A FAST AND EFFICIENT MODEL OF AN MPEG-4 VIDEO TRAFFIC BASED ON PHASE SPACE LINEARISED DECOMPOSITION

Agnieszka Chodorek
Department of Telecommunications and Photonics
Kielce University of Technology
al. Tysiąclecia Państwa Polskiego 7
25-314 Kielce
Poland
E-mail: eweach@tu.kielce.pl

KEYWORDS
traffic characterisation, linear modelling, regression analysis, phase space, MPEG-4.

ABSTRACT

The common practice in the MPEG video traffic modelling is to split video sequence into I, P and B subsequences and to model the traffic (generated by each subsequence) separately. This frame-oriented decomposition both allows avoiding non-homogeneity of modelling data and lead to linearisation of MPEG video traffic characteristics.
In this paper, the low-complexity regression model of an MPEG-4 video traffic is proposed. The model combines regression analysis and a new type of linearised decomposition, based on partitioning of the phase space into sub-regions. The model was tested as traffic predictor (for network management purposes) and traffic generator (for simulation purposes). Experimental results show good accuracy of the proposed modelling scheme.

INTRODUCTION

In modern telecommunication networks, transmission of multimedia stream has been one of the main targets of researches. Methods of video and audio signal compression are defined, among others, by MPEG standard. The family of MPEG video coding standards includes MPEG-1, MPEG-2, and MPEG-4. Many modern Internet applications use MPEG standard for video signal coding. Traffic generated by such applications (MPEG video traffic), observed at the video frame level, is characterised by high burstiness, non-homogeneity and strong periodicity (Rose 1995; Chodorek and Chodorek 2000; Fitzek and Reisslein 2001).
The common practice in the modelling of the frame-level traffic generated by MPEG video sources is to split video sequence into I, P and B subsequences and model each subsequence separately. This frame-oriented decomposition allows avoiding non-homogeneity of modelling data. From the other side, frame-oriented decomposition lead to linearisation of MPEG video traffic characteristics and, as result, it allows utilisation of linear methods of traffic modelling.
In the paper (Chodorek and Chodorek 2002) an alternative method of linearised decomposition has been proposed - decomposition related to regions in phase space. The linearised decomposition in phase space has been successfully used for construction of low-complexity, fast and effective predictors of an MPEG-1 and MPEG-2 video traffic. The aim of the paper is to prove that the linearised decomposition in phase space can be also used for MPEG-4 (Ebrahimi and Horne 2000) modelling purposes.
The paper is organised as follows. Section two contains a brief overview of an MPEG-4 video encoding. Section three describes phase space analysis of MPEG video traffic, while section four proposes an MPEG-4 video traffic model, based on above analysis. Section five addresses the evaluation of model accuracy and compares it to the model based on the traditional, frame-oriented decomposition of MPEG traffic. Section six summarizes our experiences.

OVERWIEW OF AN MPEG-4 VIDEO ENCODING

MPEG-4 (Ebrahimi and Horne 2000; Fitzek and Reisslein 2001) is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group). MPEG-4 is intended for interactive multimedia, especially for video conferencing, Internet distribution and similar applications using low bandwidths with a target rate up to 4 Mbps.
The MPEG-4 video coding is object-based (in contrast to the frame-based MPEG-1/MPEG-2 video coding). Each video frame covers a number of arbitrarily shaped image regions (so-called Video Object Planes, VOP), e.g. describing physical objects. A conventional video frame can be represented by a VOP with rectangular shape (Ebrahimi and Horne 2000).
The MPEG-4 video compression algorithm, as the other Moving Picture Expert Group video compression algorithms, reduces both the spatial and the temporal redundancy of the video data stream. The spatial redundancies are reduced by transforms and entropy coding while temporal ones are reduced by prediction of future VOPs based on motion vectors. As a result, three types of VOPs are obtained:

- I-VOP, encoded independently of any other VOP,
- P-VOP, encoded using motion compensated forward prediction,
- B-VOP, encoded with motion compensated bi-directional (forward and backward) prediction.

After coding, VOPs can be formed, optionally, in a deterministic periodic sequence, called Group of Video Object Planes (GOV). Because of its similarity to the MPEG-1/MPEG-2 Group of Pictures (GOP), the sequence is sometimes referred as GOP (Fitzek and Reisslein 2001). The GOV pattern is usually described by *I*-to-*I* VOP distance N (GOV size) and *I*-to-*P* VOP distance M.

