SAMPLING OF RANDOM DATA STREAMS

Gustav CEPCIANSKY¹, Ladislav SCHWARTZ²

¹Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Univerzitna 1, 010 26 Zilina, Slovakia

gcepciansky@zoznam.sk, schwartz@fel.uniza.sk

Abstract. Modern telecommunication networks work on the transmission method of common data streams in which data bursts consisting of packets that further consist of particular bits are multiplexed from various traffic sources. The larger amount of data is transmitted through a transmission medium (optical fibre), the more frequently bursts occur, and the lower amount of data, the more rarely they follow. If it is required to monitor how large amount of data is being transmitted in a network branch in order to find out, to which measure of this branch is occupied, it is not necessary to take each information unit (each packet or even each particular bit). It will do if information whether a data burst occurs in the transmission or does it not occur is taken in certain time intervals – with a certain sampling frequency. The paper deals with this sampling intervals.

Keywords
Sampling, traffic, data burst, passive optical network.

1. Introduction
There are many types of data bursts on a transmission route between two network nodes from various traffic sources. The data burst length is considered as a random variable. According to Fig. 1, the burst train in a common data stream can be described as a random process \( X(t) \) consisting of series of random occurring rectangular pulses with a random amplitude that only gains 2 values of \( x_0 = 0 \) and \( x_1 = 1 \) and with a random time length \( \Theta \) that gains values of \( \tau_k, k = 1, 2, ... \)

The mean level or the average value of this random process

\[
 y = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{N} \delta_k ,
\]

also expresses the average traffic use of a communication channel, e.g. with which part of the total time \( T \) theoretically infinitely long the channel was occupied by the transmitted data bursts. In other words, the value \( y \) also expresses the probability that the communication channel will be in any moment of sampling \( t_v \) found as occupied.

Further, let \( \lambda \) be the count of seizures of the communication channel by the transmitted data bursts and \( \mu \) the count of releases of the communication channel during a time unit. The ratio

\[
 a = \frac{\lambda}{\mu},
\]

represents the traffic offered to the communication channel. The inverse value of \( \mu \)

\[
 \overline{\Theta} = \frac{1}{\mu} = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \delta_k ,
\]

is the average burst length.

2. Theory

2.1 Random Process Characteristics
The occurrence of data bursts in the common data stream can be considered as the Markov random process because the features of such a process are fulfilled:

- a data burst occurrence is fully random,
- bursts last for a random time,
- the probability of a data burst occurrence or end of a burst during a very small time interval \( \Delta t \) will be proportional to the duration of this interval.
These features represent a hidden deterministic element in the random process. The consequence of this is that it is possible to explicitly derive basic characteristics – correlation and co-variation functions, dispersion and spectrum for that random process [1], [2], [3], [4].

1) Solution of Random Process

The random process gains only 2 states: \( S_0 \) – communication channel is free and \( S_1 \) – communication channel is occupied.

The process stays in the zero state \( S_0 \) if no data burst occurs during \( \Delta t \), the probability of which is:

\[
p_{00}(t, t + \Delta t) = 1 - \lambda \Delta t . \tag{4}
\]

The process transits from the state \( S_0 \) to the state \( S_1 \) (communication channel will be seized) if a data burst occurs during \( \Delta t \). The probability of this is:

\[
p_{01}(t, t + \Delta t) = \lambda \Delta t . \tag{5}
\]

The process stays in the state \( S_1 \) if a data burst does not finish during \( \Delta t \), the probability of which is:

\[
p_{11}(t, t + \Delta t) = 1 - \mu \Delta t . \tag{6}
\]

The process transits from the state \( S_1 \) to the state \( S_0 \) (communication channel will be released) if a data burst finishes during \( \Delta t \). The probability of that is:

\[
p_{10}(t, t + \Delta t) = \mu \Delta t . \tag{7}
\]

The probability that more than 1 change occurs in the state of the channel during \( \Delta t \) is equal to zero.

The probabilities \( p_{00}, p_{01}, p_{10}, p_{11} \) are arranged into the matrix of transition probabilities:

\[
P(t, t + \Delta t) = \begin{pmatrix}
p_{00}(t, t + \Delta t) & p_{01}(t, t + \Delta t) \\
p_{10}(t, t + \Delta t) & p_{11}(t, t + \Delta t)
\end{pmatrix} = \begin{pmatrix}
1 - \lambda \Delta t & \lambda \Delta t \\
\mu \Delta t & 1 - \mu \Delta t
\end{pmatrix} = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \cdot \lambda \Delta t + \begin{pmatrix}
\mu & -\lambda \\
-\mu & \mu
\end{pmatrix} \cdot \Delta t = J + A \cdot t . \tag{8}
\]

The matrix \( J \) is the unit one. The matrix \( A \) is the matrix of intensities of transition probabilities. In general, the elements of the matrix \( A \) can be time dependent. But if it is supposed the ergodicity of the random process, e.g. the Markov process is homogenous, the time dependence does not exist.

