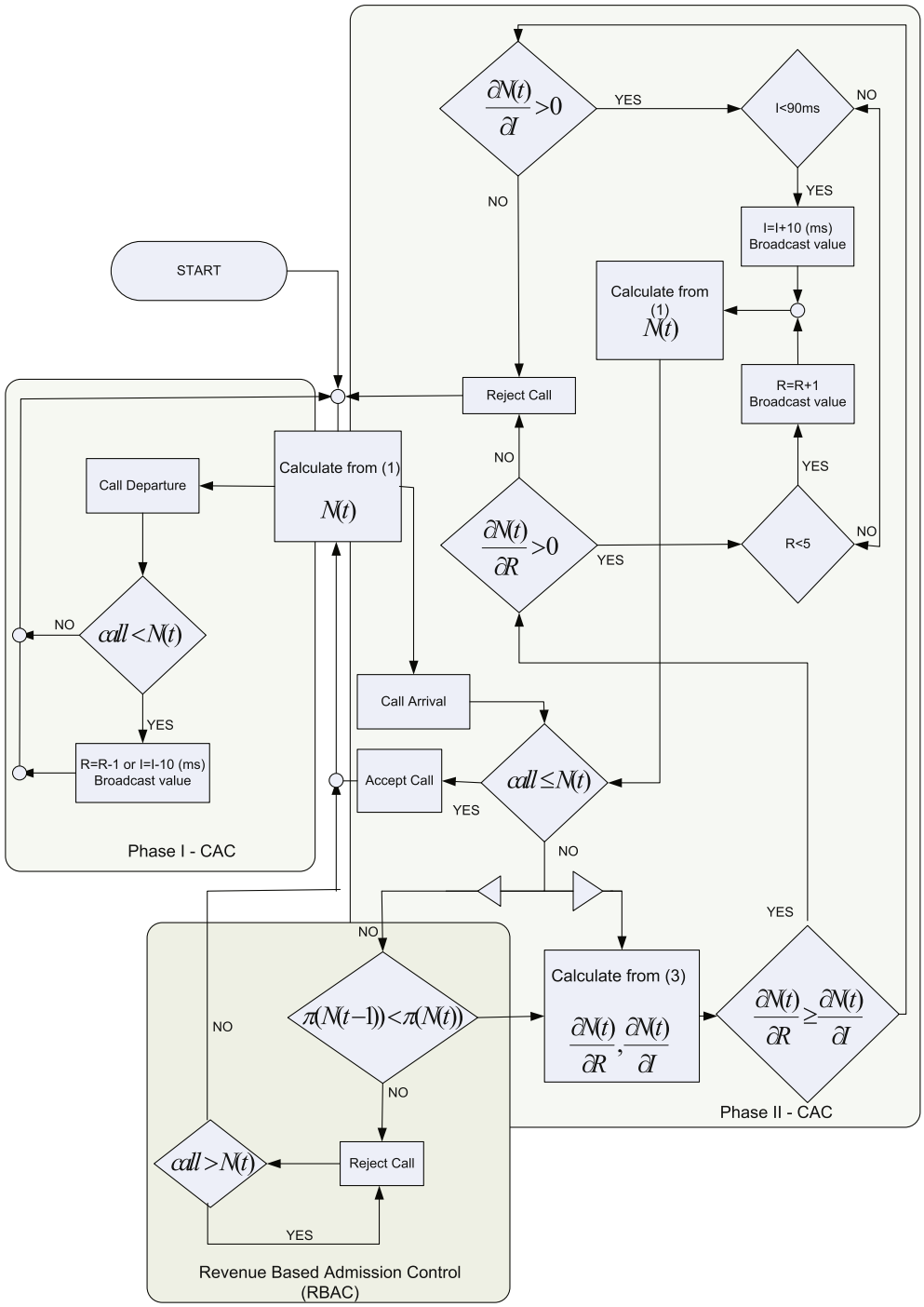

Fig. 8. Flow chart of the algorithm: optimal capacity allocator with a separate module for RBAC.

again to determine a new $N(t+2)$. On the other hand, if a call leaves the system and the number of calls in the system is less than $N(t)$, then the spare resources are removed from the system and $N(t)$ is decreased. The delay is minimized for the rest of the pre-existing calls.

In Fig. 8 the main parts of the algorithms are shown in a flow chart. Note that the metamodel parameters D and PER are always available at the AP as a property of a shared medium provided by the IEEE 802.11 standard. Parameter I is obtained from an ADDTS request, while the maximum number of retransmissions R is configured by the AP using an ADDTS response.

