



Frequency Domain Identification of Data Loss Models

László Sujbert, György Orosz

Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, Hungary

ABSTRACT

Recently measurement data loss has been of greater interest, due to the spread of sensor networks and the idea of Internet of Things. A procedure is proposed that is able to identify the most frequently employed data loss models. It is assumed that the communication protocol provides information about data loss, i.e. the so-called data availability indicator function is known. The power spectral density (PSD) of the indicator function is representative for the model, and can be used for identification. Spectral estimation is carried out by Fast Fourier Transform (FFT) based techniques. The paper introduces the identification procedure for random independent, random block-based and general Markov model-based data loss patterns. The efficiency of the proposed method is demonstrated by simulation and measurement results.

Section: RESEARCH PAPER

Keywords: data loss; measurement; FFT; PSD; system identification

Citation: László Sujbert, György Orosz, Frequency domain identification of data loss models, Acta IMEKO, vol. 6, no. 4, article 9, December 2017, identifier: IMEKO-ACTA-06 (2017)-04-09

Editor: Alexandru Salceanu, Technical University of Iasi, Romania

Received May 5, 2017; In final form July 11, 2017; Published December, 2017

Copyright: © 2017 IMEKO. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Corresponding author: L. Sujbert, e-mail: sujbert@mit.bme.hu

1. INTRODUCTION

Nowadays measurement data transfer is frequently carried out in sensor networks or on the Internet. In this case data can be corrupted or the transmission medium can be partially damaged, etc. [1], [2]. In the era of Internet of Things (IoT) the significance of the problem culminates. Data loss modeling has a great interest for a long while in the communication field. Detailed models have been developed for video transmission [3], [4], or Voice over Internet (VoIP) applications [5].

Data loss is usually modeled by a random process, where a single sample or a block of samples (a packet) is randomly available or lost. Data availability or data loss is described by the so-called data availability indicator function K_n [3], [4], [6], [7]. This function is a series of zeros and ones, and the data loss pattern depends on the data transmission system. In the simplest model the samples of K_n are independent from each other [10], but it is often supposed that the successive samples depend on each other. Comprehensive studies utilize different classes of Markov models, see, e.g., [3], [4], [11].

The identification of the data loss model is based on statistical methods. First the structure and the order of the model are chosen, then the parameters of the model are determined by measurement data [3], [4]. Model verification is an important task, as the process is not necessarily stationary

and even the model order is to be adaptively changed. To this end, a correlation-based method [3] or off-line processing [4] can be used.

The presence of data loss in measurement systems motivated the investigation of the phenomena from a signal processing point of view. Our recently published paper [8] discussed one of such problems: the handling of data loss in the case of spectrum estimation. In spite of the existing methods concentrating to the spectrum estimation, this paper dealt with the characterization of distortion caused by missing data.

Spectrum estimation based on time records with irregular sampling has been used for a long time. Records with missing data can be treated as a special irregular sampling, substituting the signal samples with zeros where the data are lost. Theoretically, such records can be synthesized by the multiplication of the original signal (without data loss) and the data availability indicator function K_n . The spectral estimator of the damaged record is the discrete convolution of the original spectrum and the spectrum of K_n .

Thus the introduction of the distorted spectra involved the calculation of the spectra of the data availability indicator functions. Three data loss models have been investigated: random independent, random independent block-based, and two-state Markov model-based data loss. All their spectra have

been determined, and quantitative connection between the data loss model parameters and the spectral parameters have been calculated.

In this paper we propose the inverse procedure: the data loss model can be identified by the Fourier transform of the data availability indicator function. First the PSD of the indicator function is calculated, then a parametric system identification method is used to get the spectral parameters. As the spectral shape is simple in most cases, this step is not critical. The last step is the calculation of the data loss parameters by the already known relations. Model order selection and verification can also easily be done by comparing the measurement spectra to the model spectra. The spectra are calculated by the Fast Fourier Transform (FFT), which makes the method computationally effective. Summing up, our method offers not only the spectral estimation of the lossy signal [8], but also provides an estimate for the data loss model along with its parameters.

In Section 2 the data loss models and their spectra are introduced in detail. The analysis of the Markov model-based data loss has been expanded to the general case. Section 3 deals with the identification procedure itself, while Section 4 presents simulation and measurement results that confirm the procedure in practice, as well.

2. SPECTRUM ESTIMATION IN THE CASE OF DATA LOSS

2.1. Power spectrum estimation

The Fourier transform of a sampled signal $x(t_n)$ can be estimated by a finite set of samples [9]. The signal $x(t)$ is usually equidistantly sampled, and the spectrum is calculated by the Discrete Fourier Transform (DFT):

$$X(f_k) = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}nk}, \quad n, k = 0 \dots N-1, \quad (1)$$

where $f_k = k/N$ and $x_n = x(t_n)$. The DFT of a signal is usually calculated by the FFT. The transformed vector $X(f_k)$ is generally complex valued, and the spectral content of the signal is expressed by the real valued Power Spectral Density (PSD) function:

$$S(f_k) = \frac{1}{N} |X(f_k)|^2. \quad (2)$$

In order to reduce the variance of the PSD, a long series of samples is recorded, and many consecutive blocks of N samples are transformed, and the estimator is obtained by averaging the individual PSDs. The mean of the individual estimates can be calculated either by linear or exponential averaging.