By the solution of the basic equation of the Markov random process

\[
\frac{dp(t)}{dt} = p(t) \cdot A , \tag{9}
\]

it can be calculated, with which probability \( p_{0k}(t) \) the communication channel will be free and with which probability \( p_{1k}(t) \) the communication channel will be occupied by a data burst in a time \( t \). The equation is solved by means of Laplace transform:

\[
L[p(t)] = F(s) = \int_0^\infty p(t) \cdot e^{-st} dt , \tag{10}
\]

\[
L\left[ \frac{dp(t)}{dt} \right] = L[p(t) \cdot A] . \tag{11}
\]

Hence, when the transform is carried out

\[
F(s) = p(0) \cdot [s \cdot J - A]^{-1} . \tag{12}
\]

The sought vector \( p(t) \) will be obtained after the inverse Laplace transform

\[
p(t) = L^{-1} [F(s)] = L^{-1} \left[ p(0) \cdot [s \cdot J - A]^{-1} \right] . \tag{13}
\]

First, let \( s \cdot J - A \) be written in:

\[
s \cdot J - A = s \cdot \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} = \begin{pmatrix}
-\lambda & \lambda \\
-\mu & \mu
\end{pmatrix} = \begin{pmatrix}
s + \lambda & -\lambda \\
-\mu & s + \lambda
\end{pmatrix} . \tag{14}
\]

The inverse matrix:

\[
[s \cdot J - A]^{-1} = \begin{pmatrix}
1 & \frac{s + \mu}{s + \lambda + \mu} & \lambda \\
\frac{s + \mu}{s + \lambda + \mu} & \frac{s + \lambda}{s + \lambda + \mu} & \frac{\lambda}{s + \lambda + \mu}
\end{pmatrix} . \tag{15}
\]

To carry out the inverse Laplace transform, the elements of the inverse matrix shall be decomposed into the sum of particle fractions. Then the basic system of equations (13) acquires the form:

\[
F(s) = p(0) \cdot \begin{pmatrix}
\lambda \\
\lambda + \mu
\end{pmatrix} + \begin{pmatrix}
\lambda \\
\lambda + \mu
\end{pmatrix} \cdot \begin{pmatrix}
\frac{1}{s + \lambda} & \frac{1}{s + \lambda + \mu}
\end{pmatrix} . \tag{16}
\]

The result will be obtained by the inverse Laplace transform:
The random variable does not depend on the value
Then:

\[ P(X(0) = 1 \cap X(\tau) = 1) = P\{X(0) = 1, P\{X(\tau) = 1 / X(0) = 1\} \} \quad (22) \]

Supposing stability of the random process with
probability \( P\{X(0) = 1\} \), that the communication channel
will be occupied at the beginning (in the time \( t = 0 \)). The
conditional probability is equal to \( P\{X(\tau) = 1 / X(0) = 1\} \),
if the communication channel was occupied at the
beginning, and if it will be also occupied after a time \( \tau \), it
can be calculated from the equation (19), in which \( p_0(0) = 0 \) and \( p_1 = 1 \):

\[ R(\tau) = P\{X(0) = 1\} \cdot P\{X(\tau) = 1 / X(0) = 1\} \]

\[ = p_1(\tau) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} \cdot e^{-(\lambda + \mu)\tau} = \]

\[ = \frac{\lambda}{\mu + 1} \left[ 1 + \frac{\mu}{\lambda} e^{-(\lambda + \mu)\tau} \right] \quad (23) \]

\[ = \frac{a}{a+1} \left[ 1 + \frac{1}{a} e^{-a(1+\frac{\tau}{a})} \right]. \quad (24) \]

thereby the equations (2) and (3) were used, too.

Then the correlation function \( R(\tau) \) using (22) and
(24) will be:

\[ R(\tau) = y \cdot \frac{a}{a+1} \left[ 1 + \frac{1}{a} e^{-a(1+\frac{\tau}{a})} \right]. \quad (25) \]

The following relationship holds between traffic
a offered to a communication channel and their traffic
use \( y \):

\[ y = (1 - p_1) \cdot a. \quad (26) \]

where \( p_1 \) is the probability that a channel is occupied in
the steady state:

\[ p_1 = p_1(\infty) = \frac{a}{a+1} \lim_{t \to \infty} \left[ 1 + \frac{1}{a} e^{-a(1+\frac{\tau}{a})} \right] = \frac{a}{a+1}. \quad (27) \]