In order to study the performance of the CAC, we performed validating simulations. The VoIP call arrivals were assumed to follow an exponential distribution with variable mean and the call duration was assumed to follow an exponential distribution with $\mu = 180$, Song et al. [25]. Our detailed simulation showed that the results are not affected by the distribution of the mean call arrivals or the call duration time. In Fig. 9, the number of calls rejected or serviced is shown by varying the inter-arrival time of the calls. As it can be observed during high congestion (low values of mean interarrival time) and $D = 11$ Mbps, more calls are being serviced than rejected. However, in case of 5.5 Mbps more calls are being rejected because the output of the metamodel $N(t)$ is smaller (as derived by Eq. (1)). Therefore, calls are being rejected effectively based on the available resources.

In Fig. 10, $N(t)$ is shown to follow the pattern of the incoming calls, which proves that CAC may dynamically change the parameters based on the arrival patterns. More specifically, the mean absolute percentage of difference between the number of incoming calls and the output of the meta-model $N(t)$ was small, therefore, most of the resources are fully utilized. Moreover, as it has been shown in the optimization formulated in Eq. (2), the CAC satisfies the requirement of not accepting more than 20 calls at any time. Finally one other issue is what happens with the calls that are being rejected. In this case the customers may easily try at a later time and get access to the network. For example, in Fig. 10 the spikes have a length of at most a second; therefore the waiting time after a rejection will not be noticed by the end-user.

5. Backround work

Cross-layer design derives from the observation that the performance of a network or other system depends on several mechanisms situated at different levels of the protocol stack interacting in a complex fashion Toumpis and Goldsmith [26]; Pollin et al. [22]; Chen et al. [5]; Lin and Shroff [15]. Vadde and Syrotiuk [27] studied the impact of different layers in order to optimize service delivery in mobile ad-hoc networks, while Granelli and Devetsikiotis [10]; Hui and Devetsikiotis [12] introduced a metamodeling approach to study cross-layer scheduling in wireless local area networks. Nevertheless, little formal characterization of the cross-layer interaction among different levels of the protocol stack is available yet. So a clear need has emerged for identifying approaches able to analyze and provide quantitative guidelines for the design of cross-layer solutions and, even more important, to decide whether cross-layering represents an effective solution or not. In this work we propose a way of quantifying those interactions through an RSM polynomial and study the effects of modifying the system parameters. We optimize over several QoS constraints and determine whether the benefits for the network provider outperforms the cost of the layer violation.

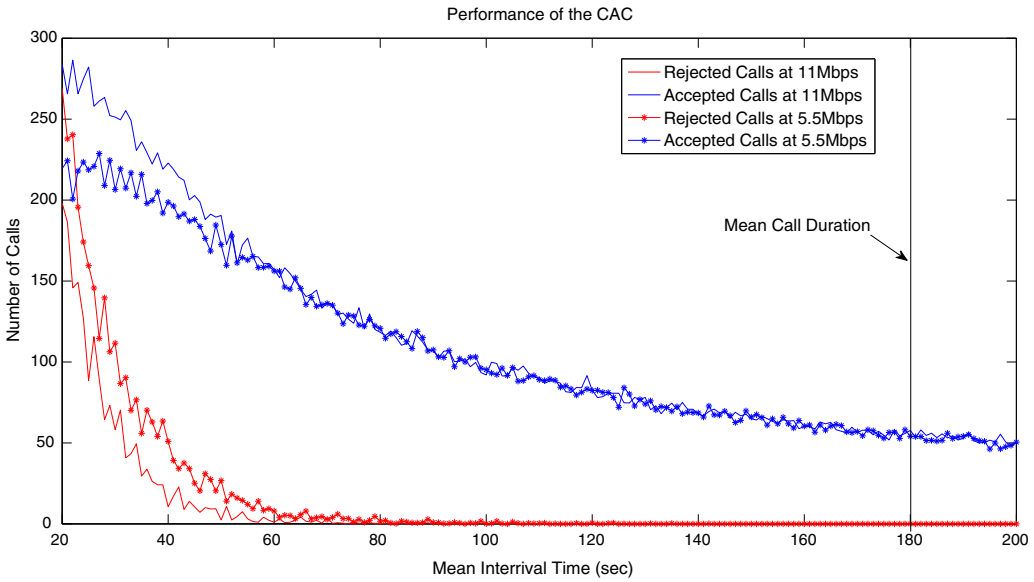

Fig. 9. Performance of the system as a function of mean interarrival time.