2.2. Formulation of data loss

In order to model the data loss, a so-called data availability indicator function, K_n , is introduced [3], [4], [6], [7]:

$$K_n = \begin{cases} 1, & \text{if the sample is processed at } n \\ 0, & \text{if the sample is lost at } n \end{cases}. \quad (3)$$

Samples which are not lost will be termed as processed or available samples. The data loss rate can be defined with K_n as:

$$\gamma = \mathbb{P}\{K_n = 0\}, \quad (4)$$

where $\mathbb{P}\{\cdot\}$ stands for the probability operator. The probability that a sample is available is $\mu = 1 - \gamma$.

2.3. Spectrum Estimation with missing data

Using the indicator function, K_n , (1) can be rewritten for the case of data loss:

$$\begin{aligned} \hat{X}(f_k) &= \text{DFT}(x_n K_n) = \sum_{n=0}^{N-1} x_n K_n e^{-j\frac{2\pi}{N}nk} \\ n, k &= 0 \dots N-1 \end{aligned} \quad (5)$$

This formula means that by incorporating K_n into the usual form of DFT, missing samples are practically substituted with zeros. Equation (5) can also be evaluated via FFT.

The spectrum of the signal containing missing samples is obtained as the convolution of the spectrum of the lossless signal and the spectrum of the data loss indicator function. Now only the latter is interesting. Let $X_K(f_k)$ denote the Fourier transform of the data loss indicator function:

$$X_K(f_k) = \text{DFT}(K_n). \quad (6)$$

Thus the PSD of the data loss indicator function is:

$$S_K(f_k) = \frac{1}{N} |X_K(f_k)|^2. \quad (7)$$

The variance of $S_K(f_k)$ can also be reduced by averaging.

2.4. Data loss models and their spectra

In [8], three data loss models have been investigated:

1. random independent data loss,
2. random block-based data loss,
3. two-state Markov model-based data loss.

The random data loss is one of the most essential data loss models; it is often used because of its simplicity [10]. Block-based data loss models are often used, e.g., when several measurement results are transmitted over packet-based communication systems. The Markov model has been proven to be useful, e.g., in the description of data loss pattern in real-time data transmission over Internet [11]. In [8] the two-state model has exhaustively been investigated, but real situations may require more states.

In the following the main features of the data loss models are recalled.

2.4.1. Random independent data loss

Random independent data loss can be defined as follows:

$$\begin{aligned} K_n &= 1, \text{ with probability } \mu = 1 - \gamma \text{ for } \forall n. \\ K_n &= 0, \text{ with probability } \gamma \end{aligned} \quad (8)$$

The definition means that each sample is lost with probability γ , and data losses at different time instants are independent of each other. The PSD of the data loss pattern is [8]:

$$S_K(f_k) = \frac{\mu(1-\mu)}{N} + \mu^2 \delta(f_k), \quad (9)$$

where $\delta(f_k)$ stands for the Dirac-delta function. The PSD is white, which is represented by the first term, while the second term represents the power of the mean value μ of the data loss pattern.

2.4.2. Random block-based data loss

To define the random block-based data loss, the indicator function is given as:

$$\begin{aligned} \{K_{kM} \dots K_{(k+1)M-1}\} &= 1, \text{ with probability } \mu \text{ for } \forall k. \\ \{K_{kM} \dots K_{(k+1)M-1}\} &= 0, \text{ with probability } \gamma \end{aligned} \quad (10)$$

The definition means that each block of length M is lost with probability γ , and the data loss in different blocks are independent of each other. The power spectral density of the data loss pattern is [8]:

$$S_K(f_k) = \frac{\mu(1-\mu)}{MN} \left| \frac{\sin(f_k \pi M)}{\sin(f_k \pi)} \right|^2 + \mu^2 \delta(f_k). \quad (11)$$

2.4.3. Two state Markov model-based data loss

The first Markov model-based data loss is described by the two-state Markov chain shown in Figure 1. This is the simplest stochastic model that describes the dependency of the consecutive samples of K_n . The states of the Markov chain represent the value of the indicator function K_n . If a sample is available at time instant n , the next sample will be available with probability p , and will be lost with probability $1-p$. If a sample is missing at time instant n , the next sample will be available with probability $1-q$, and will be lost with probability q . The data availability rate μ is the following [12]:

$$\mu = \frac{q-1}{p+q-2}. \quad (12)$$

Note that the parameters p , q , and μ cannot be prescribed simultaneously.