Proceedings 14th European Simulation Symposium
A. Verbraeck, W. Krug, eds. (c) SCS Europe BVBA, 2002

LINEARISED DECOMPOSITION OF AN MPEG-4 VIDEO TRAFFIC IN THE PHASE-SPACE

In this section the phase space analysis of MPEG-4 video traffic is presented. Special emphasis is placed on the linearised decomposition of the MPEG-4 video traffic according to sub-regions in phase space.

The Phase Space

The term "phase space" was introduced by H.Poincaré and can be understood as an n-dimensional space, which coordinate axes are specified by n variables $x_1, x_2, \ldots, x_n$ (continuous or discrete). Relationships between discrete variables are given by the following expression:

$$x_1(k),\ x_2 = x_1(k+1), x_3 = x_2(k+1), \ldots, x_n = x_{n-1}(k+1) \qquad (1)$$

where:

$x_1(k), x_2(k), x_3(k), \ldots, x_n(k)$ – discrete variable,

k – time step.

Phase space analysis describes the behaviour of the system in dependence of time. Values of $x_1, x_2, \ldots, x_n$ variables, evaluated for given time τ, define a point q in the phase space, which represents an instantaneous state of the system. These time-dependent points, plotted sequentially in the n-dimensional phase space, construct so-called phase-space trajectory.

Phase space approach was introduced to time series analysis by several authors (e.g. Packard et al. 1980).

Let $X = \{x(k), k = 1,2,\ldots\}$ be a time series, obtained by a sampling of a single variable of a dynamical system over time. Trajectories of the series in an m-dimensional phase space are defined by expressing each point q in the phase space by an embedding vector, created by treating each time-dependent coordinate of the point q as a component of a vector $\vec{v}$. The vector $\vec{v}$ depends on two parameters: embedded lag j and embedding dimension m.

Phase Space Analysis of an MPEG-4 Video Traffic

Time series of the MPEG video traffic were embedded in phase space described by following parameters: j=1, m=3. Embedding vector is formed as:

$$\vec{v}(k) = \begin{bmatrix} x(k+1) \\ x(k) \\ x(k-1) \end{bmatrix} \qquad (2)$$

Components of embedding vector represent the next, the current, and the previous VOP, respectively. Geometrical properties of phase space trajectory describe the evolution of the state of the system (represented by given time series) over time. Trajectories of the MPEG-4 video traffic embedded in phase space shows that points create an object confined over time to a sub-region of the phase space.

Against the different encoding techniques, phase space trajectories of the MPEG-4 video traffic show the same properties, as previously observed in the case of the MPEG-1 and MPEG-2 video traffic (Chodorek and Chodorek 2002). Above properties can be described as follows:

- Point collections are visually observed as 6 clusters.
- Phase space trajectory shows global nonlinearity.
- Each cluster shows pseudo-linear properties.

The local linearity, observed as the pseudo-linear character of six data clusters, lead to linearised decomposition of the MPEG-4 video traffic according to sub-regions in phase space.

Linearised decomposition of an MPEG video traffic

Linearised decomposition of an MPEG-4 video traffic consists in partitioning of the phase space into six regions, associated with the 6 clusters mentioned above.