Substituting into (24) we obtain:

\[ R(\tau) = (1 - p_1) \cdot a \cdot \frac{a}{a+1} \left[ 1 + \frac{1}{a} e^{-a(1+\frac{\tau}{a})} \right] = \]

\[ \left( \frac{a}{a+1} \right)^2 \left[ 1 + \frac{1}{a} e^{-a(1+\frac{\tau}{a})} \right]. \quad (28) \]
Co-variation function is:

\[ K(\tau) = R(\tau) - m^2. \quad (29) \]

where \( m \) is the mean value (mean level) of the random process. Then

\[ m^2 = \mathbb{E}[1 - p_1 \cdot a]^2 = \left[ \frac{1 - \frac{a}{a + 1}}{a + 1} \right]^2 = \left( \frac{a}{a + 1} \right)^2. \quad (30) \]

So the co-variation function will be:

\[ K(\tau) = \frac{a}{a + 1} \left[ 1 + \frac{1}{a} e^{-(a + 1)\frac{\tau}{\pi}} \right] \left( \frac{a}{a + 1} \right)^2 = \frac{a}{(a + 1)^2} e^{-(a + 1)\frac{\tau}{\pi}}. \quad (31) \]

If \( a \) will be expressed by means of \( y \) from (26) and (27), we obtain:

\[ a = \frac{y}{1 - y}, \quad y < 0, \]

and then

\[ K(\tau) = \frac{y}{1 - y} \left[ 1 + \frac{1}{y} e^{\frac{\tau}{\pi}} \right] = y \cdot \left( 1 - y \right) e^{\frac{\tau}{\pi}}. \quad (32) \]

Dispersion is:

\[ \sigma^2 = K(0) = y(1 - y). \quad (33) \]

3) Spectrum

The spectrum of a random process can be calculated by means of the Wiener-Khintchine transform which is the Fourier transform of a co-variation function:

\[ S(\omega) = \int_{-\infty}^{\infty} K(\tau) \cdot e^{-j\omega \tau} d\tau. \quad (34) \]

Let in the equation (32)

\[ \alpha = \frac{1}{(1 - y)\cdot \frac{\pi}{\sqrt{y}}}, \quad (35) \]

and \( \sigma^2 \) is given by (33). Then

\[ S(\omega) = \int_{-\infty}^{\infty} \sigma^2 \cdot e^{-\frac{\omega}{\sqrt{y}}} \cdot e^{-j\omega \tau} d\tau = \]

\[ = \sigma^2 \left[ \int_{0}^{\infty} e^{-(\alpha + j)\omega} d\tau + \int_{0}^{-\infty} e^{(\alpha - j)\omega} d\tau \right] = \]

\[ = \sigma^2 \left( \frac{1}{\alpha - j\omega} + \frac{1}{\alpha + j\omega} \right) = \sigma^2 \cdot \frac{2\alpha}{\alpha^2 + \omega^2}. \quad (36) \]

Fig. 2: Shape, co-variation function and spectrum of the random process.

The data bursts train, the shape of the co-variation function and the spectrum belonging to it are drawn on figure 2 for the comparison. The spectrum on figure 2 c) has a shape of a probability distribution which can advantageously be used for determination of the sampling frequency.

2.2 Sampling Frequency

If the spectrum \( S(\omega) \) is to be considered for a probability distribution, it must be fulfilled:

\[ c \cdot \int_{-\infty}^{\infty} S(\omega) d\omega = 1, \quad (37) \]

hence

\[ c = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \frac{2\omega^2}{\alpha^2 + \omega^2} d\omega = \frac{2\alpha^2}{\sigma^2} \int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \frac{1}{2\pi \sigma^2}. \quad (38) \]

So the distribution will have the shape:

\[ f(\omega) = \frac{\alpha}{\pi} \cdot \frac{1}{\alpha^2 + \omega^2}, \quad (39) \]

which is the Cauchy distribution and it expresses the probability density of occurrence of amplitudes of particular spectral components. Then the elementary rectangle \( f(\omega) \cdot \Delta \omega \) in figure 3 gives the probability with which the components in the interval from \( \omega \) to \( \omega + \Delta \omega \) will occur in the spectrum. All spectral components in the range from \(-\infty\) to \(+\infty\) occur with the probability \( P = 1 \) because the integral from \( f(\omega) \) in this range equals to 1. That means if we want to catch all changes in the course of the random process with the total certainty, it must be sampled with the infinite high frequency. Therefore, let the sampling frequency is limited by the frequency \( \omega_v \) and then the probability \( P \) with which each change will be caught in the random process at this sampling frequency will be calculated:

\[ f(\omega) = \frac{\alpha}{\pi} \cdot \frac{1}{\alpha^2 + \omega^2}, \quad (39) \]
The sampling frequency or the sampling period can be calculated from this equation for the given probability $P$:

$$f_s = \frac{1}{2\pi} \alpha \cdot \text{tg} \frac{\pi P}{2} = \frac{1}{2\pi(1 - y)} \frac{\pi P}{2},$$

$$T_s = 2\pi(1 - y) \frac{\pi P}{2}.$$  \hspace{1cm} (41)

The equations (41) and (42) give the sampling frequency or sampling interval, respectively in the case data bursts occur totally randomly and their length is totally random, too. It is also the case when the limit situation can occur when data bursts follow consecutively each other and the gap between them is infinitely small ($y \to 1$), or the data bursts length is infinitely short ($v \to 0$), or both cases occur simultaneously.