The spectral property of a data loss sequence generated by the Markov chain shown in Figure 2 can be determined according to [12]. The PSD of K_n is a first-order, low-pass type spectrum defined as [8]:

$$S_K(f_k) = \frac{1-a^2}{N(1-a^{2N})} \cdot \frac{1}{|1-az^{-1}|^2} + \mu^2 \delta(f_k), \quad (13)$$

where $z^{-1} = e^{-j2\pi f_k}$ and

$$a = p + q - 1. \quad (14)$$

2.4.4. General Markov model-based data loss

The two-state Markov model describes the relation of the actual and only the previous sample of K_n . Generally, the actual state could be determined by the last r samples. As each sample

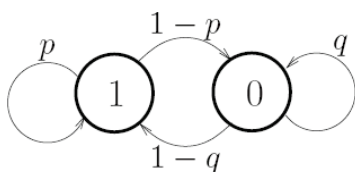


Figure 1. A two-state Markov model of data loss. State "1": actual sample is available ($K_n = 1$). State "0": actual sample is lost ($K_n = 0$).

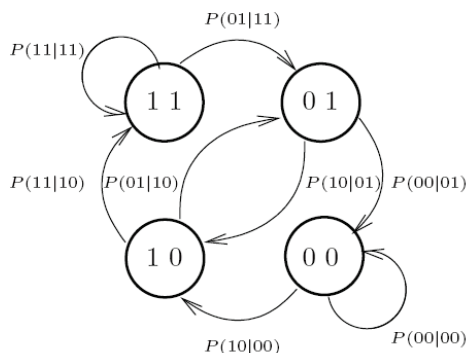


Figure 2. A four-state Markov model of data loss.

could have two distinct values (zero or one), a general model could have 2^r states. The parameter r is the order of the model. The states and the transitions can be seen in Figure 2 for the case $r = 2$. Note that there are impossible state transitions that have formally zero probability in the model. The relation between the transition probabilities and the PSD of the model can also be determined [12]. Let the PSD be as follows:

$$S_K(f_k) = \frac{A}{|1 - \sum_{m=1}^r a_m z^{-m}|^2} + \mu^2 \delta(f_k), \quad (15)$$

where $z^{-1} = e^{-j2\pi f_k}$. The constant A is a function of the vector a_m , $m = 1, \dots, r$ and the number of the FFT points N . The state transition probabilities are:

$$\mathbf{P} = [P_{ij}] = \begin{cases} \mu + \sum_{m=1}^r a_m (K_{n-m} - \mu), & \text{if } K_n = 1 \\ 1 - \left(\mu + \sum_{m=1}^r a_m (K_{n-m} - \mu) \right), & \text{if } K_n = 0 \end{cases}, \quad (16)$$

where

$$[P_{ij}] = \mathbb{P}(s_i | s_j), \quad (17)$$

i.e., the probability of the state s_i with a condition of the previous state s_j . The s_i and s_j states are the following:

$$\begin{aligned} s_i &= [K_n, K_{n-1}, \dots, K_{n-r+1}], & i &= 1 \dots 2^r \\ s_j &= [K_{n-1}, K_{n-2}, \dots, K_{n-r}], & j &= 1 \dots 2^r \end{aligned} \quad (18)$$

Table 1 summarizes the PSDs of the data loss indicator functions for the most important data loss models. The small figures in the table illustrate the typical shapes of PSD functions. General Markov models can have diverse spectra, thus they are not represented in the table.

3. IDENTIFICATION OF DATA LOSS MODELS

Data loss model identification consists of the model selection and the determination of the model parameters [3], [4]. The procedure utilizes some direct parameters of K_n and is completed by the evaluation of $S_K(f_k)$. The theoretical basis for this method is the exact relation between spectral leakage and the parameters of different data loss models presented in the previous section. The identification process is summarized in Figure 3.

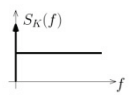
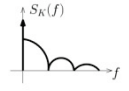
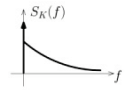
An essential requirement is that the communication protocol provides information about each sample whether its transfer was successful or not. Without such information only qualitative assessment of the data loss can be done. In such a case the shapes depicted in Table 1 can be used for identification.

If K_n is available, it is also known whether the protocol is block-based. In the latter case, one sample of K_n is enough for each block representing the data loss. It is a kind of decimation. The block length M is obviously available.

The data availability rate μ can easily be estimated as the mean value or DC level of K_n . This DC level should be subtracted from K_n , in order to remove $\delta(f)$ from the PSD, as its presence can impair the transfer function fitting.

The next step is the calculation of $S_K(f_k)$. It can be done by the averaging of the FFTs of consecutive (possibly overlapping) blocks of N samples of K_n . As the DC component is removed from the PSD, windowing is not necessary.