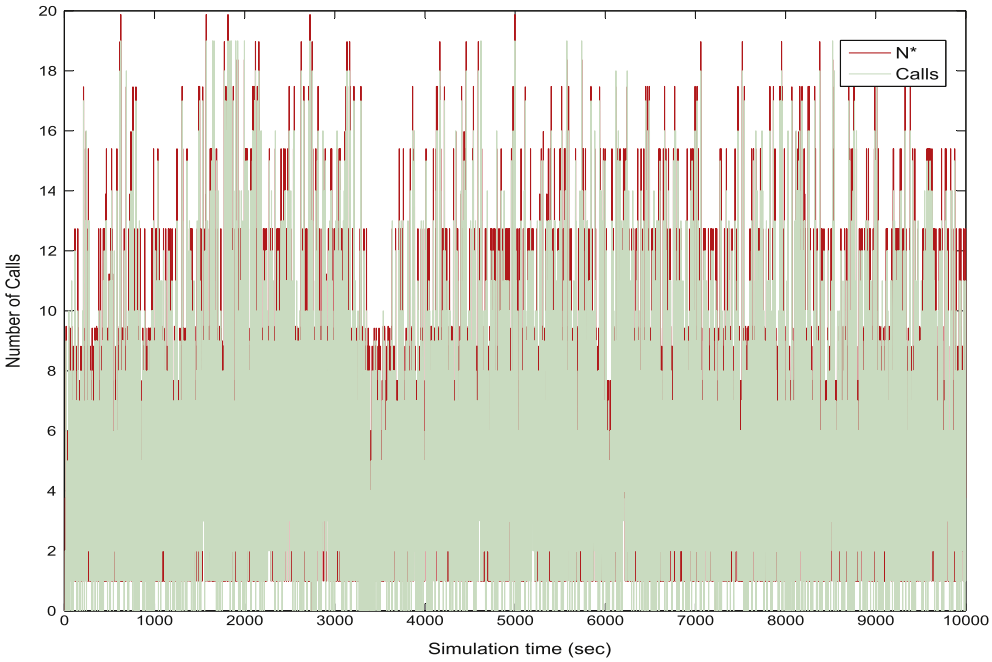

Fig. 10. Dynamic performance of the Call Admission Control.

Several other works have focused [16,2,21,28,23,11,8,7]. Those schemes can be classified into the following classes based on the design principle, Ahmed [1]: (A) centralized or distributed schemes based on the decision making principle; (B) complete or local knowledge schemes; (C) single or multiple services/classes support schemes; (D) proactive or reactive based on the type of QoS and performance analysis; and (E) schemes for uplink, downlink, or both. Despite the design principles classified above, most of the available CAC schemes limit the access of real-time traffic flows into the network based on predefined criteria, typically optimizing the QoS of the flows taking into consideration network load, signal quality, level of interference, terminal power resources, and other parameters. Thus differently from what is known in the literature, we propose a CAC whose decision process is correlated to the system model, and the modifiable parameters are not only from a single layer, but from multiple layers.

6. Conclusion

This paper conducts a detailed quantitative study of cross-layer performance optimization applied to a Voice over WiFi scenario. The proposed methodology enables us to analyze and quantify interlayer dependencies, and to identify the optimal operating point of the system. We call this second level of modeling as *metamodeling*. Based on the results of the metamodel, a profit based utility function is proposed, and the parameters that lead to an optimum profit operating point are identified.

By taking into account those two optimal points (system and profit), we propose a two stage Call Admission Control scheme. One that will certify that the performance of the system is under specific SLAs, and an additional module that will certify that the service provider is getting the maximum profit at the current system configurations.