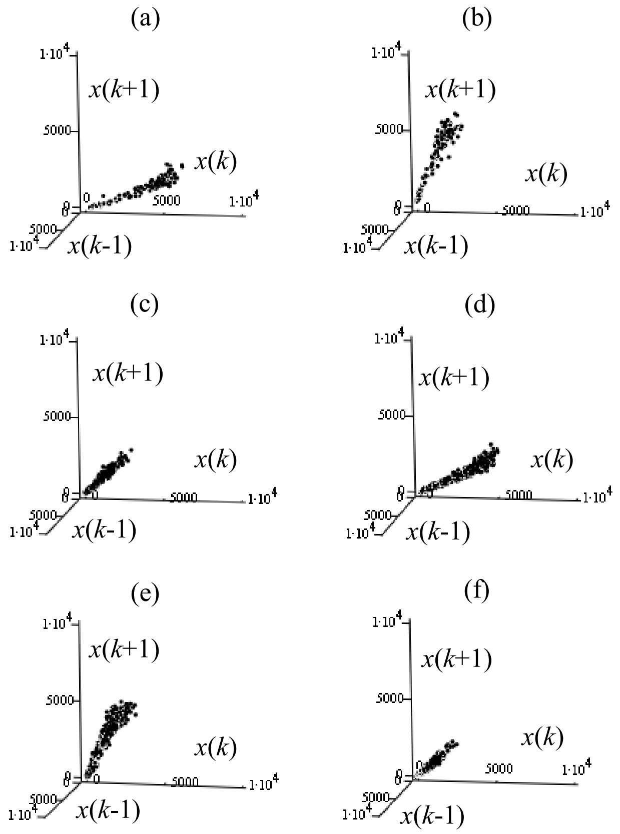

Figure 1: An Example of the Phase Space Decomposition of an MPEG-4 Video Stream: a) Cluster S_1, b) Cluster S_2, c) Cluster S_3, d) Cluster S_4, e) Cluster S_5, f) Cluster S_6.

The algorithm of linearised decomposition of MPEG-4 video traffic characterised by GOV pattern M=3, N=12, can be described as follows:

- If the current VOP is an I-VOP, then the point $q(k)$ will belong to cluster S_1. According to established GOV structure, the next and the previous VOP will be Bi-directional (Figure 1a).
- If the current VOP is a B-VOP and the previous VOP is an I-VOP, then the point $q(k)$ will belong to cluster S_2. According to established GOV structure, the next VOP will be Bi-directional (Figure 1b).
- If the current VOP is a B- VOP and the next VOP is a P-VOP, then the point $q(k)$ will belong to cluster S_3.

According to established GOV structure, the previous VOP will be Bi-directional (Figure 1c).

- If the current VOP is a P- VOP then the point $q(k)$ will belong to cluster S_4. According to established GOV structure, the next and the previous VOP will be Bi-directional (Figure 1d).
- If the current VOP is a B- VOP and the previous VOP is a P- VOP, then the point $q(k)$ will belong to cluster S_5. According to established GOV structure, the next VOP will be Bi-directional (Figure 1e).
- If the current VOP is a B-VOP and the next is an I-VOP, then the point $q(k)$ will belong to cluster S_6. According to established GOV structure, the previous VOP will be Bi-directional (Figure 1f).

THE REGRESSION MODEL OF AN MPEG-4 VIDEO TRAFFIC

The proposed MPEG-4 video traffic piecewise linear modeling, based on the linearised decomposition in the phase space, consists in partition the space into 6 regions, each with different linear regressive model. Therefore modelling of a time series representing MPEG-4 video traffic is carried out using six locally linear, sequentially switched models. Each local linear model can by its linear transfer function $f_j(\vec{x})$. Switching between local linear models may be described formally using sequential finite state automaton.

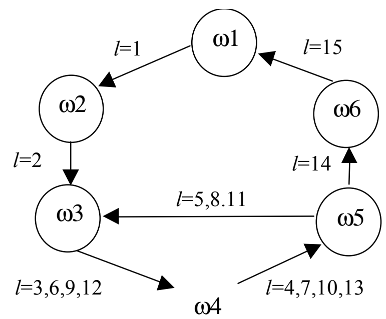

Figure 2: Finite State Automaton of the MPEG Video Traffic Defined for GOV Parameters N=12, M=3 (l - a Sequential Number of a VOP in GOV Structure).

Finite state automaton of the MPEG-4 video traffic (Figure 2) describes movement of the point q over phase space. The automaton consists of 6 states ω1...ω6. Each state ω1...ω6 is assigned to a cluster S_1, S_2, …, S_6, respectively and represents an autonomous sub-model characterized by linear transfer function $f_j(\vec{x})$. The linear transfer function can be described as the linear combination of p values $x_{k-1}, x_{k-2}, \ldots, x_{k-p}$, which models x_k with minimum mean square error.

A p^{th}-order auto-regressive process, denoted AR(p), satisfies

$$x_k = a_0 + a_1 x_{k-1} + a_2 x_{k-2} + \ldots + a_p x_{k-p} + \varepsilon_k , \qquad (3)$$

where $a_1, a_2, \ldots, a_p$ are so-called regression coefficients (real numbers) and ε_k is an regression error.

