STOCHASTIC MODELING OF THE NETWORK INTRUSION DETECTION THRESHOLD

Vladica Stojanović, PhD[1]
University of Criminal Investigation and Police Studies, Belgrade, Serbia

Mihailo Jovanović, PhD
Office for Information Technologies and eGovernment, Belgrade, Serbia

Brankica Popović, PhD
University of Criminal Investigation and Police Studies, Belgrade, Serbia

Petar Čisar, PhD
University of Criminal Investigation and Police Studies, Belgrade, Serbia

Kristijan Kuk, PhD
University of Criminal Investigation and Police Studies, Belgrade, Serbia

Introduction

Intrusion detection systems (IDSs) are primarily focused on identifying potential incidents, i.e. unauthorized access, information retrieval, and log-in attempts. In addition, IDSs are also used for other purposes, such as identifying various problems related to certain security policies, recording all existing and potential threats, and deterring users from violating security policies (Eskin, 2000; Fengmin, 2003). Therefore, it can be said that the basic role of IDSs is to register all the important information and knowledge related to the observed events, to inform the security administrators about the recorded events and to generate the corresponding reports. These systems use several different response techniques, which include, among other things, detecting the intrusion and stopping the attack itself, changing the security environment, or changing the content of the attack (Čisar P. , Maravić-Čisar, Popović, Kuk, & Vuković, 2022; Čisar, Popović, Kuk, & Vuković, 2022).

In anomaly detection, two approaches are commonly used: algorithms with adaptive and non-adaptive (fixed) thresholds. Most authors (Sorensen, 2004; Spathoulas & S. Katsikas, 2010) are of the opinion that systems with fixed thresholds are not robust enough and that tests with such thresholds will give unsatisfactory results with normal network traffic variations. On the other hand, the adaptive approach is a way that has a positive effect on reducing the number of false positives (Jovanović et al., 2018). However, the adaptive approach also has the disadvantage that the security system based on this approach can be fooled by applying an adequate attack strategy. Otherwise, several different versions of adaptive algorithms are used in practice, developed with the intention of improving the efficiency of intrusion detection (Čisar & Maravić-Čisar, 2010a; 2010b; 2012)

The basic assumption in this article is the interpretation of network traffic in the form of time series, whose pronounced fluctuations show potential intrusions into the system. Therefore, in order to successfully detect them, an adaptive threshold in the form of a random variable (RV) is introduced, with which such intrusions can be registered. Also, it is easy to see that the observed time series have

1 vladica.stojanovic@kpu.edu.rs

non-linear and non-stationary dynamics, which is usually reflected in the increasing complexity of their stochastic structure. In order to describe their dynamics, the so-called General Split-BREAK (GSB) process is proposed here, which has already been applied in modelling various time series with persistent and accentuated fluctuations. The basic form of this stochastic model was introduced by (Stojanović, Popović, & Popović, 2011; 2014; 2015), where RVs with Gaussian distribution are used as an innovation series. Then, some recently and more general results of the Split-BREAK process, with Laplacian and Cauchy distributions, respectively, are reported by (Jovanović, Stojanović, Kuk, Popović, & Čisar, 2022), as well as (Ljajko, Stojanović, Tošić, & Božović, 2023). It is worth to point out that, for the purpose of practical application to intrusion detection, some of these forms of GSB processes are considered and then compared for their effectiveness.

Definition and Main Properties of the GSB Process

Basic assumptions about the GSB process can be made based on its corresponding time series, as is given below. Specifically, the GSB process consists of the following three components:

i) (ε_t) is an innovation series, that is, independent identically distributed (IID) random variables (RVs) with some stochastic distribution of absolute-continuous type;

ii) (m_t) is a series of martingale means given by recurrence relation:

$$m_t = m_{t-1} + q_{t-1}\,\varepsilon_{t-1} = m_0 + \sum_{j=0}^{t-1} q_j \varepsilon_j, \tag{1}$$

where is almost surely (as) $m_0 \overset{as}{=} \mu\ (const)$, $\varepsilon_{-1} = \varepsilon_0 \overset{as}{=} 0$, and

$$q_t = I(\varepsilon_{t-1}^2 > c) = \begin{cases} 1, & \varepsilon_{t-1}^2 > c \\ 0, & \varepsilon_{t-1}^2 \leq c \end{cases} \tag{2}$$

is the Noise-Indicator with the critical value $c > 0$;

iii) (y_t) is a the basic GSB series given as an adaptive decomposition:

$$y_t = m_t + \varepsilon_t. \tag{3}$$

In practice, the innovations (ε_t) are interpreted as a noise series, that is represent deviation (fluctuating) component of the GSB process. Depending on the choice of stochastic distribution, we further distinguish three forms of these innovations, with Gaussian, Laplacian and Cauchy distributions. In the following Table 1 are shown their basic stochastic characteristics: density functions, denoted with $f_\varepsilon(x)$, as well as means and variances, which are well-known statistical measures of average and dispersion, respectively. It can be easily noticed that all these distributions depend on the corresponding parameters that can also be observed in Table 1. At the same time, it is worth emphasizing that the symmetry of these distributions is assumed, so that all of them, expecting a Cauchy distribution, have zero-means. Finally, variances of the Gaussian and Laplace distributions also depend on these parameters, but the variance of the Cauchy distribution is infinite.

Table 1 *Key stochastic characteristics of innovations* (ε_t) *of the GSB process*

Distribution	**Gaussian**	**Laplacian**	**Cauchy**
Density $f_\varepsilon(x)$	$\frac{1}{2\sigma\sqrt{\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$	$\frac{1}{2\lambda}\exp\left(-\frac{\lvert x\rvert}{\lambda}\right)$	$\frac{\lambda}{\pi(x^2+\lambda^2)}$
Mean	0	0	∞
Variance	σ^2	$2\lambda^2$	∞

The second component, the martingale mean series (m_t), represents predictive and stability component of the GSB process, that is the values without emphatic fluctuations. The main role in this has the parameter $c > 0$, named *the critical value of reaction*, which indicates significance of previous realizations of innovations (ε_t) to be included in Equation (1). More precisely, when $q_{t-1} = 0$, the martingale mean m_t is equal to its previous value m_{t-1}, and the main GSB series (y_t), given by Equation (3), is then realized with 'low' fluctuation. This means that there is no intrusion into a specific security system. Otherwise, the case $q_{t-1} = 1$ indicates a pronounced fluctuation of the series (y_t), and thus there is a reason that there was an intrusion and the activation of the corresponding alarm. It is worth pointing out that the series (m_t) and (y_t) depend on the time moment $t \in T$ in which they are observed. In that way, using the previously obtained results on the GSB process, the basic distributional properties of these series (except in the case of Cauchy distributed innovations) can be shown as follows:

ι) Both series (m_t) and (y_t) have the constant and equal mean $E(m_t) = E(y_t) = \mu$.

ιι) The variances of series (m_t) and (y_t) are, respectively,

$$Var(m_t) = 2ta_c\lambda^2,\ Var(y_t) = 2(ta_c + 1)\lambda^2, \quad t \geq 0, \tag{4}$$

where $a_c = E(q_t) = E(q_t^2) = P\{\varepsilon_t^2 > c\}$.

ιιι) The correlation functions of series (m_t) and (y_t) are, respectively,

$$\rho_m(s,t) = \frac{\min(s,t)}{\sqrt{s \cdot t}},\ \rho_y(s,t) = \frac{a_c \min(s,t) + 1}{\sqrt{(a_c s + 1)\cdot(a_c t + 1)}}.$$

According to the presented results, it follows that series (m_t) and (y_t) have the non-constant variances, dependent on the time (t) in which they are observed. Therefore, the correlation functions $\rho_m(s,t)$ and $\rho_y(s,t)$ depend on both time variables t, s, which confirms the non-stationarity of these time series. We emphasize that the condition of non-stationarity is fully consistent with the characteristics of real traffic flows in a certain information system. On the contrary, the stationarity is very important and useful property of time series. It enables, among other things, a simple estimation of the parameters of the corresponding stochastic model. For these reasons, we define another important GSB series, the so-called increment series:

$$X_t = y_t - y_{t-1}, \qquad t = 1, \ldots, T. \tag{5}$$

Using Equations (1)-(3), increments can be represented as follows:

$$X_t = \varepsilon_t - \theta_{t-1}\varepsilon_{t-1}, \tag{6}$$

where $\theta_t = 1 - q_t = I(\varepsilon_{t-1}^2 \leq c)$. In that way, the series (X_t) is stationary stochastic process that operates in two regimes:

ι) Emphasized fluctuations of innovations (ε_t) in the previous moment of time imply equality $\theta_{t-1} = 0$. Thus, Equation (6) becomes $X_t = \varepsilon_t$.