Table 1. Summary of PSDs belonging to different data loss models.

Model	$S_K(f_k)$	Shape of $S_K(f_k)$
Random independent data loss	$\frac{\mu(1-\mu)}{N} + \mu^2\delta(f_k)$	
Block-based data loss	$\frac{\mu(1-\mu)}{MN} \left \frac{\sin(f_k\pi M)}{\sin(f_k\pi)} \right ^2 + \mu^2\delta(f_k)$	
Two-state Markov model-based data loss	$\frac{1-a^2}{N(1-a^{2N})} \cdot \frac{1}{ 1-az^{-1} ^2} + \mu^2\delta(f_k)$	

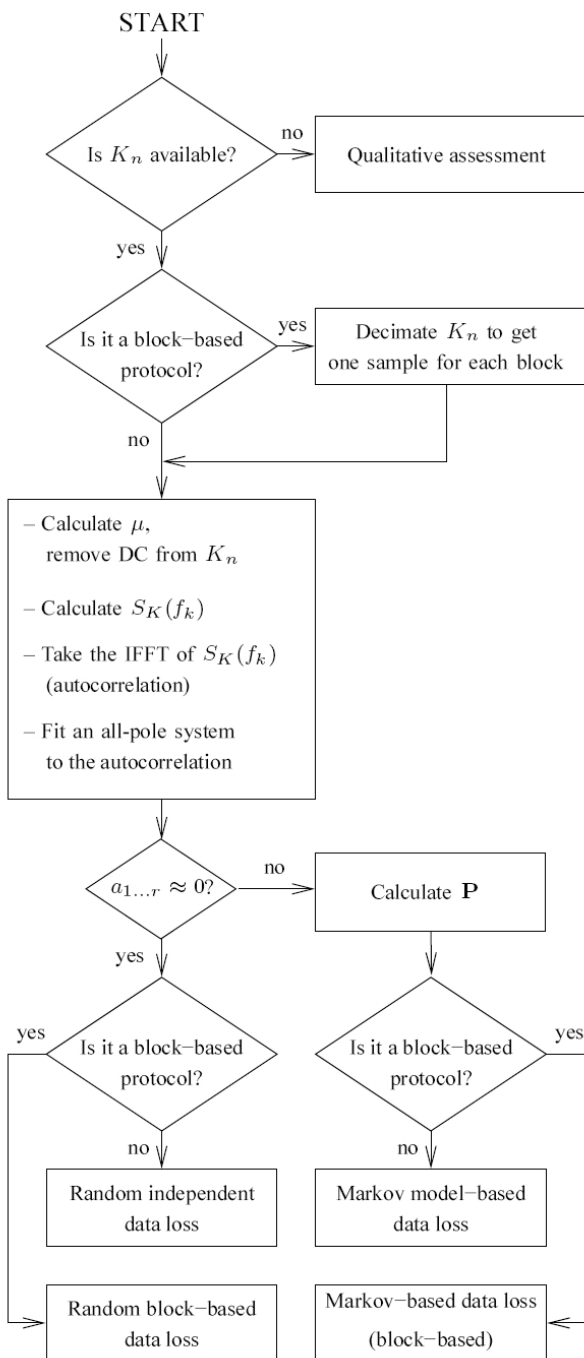


Figure 3. Data loss model identification.

The inverse Fourier transform (IFFT) of $S_K(f_k)$ provides the autocorrelation function of K_n . The FFT block size should be greater than the length of the autocorrelation function. It is not a hard requirement, as the usual FFT block size is much greater than required by K_n .

The main part of the identification is the approximation of $S_K(f_k)$. As the previous investigations have shown, the transfer function can be well approached by an all pole or autoregressive (AR) system. Theoretically it has no pole if K_n is random independent, and has only one pole if K_n is Markov model-based with two states. Generally a Markov model with 2^r states results in an AR system with r poles. The PSD $S_K(f_k)$ of a block based data loss has zeros, but after decimation K_n is either random independent or Markov model-based. As the system is quite simple, there are no special requirements for the identification. We have used a linear prediction filter (LPC) which determines the coefficients of a forward linear predictor by minimizing the prediction error in the least squares sense [14]. To this end, the `lpc` function of Matlab [15] has been applied.

There are $r + 1$ LPC coefficients, where $a_0 = 1$. If the LPC coefficients $a_1, a_2, \dots \approx 0$, the data loss can be handled as random independent. Its only parameter μ has already been calculated. However, if $a_1, a_2, \dots \neq 0$, Markov model-based data loss has been happened.

In the general case of a r th order Markov model, the transition probabilities are to be determined by the formula (16). In the special case of the two-state Markov model-based data loss the parameters p and q can be estimated using the relations (12) and (14):

$$\hat{p} = \hat{\mu}(1 - \hat{a}) + \hat{a}, \quad \hat{q} = \hat{\mu}(\hat{a} - 1) + 1, \quad (19)$$

where

$$\hat{a} = -a_1. \quad (20)$$

In (19) and (20) the hat operator indicates the estimation.

