SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Tibor Csóka

Autoreferát dizertačnej práce

STOCHASTIC BINARY ERROR BURST MODELING IN WIRELESS CHANNELS

STOCHASTICKÉ BINÁRNE MODELY ZHLUKOVEJ CHYBOVOSTI V BEZDRÔTOVÝCH KANÁLOV

na získanie akademickej hodnosti doktor (philosophiae doctor, PhD.)

v doktorandskom študijnom programe: **Telekomunikácie**
v študijnom odbore 5.2.15 Telekomunikácie

Miesto a dátum: Bratislava, 31.5.2016

SLOVENSKÁ TECHNICKÁ UNIVERZITA
V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Tibor Csóka

Autoreferát dizertačnej práce

STOCHASTIC BINARY ERROR BURST MODELING IN WIRELESS CHANNELS

STOCHASTICKÉ BINÁRNE MODELY ZHLUKOVEJ CHYBOVOSTI V BEZDRÔTOVÝCH KANÁLOV

na získanie akademickej hodnosti doktor (philosophiae doctor, PhD.)

v doktorandskom študijnom programe:
Telekomunikácie

Miesto a dátum: Bratislava, 31.5.2016

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia

Na Ústave telekomunikácií FEI STU v Bratislave

Predkladateľ: Ing. Tibor Csóka
Ústav telekomunikácií, Fakulta elektrotechniky a informatiky, Slovenská technická unverzita v Bratislave, Ilkovičova 3, 812 19 Bratislava

Školiteľ: prof. Ing. Jaroslav Polec, PhD.
Ústav telekomunikácií, Fakulta elektrotechniky a informatiky,
Slovenská technická unverzita v Bratislave, Ilkovičova 3, 812 19 Bratislava

Oponenti: prof. Ing. Stanislav Marchevský, CSc.
Katedra elektroniky a multimediálnych telekomunikácií, Technická univerzita v Košiciach
Park Komenského 13, 04120 Košice

Ing. Juraj Štefanovič, PhD.
Fakulta informatiky, Paneurópska vysoká škola
Tematínska 10, 851 05 Bratislava

Autoreferát bol rozoslaný: ....

Obhajoba dizertačnej práce sa koná: 31.08.2016 o hod.

Na FEI STU, Ilkovičova 3, 812 19 Bratislava, v zasadacej miestnosti ÚT, B-blok, 6.p., miestnosť B602

prof. Dr. Ing. Miloš Oravec

Index

1 Abstract (in Slovak) 4

2 Related work 5

2.1 Markov models 5

2.2 Empirical models 5

2.3 Chaos theory models 5

2.4 Discrete process based generative model 5

2.5 Other approaches 5

3 Specification of research goals 6

4 Trace capture and analysis 6

4.1 Available software and hardware 6

4.1.1 GNU Radio 6

4.1.2 Universal software radio peripheral (USRP) 6

4.2 Proposed implementation 6

4.2.1 Data capture transceiver 7

4.3 Proposed measurement scenario 7

4.4 Trace analysis 7

4.4.1 Propagation channel 7

4.4.2 Logical channel 7

4.4.3 Error control channel 8

5 Proposed models 8

5.1 Classification based model for the Physical Layer (PHY-CBM) 8

5.1.1 Parameterization 8

5.1.2 Generation 9

5.2 Novel hyper-exponential distribution parameterization using MMPP-2 9

5.3 Hierarchical Vector Quantization model (HVQ) 9

5.3.1 Parameterization 10

5.3.2 Generation 10

5.4 Classification Based VQ Model (CBVQM) 11

5.4.1 Parameterization 11

5.4.2 Generation 11

6 Evaluation of results 12

6.1 Reference statistical distances 12

6.2 Cluster error probability $\boldsymbol{p(n)}$ analysis 13

6.3 Error burst structure analysis 14

6.4 Error gap structure analysis 15

6.5 Summary 16

7 Original scientific contributions 17

8 Bibliography 18

9 The list of published works related to the research topic 20

1 Abstract (in Slovak)

V telekomunikáciách často používané digitálne bezdrôtové kanály sú postihované chybami vykazujúcimi zhlukový charakter, jav známy v odbornej literatúre aj ako „pamäť kanála“. Je preukázané, že práve táto vlastnosť bezdrôtových kanálov je kritickým faktorom pri návrhu telekomunikačných systémov, schém pre detekciu a korekciu chýb, ako aj medzivrstvovú optimalizáciu. Generatívne kanálové modely chybovosti sú schopné zachytiť vhodne navrhnutým matematickým opisom kľúčové charakteristiky procesu generovania chýb do skupiny parametrov a následne reprodukovať štatisticky podobné dáta pre simulačné účely. Ich použitie je obmedzené len matematickými alebo štatistickými obmedzeniami navrhnutého riešenia. Výskum prezentovaný v tejto práci si kladie za cieľ na vybranom bezdrôtovom kanáli vykonať merania charakteristík skutočného binárneho chybového procesu za účelom identifikovania jeho štatistických vlastností a následne navrhnúť a realizovať také modely, ktoré by boli schopné s dostatočnou presnosťou tieto významné charakteristiky modelovať. V práci dokumentované navrhnuté modely rozumne kombinujú rôzne matematické a štatistické postupy, vrátane vektorovej kvantizácie, klasifikácie a Markovovského procesu, a tým vytvárajú unikátne riešenia prevyšujúce možnosti existujúcich alternatív. Výsledky zhrnuté v hodnotiacej kapitole zahŕňajúcej výsledky kvalitatívnych porovnaní navrhnutých modelov vzhľadom na referenčný model aj zozbierané dáta, toto tvrdenie potvrdzujú.

2 Related work

The most widely accepted classification of error models is proposed by Kanal and Sastry in their review of channel error models [1] using a classification system based on the model's inner modeling principle as: **generative** (utilizing a generating "underlying mechanism" to obtain the output) or **descriptive** (fit specific stochastic properties of the observed trace using empirical functions).

More recent classifications have typically used the classification based on model's employed mathematical concept to classify them as: **pure** (using only one mathematical method or a single principle) and **extended** (various model configurations primarily using *cascading* or *modulating*).

2.1 Markov models

Markov models were originally used to define the so called generative model group and have since the first applications become centric to multiple different model branches.

The original proposals were based on the discrete time Markov chain, with the pioneering *Gilbert's model* [2]. *Elliot* suggested a modification [3] of Gilbert's error model by introducing the probability of generating an error also in the model's good state. Long after these, the next revolution came with *Fritchman's* model and particularly its simplification, the *Simplified Fritchman's model*, widely applied in high-frequency channel error modeling. More recent alternatives include the *bipartite model* [4], *hierarchical Markov model* [5] and extended models, such as cascaded Markov model [6] that employs parallel Gilbert's and Elliot's generators.

