Forecasting WiMAX BS Traffic by Statistical Processing in the Wavelet Domain

Cristina Stolojescu[1], Alina Cusnir[2], Sorin Moga[3], Alexandru Isar[1]
[1] Politehnica University, Electronics and Telecommunications Faculty, Communications Department, Timisoara, Romania, c.stolojescu@etc.upt.ro, alexandru.isar@etc.upt.ro,
[2] Alcatel-Lucent, Timisoara, Romania, alina.cusnir@alcatel-lucent.ro,
[3]Sorin Moga, Telecom Bretagne, Brest, France, Sorin.Moga@telecom-bretagne.eu,

Abstract—**The goal of this paper is to adapt a already known traffic forecasting methodology, based on statistical data processing in the field of wavelets, to the case of WiMAX networks, to predict where and when upgrading must take place.**

Keywords-forecasting, WiMAX, traffic, wavelets, statistical model.

I. Introduction

One of the goals of the WiMAX technology is to support advanced IP applications, such as voice, video, and multimedia. A fundamental question about this technology is: How do the technologies compare in terms of prioritizing traffic and controlling quality?

The WiMAX media access control layer is built from the ground up to support a variety of traffic mixes, including real-time and non-real-time constant bit rate and variable bit rate traffic, prioritized data, and best-effort data. A partial response to the last question can be given by studying the traffic forecasting methodology for WiMAX.

Tables 7 and 8 in [1] simply describe the capacity of an 802.16 base station (BS) for various channel bandwidths and coding/modulation schemes. By assuming a deployment scenario – e.g., available bandwidth and MHz per cell, distribution of various user types, and application breakdown - it is possible to calculate the total traffic volume of a BS.

Few months ago a new WiMAX network was installed by Alcatel-Lucent. At each BS the traffic was measured and its evolution on 11 weeks was recorded. It can be observed, analyzing these traces, that the BS traffic exhibits visible long term trends, strong periodicities and variability at multiple time scales (see for example figure 3). This variability determines some of the BS to be more exposed at the risk of saturation than others. So, the points where and the moments when future upgrades are necessary must be identified. This identification can be realized based on the traffic forecasting at the level of each BS.

So far the best practice in the area of network capacity planning is based on the experience and the intuition of the network operators. Moreover, it usually relies on marketing information regarding the projected number of customers at different locations within the network. Being given provider-specific oversubscription ratios and traffic assumptions, the operators estimate the effect that the additional customers may have on the network-wide load. The points where link upgrades will take place are selected based on experience, and/or current network state. For instance BSs that currently carry larger volumes of traffic are likely to be upgraded first. **Our goal is to enhance the above practices using historical network measurements for each BS. The intuition behind this approach is to use mathematical tools to process historical information and to extract trends in the traffic evolution at different time scales. This approach requires the collection of network measurements over long periods of time.**

We introduce a methodology to predict **when and where** upgrades have to take place in a WiMAX network. Using statistics, collected continuously, we compute an aggregate demand for each BS and we look at its evolution at time scales larger than one hour. Our methodology, inspired by [2] relies on the wavelet multiresolution analysis and linear time series models. We model the BS aggregate demand as a multiple AutoRegressive Integrated Moving Average (ARIMA) model. We show that forecasting the long term trend and the fluctuations of the traffic yields accurate estimates for the future. The “capacity planning” process consists of many tasks, such as addition or upgrade of specific nodes, addition of BSs, and expansion of already existing BSs. For the purposes of this work, we use the term “capacity planning” only to refer to the process of upgrading or adding BSs in the core of a WiMAX network. In this work, we use historical information collected continuously during a period of 11 weeks between March 17 and June 02, 2008 at the level of each BS composing a WiMAX network.

The structure of this paper is the following. In Section II we described our data. Section III presents some initial observations. Section IV highlights how the wavelets theory can be used to extract the overall trend of the traffic. In section V we propose the traffic statistical model. The aim of the section VI is to present the proposed traffic forecasting method. Finally, Section VII is dedicated to some concluding remarks.