It is assumed, that time series $\varepsilon(k)$, formed as a sequence of regression errors, is a white Gaussian noise, with expectation zero and finite variance σ_ε^2. Evaluation of regression coefficient requires assumption, that the time series is a zero-mean, stationary series with known auto-correlation function.

The equation (3) defines the linear transfer function, $f_j(\vec{x})$ which describes local linear model. The simplicity of the model (only p summations and p multiplications) implies that it can be easy implemented in software as well as in VLSI circuits.

EXPERIMENTAL RESULTS

The model validation and verification was carried out using 30 empirical MPEG-4 video traces, described in the second section. Empirical MPEG-4 video traces were used for investigations. The Akaike information criterion (AIC) was used to determine the best prediction order, p=12.

The model was tested as a traffic predictor (for network management purposes) and a traffic generator (for simulation purposes).

The Test video sequences

Validation and verification of the model was carried out using empirical MPEG-4 video traces, collected in a publicly available library of VOP size traces of MPEG-4, at the TU Berlin site. Properties of traces are described in (Fitzek and Reisslein 2001).

From the set of above 60 video traces, about the half was chosen for investigations. Number of predicted VOPs, n, should be large enough to assure the obtained results do not depend on dynamics of a single scene. Each video trace, used in the experiment, consists of 40 000 VOPs, grabbed at 25 Hz and represents about 30 min of video. Parameters of a GOV pattern are N = 12 and M = 3.

Table 1: MPEG-4 Video Traces Used for Analysis.

content type	video source	description
slow	cam1	static camera (lecture room)
	cam2	static camera (office)
	cam3	parking security camera
medium	lambs	movie (*Silence Of The Lambs*, 1991)
	bean	TV series (*Mr. Bean*)
dynamic	diehard	movie (*Die Hard III*, 1995)
	startrek	movie (*Star Trek: First Contact*, 1996)
very dynamic	form1	sport event (race in Japan, 2000)
	ski	sport event (Alpin ski, 2000)
	vclips	video clips (2000)

The original sequences are encoded at three different quality levels (low, medium, high) and can be collected into four groups, different in content type (Table 1).

Number of predicted VOPs, n, should be large enough to assure independence of prediction results on dynamics of a single scene. In the paper, MPEG-4 video traces, 40.000 VOPs each, were predicted.

The MPEG-4 Video Traffic Predictor

The accurate prediction denotes, that the value of the predictor's output quantity $\hat{x}_{k+1}$ should be as close as possible (in sense of assumed criterion of goodness) to the size x_{k+1} of the next VOP (generated during the next 40 milliseconds), $\hat{x}_{k+1} \approx x_{k+1}$. Inverse signal-to-noise ratio SNR^{-1} was used to determine predictor accuracy:

$$\mathrm{SNR}^{-1} = \frac{\sum_{k=1}^{n} [\hat{x}_{k+1} - x_{k+1}]^2}{\sum_{k=1}^{n} x_{k+1}^2} = \frac{\sum_{k=1}^{n} \varepsilon_{k+1}^2}{\sum_{k=1}^{n} x_{k+1}^2}, \qquad (4)$$

(where x_{k+1} is the real value of $(k+1)^{\text{th}}$ MPEG VOP size, $\hat{x}_{k+1}$ is the predicted value of $(k+1)^{\text{th}}$ MPEG VOP size and n is the number of predicted VOPs).
Inverse signal-to-noise ratio SNR^{-1} may be treat as relative mean square error. Therefore, the smaller SNR^{-1}, the better the prediction.

Table 2: Quality of MPEG-4 Video Traffic Prediction.

video source	picture quality		
	low	medium	high
bean	0.048	0.054	0.031
cam1	0.0023	0.0039	0.0025
cam2	0.0003	0.0007	0.0010
cam3	0.0003	0.0011	0.0013
diehard	0.045	0.043	0.025
form1	0.027	0.026	0.015
lambs	0.041	0.035	0.019
ski	0.024	0.022	0.015
startrek	0.041	0.040	0.025
vclips	0.055	0.043	0.022

Table 3: The Relative Improvement of MPEG-4 Prediction Accuracy.