At the end, the information whether the data loss is block based has to be incorporated. If so, the estimators $\hat{\mu}$, \hat{p} , and \hat{q} do not change, but the parameter set has to be completed by the block size M .

4. RESULTS

The procedure presented above has been intensively tested by simulations and measurements. In this section results of both tests are presented. The data processing has followed the procedure given in Figure 3.

4.1. Simulation results

First a random independent data loss pattern has been generated, then it has been identified by the proposed method. The parameters of the simulation are summarized in Table 2, where L is the total length of the record in samples, N is the FFT size. The spectra have been exponentially averaged, with a smoothing factor α . The constant μ is the parameter of the data loss model. At the end a 10th order LPC model has been fitted, in order to check the dynamic behavior of the data loss. The identified model parameters are the following:

$$\hat{\mu} = 0.9900, \quad \hat{a} = 0.0013. \quad (21)$$

Note that $\hat{a} = -a_1$ is the spectral parameter of the two-state Markov model, according to (20). As already $a_1 \approx 0$, the random independent data loss has been verified. The PSD of the model has also been calculated. The upper plot of Figure 4 shows the PSD of the data loss pattern (green line), and the PSD of the model (blue line). The coefficients of the 10th order AR system are depicted in the lower plot. It can be seen that the fitted PSD is in good accordance with the generated one. All the AR coefficients equal approximately zero except the first one.

The second simulation example deals with two-state Markov-based data loss. The parameters of the simulation are summarized in Table 3, where L is the total length of the record in samples, N is the FFT size, and α is the smoothing factor, again. The constants p and q are the parameters of the Markov model. A 10th order LPC model has been fitted as before. The identified model parameters are the following:

$$\begin{aligned} \hat{\mu} &= 0.9118, & \hat{a} &= 0.8872, \\ \hat{p} &= 0.9900, & \hat{q} &= 0.8971. \end{aligned} \quad (22)$$

Table 2. Main data of the first simulation.

Record length L	FFT length N	Smoothing factor α	μ
10^6	1024	0.01	0.9900

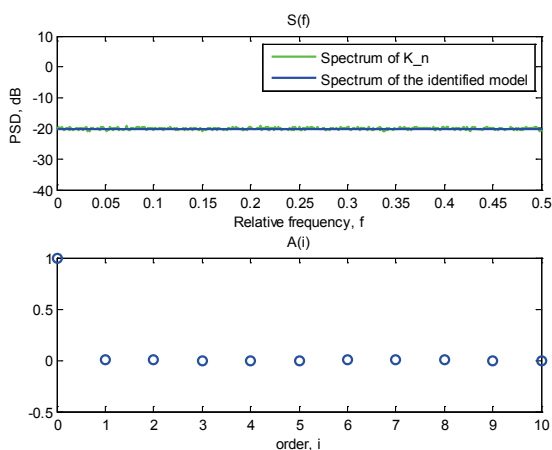


Figure 4. Identification of a random independent data loss: PSD of the data loss pattern, and the PSD of the model (upper plot). Coefficients of the identified AR system (lower plot).

Table 3. Main data of the second simulation.

Record length L	FFT length N	Smoothing factor α	p	q
10^6	1024	0.01	0.9900	0.9000

The original parameters of the model are p and q , therefore the data availability rate μ is a resulted constant. The second LPC coefficient a_1 is non-zero, but the rest of the coefficients are close to zero. The main result of the identification \hat{p} and \hat{q} are really close to the initial parameters given in Table 3. The PSD of the model has also been calculated. The upper plot of Figure 5 shows the PSD of the data loss pattern (green line), and the PSD of the model (blue line). The coefficients of the 10th order AR system are depicted in the lower plot. Both the spectra and the LPC coefficients verify that the calculated first order model is appropriate for the two-state Markov-based data loss.

4.2. Measurement results

Measurements were carried out by a test system introduced in [13]. In this testbed, wireless sensors perform real-time data collection, and they transmit the data to a PC through a gateway node. In this measurement we used only one sensor. The data sent by the sensor are recorded and processed on the PC. Since data transmission and collection is performed in a hard real-time manner, there is no possibility to apply any acknowledge mechanism for the indication and retransmission of lost packets, hence data loss is inevitable. The data loss is recognized by a time-out mechanism. The sampling frequency is $f_s = 1800$ Hz, and the sensor transmits data packets of $M = 25$ samples. If data loss occurs, it can be described by the block-based model.