II. Data Structure

We collected values for two particular MIB (Management Information Base) objects, incoming (uplink) and outgoing (downlink). These values represent the amount of traffic measured in bytes or in packets, for all the links of all the BSs in the WiMAX network during a period that spans from March 17 until June 02, 2008. The data is collected with a period of

15 minutes. So, the BS capacity can be expressed in bytes/s or in packets/s. These values are organized in eleven xls files, corresponding to a particular week, meaning that the first file corresponds to the first of the eleven weeks and the last file corresponds to the last week. Each of these files contains information about 64 BSs. So we deal with 64 traces 11 weeks long.

III. Initial Observations

A traffic curve with the duration of one week arbitrarily selected is represented in figure 1. The curve contains large spikes and valleys indicating periodicities in the traffic. In order to verify the existence of those periodicities we calculate the Fourier transform of the signal. In figure 2 is represented the power spectral density of the signal represented in figure 1. We have represented only 64 harmonics, starting with the third one. We have not represented the continuous component or the fundamental because we wanted to see more clearly the relative strength of the harmonics. It can be remarked the eighth harmonic. The sampling step used has a value of 15 minutes. It corresponds to a sampling frequency of 1.1 mHz. So, the maximal frequency contained in the analyzed power spectral density equals 0.55 mHz. The representation contains 670 values. Hence the fundamental frequency of the representation equals 0.55 /335 mHz. The frequency of the eighth harmonic equals 0.013 mHz. The corresponding period is equal with 76923.07 s, or 1282.05 minutes or 21.36 hours (near 24 hours). Our results indicate that the most dominant period across all traces is the 24 hour one. However, depending on the trace, such periods may not even be present. Analyzing the first week of the considered period for all the 64 BSs we have found a periodicity of 24 hours in 77% of cases.

Let us analyze now a traffic curve with the duration of eleven weeks (between March 17 and June 02, 2008) arbitrarily selected. It corresponds to a particular BS (randomly selected) in uplink and it is represented in figure 3.

The curve contains specific underlying overall trends. All the following examples will be related to the traffic represented in figure 3.

The overall trends were extracted with the aid of the methodology proposed in [2], based on multi-timescale analysis.

IV. Traffic Overall Trend Extraction

The multi-resolution analysis (MRA) is a signal processing technique that takes into account the representation of a signal $s(t)$ at multiple time resolutions. When the original signal $s(t)$ is involved, the maximal resolution is exploited. When a variant of the original signal, $s(2t)$, is used then a poorer resolution is exploited. Combining several analyses realized at different resolutions, a MRA is obtained. The temporal resolution of the traffic curves represented in figure 3 is of 15 minutes. By temporal decimation with a factor of 6 these time series can be transformed in signals at a temporal resolution of 1.5 hours.

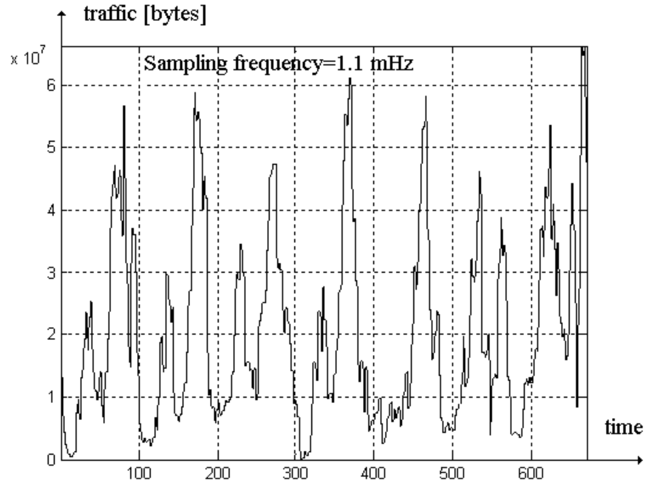

Figure 1. A curve describing the traffic evolution for a BS arbitrarily selected.