Hidden Markov Models generate output trace using the same generative and mathematical principle, but the internal structure of the model is considered hidden, and most approaches estimate its parameters using algorithms such as Baum-Welch or Turin-Sondhi [7]. Among the more most recent models from this group belongs the Double Embedded Processes based Hidden Markov Model [8].

Semi-Markov models were promoted for error modeling after [9] showed that packet loss can only be modeled using a time-inhomogeneous Markov chain.

2.2 Empirical models

Feasibility of Pareto distribution for error process modeling was successfully explored in a study by Ilyas and Radha [10] in their extensive research of errors on IEEE 802.15.4 LR-WPAN. Nogueira et al. [11] offered a new perspective on empirical approach utilizing Markov concepts, a subgroup of Markov Arrival Processes (MAP) called Markov Modulated Poisson Process (MMPP) producing a a hyper-exponentially distributed Random Variable. The problem of parametrizing MMPP models was addressed in [11], [12], [13] and [14].

2.3 Chaos theory models

[15] demonstrated how heavy-tailed self-similar traffic is related to ON/OFF sources. Strong short-term memory channels were modeled using nonlinear chaos equations in [16] and [17]. A more complex approach to chaos modeling providing a comparison with other commonly used methods is presented by Kopke et al. in [18]

2.4 Discrete process based generative model

[19] and [20] validate the model proposed in [21] on a real EGPRS channel trace and confirm the reliability of their DPBGM proposal based on the principle of Rice's sum of sinusoids.

2.5 Other approaches

Other approaches include the Stochastic context free grammars, fractal models, multi-fractal wavelet model presented in [22] and improvements of existing models by new concepts, such as genetic algorithms. [23] proposes a GA based search for the optimum model transition matrix identification.

3 Specification of research goals

Apart from only a handful of documented models, it is unclear what binary error trace was used to fit the model to, how it was extracted, what type of channel it represents and how statistically likely are the models to fit stochastic properties of the error trace for different applications. Sometimes the data trace is just a substandard and unrealistic output of a simulation software, often based on similar mathematical concepts with similar limitations as the proposed models.

Therefore, in order to provide a proper analysis, it is imperative to first select a single technological realization of a real wireless digital channel and identify its basic stochastic properties by analyzing the captured trace under realistic transmission scenarios. Secondly, only after the first step was performed, is it appropriate to propose novel error modeling solutions or modifications of already existing models.

Considering these presented facts and requirements based on the current state-of-the-art in channel error modeling, I hereby set research goals for my dissertation thesis as:

1. Choose an appropriate standardized digital wireless channel, propose realistic measurement scenario and realize trace collection.
2. Analyze the collected traces considering different presented concepts of channel and for each trace establish stochastic properties allowing binary error trace modeling.
3. Propose novel models for binary error modeling.
4. Evaluate the results by comparing them to a chosen reference model and captured binary trace using relevant statistical distance metrics and a goodness-of-fit test.

4 Trace capture and analysis

The desired characteristics can be obtained on both the physical and datra-link layer by using a software defined radio (SDR).

4.1 Available software and hardware

The trace collection was realized in the GNU Radio suite, which proved an efficient tool, but at that point still suffered from minor programming bugs that needed attention.

4.1.1 GNU Radio

GNU Radio [24] is an open-source digital signal processing kit providing tools for implementations of software radios and can be used with a variety of radio hardware or even as a standalone simulation environment. It offers a wide variety of DSP blocks necessary for construction of simple or more complex signal processing systems.

Used version of the GNU Radio (v3.7) supports a wide variety of different modulation, demodulation and processing blocks and it is possible to combine existing blocks into a functional IEEE 802.15.4 compliant transceiver, with the exception of spreading and despreading.

4.1.2 Universal software radio peripheral (USRP)

USRP is a flexible platform for software defined radios. In order to communicate with the connected computer, the USRP device family utilizes the USRP hardware driver (UHD) that supports operating systems Linux, MacOS X and Windows.

4.2 Proposed implementation

Software defined radio based realization of the IEEE 802.15.4 transceiver was chosen as the most efficient method for obtaining the data necessary for a proper noise and error process analysis followed by error process modeling.

Therefore, a new transceiver realization became a necessity, one that would be able to extract all desired characteristics for the propagation, logical and error control channels.

4.2.1 Data capture transceiver

Newer versions of GNU Radio offered many blocks that could be used to recreate a complete IEEE 802.15.4 communication link between the transmitter and the receiver, but it lacked an efficient despreading block. I decided to realize the spreading and despreading procedures separately in a C program and have the GNU Radio communication link script extract only the transmitted bits and noise/interference I/Q samples as a separate transmitter and a receiver.

4.3 Proposed measurement scenario

The measurement scenario and the real measurement setup are depicted in Fig. 1. NLOS transmission was measured with the TX and RX antennas deployed in the red circle positions, whereas the LOS measurements were performed with TX and RX located in the purple circle positions. 868/915 PHY band was used, for which the suitable interferers are:

- WSN node interferers (black circles)
- GSM interferer (purple square)
- GSM pulse jammer (purple square)
- Handheld radio interferer (cyan squares)
- Long range radio interferer

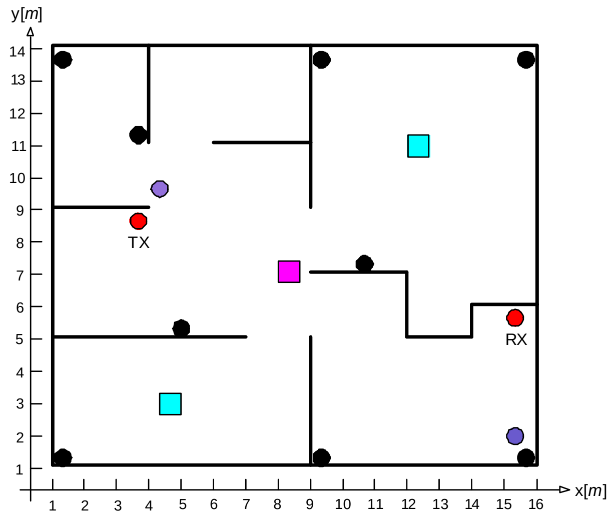

Fig. 1: Measurement scenario layout.

4.4 Trace analysis

The trace for different concepts of channel was captured in order to establish basic real WSN transmission channel stochastic properties and to provide a reference baseline for further verification of the proposed model's application and goodness of fit.

4.4.1 Propagation channel

An analysis of noise trace, encompassing as many realistic interferers as possible, is a necessary first step required to identify the specific indoor WSN noise/error process characteristics.

The feasability of applying a linear model in order to model the physical characteristics that could be transformed into the binary error characteristics after demodulation, was examined. The results of this analysis confirmed that certain realizations of the noise process could be modeled using the MA or ARMA models, but others are beyond the capabilities of linear models. That is why a new model proposal that addresses the binary error generation using the knowledge obtained in interference trace analysis could be designed.

4.4.2 Logical channel

The analysis of the binary trace is limited to observing the error burst and error gap process and the changing nature of their characteristics.