video source	low quality	medium quality	high quality
bean	9.2%	12.5%	14.6%
cam1	0.5%	1.3%	20.2%
cam2	11.6%	29.1%	45.2%
cam3	2.1%	36.6%	32.2%
diehard	16.0 %	17.1%	20.7%
form1	17.9%	17.8%	19.4%
lambs	14.6%	15.0%	19.2%
simpsons	14.4%	18.2%	18.9%
ski	18.5%	16.8%	19.0%
startrek	16.2%	20.6%	22.9%
vclips	24.6%	26.1%	29.8%

Quality of MPEG-4 video traffic prediction is summarised in Table 2. The inverse signal-to-noise ratio, SNR^{-1}, ranges from 0,0003 (*cam*1 and *cam*2 video source, low picture quality) to 0,055 (*vclips* video source, low picture quality). The predictor estimates the empirical VOP sizes perfectly for very slow video content (*cam*1, *cam*2, *cam*3). Good prediction was also obtained in the case of prediction of sport events (*form*1, *ski*) characterized by very dynamic content. Therefore it can be assumed that prediction accuracy is larger for more uniform video content, while videos consist of short scenes (as movies or video clip) are less predictable.
Generally, the prediction accuracy is the best in the case of high quality video, while the worst prediction and largest SNR^{-1} was detected for the video traffic generated from low quality stream.
In Table 3 is shown the relative improvement of accuracy of prediction, which can be achieved when the linearised decomposition in the phase space is used instead of typical, VOP-oriented decomposition. Differences in prediction accuracy are largest for most uniform video content (*cam*1, *cam*2, *cam*3, *vclips*). For high quality *cam*2 video source prediction based on phase space decomposition is near two times better (in sense of assumed measure of accuracy) than VOP-oriented linear regressive prediction.

The MPEG-4 video traffic generator

In the paper, a chi-square test will be carried out to determine goodness of fit between observed VOP sizes and VOP sizes received from a model. The null hypothesis is that results obtained from a model fit the observations. Obtained results were also compared visually, using Quantile-Quantile plot (Q-Q plot). Because of high burstiness, suggesting strong selfsimilar behaviour, autocorrelation function of the model and video trace also were compared.
The proposed model estimates the empirical VOP sizes distribution properly, and usually is more accurate than a VOP-oriented regressive model does (Figure 3, Figure 4). In the case of each analysed video trace, an approximation was good enough to pass the chi-square test.

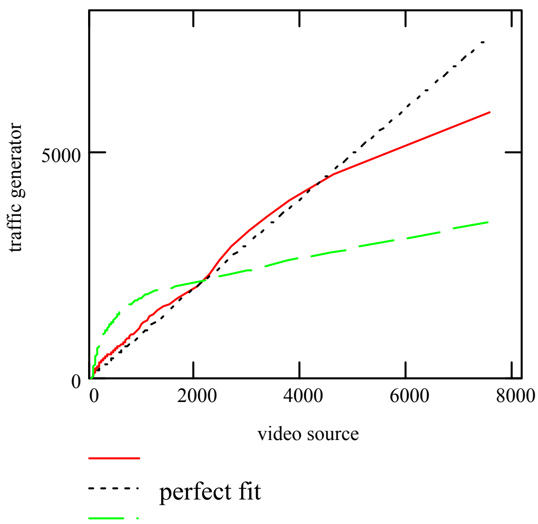

Figure 3: Q-Q plot (*diehard*, low picture quality)

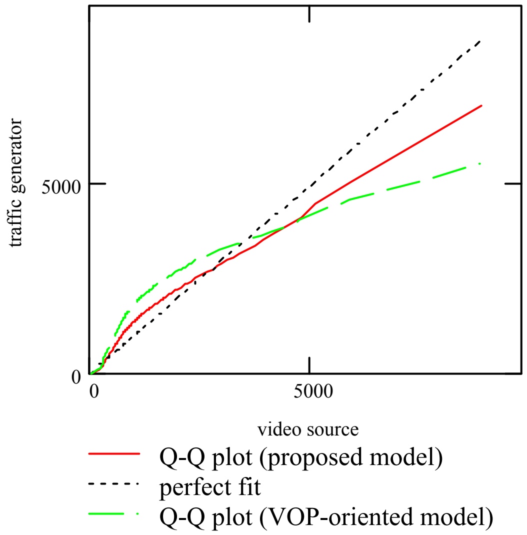

Figure 4: Q-Q plot. (*startrek*, high picture quality).