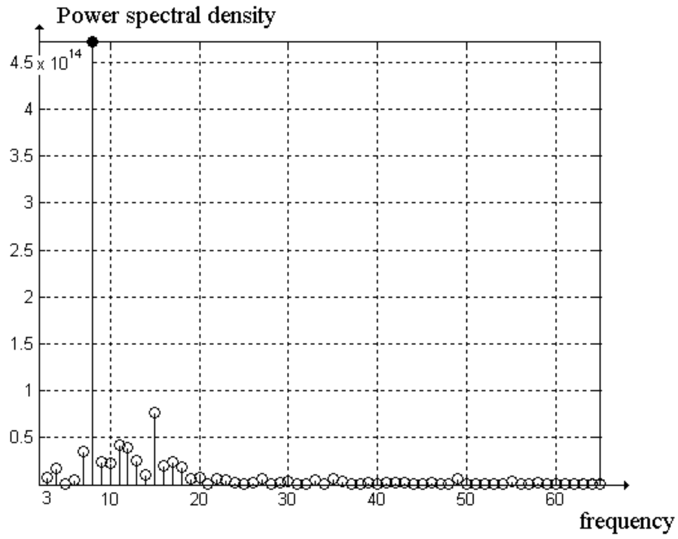

Figure 2. The power spectral density of the signal from figure 1.

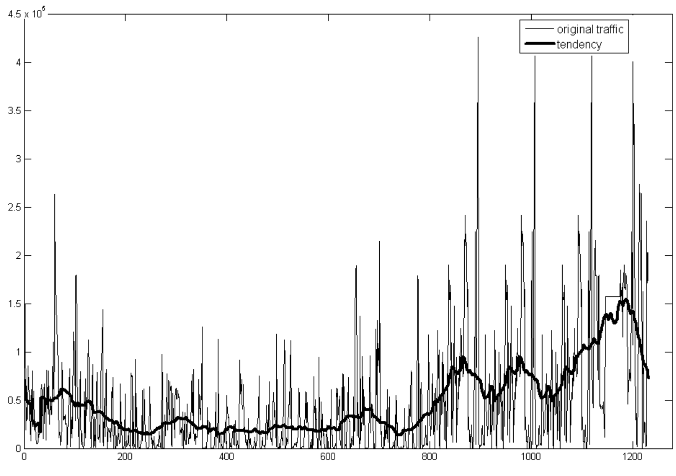

Figure 3. A curve describing the uplink traffic evolution measured in number of packets/s for a BS, arbitrarily selected, during 11 weeks and its overall tendency extracted using wavelets.

In the following these temporal series will be denoted by $ca_{sd}(t)$. The derived temporal series $ca_{sd}(2^p t)$ have a temporal resolution of 2^px1,5 hours. To extract the overall trends of the traffic time series we performed the MRA of the temporal series $ca_{sd}(t)$ using temporal resolutions between 1.5 and 96 hours. At each temporal resolution two categories of coefficients are obtained: approximation coefficients and detail coefficients.

Generally, the MRA are implemented based on the algorithm of Mallat (which corresponds to the computation of the Discrete Wavelet Transform (DWT)). The disadvantage of this algorithm is the decreasing of the length of the coefficient sequences with the increasing of the iteration index due to the use of decimators. Another way to implement a MRA is to use Shensa's algorithm (which corresponds to the computation of the Stationary Wavelet Transform (SWT)). In this case the use of decimators is avoided but at each iteration different low-pass and high-pass filters are used. In this work we used the MRA with the following purposes:

- to extract the overall trend of the temporal series that describes the traffic under analysis with the aid of the approximation coefficients,
- to extract the variability around the overall trend with the aid of some detail coefficients.

Preliminary simulations prove that the overall trend of the traffic time series is better highlighted by the approximation coefficients obtained at the time resolution of 96 hours (corresponding to the sixth decomposition level) c_6 (the smooth line in figure 3). The energy of those coefficients represents 56,4 % of the overall energy that corresponds to the analyzed time series in the considered example. In the following the detail sequences corresponding to time resolutions between 1.5 hours and 96 hours will be denoted by d_1-d_6. The equation describing the proposed MRA is:

$$ca_{sd}(t) = c_6(t) + \sum_{p=1}^{6} d_p(t) \quad (1)$$

Computing the energies of the detail sequences corresponding to our example we have obtained the following results: E_d_1 =7.1028e+011; E_d_2=6.6548e+011; E_d_3=7.3784e+011; E_d_4=6.8544e+011; E_d_5=3.2895e+011; E_d_6 =2.3547e+011. The highest energy is E_d_3 (corresponding to a time resolution of 12 hours). The energy of the coefficients c_6 and d_3 represents 65.25 % of the overall energy of the analyzed time series. The next detail energy value in decreasing order is E_d_4 (corresponding to a time resolution of 24 hours-where the higher periodicities of the time series were observed). The energies of the coefficients c_6, d_3 and d_4 represent 74.83 % of the overall energy of the analyzed time series. Hence, we have decided to keep in our MRA only the details d_3 and d_4:

$$ca_{sd}(t) = c_6(t) + \beta \cdot d_3(t) + \gamma \cdot d_4(t) \quad (2)$$

These details explain the deviation of the time series around its overall trend.