Results of an intra-packet error burst and error gap process analysis demonstrate that arbitrarily selected packets throughout the entire chosen measurement realization of available binary trace (20000 packets) exhibited similar stationary behavior and error burst/gap distributions. The reason for such a stationarity most probably lies in the high BPSK demodulator noise resilience. The gap distributions exhibit stationary behavior over a significantly longer sequence observation and demonstrate an exponentially decreasing curve resembling an Exponential PDF.

Examination of the packet structure leads to the conclusion that at the logical channel each received packet is corrupted by errors to some degree, there are no packets received without any errors.

4.4.3 Error control channel

Error control channel applies different techniques with the goal of eliminating as many errors as possible and at the same time applying counter measures that would decrease the probability of long error burst occurrence in the received binary sequence.

Due to the decrease of the total number of error bursts within the packet, uninterrupted error gap runs are much longer than in the logical channel trace.

The packet structure confirms the expectations. The observed trace contains large amounts of correctly received packets without any errors and the overall gap and burst average size has decreased. In case of bursts, this is caused by the despreading process' error correcting ability, in case of gaps it is because the corrected errors enabled gaps to be connected into larger gap length units that formed correct packets.

Modeling at the error control channel is the most common type of binary error modeling, although the applications of such models in real systems are questionable (with the exception of ARQ and HARQ design), because such models in fact do not model the actual binary error process itself, but the deficiencies of the applied error control techniques and the errors that pass through their limitations.

5 Proposed models

Multiple models were proposed based on the knowledge obtained from the analysis of trace at different channel representations.

5.1 Classification based model for the Physical Layer (PHY-CBM)

I proposed a novel concept of a classification based model centered around the philosophy of exploiting inner texture structure inherent to both non-stationary and stationary interference and noise. The proposed classification model is similar to and was inspired by texture feature extraction from 2D images (e.g. [25]). Instead of working with a $2D$ signal, it works with a $1D$ vector sequence of N complex values partitioned into a set of M independent vectors $\mathbf{Y} = \{y_1, y_2, \dots, y_M\}$ of size l representing individual signal textures. Let us also assume that the vector length is sufficiently long to capture the mean statistical behavior of the observed process.

If the set contains no more than k distinct textures, a classification algorithm (e.g. *kmeans* [26]) can be used to identify these textures as centroids, thereby simplifying the noise process modeling to observation of k individual alternating classes of data with a defined optimal distance to the closest centroid.

Transitions of these classes are consequently captured into a stochastic transition matrix of an abstracted DTMC, thereby effectively creating a Centroid space modulating a DTMC.

5.1.1 Parameterization

First, the entire noise values time series is partitioned into a set of shorter sub-series, each with a length n. Each of the sub-series of the complete noise sequence of complex values is transformed into a feature vector (FV) capable of capturing a specific characteristic of the observed time series process.

FVs are construced in such a way, that they produce centroids that enable the generation process to design a truly random source, therefore, in order to retain both, the information about magnitudes and phases, the FV is constructed in a following manner:

1. Calculate magnitude histogram limited to interval $[0, UB]$ from the values contained in the analyzed sub-series (where UB denotes an upper bound parameter)
2. Calculate histogram limited to interval $[-\pi, \pi]$ of phase values
3. Transform the histograms to relative values
4. Append the phase histogram after the magnitude histogram

It should be noted that values generated from this FV will have de-correlated magnitude and phase components, however, no correlation between these chatacteristics has been identified in the observed noise/interference trace.

Once the entire FV set is constructed from all the different sub-series of the analyzed times series, classification (e.g. using kmeans) is invoked to identify the trace centroids.

Assuming that the output of the classification produced k distinct classes, time series vectors are replaced by the corresponding class to which they were assigned and transitions probabilities for the classes of an equivalent abstracted DTMC are recorded.

5.1.2 Generation

Generation process uses the higher layer Markov chain to introduce randomness into class generation, while the intra-class statistics are obtained from the lower layer generation process within each class. During initialization an arbitrary state is chosen and the histograms contained in centroids are interpolated using a cubic spline function to enable generation of arbitrary values from the range $[0, UB]$. Due to the Markovian nature of the generating process, each iteration starts by using the random number generator to select the next state transition. Once the transition was identified, the model transitions to that state and loads the centroid for the class represented by that state. Within each state the centroid vector is separated to a magnitude and phase histogram and the desired number of n values are generated from cubic spline interpolated version of each of these histograms. Generated values of magnitudes and phases are combined to create a single complex time series sample.

5.2 Novel hyper-exponential distribution parameterization using MMPP-2

I proposed a novel approach eliminating complex calculations necessary for parameter extraction for purposes of the presented research; all 4 MMPP-2 parameters were obtained directly from the observation of the trace without the need to construct histograms or implement non-linear optimization.

Consider an error burst sequence (Fig. 2) that consists of error burst only or error gap only lengths detected in the trace recorded precisely as they occur in the trace.

*000110***11110***101***000000***101***00000***11101***0000***11111***0***11101*

↓

*321***4***11111***6***111***5***311***4***5***1311*

Fig. 2: Explanation for MMPP-2 parameter extraction algorithm.

Because the states in Markovian representation can be identified with individual intra process transitions, the first step of the proposed algorithm is to choose a threshold value that determined which Markov state was used for generation. MMPP-2 transitions between two Poisson processes and the threshold determines which one should be assigned generation of the currently processed burst or gap length. For typical applications the mean value is sufficient as a viable threshold, if a balanced transitioning is required, then median value should be chosen as a threshold. It is imperative to use such threshold that the system transitions from one Poisson process to another and both (HE) component processes take part on the generation process as equally as possible.

Parameters λ_1 and λ_2 representing the exponential distribution parameter can be obtained as an inverse value to mean of all lengths smaller than or equal to the threshold in the first case and longer than the threshold in the second case.

Transitions between these Poisson processes are then given by cumulative lengths τ_i, where τ_i represents the sum of all uninterrupted elements generated by the i-th Poisson process. Transition rates r_1 and r_2 can be obtained as an inverse value to the mean of the corresponding τ_i.

5.3 Hierarchical Vector Quantization model (HVQ)

The proposed HVQ model uses a well known VQ concept and introduces a novel approach to codebook construction for binary vector generation.

The codebook's vectors are constructed from an arbitrary basis that allows multiresolution by using vectors with different sizes.

A novel idea introduced for modeling arbitrary binary traces using a HVQ model is the codebook construction exploiting the properties of binary sets and maximizing both parameterization and generation computation resource efficiency. Assuming that the burst and gap processes do not necessarily have to be independent leads to formulation of claim, that there are binary runs of higher than minimal order with having a higher probability of occurrence in the observed binary trace than other runs. It is preferable to construct such codebook that contains not only as small basis set of a chosen binary space, as possible, but also the frequently occurring identified specific runs.