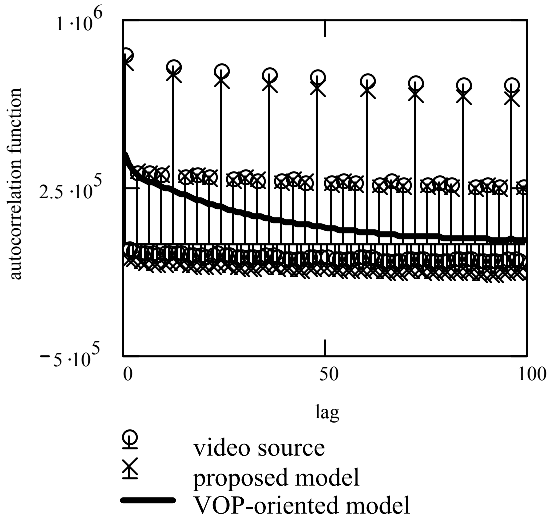

Figure 5 :Autocorrelation function (*diehard*, low quality).

MPEG-4 video traces are superposition of three traces corresponding to three types of VOPs and repeated by the GOV pattern. Due to a cyclic multiplication, the multiplexed stream presents a strong pseudo-periodic property. This property has great impact on a shape of a VOP-level correlation.

The empirical autocorrelation function is approximated better than by the VOP-oriented regressive model (Figure 5). The pseudo-periodic shape of of a VOP-level correlation is always remained.

CONCLUSIONS

Geometrical properties of phase space trajectory describe the evolution of the state of the system (represented by given time series) over time. Trajectories of the MPEG-4 video traffic embedded in phase space are visually observed as 6 pseudo-linear clusters.

An MPEG-4 video traffic piecewise linear model, based on linearised decomposition related to regions in phase space, consists of six locally linear, sequentially switched models. Validation of the model was carried out using about 30 empirical MPEG-4 video traces, collected in a publicly available library of VOP size traces of MPEG-4, at the TU Berlin site. The model was tested as a traffic predictor and traffic generator. Two modeling schemes were compared: the proposed one, based on phase-space decomposition, and the commonly used one, based on VOP-oriented decomposition. Experimental results show good accuracy of the proposed modelling scheme. Proposed model can improve prediction accuracy up to 45%, when act as predictor. Proposed model gives better approximation of empirical distribution and correlation behaviour, when act as traffic generator.

ACKNOWLEDGEMENTS

This research was supported by State Committee for Scientific Research (KBN) under grant 7 T11D 012 20.

REFERENCES

Chodorek, A.; and R.R. Chodorek. 2000. "Characterization of the MPEG-2 video traffic generated by DVD applications". In *Proceedings of. 1st IEEE European Conference. on Universal Multiservice Networks* (Colmar, France, Oct.2-4). IEEE, Piscataway, N.J., 62-70.

Chodorek, A.; and R.R. Chodorek. 2002. "An MPEG-2 video traffic prediction based on phase space analysis and its application to on-line dynamic bandwidth allocation". In *Proceedings of. 2nd IEEE European Conference. on Universal Multiservice Networks* (Colmar, France, Apr.8-10). IEEE, Piscataway, N.J., 44-55.

Ebrahimi, T.; and C. Horne. 2000. "MPEG-4 natural video coding - An overview". *Signal Processing: Image Communication* 15, No.4-5 (Jan), 365-385.

Fitzek, F.H.P. and M. Reisslein. 2001. "MPEG-4 and H.263 Video Traces for Network Performance Evaluation". *IEEE Network* 15, No.6 (Nov./Dec.), 40-54.

Packard, N. H.; J. P. Crutchfield; J. D. Farmer; and R. S. Shaw. 1980. "Geometry from a time series". *Phys. Rev. Lett.* 45, No.9, 712-716.

Rose, O. 1995. "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems". Research Report No.101. Institute of Computer Science, University of Wuerzburg, Germany. (Feb).

AUTHOR BIOGRAPHY

AGNIESZKA CHODOREK was born in Końskie, Poland, in November 1967. She received the M.S. degree in electrical engineering from the Kielce University of Technology, Kielce, Poland, in 1991, and the Ph.D. degree in telecommunications from the AGH University of Technology, Cracow, Poland, in 2001. She is currently an assistant professor at the Kielce University of Technology. Her current research interests lie in the area of telecommunication networks, with emphasis on QoS assurance and traffic modeling.