V. Traffic Statistical Model

The model in (2) represents the new statistical model for the traffic time-series which we want to forecast. It reduces the multiple linear regression models in (1) at two parameters only: the overall trend of the traffic (described by c_6) and the variability (described by the detail coefficients d_3 and d_4). For the utilization of the new statistical model, the weights β and γ must be known. First, the contribution of the coefficients d_4 is neglected. So, the new statistical model will be expressed by:

$$ca_{sd}(t) = c_6(t) + \beta \cdot d_3(t) + e(t) \quad (3)$$

The parameter β can be found by the minimization of the mean square of e:

$$\beta_{opt} = \arg\min_{\beta} \left\{ \left\| ca_{sd}(t) - c_6(t) - \beta d_3(t) \right\|^2 \right\} \quad (4)$$

We have implemented this search procedure in Matlab and for our example (the signal represented in figure 3) we have obtained: beta =1.2200 and mean_square_error =1.4071e+006. The search procedure already mentioned can be used for the computation of γ also. This time, the contribution of the coefficients d_4 is taken into account. The new statistical model will be expressed by:

$$ca_{sd}(t) = c_6(t) + \beta_{\text{opt}} \cdot d_3(t) + \gamma \cdot d_4(t) + e(t) \quad (5)$$

The parameter γ can be found by the minimization of the new mean square of e:

$$\gamma_{\text{opt}} = \arg\min_{\gamma} \left\{ \left\| ca_{sd}(t) - c_6(t) - \beta_{\text{opt}} d_3(t) - \gamma d_4(t) \right\|^2 \right\} \quad (6)$$

We have implemented this search procedure in Matlab and for our example we have obtained: gama=0.6500, mean_square_error=1.2959e+006.

VI. Traffic Forecasting Methodology

Constructing a time series model implies expressing X_t in terms of previous observations X_{t-j}, and noise terms Z_t which typically correspond to external events. The noise processes Z_t are assumed to be uncorrelated with a zero mean and finite variance. Such processes are the simplest processes, and are said to have "no memory", since their value at time t is uncorrelated with all the past values up to time $t-1$. A time series X_t is an ARMA(p,q) process if X_t is stationary and if for every t:

$$\varphi(B) X_t = \theta(B) Z_t \quad (7)$$

where φ(·), and θ(·) are p^{th} and q^{th} degree polynomials, and B is the backward shift operator ($B_j X_t = X_{t-j}$, $B_j Z_t = Z_{t-j}$, $j = 0, \pm 1, \cdots$). The ARMA model fitting procedure assumes the data to be stationary. If the time series exhibits variations that violate the stationary assumption, then there are specific approaches that could be used to render the time series stationary. The most common one is the differencing operation. If the non stationary part of a time series is a polynomial function of time, then differencing finitely many times can reduce the time series to an ARMA process. An ARIMA(p,d,q) model is an ARMA(p,q) model that has been differenced d times. Thus it has the form:

$$\varphi(B)(1-B)^d X_t = \theta(B) Z_t, \quad Z_t \sim WN(0,\sigma^2) \qquad (8)$$

If the time series has a non-zero average value through time, then the previous equation also features a constant term μ on its right hand side. Let us denote the terms describing the variance in equation (5) with:

$$dt_3(t) = \beta_{opt} \cdot d_3(t) + \gamma_{opt} \cdot d_4(t) \qquad (9)$$

We use the Box-Jenkins methodology [3] to fit linear time series models, separately for the overall trend and for the variability. Such a procedure involves the following steps:

i) determine the number of differencing operations needed to render the time series stationary,
ii) determine the values of p and q,
iii) estimate the polynomials φ, and θ, and
iv) evaluate how well the derived model fits the data.

The identification of model parameters is done using Maximum Likelihood Estimation. The best model is chosen as the one that provides the smallest AICC, BIC, and **FPE** measures [4], while offering the smallest mean square prediction error four weeks ahead.