Such a set that contains basis vectors from a particular binary space and additional vectors from the same or other binary spaces exhibits overcompleteness. The significant advantage of implementing the proposed codebook lies in 2 important aspects: ability to represent all binary vectors of specific lengths with the basis vectors contained in the codebook and ability to represent specific longer binary sequences using the additional codebook vectors.

Definition of a suitable codebook does not guarantee an efficient model. However, a combination of random process selecting vectors from the codebook in the generation process is enabled by the abstract DTMC, whose states represent different groups of vectors within the codebook. Particularly interesting is the case, in which the vectors are organized into groups based on their length or focus. Binary vector assignment in the parameterization phase is equivalent to transition to the first state of the multiresolution chain. Within each state (group of vectors), vectors are assigned different generating probabilities based on their occurrence in trace, effectively producing an intra-state generating process that can be labelled VQ modulated Markov chain.

5.3.1 Parameterization

Codebook can be constructed using different bases, however one that lends itself due to multiple documented advantageous properties is Hadamard base. En axample codebook is constructed as: $BE = \{H_3, H_2, H_1, -H_1\}$, where H_1 is the Hadamard matrix with order 2 (lower order not considered).

Each of the vectors in the particular component Hadamard matrix is represented by a state in the abstract Markov chain that introduces randomness into the model. The transitions among the states are strictly limited to transitions from the state with higher priority to the nearest state with lower priority, unless a precise match of the analyzed binary vector is identified, at which point the system returns to the state with the highest priority. Priority is assigned by the order of vector groups, with the leftmost element of each codebook having the highest priority and rightmost element having the lowest priority. This principle is depicted for the vector codebook BE in Fig. 3.

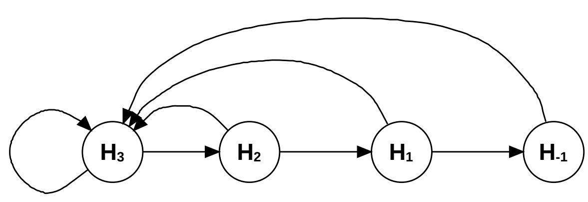

Fig. 3: Abstract DTMC for the codebook defined by *BE*.

The binary set is sequentially and iteratively compared on the binary vector basis to the vectors from each of the groups contained in the codebook. A match is followed with increasing the probability of the state corresponding to the group containing the vector and increasing individual probability of identified vector's generation within the group (state). If there is no match for the vectors of the current group, a lower priority group is selected and the sample from the trace is compared to the new group.

5.3.2 Generation

The generation process is controlled by the Markovian transition matrix. Each iteration the random number generator produces a value used by the current state (representing a group of codebook vectors) to establish its transition to the next state. If and only if this next transition transitions to the highest priority state, the group is used to generate a binary vector. It does so by using a new RNG value to produce a vector based on the intra-group (intra-state) vector probabilities established in the parameterization process. Regardless of whether the generation in the current iteration occurred or not, the state transition defined by the first random number generator occurs at the end of the iteration. The process is repeated until the desired number of bits are generated.

5.4 Classification Based VQ Model (CBVQM)

Classification based binary VQ model represents novel binary extension of the proposed concept presented as the classification model for the propagation channel.

As in case of HVQ model, error burst and gap processes are considered in the general case, when they can be both dependent or independent, meaning that generation of binary bursts and gaps is not separated, both are being generated together in the same iteration and instance. In order to achieve maximal precision, the total number of different dominant stochastic sub-processes forming the error bursts and gaps should be less than or equal to the total number of desired classes. As is the case with propagation channel classification model, higher efficiency is achieved by recording the histogram characteristic (relative number of runs) of the observed burst and gap processes.

Utilizing a classification approach allows great variability in applying different classification techniques (e.g. *kmeans*) and sorting techniques (e.g. *KNN*). As in the case of the PHY-CBM, the key issue is establishing the optimal number of classes, which due to the nature of combined binary burst and gap processes is most reasonably performed by running classification process with different settings and choosing the results with the best fit.

The choice of distance metric can also be a factor affecting the precision of the resulting model. Assuming that part of the process invariance is removed by employing its histogram representation in feature vector construction, Euclidean metric is considered a sufficient distance measure for the proposed classification process of the model.

Randomness is added to the CBVQM model the same way as in PHY-CBM and HVQ - by abstracting the identified classes as states of a DTMC, thus limiting the class transitioning process to geometrical distribution. This can, of course, be improved by using a different stochastic concept for state transitions, but experiments have demonstrated, that it is sufficient for binary error channel modeling purposes.

5.4.1 Parameterization

The CBVQM model parameterization process starts with input variable initialization. Then, vectors of length l are taken and their burst and gap run lengths are used to produce histogram of each component (burst and gap) process. These histograms are thereafter organized into a single FV in the same manner as in PHY-CBM. A sequence of n binary trace bits produces $\lfloor n/l \rfloor$ FVs serving as a classification input.

The chosen classification method is then used to sort feature vectors into the desired number of classes using the specified metric. Each class is uniquely defined by its centroid value representation identified during the classification procedure.

The vector sequences of class representations are then transformed into transition probabilities of the abstracted Markovian transition matrix.

5.4.2 Generation

The initialization within the generation process (Fig. 40) starts with the creation of an entire set of vectors from the binary set of order l. By assigning each of the vectors from the create complete binary alphabet (all binary vectors of length l) into one of the classes defined by the centroid using the same metric, the model is capable of producing any binary sequence of the specified length. A classifier such as *KNN* can be used for this purpose.

The data set could exhibit high imprecision, if the previous step was not compensated to reflect the actual probability of generation for each of the vectors within a single class. One of the relevant factors that could be used to generate the probability of generation for an individual vector is its distance to the centroid. Because all binary vectors of length l are represented in the binary alphabet, it is possible that a number of vectors will not be within a reasonable distance to any of the centroid, yet will be assigned to a particular group because of its biggest proximity to it. To increase the generating probability of those vectors that are closer to the centroid and decrease the generating probability for vectors that are further away (distance penalization). Thus, the generating probability of the i-th vector in the j-th class depends on the second power of the inverse distance d_i of the vector from the centroid:

The distance proposed by this metric proved sufficient for binary error burst and gap modeling, but is subject to further discussion and change for another metric that could prove to produce even more precise results.

All probabilities p_i for each state are used to produce a vector generating histogram that can be transformed into the CDF of the intra-state vector generating process.

A distinct complication arises, particularly if the total number of classes is higher than the number of components of an error burst and gap process. In such a case multiple centroids close to each other are identified in the trace, but the process of assigning the vectors from a complete set based only on the nearest neighbor would assign the binary vector to the nearest class. That would, however, undermine the distance concept of generating probability calculation. Two readily available solutions could rectify this problem: reduction of total number of classes or assignment of the same vector to multiple classes.

Multiple vector class assignment is a faster solution that retains the proposed number of classes, where therefore any distance shorter than the nearest neighbor for all vector assignments must be considered resulting in the possible presence of a single vector in multiple classes with different generating probabilities.