ιι) If the square of fluctuations ε_{t-1} do not exceed the critical value c, it follows $\theta_{t-1} = 1$. Thus, the value of X_t is given as a linear integrated MA(1) process:

$$X_t = \varepsilon_t - \varepsilon_{t-1}.$$

Obviously, the series (X_t) has a structure similar to the ordinary first-order moving average (abbr. MA(1)) processes, which can be applied in their examination. With earlier assumptions, the basic properties of this series, obtained by some simple computation, can be expressed as follows:

Theorem 1. Let (X_t) be the time series defined by Equations (5) and (6). The mean and the variance of this series are, respectively,

$$E(X_t) = 0, \qquad \mathrm{Var}(X_t) = E(X_t^2) = \begin{cases} \sigma^2(b_c + 1), & (\varepsilon_t) \text{ is Gaussian} \\ 2\lambda^2(b_c + 1), & (\varepsilon_t) \text{ is Laplacian} \end{cases},$$

where $b_c = 1 - a_c = P(\varepsilon_{t-1}^2 \leq c)$. In addition, the correlation function of this series is given as follows:

$$\rho_X(h) := \frac{\mathrm{Cov}(X_t, X_{t+h})}{\mathrm{Var}(X_t)} = \begin{cases} 1, & h = 0 \\ -\dfrac{b_c}{b_c + 1}, & h = \pm 1 \\ 0, & \text{otherwise.} \end{cases}$$

As will be seen below, the specific structure of the stationary series (X_t) is important in practical application of the GSB process, as well as in estimating its parameters. As an illustration, Figure 1 shows realizations of the above-mentioned GSB time series, obtained by Monte Carlo simulations of the series (ε_t) with Gaussian distribution.

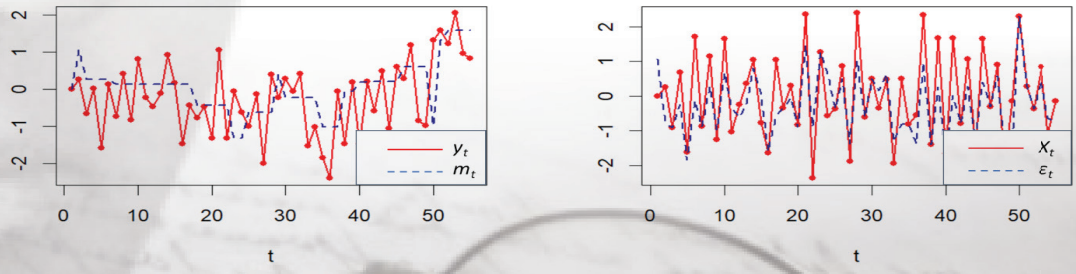


Figure 1 *Dynamics of the GSB series with Gaussian innovations. (Parameters values are: $\mu = 0$ and $c = \sigma = 1$.)*

Parameters Estimation Procedures

In this part, we are discussed the procedures for estimating (unknown) parameters of the GSB process, primarily the critical value $c > 0$ which determines the stochastic threshold of our model. In that cause, an increment series (X_t) is commonly used, because it represents the only observable and

stationary series of GSB processes. Since the series (X_t) is close to the standard MA models, similar estimation procedures are used here. Nevertheless, the specificity of the series (X_t) requires some additional estimation procedures, as well as the examination of the quality of the estimators thus obtained. In the following, we will briefly describe these procedures, first for the Gaussian and Laplace innovations (ε_t).

The first form of estimation give the so-called *moments-based estimators*, obtained from the theoretical moments of the basic GSB series. Using Theorem 1, that is, by solving the expression for the first correlation of the series (X_t) on $b_c \in (0,1)$, the estimator of this parameter is easily obtained as follows:

$$\tilde{b}_c = -\frac{\hat{\rho}_X(1)}{1+\hat{\rho}_X(1)}. \tag{7}$$

Here, $\hat{\rho}_X(1) = (\sum_{t=1}^{T} X_t X_{t-1})/(\sum_{t=1}^{T} X_t^2)$ is the sample autocorrelation of (X_t), and according to Equation (7) it is easy to see that b_c is the appropriate estimate if and only if $0 < \tilde{b}_c < 1$, that is, $-0.5 < \hat{\rho}_X(1) < 0$.