In the first measurement setup, the sensor and the gateway were placed 4 m away from each other in a room, and the sensor was placed near to a big wardrobe made of metal, in order to degrade the radio transmission properties. The extensive metal surfaces of the wardrobe without any special arrangement (e.g., connection to the earth) produced data loss in the communication. Depending on the physical circumstances, data loss rates in the range of [0.1 ... 30]% could be detected. Now the analysis results for the 3.75% case are introduced. The parameters of the measurement are summarized in Table 4, where L_B is the length of the record in

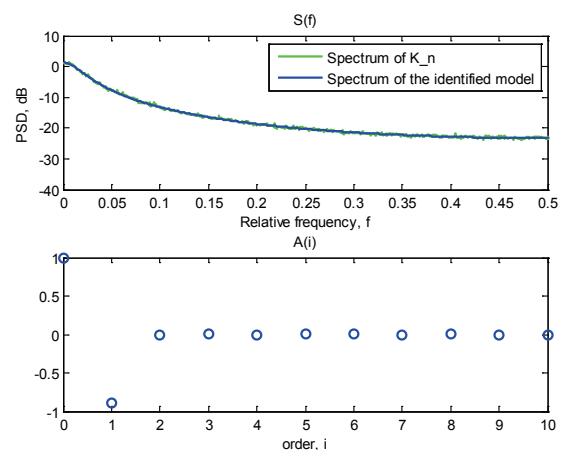


Figure 5. Identification of a Markov-model based data loss: PSD of the data loss pattern, and the PSD of the model (upper plot). Coefficients of the identified AR system (lower plot).

Table 4. Main data of the first experiment.

Record length in blocks, L_B	Duration of the record, t_m	FFT length, N	Smoothing factor, α
4638	64.4 sec	1024	0.01

blocks, while t_m is its duration. The parameter N denotes the FFT size, and α is the smoothing factor.

The result of the identification can be seen in Figure 6. The estimated data loss parameters are the following:

$$\begin{aligned} \hat{\mu} &= 0.9625, & \hat{a} &= 0.0201, \\ \hat{p} &= 0.9632, & \hat{q} &= 0.0569. \end{aligned} \quad (23)$$

In order to check our assumptions, the estimators \hat{p} and \hat{q} of a two-state Markov-model have also been calculated. The system is usually in the '1' state, and if it moves to '0', it has a small probability that stays also in '0' for the next time instant. The upper plot of Figure 6 shows the PSD of the measured data loss pattern (green line), and the PSD of the model (blue line). The coefficients of the 10th order AR system are depicted in the lower plot.

Both the graphical result and the estimated parameters imply that this radio communication suffers from random independent block-based data loss.

The second measurement aimed the investigation of a different data loss mode. A mobile phone has been placed next to the wireless sensor, and the WiFi function of the phone has been activated by playing an on-line media stream. As both devices use the same 2.4 GHz radio band, the communication of the phone causes a disturbance for the wireless sensor. The parameters of the measurement are summarized in Table 5, where L_B is the length of the record in blocks, while t_m is its duration. The parameter N denotes the FFT size, and α is the smoothing factor.

The result of the identification can be seen in Figure 7. The estimated data loss parameters are the following:

$$\begin{aligned} \hat{\mu} &= 0.9533, & \hat{a} &= 0.2529, \\ \hat{p} &= 0.9651, & \hat{q} &= 0.2878. \end{aligned} \quad (24)$$

As $a_1 \neq 0$, the data loss introduced by the WiFi function of the mobile phone cannot be random independent. Nevertheless, a two-state Markov-based model can be well fitted to this data loss pattern, as the rest of the coefficients are

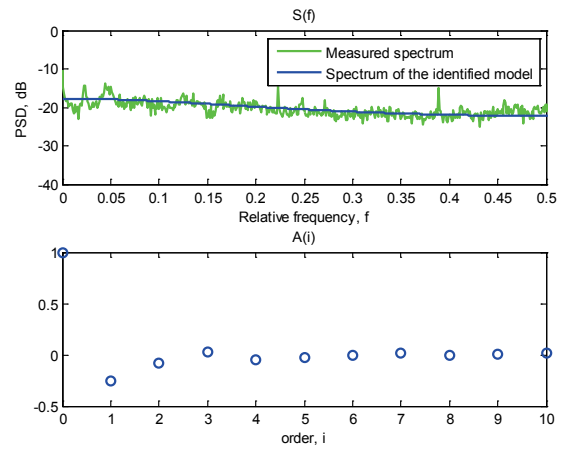


Figure 7. Identification results of the second experiment: PSD of the measured indicator function, and the PSD of the model (upper plot). Coefficients of the identified AR system (lower plot).

sufficiently small ($a_2, \dots \approx 0$). The PSD of the model has also been calculated. The upper plot of Figure 7 shows the PSD of the measured data loss pattern (green line), and the PSD of the model (blue line). The coefficients of the 10th order AR system are depicted in the lower plot.