Lastly, an arbitrary state from the abstracted Markov chain representing the stochastic transitioning process is, is selected as the starting state.

Once the generating set has been configured to reflect the binary data trace parameters, the generation process is used to produce a state transition based on the output of the RNG. After each state transition, the destination state's intrastate generating CDF is used to produce a binary vector from that state using the inverse method, repeating the process for as many bits as need to be generated.

6 Evaluation of results

Both of the commonly modeled and observed binary channel characteristics are essential in characterizing the digital binary channel and therefore these statistics were captured using the proposed data harvesting scheme, they include:

- Binary error burst $B_l(l)$ and error gap $G_l(l)$ distributions as functions of the observed burst and gap length l
- First ($E(B_l(l))$, $E(G_l(l))$) and second ($var(B_l(l))$, $var(G_l(l))$ order moment characteristics of the burst error and gap process
- Cluster error probability $p(n)$ of 0 errors within a cluster of length n

The results of the data generated by proposed models are further analyzed with statistical methods and metrics related to the captured binary trace.

6.1 Reference statistical distances

In order to establish the closeness of fit of the modeled data to the oberved origianl channel binary error trace, statistical distances are used. Among the possible variations of statistical distances, after considering the specific discrete nature of the resulting modeled data, I decided to implement the following statistical distances based on the f-divergence (lower value indicates a better fit):

- Hellinger distance (HD)
- Jeffrey divergence (JD)
  - Used to only compare those bins, where BOTH histograms are represented with data
- Mean squared Error

To verify, whether the data set obtained from the model could be considered a suitable statistical fit to the observed binary trace, the **Pearson χ^2 goodness of fit test** (fail is in *red* italics, **pass** in green bold) for large data sets was performed. A visual method for verifying the goodness of observed and modeled PDF fit in the form of a Probability-Probability plot available in Appendix B of the presented thesis.

6.2 Cluster error probability $p(n)$ analysis

Analysis of cluster probability $p(n)$ is essential to understand the range nature and similarity of the observed process. Results of statistical distance analysis of the generated model data for all proposed and reference models can be found in Tab. 1.

Tab. 1: Statistical distance of cluster probability $p(n)$ for all observed models.

	Cluster error probability		
	JD (D_J)	**HD (D_H)**	**MSE**
HE(MMPP-2)	0.00021	0.01549	0.00125
Gamma	0.00039	0.02126	0.00200
Elliot's model	0.00005	0.00746	0.00037
HVQ-BE	0.02726	0.18014	0.04886
HVQ-BG	0.00044	0.02247	0.00051
HVQ-D	**0.00004**	**0.00708**	**0.00008**
CBVQM(5,8)	0.00022	0.01584	0.00148
CBVQM(5,10)	0.00008	0.00988	0.00047
CBVQM(5,15)	0.00020	0.01528	0.00110
PHY-CBM	0.01165	0.11664	0.02193

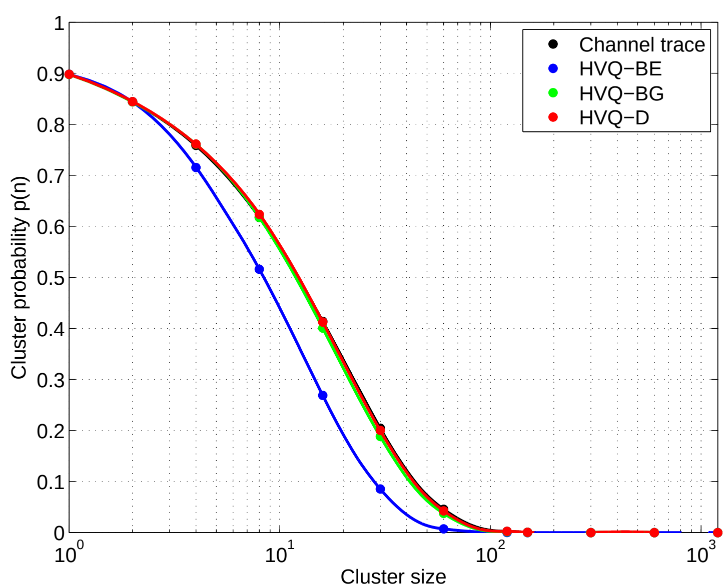

Fig. 4: Cluster error probability $p(n)$ of the best model.

As can be observed from the cluster error probability analysis (Tab. 1), the best fit of the modeled data into the observed trace was obtained from the proposed Hierarchical Vector quantization model (specifically its variant D). This even despite the fact, that the reference Generalized Elliot's model is parameterized to fit the cluster error probability statistic precisely, even if it fails to model the internal error burst and gap characteristic distribution properly.

Visual confirmation of the results can be seen in Fig. 4, where all HVQ model's cluster error probability curves are plotted along with the reference trace. The HVQ-D model curve completely covered the original trace plot.

6.3 Error burst structure analysis

Error burst histogram distribution and a corresponding CDF are used to establish important aspects of the analyzed models, such as the ability to capture individual characteristics of the error burst sub-process within the overall data. Analyses of the distance metrics are not conclusive in that not all distance statistics results agree on the best proposed model variant, but the most statistical distance metrics (2/3) are in favor of the proposed HVQ model (Tab. 2).

Tab. 2: Analysis of statistical distance of error burst characteristic for all observed models.

	Error bursts			
	JD (D_J)	**HD (D_H)**	**MSE**	**C(χ^2)**
HE(MMPP-2)	0.12643	0.39947	0.29945	*60.6121*
Gamma	0.10251	0.36269	0.15835	*1.94607*
Elliot's model	0.14627	0.41695	0.56201	*1.80611*
HVQ-BE	0.09965	0.34254	0.40329	*1.29292*
HVQ-BG	**0.00093**	**0.03277**	0.00174	**0.00874**
HVQ-D	0.00168	0.04441	0.00341	**0.00335**
CBVQM(5,8)	0.00903	0.10331	0.01281	**0.01259**
CBVQM(5,10)	0.02559	0.19682	0.00525	*0.09887*
CBVQM(5,15)	0.00507	0.14562	0.01813	**0.01308**
PHY-CBM	0.00382	0.06732	**0.00130**	**0.04247**