Thereafter, using the estimator $\tilde{b}_c$, in the cases of Gaussian and Laplacian innovations (ε_t), the appropriate parameters σ^2 and λ^2 are, respectively:

$$\tilde{\sigma}^2 = \frac{1}{T(\tilde{b}_c+1)}\sum_{t=1}^{T} X_t^2, \tilde{\lambda}^2 = \frac{1}{2T(\tilde{b}_c+1)}\sum_{t=1}^{T} X_t^2. \tag{8}$$

It is worth pointing out that in (Jovanović, Stojanović, Kuk, Popović, & Čisar, 2022) and (Stojanović, Bakouch, Ljajko, & Božović, 2023) is proven that the estimators $\tilde{b}_c$, $\tilde{\sigma}^2$ and $\tilde{\lambda}^2$ are strictly consistent and asymptotically normal. Moreover, in the case of Gaussian innovations the RVs $(\varepsilon_t/\sigma)^2$ have a chi-square distribution χ_1^2. Thus, the estimate of the critical value $\tilde{c}$ can be simply obtained from the equality:

$$\tilde{c} = \tilde{\sigma}^2 \cdot F_{\chi_1^2}^{-1}(\tilde{b}_c), \tag{9}$$

where $F_{\chi_1^2}(x)$ is the corresponding cumulative distribution function (CDF). Similarly, when (ε_t) are with Laplacian $(0,\lambda^2)$ distribution, the critical value estimator $c = \tilde{c}$ can be easily found from the equation:

$$P\{\varepsilon_t^2 \le c\} = \tilde{b}_c,$$

whose solution is

$$\tilde{c} = \left[F_{\varepsilon}^{-1}\left(\frac{\tilde{b}_c+1}{2}\right)\right]^2, \tag{10}$$

and $F_{\varepsilon}(x)$ is the CDF of the Laplacian innovations (ε_t).

It can be shown that moment-based estimators are not the most efficient estimators of GSB process parameters. In order to obtain more efficient estimators of the observed parameters, the so-called modified Gauss-Newton estimation procedure for nonlinear functions can be used. First, we can write Equation (6) as:

$$\varepsilon_t = X_t + \theta_{t-1}\varepsilon_{t-1}, t = 1, \dots, T,$$

or, in the functional form,

$$\varepsilon_t(X,\theta) = X_t + \theta_{t-1}\varepsilon_{t-1}(X,\theta).$$

Using the estimator $\tilde{b}_c$, obtained according to the previously mentioned procedure, for an arbitrary $t = 1,2,\dots,T$ can be recursively computed values:

$$\tilde{\theta}_t := I\left(\varepsilon_{t-1}^2(X,\tilde{\theta}) \le \tilde{c}\right),$$
$$\varepsilon_t(X,\tilde{\theta}) := X_t + \tilde{\theta}_{t-1}\varepsilon_{t-1}(X,\tilde{\theta}), \tag{11}$$

where the initial values are $\tilde{\theta}_0 = 1$ and $\varepsilon_0(X,\tilde{\theta}) = \varepsilon_{-1}(X,\tilde{\theta}) = 0$. Additionally, if we define a series:

$$W_t(X,\theta) = \theta_t W_{t-1}(X,\theta) + \varepsilon_{t-1}(X,\theta), \tag{12}$$

where $t = 1, 2, \dots, T$, and $W_0(X,\tilde{\theta}) = 0$, using the mentioned properties of RVs (θ_t) and (ε_t), it follows that $(W_t(X,\theta))$ is a stationary and ergodic series of RVs with:

$$E\left(W_t(X,\theta)\right) = 0,$$
$$\mathrm{Var}\left(W_t(X,\theta)\right) = E\left(W_t(X,\theta)\right)^2 = \begin{cases} \sigma^2/(1-b_c), & (\varepsilon_t) \text{ is Gaussian} \\ 2\lambda^2/(1-b_c), & (\varepsilon_t) \text{ is Laplacian} \end{cases},$$

and correlation function $\rho_W(h) = b_c^{|h|}, h = 0, \pm 1, \dots$ Now if we define the so-called residual series:

$$R_t(X,\theta) = W_t(X,\theta) - b_c W_{t-1}(X,\theta), \tag{13}$$

it can be shown (see, c.f. (Jovanović, Stojanović, Kuk, Popović, & Čisar, 2022)) that $R_t(X,\theta)$ are mutually non-correlated RVs. Thus, Equation (13) represents a linear autoregressive (AR) process $(W_t(X,\theta))$ with innovations $(R_t(X,\theta))$. From here, by applying the regression procedure, another estimator of the threshold parameter $b_c \in (0,1)$ can be obtained:

$$\hat{b}_c = \left(\sum_{t=0}^{T-1} W_t(X,\tilde{\theta}) W_{t+1}(X,\tilde{\theta})\right)\left(\sum_{t=0}^{T-1} W_t^2(X,\tilde{\theta})\right)^{-1}. \tag{14}$$