Applying the Box-Jenkins methodology for the first difference of the time series c_6 in our example we have obtained an ARIMA(011) model for the overall tendency:

$X(t) = 0.7645X(t-1) + Z(t) - 0.184Z(t-1)$ with μ_{ot} = 4.8138e+004.

The coefficients dt_3 are used to appreciate the variability of the traffic. They are treated following a similar procedure based on the Box-Jenkins methodology. We have obtained an ARIMA(011) model for the first difference of the variability of the traffic in our example:

$X(t) = 0.4668X(t-1) + Z(t) + 0.0691Z(t-1)$ with μ_d =214.6956.

The computed models for the long term trend c_6 indicate that the first difference of those time series (i.e. the time series of their changes) is consistent with a simple MA model with one or two terms (i.e, $d = 1$, $q = 1$ or $d = 1$, $q = 2$), plus a constant value μ_{ot}. The need for one differencing operation at lag 1, and the existence of term μ_{ot} across the model indicate that the long-term trend is a simple exponential smoothing with growth. The trajectory for the long-term forecasts will be a sloping line, whose slope is equal to μ_{ot}. For the trace in our model the long term forecast will correspond to a weekly increase of 48.138 kb. This forecast corresponds to the considered BS. Applying the Box-Jenkins methodology on the deviation measurements (that reflect the variability of the traffic), we see that for some traces the deviation dt_3 can be expressed with simple MA processes after one differencing operation. Therefore, for the considered example, the deviation increases with time at a rate of 0.2 kb per week (two orders of magnitude smaller than the increase in their long term trends).

Given the estimates of μ_{ot} across all models corresponding to all BSs we can conclude that all traces exhibit upward trends, but grow at different rates. In the following table are presented the BSs with the higher values of μ_{ot} in downlink (higher then 10^8). These BSs represent those with the highest saturation risk from the considered WiMAX network.

TABLE I. THE LIST OF BSS WITH THE HIGHEST RISK OF SATURATION

BS	μ_{ot}	BS	μ_{ot}	BS	μ_{ot}
61	1.7382e+008	49	1.4356e+008	57	1.3163e+008
63	1.6927e+008	59	1.3570e+008	51	1.2621e+008
62	1.5023e+008	3	1.3335e+008	54	1.1512e+008
48	1.4790e+008	58	1.3215e+008	4	1.0655e+008
				56	1.0535e+008

Using these values, the corresponding weekly increases can be found and the moment when a traffic threshold (of 80% for example) is attained can be estimated. This is the moment when the considered BS (whose location is known) must be upgraded.

We cannot come up with a single WiMAX network-wide forecasting model for the aggregate demand.

VII. CONCLUSION

We have proved that the traffic forecasting methodology proposed in [2] can be adapted for the WiMAX traffic prediction. This methodology extracts the traffic trends from historical measurements and can identify the BSs which exhibit higher growth rates and thus may require additional capacity in the future. It is capable of isolating the overall long term trend and identifying those components that significantly contribute to its variability. Predictions based on approximations of those components provide accurate estimates with a minimal computational overhead. All our forecasts were obtained in seconds. All the procedures described are implemented as Matlab functions. We have found that the BSs of the considered network are more charged in the downlink cycles then in the uplink cycles. This unbalanced comportment gives some indications about the needs of the users of the considered WiMAX network and its analysis can give useful information regarding the projected number of customers at different locations within the network for the local operators. As further research directions we intend to search forecasting WiMAX BS traffic methodologies based on data mining able to make the prediction based on a reduced volume of data.

REFERENCES

[1] IEEE L802.16-04/37r2, “Response to ITU-R WP 8F Questionnaire on the Services and Market for the Future Development of IMT-2000 and Systems Beyond IMT-2000”, ITU, Radio communications Study Group, December 2004,

[2] Konstantina Papagiannaki, Nina Taft, Zi-Li Zhang, Cristophe Diot, “Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models”, *IEEE Infocom*. San Francisco. March 2003.

[3] Box-Jenkins Methodology, http://web.ntpu.edu.tw/~tsair/1Teaching/1semester/TimeSeriesUnder/Box.doc

[4] Matlab toolbox System Identification, the function BJ.m.