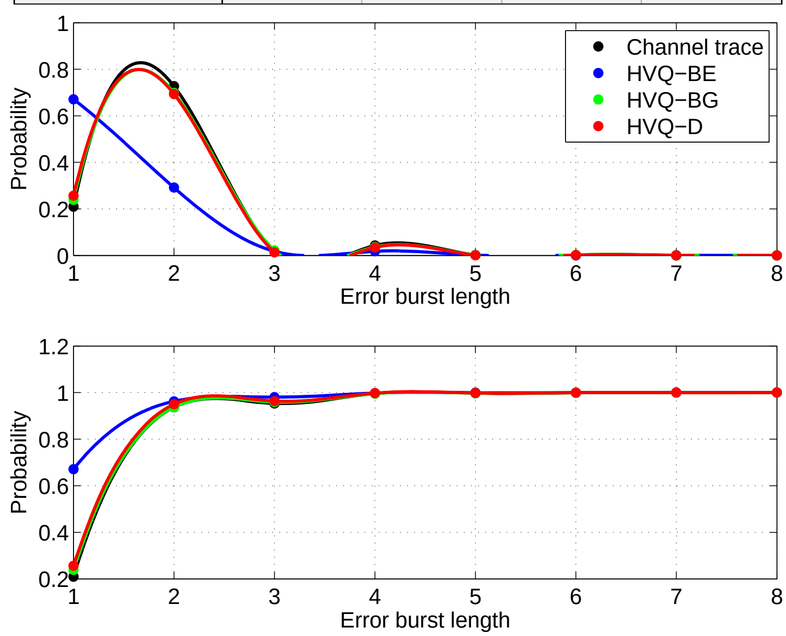

Fig. 5: Error burst characteristic fit of the best model (HVQ-BE) (for burst lengths up to 8)

Tab. 2 offers interesting results. Altogether 5 different models did not pass the Pearson goodness of fit test, including all empirical models and the reference Elliot's model.

PDF and CDF of error bursts for the best fitting HVQ-BG model data is located in Fig. 5. As witnessed by the closeness of the fit, HVQ-BG model produces an excellent fit that passes the goodness of fit tests. It should be, that the HVQ-D alternative also present in the figure produces qualitatively excellent results, but I'm purposefully identifying the best fit only for the particular characteristic. One of the novel models based on the proposed generating methods based on the HVQ (HVQ-BE) that did not pass the goodness of fit test can be seen along with the best fitting model's (HVQ-BG) characteristic in Fig. 5.

6.4 Error gap structure analysis

Results of the statistical distance comparisons of the generated and refererence model characteristics (Tab. 3) suggests, that the best fit on the specified observation inverval of 5000 packets is produced by the hyper-exponential distribution parameterized using the novel parameterization algorithm proposal.

Tab. 3: Analysis of statistical distance of error gap characteristic for all observed models.

	Error bursts			
	JD (D_J)	**HD (D_H)**	**MSE**	**C(χ^2)**
HE(MMPP-2)	**0.00170**	**0.04489**	**0.00049**	**0.00018**
Gamma	0.00473	0.07429	0.00419	**0.00067**
Elliot's model	0.00535	0.07879	0.00304	**0.00049**
HVQ-BE	0.02529	0.17391	0.01088	**0.01408**
HVQ-BG	0.02616	0.17542	0.00788	**0.00073**
HVQ-D	0.02182	0.15993	0.00791	**0.00031**
CBVQM(5,8)	0.00826	0.09822	0.00383	**0.00024**
CBVQM(5,10)	0.02467	0.16964	0.01274	**0.00043**
CBVQM(5,15)	0.08792	0.33198	0.02020	**0.01251**
PHY-CBM	0.00417	0.07152	0.00054	**0.04703**

There are no fails at passing the goodness of fit tests for any of the models and the visual inspection of table also uncovers multiple suitable alternatives to the best HE(MMPP-2) fit, e.g. the classification based model realized on the physical layer or the binary classification model CBVQM(5,8). HVQ models produce satisfactory results, however they are slightly behing the HE(MMPP-2) model regarding precision, when compared to the best solutions.

The resulting best fit of the HE(MMPP-2) model can be seen in Fig. 6. Close fit of other empirical models and also the Elliot's model can be observed as well. Both were, however, expected to give good results not only for gaps, but any process exhibiting Poission shaped distribution.

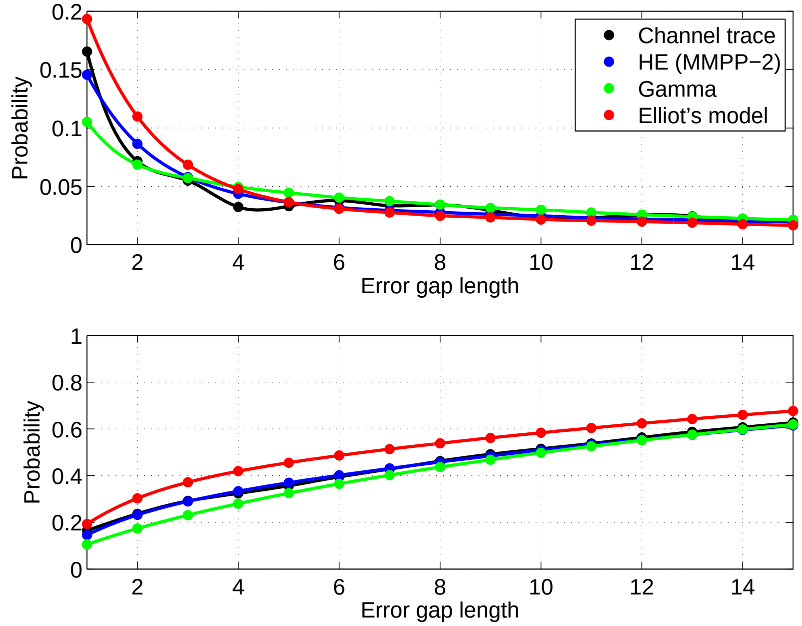

Fig. 6: Error gap characteristic fit of the winning model (HE(MMPP-2)) (for gap lengths up to 15)

6.5 Summary

The proposed models proved to be more than a suitable alternative to the already established models and the reference Elliot's model and the used innovative concepts significantly improve on adverse properties observed in other models, e.g. high parameterization complexity (e.g. Elliot's model). Unlike the reference Elliot's model and its prameters obtained from analytical parameterization that do not necessarily produce the best possible results from the model realization, the proposed HVQ as well as the proposed CBVQM and PHY-CBM models offer a significantly simpler, algorithmic parameterization which always leads to the best possible estimate of model parameters.

Furthermore, it was shown, that by modeling the processes as burst and gap dependent using the proposed concepts, the model does not lose accuracy in any of the observed statistical characteristics (as witnessed by the proposed model statistical distances and goodness of fit test results).

There are multiple possible representations of the result. Assuming that the best possible model has to be chosen for a particular binary error modeling task at any given binary layer, the results presented in Tab. 1, Tab. 2 and Tab. 3 conclusively show that the undisputed first choice, when considering all results together, is either the HVQ-BG or HVQ-D version of the HVQ model. Both produce statistically precise error burst and error gap sequences with extremely good precision in all observed statistical characteristics. Furthermore, they are efficient and easy to implement, both for parameterization and generation.