Here, according to Equations (12) and (13), the above values can be iteratively calculated:

$$W_t(X,\tilde{\theta}) := W_t(X,\theta)|_{\theta=\tilde{\theta}} = \tilde{\theta}_t W_{t-1}(X,\tilde{\theta}) + \varepsilon_{t-1}(X,\tilde{\theta}),$$
$$R_t(X,\tilde{\theta}) := R_t(X,\theta)|_{\theta=\tilde{\theta}} = W_t(X,\tilde{\text{ë}}) - b_c W_{t-1}(X,\tilde{\text{ë}}).$$

Similar to previous moment-based estimators, the estimator $\hat{b}_c$ allows to obtain an estimator of the critical value $c = \hat{c}$, as a solution of the equation:

$$P\{\mathring{\text{a}}_t^2 \le c\} = \hat{b}_c. \tag{15}$$

Let us notice that the previously obtained estimators $\tilde{b}_c$ (as well as $\hat{b}_c$), that is, the modelled values of innovations $(\mathring{\text{a}}_t)$, defined by Equation (11), can be used to obtain the estimators of parameter ë. Thereby, it is well-known that for the Gaussian, as well as the Laplace innovations $(\mathring{\text{a}}_t)$, the so-called maximum likelihood (ML) method provides the most efficient estimates of the parameter ë. In the case of our GSB process with Gaussian innovations, the ML estimator can be obtained according to Equations (1) and (2), that is, based on the maximization of the log-likelihood function:

$$L(y_1,\dots,y_T;\sigma^2) = \ln\prod_{t=1}^{T}\left[\frac{1}{2\sigma\sqrt{\pi}}\exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right)\right] = -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

By solving equation $\partial L(y_1,\dots,y_T;\acute{o}^2)/\partial\acute{o}^2 = 0$ the estimator of the parameter $\acute{o}^2$ is obtained as the sample variance of the series $(\mathring{a}_t)$:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(X,\hat{\theta}). \qquad (16)$$

In the case of Laplacian innovations $(\mathring{a}_t)$, the ML estimator can be obtained by maximization of the log-likelihood function:

$$\ell(y_1,\dots,y_T;\lambda) = \ln\prod_{t=1}^{T}\left[\frac{1}{2\lambda}\exp\left(-\frac{|y_t - m_t|}{\lambda}\right)\right] = -T\ln(2\lambda) - \frac{1}{\lambda}\sum_{t=1}^{T}|\varepsilon_t(X,\tilde{\theta})$$

In the same way as in the previous one, solving equation $\ell(y_1,\dots,y_T;\ddot{e})/\partial\ddot{e} = 0$, as well as Equation (3), give the estimator of scale parameter $\ddot{e}$ as the mean absolute deviation (MAD):

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}|\varepsilon_t(X,\tilde{\theta})|. \qquad (17)$$

Finally, estimators of the mean of the GSB process, that is, the parameter $\grave{\imath} = E(y_t)$ can be obtained as the sample mean of series (y_t):

$$\tilde{\mu} := \bar{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t. \qquad (18)$$

This is unbiased estimator, that is, $E(\tilde{\mu}) = E(\bar{y}_T) = \mu$, but it can be easily shown that its variance is unbounded. In order to obtain a more efficient estimator for the parameter μ, a sample mean of the mean series $\bar{y}_t$, when $t = 1,\dots,T$, can be taken, i.e., the following estimator can be used:

$$\hat{\mu} := \frac{1}{T}\sum_{t=1}^{T}\bar{y}_t = \frac{1}{T}\sum_{t=1}^{T}\omega_t y_t. \qquad (19) \qquad (19)$$

Here, $\omega_t := H(T) - H(t-1)$ and $H(t) := \sum_{j=1}^{t} j^{-1}, t = 1,\dots,T$ are the harmonic numbers, with $H(0) = 0$. The estimator $\hat{\mu}$ is also unbiased for the parameter μ, but its weights are more pronounced at the 'older' time points of the realization of the series (y_t). This is also consistent with the fact that the covariances of RVs y_t depend on these 'older' time indices. For these reasons, the estimator $\hat{\mu}$ is more efficient than $\tilde{\mu}$, which can be shown using a procedure similar as in (Jovanović, Stojanović, Kuk, Popović, & Čisar, 2022) and (Stojanović, Bakouch, Ljajko, & Božović, 2023). It is shown here that the asymptotic values of variances of the estimators $\tilde{\mu}$ and $\hat{\mu}$ are, respectively:

$$\tilde{V} := V(\tilde{\mu}) = \frac{a_c Var(\varepsilon_t)T}{3} + O(T^{-1}) \longrightarrow +\infty, T \to +\infty,$$
$$\hat{V} := V(\hat{\mu}) = a_c Var(\varepsilon_t)H^2(T) + o\left(H^{-2}(T)\right) \longrightarrow +\infty, T \to +\infty.$$