The last example demonstrates yet another lossy measurement. The wireless sensor has been carried by a walking person in the building, resulting in an occasionally very large distance from the base station. The parameters of the measurement are summarized in Table 6, where the parameters L_B , t_m , N , and α have the same meaning as before.

The result of the identification can be seen in Figure 8. Based on the PSD of the data loss pattern, we have tried to fit a two-state Markov model. The estimated data loss parameters are the following:

$$\begin{aligned} \hat{\mu} &= 0.8829, & \hat{a} &= 0.5915, \\ \hat{p} &= 0.9522, & \hat{q} &= 0.6393. \end{aligned} \quad (25)$$

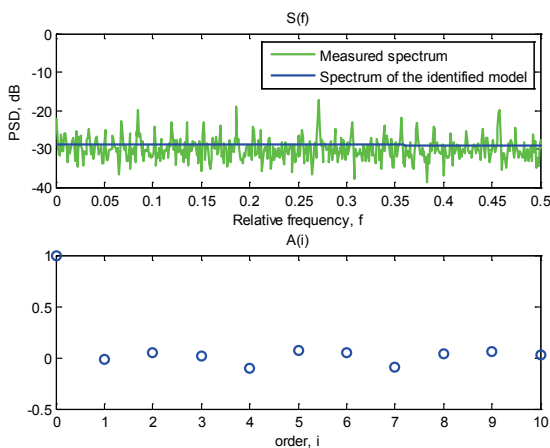


Figure 6. Identification results of the first experiment: PSD of the measured indicator function, and the PSD of the model (upper plot). Coefficients of the identified AR system (lower plot).

Table 5. Main data of the second experiment.

Record length in blocks, L_B	Duration of the record, t_m	FFT length, N	Smoothing factor, α
25988	361 sec	1024	0.01

Table 6. Main data of the third experiment.

Record length in blocks, L_B	Duration of the record, t_m	FFT length, N	Smoothing factor, α
10200	142 sec	1024	0.01

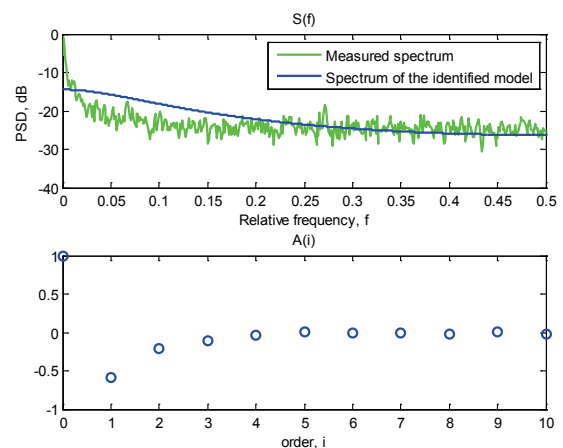


Figure 8. Identification results of the third experiment: PSD of the measured indicator function, and the PSD of the model (upper plot). Coefficients of the identified AR system (lower plot).

The PSD of the model has also been calculated. The upper plot of Figure 8 shows the PSD of the measured data loss pattern (green line), and the PSD of the model (blue line). The coefficients of the 10th order AR system are depicted in the lower plot. In this case the PSD of the model does not fit well the measured PSD. It can also be observed that the higher-order LPC parameters are non-zero, as well.

Finally we have found that a third-order Markov model describes the data loss well. The estimated LPC parameters are the following:

$$\mathbf{a} = [1, -0.6031, -0.2235, -0.1483]. \quad (26)$$

The PSD of the model has also been calculated. Figure 9 shows the PSD of the measured data loss pattern (green line), and the PSD of the model (blue line). Here the graphical representation of the spectral coefficients has already been omitted. The model already fits to the measured spectrum, the identification is successful. Now the state transition probabilities can be calculated using the formula (16):

$$\mathbf{P} = \begin{bmatrix} 0.978 & 0 & 0.83 & 0 & 0 & 0 & 0 & 0 \\ 0.022 & 0 & 0.17 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & 0 & 0.61 & 0 \\ 0 & 0 & 0 & 0 & 0.25 & 0 & 0.39 & 0 \\ 0 & 0.37 & 0 & 0.23 & 0 & 0 & 0 & 0 \\ 0 & 0.63 & 0 & 0.23 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0.003 \\ 0 & 0 & 0 & 0 & 0.85 & 0 & 0 & 0.997 \end{bmatrix}. \quad (27)$$

Unfortunately, the probability matrix \mathbf{P} has a size of 8 by 8 as the model has $2^3 = 8$ states. Note that the zero probabilities belong to impossible state transitions. Here the probabilities $P_{11} = 0.978$ and $P_{88} = 0.997$ are heightened. They can be interpreted as the probability that the system remains in the 'data loss' and the 'data available' state, respectively.