CBVQM and PHY-CBM also produce satisfying results, although they are not as dominantly superior to other modeling approaches, as the HVQ model. This is due to several factors, one of them the classification nature of the CBVQM introducing ambiquity in the generation phase by introducing also binary sequences not present in the original trace. The PHY-CBM, on the other hand, proved viability of modeling the binary error process directly in the physical channel as noise, followed by combination with arbitrary known transmission signal and consequent demodulation. The statistical characteristics of the trace produced by the PHY-CBM predispose this model to further research of binary errror process modeling, as well as other possible applications.

Excellent results obtained from fitting the exponentially shaped gap process using the HE(MMPP-2) model suggests that this model can be used for a variety of modeling scenarios, including the binary error control channel modeling of independent process, but also any generator-based application that needs to produce the desired stochastic behavior on intervals consisting of relatively small number of elements. Furthermore, not only is the proposed novel HE(MMPP-2) parameterization algorithm fast and efficient, it is easy to implement and for a particular desired input threshold will produce a unique best possible solution.

7 Original scientific contributions

The goal of the research presented in the dissertation thesis was twofold: to choose a real transmission channel for analysis of noise/error processes and proposal of novel or improvement of existing models capable of proper modeling of chosen specific stochastic properties of the observed binary error process. The proposed goals of the dissertation thesis were fulfilled, various real channel traces were captured and novel models were proposed and compared with the trace and reference model.

I hereby establish the original contributions presented in the dissertation thesis as:

1. Design and implementation of an extensible toolset capable of extracting real binary and I/Q traces from the wireless channel in a Wireless Sensor Network.

2. Identification of specific stochastic behavior exploitable in modeling of the captured trace in three conceptual representations of the channel: propagation channel, logical channel and error control channel, followed by feasibility verification of modeling the interference at the physical channel using linear models.

3. Proposal and implementation of VQ based model using a codebook constructed on the basis of different order Hadamard matrices and utilizing a Markov transition scheme for generating vector selection. Model applicability to real binary traces was verified using statistical distances and Pearson's statistical test.

4. Proposal and implementation of classification based propagation channel model capable of modeling the interference source of the propagation channel with the goal of modeling the desired binary error process directly in the physical layer. Model applicability to real binary traces was verified using statistical distances and Pearson's statistical test.

5. Proposal and implementation of binary VQ model utilizing a classification based codebook construction with distance punishment rule for optimal codebook vector probability weighing. Model applicability to real binary traces was verified using statistical distances and Pearson's statistical test.

6. Proposal and implementation of novel parameterization technique for a two-state hyper-exponential distribution based on Markov Modulated Poisson Process-2.

8 Bibliography

[1] L. N. Kanal and A. R. K. Sastry, "Models for Channels with Memory and Their Applications to Error Control," *Proceedings of the IEEE,* vol. 66, no. 7, pp. 724-744, July 1978.

[2] E. N. Gilbert, "Capacity of a Burst-Noise Channe," *Bell System Technical Journal,* vol. 39, no. 5, pp. 1253-1265, September 1960.

[3] E. O. Elliot, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell System Technical Journal,* vol. 42, no. 5, pp. 1977-1997, September 1963.

[4] A. Willig, "A New Class Of Packet- And Bit-Level Models ForWireless Channels," *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications,* vol. 5, pp. 2434-2440, 2002.

[5] T. Tao, J. Lu and J. Chuang, "Hierarchical Markov Model for Burst Error Analysis in Wireless Communications," *IEEE VTS 53rd Vehicular Technology Conference,* vol. 4, pp. 2843-2847, 2001.

[6] J. Polec, V. Hirner, M. Martinovič and K. Kotuliaková, "A Generator from Cascade Markov Model for Packet Loss and Subsequent Bit Error Description," *World Academy of Science, Engineering and Technology,* vol. 7, no. 4, 2013.

[7] W. Turin and M. Sondhi, "Modeling error sources in digital channels," *IEEE Journal on Selected Areas in Communications,* vol. 11, no. 3, pp. 340-347, 1993.

[8] O. S. Salih, C.-X. Wang and D. I. Laurenson, "Double Embedded Processes Based Hidden Markov Models for Binary Digital Wireless Channels," *IEEE International Symposium on Wireless Communication Systems,* pp. 219-223, 2008.

[9] C. Jiao, L. Schwiebert and B. Xu, "On Modeling the Packet Error Statistics in Bursty Channels," *27th Annual IEEE Conference on Local Computer Networks,* pp. 534-541, 2002.

[10] M. U. Ilyas and H. Radha, "A channel model for the bit error rate process in 802.15. 4 LR-WPAN wireless channels," *Proceedings of IEEE International Conference on Communications,* pp. 257-261, 2008.

[11] A. Nogueira, P. Salvador and R. Valadas, "Fitting Algorithms for MMPP ATM Traffic Models," *Proceedings of the Broadband access conference,* 1999.

[12] A. Nogueira and R. Valadas, "Analysing the versatility of the 2-MMPP traffic model," *Proceedings of the Second International Symposium on Communication Systems Networks and Digital Signal Processing, ,* 2000.

[13] S. H. Kang, Y. Kim, D. K. Sung and B. D. Choi, "An Application of Markovian Arrival Process (MAP) to Modeling Superposed ATM Cell Streams," *IEEE Transactions on Communications,* vol. 50, no. 4, pp. 633-642, 2002.

[14] H. Okamura, T. Dohi and K. S. Trivedi, "Markovian arrival process parameter estimation with group data," *IEEE/ACM Transactions on Networking (TON),* vol. 17, no. 4, pp. 1326-1339, 2009.

[15] P. Pruthi and A. Erramilli, "Heavy-tailed on/off source behavior and self-similar traffic," *IEEE International Conference on Communications,* vol. 1, pp. 445-450, 1995.

[16] E. Costamagna, L. Favalli, P. Gamba and P. Savazzi, "Block-error probabilities for mobile radio channels derived from chaos equations," *IEEE Communications Letters,* vol. 3, no. 3, pp. 66-68, 1999.

[17] E. Costamagna, A. Fanni, L. Favalli and P. Gamba, "Experiments in modeling the parameters of chaos equation models for mobile radio channels," *IEEE-APS Conference on Antennas and Propagation for Wireless Communications,* pp. 103-106, 2000.

[18] A. Kopke, A. Willig and H. Karl, "Chaotic Maps as Parsimonious Bit Error Models of Wireless Channels," vol. 1, pp. 513-523, 2003.

[19] C. X. Wang, W. Xu and M. Pätzold, "Error Models for Evaluating Error Control Strategies in

EGPRS System," *IEEE 60th Vehicular Technology Conference,* vol. 6, pp. 4238-4244, 2004.

[20] C. X. Wang and W. Xu, "A New Class of Generative Models for Burst-Error Characterization in Digital Wireless Channels," *IEEE Transactions on Communications,* vol. 55, no. 3, pp. 453-462, 2007.

[21] C. X. Wang and M. Pätzold, "A new deterministic process based generative model for characterizing bursty error sequences," *Proceedings IEEE,* 2004.