References

[1] M. Ahmed, Call admission control in wireless networks: a comprehensive survey, IEEE Communications Surveys and Tutorials 7 (2005) 49–68.
[2] F. Anjum, M. Elaoud, D. Famlari, A. Ghosh, R. Vaidyantathan, A. Dutta, R. Agrawal, Voice performance in WLAN networks – an experimental study, in: Proc. of IEEE GLOBECOM, 2003, pp. 3504–3508.
[3] G. Box, B. Draper, Empirical Model Building and Response Surfaces, Wiley Publications, 1987.
[4] C. Brouzioutis, V. Vitsas, P. Chatzimisios, Studying the impact of data traffic on voice capacity in IEEE 802.11 WLANs, in: Proc. of IEEE ICC. IEEE, 2010, pp. 1–6.
[5] L. Chen, S. Low, J. Doyle, Joint congestion control and media access control design for ad hoc wireless networks, in: Proc. of IEEE INFOCOM, 2005, pp. 2212–2222.
[6] L. Ding, R. Goubran, Speech quality prediction in VoIP using the extended E-model, in: Proc. of IEEE GLOBECOM, 2003, pp. 3974–3978.
[7] M. Ergen, P. Varaiya, Throughput analysis and admission control for IEEE 802.11a, Springer MONET Special Issue on WLAN optimization at the MAC and Network Levels 10 (2005) 705–716.
[8] Y. Fang, Y. Zhang, Call admission control schemes and performance analysis in wireless mobile networks, IEEE Transactions on Vehicular Technology 51 (2002) 371–382.
[9] P. Gill, W. Murray, M. Wright, Practical Optimization, Academic Press, London, 1981.

[10] F. Granelli, M. Devetsikiotis, Designing cross-layering solutions for wireless networks: a general framework and its application to a voice-over-wifi scenario, in: Proc. of IEEE CAMAD, 2006, pp. 1–7.
[11] D. Hole, F. Tobagi, Capacity of an IEEE 802.11b wireless LAN supporting VoIP, in: Proc. of IEEE ICC, 2004, pp. 196–201.
[12] J. Hui, M. Devetsikiotis, The use of metamodeling for voip over wifi capacity evaluation, IEEE Transaction of Wireless Communications 7 (2008) 1–5.
[13] JMP, 2010. JMP Desktop Statistical Software from Sas. <http://www.jmp.com>.
[14] P. Kleijnen, Experimental design for sensitivity analysis, optimization, and validation of simulation models, Handbook of Simulations, Wiley Publications, 1998.
[15] X. Lin, N. Shroff, The impact of imperfect scheduling on cross layer rate control in wireless networks, in: Proc. of IEEE INFOCOM, vol. 2, 2005, pp. 302–315.
[16] P. Medepalli, P. Gopolaksrishnan, D. Famolari, T. Kodama, Voice capacity of IEEE 802.11b, 802.11a and 802.11g wireless LANs, in: Proc. of IEEE GLOBECOM, 2004, pp. 1549–1553.
[17] Q. Ni, T. Li, T. Turletti, Y. Xiao, Saturation throughput analysis of error-prone 802.11 wireless networks, Wireless Communications and Mobile Computing 5 (8) (2005) 945–956.
[18] NS-2, 2010. <http://www.isi.edu/nsnam/ns/>.
[19] P.800, I.-T.R., 2003. Methods for Subjective Determination of Speech Quality. International Telecommunication Union.
[20] I. Papapanagiotou, G. Paschos, S. Kotsopoulos, M. Devetsikiotis, Extensions and comparison of QoS enabled wifi models in the presence of errors, in: Proc. of GLOBECOM, 2007, pp. 2530–2535.
[21] G.S. Paschos, I. Papapanagiotou, E. Vagenas, S. Kotsopoulos, Performance analysis of a new admission control for 802.11 WLANnetworks in ad hoc mode, in: Proc. of CSNDSP, 2006, pp. 215–219.
[22] S. Pollin, B. Bougard, G. Lenoir, L. Van Der Perre, B. Van Poucke, F. Cathoor, I. Moerman, Cross-layer exploration of link adaptation in wireless LANs with TCP traffic, in: Proc. of Symposium IEEE Benelux Chapter on Communications and Vehicular Technology, 2003.
[23] D. Pong, T. Moors, Call admission control for IEEE 802.11 contention access mechanism, in: Proc. of IEEE GLOBECOM, 2003, pp. 174–178.
[24] H. Schulzrinne, S. Cashner, R. Frederick, V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, RFC 1889, 1996.
[25] Song, W., ad W. Zhuang, H. J., Shen, X., 2005. Resource management for QoS support in cellular/wlan interworking. IEEE Networks vol. 19, pp. 12–18.
[26] S. Toumpis, A. Goldsmith, Capacity regions for wireless ad hoc networks, IEEE Transactions on Wireless Communications 2 (2003) 736–748.
[27] K. Vadde, V. Syrotiuk, Factor interaction on service delivery in mobile ad hoc networks, IEEE Journal on Selected Areas on Communication 22 (2004) 1335–1346.
[28] W. Wang, S. Liew, V. Li, Solutions to performance problems in VoIP over a 802.11 wireless LAN, IEEE Transactions on Vehicular Technology 54 (2005) 366–384.