During the walk the distance between the sensor node and the PC has been continuously increased and decreased, resulting in continuously increasing and decreasing data loss. It means that the data loss is not stationary which explains the necessity of the higher order Markov-model.

5. CONCLUSIONS

Recently the analysis of measurement data loss by the spread of sensor networks and Internet-based technology has gained

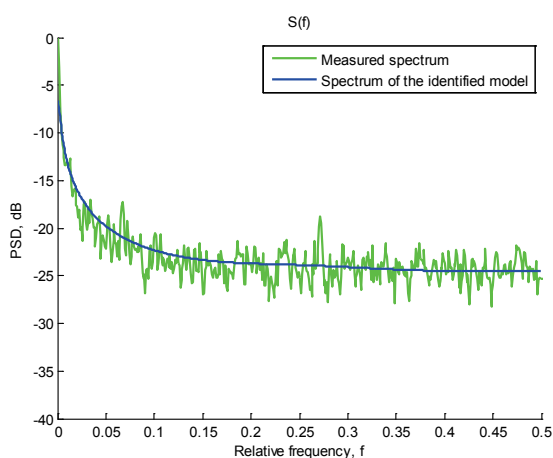


Figure 9. Identification of a third order Markov-model in the third experiment: PSD of the measured indicator function, and the PSD of the model.

importance. The investigation of the FFT based PSD estimation in the case of data loss has discovered the exact relation between spectral leakage and some data loss models. This paper introduced the inverse procedure: the data loss model can be identified by the Fourier transform of the so-called data availability indicator function. The identification procedure has been elaborated for random independent, random block-based, and Markov model-based data loss. Frequency domain identification also supports the verification process, as model fitting and checking can be accomplished simultaneously. The use of the Fourier transform is the main novelty of our approach. By its use not only the parameters of the data loss model can be calculated, but the model order can be determined, as well.

The method has been intensively tested by simulations and measurements. Based on the experiences, the proposed procedure is a promising solution for data loss model identification. Further research is required if the data availability function is not stationary or it is not directly available.

REFERENCES

- [1] L. Kong et al., "Data Loss and Reconstruction in Wireless Sensor Networks", Proc. IEEE Infocom 2013, Turin, Apr. 14-19, 2013, pp. 1654-1662.
- [2] M. Mathiesen, G. Thonet, N. Aakwaag, "Wireless ad-hoc networks for industrial automation: current trends and future prospects", Proc. of the IFAC World Congr., Prague, Czech Republic, July 4-8, 2005, pp. 89-100.
- [3] M. Yajnik, et al., "Measurement and modelling of the temporal dependence in packet loss", in: Proc. IEEE Infocom 1999, pp. 345-352.
- [4] M. Ellis et al., A two-level Markov model for packet loss in UDP/IP-based real-time video applications targeting residential users, Computer Networks 70 (2014), pp. 384-399.
- [5] J. Rachwalski, "Analysis of packet loss pattern for concatenated transmission channels using burst ratio parameter", PhD Dissertation, AGH University of Science and Technology, 2016.
- [6] H. Sanneck, G. Carle, and R. Koodli. "Framework model for packet loss metrics based on loss runlengths", PROC SPIE INT SOC OPT ENG. 3969 (2000).
- [7] G. Sinopoli et. al, Kalman Filtering with Intermittent Observations, IEEE Trans. Autom. Control 49 (2004) pp. 1453-1464.
- [8] L. Sujbert, Gy. Orosz, FFT-based Spectrum Analysis in the Case of Data Loss, IEEE Trans. Instrum. Meas. 65 (2016) pp. 968-976.
- [9] J. S. Bendat, A. G. Piersol, Random Data: Analysis and Measurement Procedures, John Wiley and Sons, Inc., New York, London, Sidney, Toronto, 1971.
- [10] T. Nagayama, B. F. Spencer, G. Agha, and K. Mechitov, "Model-based Data Aggregation for Structural Monitoring Employing Smart Sensors", 3rd Int. Conf. on Networked Sensing Systems (INSS).
- [11] O. Hohlfeld, R. Geib, G. Hasslinger, "Packet Loss in Real-Time Services: Markovian Models Generating QoE Impairments", 16th Int. Workshop on Quality of Service, Enschede, June 2008, pp. 239-248.
- [12] P. Boufounos, "Generating Binary Processes with All-pole Spectra", IEEE Int. Conf. Acoustics, Speech and Signal Processing, Honolulu, HI, vol. 3, Apr. 15-20, 2007, pp. 981-984.
- [13] Gy. Orosz, L. Sujbert, G. Péceli, "Testbed for wireless adaptive signal processing systems", Proc. IEEE Instrum. and Meas. Conf., Warsaw, Poland, May 1-3, 2007, pp. 123-128.
- [14] Jackson, L.B., Digital Filters and Signal Processing, 2nd ed., Kluwer Academic Publishers, 1989.
- [15] MATLAB (2010), www.mathworks.com