[22] S. A. Khayam, H. Radha, S. Aviyente and J. R. Deller Jr., "Markov and multifractal wavelet models for wireless MAC-to-MAC channels," *Performance Evaluation,* vol. 64, no. 4, pp. 298-314, 2007.

[23] R. Ranjan, D. Bepari and D. Mitra, "Genetic algorithm based Finite State Markov Channel modeling," *International Journal of Wireless Communications and Mobile Computing,* vol. 1, no. 4, pp. 96-102, 2013.

[24] "GNU Radio," [Online]. Available: http://gnuradio.org/. [Accessed 20 January 2015].

[25] I. Uhlikova, J. Polec, W. Benesova and T. Csoka, "Texture aware image error concealment," *EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE,* pp. 1-6, 8-11 September 2015.

[26] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory,* vol. 28, pp. 129-137, 1982.

9 The list of published works related to the research topic

ADM Vedecké práce v zahraničných časopisoch registrovaných v databázach Web of Science alebo SCOPUS

CSÓKA, Tibor [50 %] - POLEC, Jaroslav [50 %]. Modeling poisson error process on wireless channels. In International Journal of Communication Networks and Information Security. Vol. 7, No. 1 (2015), s. 1-7. ISSN 2073-607X. V databáze: SCOPUS. **Kategória B.**

AFC Publikované príspevky na zahraničných vedeckých konferenciách

CSÓKA, Tibor [50 %] - POLEC, Jaroslav [25 %] - KOTULIAKOVÁ, Kvetoslava [25 %]. Heavy-tailed Error Process Modeling. In RTT 2014. Research in Telecommunication Technology [elektronický zdroj] : 16th International Conference RTT 2014; Frymburk, Czech Republic; 10-12 Sept. 2014. Prague : Czech Technical University in Prague, 2014, CD ROM, s. 45-50. ISBN 978-80-01-05540-3.

CSÓKA, Tibor [75 %] - POLEC, Jaroslav [25 %]. Analysis of additive noise characteristics in indoor wireless sensor networks. In EUROCON 2015 : International conference on computer as a tool. Salamanca, Spain. 8-11 September, 2015. Danvers : IEEE, 2015, S. 60-65. ISBN 978-1-4799-8569-2. V databáze: IEEE. **Kategória B.**

UHLÍKOVÁ, Ivana [40 %] - BENEŠOVÁ, Vanda [20 %] - POLEC, Jaroslav [20 %] - CSÓKA, Tibor [20 %]. Texture aware image error concealment. In EUROCON 2015 : International conference on computer as a tool. Salamanca, Spain. 8-11 September, 2015. Danvers : IEEE, 2015, S. 88-93. ISBN 978-1-4799-8569-2. V databáze: SCOPUS: 2-s2.0-84961741405. **Kategória B.**

Cited by: HSIA, Shih-Chang – HSIAO, Cheng Hung. A Fast-Efficient Shape Error Concealment Technique based on Block Classification. IET Image Processing. doi: 10.1049/iet-ipr.2016.0043.

AFD Publikované príspevky na domácich vedeckých konferenciách

CSÓKA, Tibor [70 %] - POLEC, Jaroslav [30 %]. Analysis and comparison of common error burst modeling techniques. In ELITECH´14 [elektronický zdroj] : 16th Conference of Doctoral Students; Bratislava, Slovakia, 4 June 2014. 1.vyd. Bratislava : Nakladateľstvo STU, 2014, CD-ROM, [6] p. ISBN 978-80-227-4171-2.

ILČÍKOVÁ, Ivana [70 %] - POLEC, Jaroslav [10 %] - BENEŠOVÁ, Wanda [10 %] - CSÓKA, Tibor [10 %]. Fuzzy segmentation for texture aware image error concealment. In Aplimat 2016 [elektronický zdroj] : proceedings of the 15th conference on applied mathematics 2016. Bratislava, 2.-4. 2. 2016. 1st. ed. Bratislava : Vydavateľstvo STU, 2016, CD-ROM, S. 541-556. ISBN 978-80-227-4531-4.

POLEC, Jaroslav [25 %] – KIŠŠ, Pavol [25 %] - CSÓKA, Tibor [25 %] - KOTULIAKOVÁ, Kvetoslava [25 %]. An Uniform-Markov Cascade Generator for Packet Loss and Subsequent Bit Error Description. In REDŽÚR 2016 [elektronický zdroj] : 10th International workshop on multimedia and signal processing. Smolenice, Slovakia. May 24, 2016. 1. vyd. Bratislava : Nakladateľstvo STU, 2016.

POLEC, Jaroslav [25 %] – VARGIC, Radoslav [25 %], ILČÍKOVÁ, Ivana [25 %] - CSÓKA, Tibor [25 %]. Feature Extraction in Corrupted Images in Various Color Spaces. In REDŽÚR 2016 [elektronický zdroj] : 10th International workshop on multimedia and signal processing. Smolenice, Slovakia. May 24, 2016. 1. vyd. Bratislava : Nakladateľstvo STU, 2016.

ILČÍKOVÁ, Ivana [70 %] - BENEŠOVÁ, Wanda [10 %] - POLEC, Jaroslav [10 %] - CSÓKA, Tibor [10 %]. Texture Aware Image Error Concealment with Fuzzy Segmentation. In IWSSIP 2016 : The 23rd International Conference on Systems, Signals and Image Processing. 23-25 May 2016, Bratislava, Slovakia. Danvers : IEEE, 2016. ISBN 978-1-4673-9554-0. V databáze: IEEE. **Kategória B.**

CSÓKA, Tibor [70 %] - POLEC, Jaroslav [10 %] - ILČÍKOVÁ, Ivana [10 %] – DOBOŠ, Ján [10 %]. Binary error models for Wireless Sensor Networks. In IWSSIP 2016 : The 23rd International Conference on Systems, Signals and Image Processing. 23-25 May 2016, Bratislava, Slovakia. Danvers : IEEE, 2016. ISBN 978-1-4673-9554-0. V databáze: IEEE. **Kategória B.**

CSÓKA, Tibor [70 %] - POLEC, Jaroslav [10 %] - CSÓKA, Filip [10 %] - ILČÍKOVÁ, Ivana [10 %]. SDR based capture of physical and data-link layer error process characteristics in low-rate WSN. In ELITECH´16 [elektronický zdroj] : 18th Conference of Doctoral Students; Bratislava, Slovakia, 8 June 2016. 1.vyd. Bratislava : Nakladateľstvo STU, 2016. [ACCEPTED]

ILČÍKOVÁ, Ivana [70 %] - BENEŠOVÁ, Wanda [10 %] - POLEC, Jaroslav [10 %] - CSÓKA, Tibor [10 %]. Texture Aware Video and Image Error Concealment. In ELITECH´16 [elektronický zdroj] : 18th Conference of Doctoral Students; Bratislava, Slovakia, 8 June 2016. 1.vyd. Bratislava : Nakladateľstvo STU, 2016. [ACCEPTED]