### 3. Experimental

#### 3.1 Analysis of Results

The equations (41) and (42) give the sampling frequency or sampling interval, respectively in the case data bursts occur totally randomly and their length is totally random, too. It is also the case when the limit situation can occur when data bursts follow consecutively each other and the gap between them is infinitely small ($y \to 1$), or the data bursts length is infinitely short ($v \to 0$), or both cases occur simultaneously.

#### 3.2 Practical Application

An example where this theory could be applied is the gigabit passive optical network (GPON) [5], [6], [7], [8], [9]. The GPON is an access network where more subscribers (up to 128) share the same optical fibre. The upstream communication between an optical network termination (ONT) on the subscriber side and the broadband node – optical line termination (OLT) on the network side goes through optical pulses bursts called transmission containers (T-CONTs) carrying data streams. It is necessary to hinder overlapping T-CONTs of simultaneously communicating ONTs. Therefore a time window is matched to each T-CONT from OLT.

This window defines the T-CONT length and the time instants in which an ONT may send T-CONTs with data. To determine these properly, The OLT either receives and decodes dynamic bandwidth assignment (DBA) reports coming from ONT, or monitors the traffic load on each branch in the optical network. The first method is called the status DBA and the second the non-status DBA.

The status DBA requires exchange of various control information between an OLT and an ONT in headers of data batches. The non-status DBA does not require any information exchange but it is necessary to know how often the traffic status shall be sampled on a network branch. Here the formula (41) or (42) can be helpful.

The GPON bit rate is 1 244,16 Mbit/s in the upstream. The shortest data unit is one 125 µs long frame which is performed by 19 440 bytes. The longest data unit in the upstream is $2^{16} = 65 536$ bytes which lasts 421,4 µs. Let’s take the mean value of these 2 figures for the average burst length:

$$\bar{T} = \frac{125 + 421,4}{2} = 273,2 \mu s.$$ \hspace{1cm} (43)

Considering $P = 0.95$, we obtain for sampling interval from (42):

$$T_s = 2\pi(1 - y) \cdot 273,2 \cdot \text{cotg} \frac{0.95\pi}{2} \approx 65 893 \cdot (1 - y) \text{[ms]},$$ \hspace{1cm} (44)

$$T_s \approx 66 \cdot (1 - y) \text{[ms]}. \hspace{1cm} (45)$$

The longest sampling interval will be when there is no traffic load ($y = 0$):

$$T_{s, \text{max}} = 66 \text{ ms}. \hspace{1cm} (46)$$

In order to avoid collision of 2 consecutive T-CONTs, the guard interval consisting of 32 bits is introduced in the upstream. They perform the guard time $32/1 244 160 000 = = 25,72$ ns. During this time, ONT emits no energy into the optical fibre. That means there are 4-byte gaps in the 65 536-bytes bursts. That represents the traffic load of

$$y = \frac{65 536 - 4}{65 536} = 0.99994. \hspace{1cm} (47)$$

and so the shortest sampling interval will be:

$$T_{s, \text{min}} = 66 \cdot (1 - 0.9994) = 0.004 \text{ms} = 4 \mu s. \hspace{1cm} (48)$$

It is necessary to take into consideration that the maximum and minimum sampling intervals $T_{s, \text{max}}$ and $T_{s, \text{min}}$ are calculated on condition that the data bursts occur randomly on the optical fiber and their length is also random in the range from 125 to 421,4 µs.

Even the minimum sampling interval of 4 µs is 5000-times longer than if each particular bit should be
sampled which would give the sampling interval of $1/1244160000 = 0.8$ ns.

4. Conclusion

Application of this theory on data streams control may save many working operations of digital microcontrollers and thus enhance their performance.

Acknowledgments

The authors gratefully acknowledge support from the VEGA project No. 1/0375/08 “The Analytic Model Home Access Network New Generation”.

References


About Authors


Ladislav SCHWARTZ was born on 1950 in Zilina. In 1974 graduated from the University of Zilina, Department of Telecommunications with the Master’s degree (M.Sc). From 1974 to 1991 he worked in the Research Institute of Computer Technology in Zilina as the head of the Department of Data Communications. In 1986 he was awarded the title Ph.D. and in 1991 he became a full time teacher and research worker of the University of Zilina. In 1999 he was appointed an associate professor. His main focuses are functions, reliability, security telecommunication and computer networks and data communications.