Thus, the estimator $\hat{\mu}$ is (asymptotically) more efficient than $\tilde{\mu}$, because it is valid:

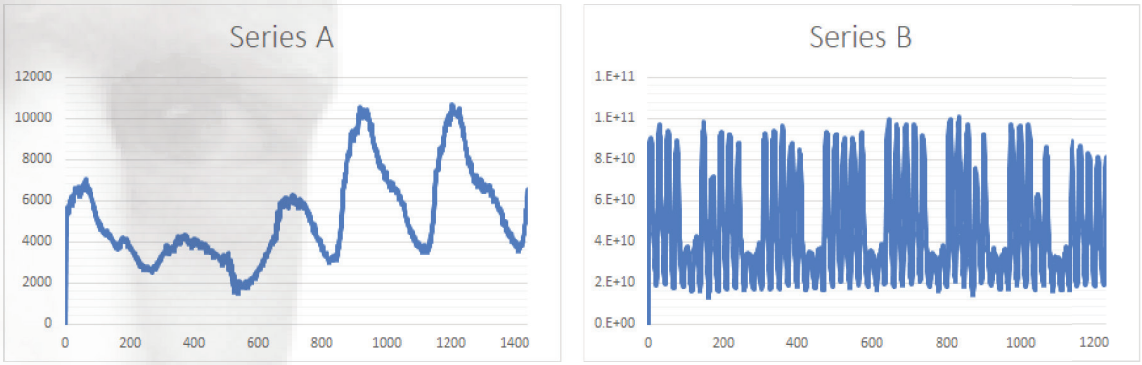
$$\lim_{T\to+\infty}\frac{V(\hat{\mu})}{V(\tilde{\mu})} = \lim_{T\to+\infty}\frac{H^2(T)}{T} = 0.$$

Finally, let us point out that the Cauchy distribution has a special problem of its "heavy tail", which prevents the previously described estimations of the parameters of the GSB process. For instance, estimation procedures based on moments cannot be applied in this case, because the mean and variance

of the Cauchy distribution do not exist, while maximum likelihood estimation (MLE) requires some complex calculations. For these reasons, the so-called empirical characteristic function (ECF) method will be used here (see, e.g. (Ljajko, Stojanović, Tošić, & Božović, 2023)), but due to the certain complexity of this method, it will not be described here in detail.

Application in the Network Intrusion Detection

In the following is given the application of the GSB process in the dynamic analysis of aggregate traffic and potential detection of unusual amounts of traffic that may be a sign of unauthorized access. Internet traffic data (in bits) used for analysis in this paper comes from an ISP and represents open data set from the web site Datamarket.com. In this analysis, these data are divided into two subsets. First, named Series A, represents aggregate traffic (in bits) in the backbone of the UK academic network. Data are collected at five-minute intervals, during exactly five days of observation, between 19 November 2004 at 09:30 and 24 November 2005 at 09:25. Second one, named Series B, corresponds to the hourly data of the transatlantic link with centres in 11 European cities, collected from 06:57 on 7 June to 11:17 on 31 July 2005. In this way, traffic samples of length $T_1 = 1440$ and $T_2 = 1330$, respectively, were obtained, the dynamics of which can be seen in the following Figure 2.



(a) (b)

Figure 2 *Dynamics of aggregate traffics (in bits):* **(a)** *Series A;* **(b)** *Series B.*

Based on them, it can be clearly concluded that there are pronounced and permanent fluctuations in both series. They are especially emphasized in Series A, where there is a distinct aperiodicity in its dynamics. On the other hand, in the case of Series B, there is a periodic dynamic of the volume of traffic, which is expected with this kind of data. The basic statistical indicators of both series are given in the following Table 2.

Table 2 *Basic statistical indicators of aggregate network traffic*

Statistics	Series A		Series B	
	Traffic	Log-volumes	Traffic	Log-volumes
Mean	5188.7	8.4672	4.57E+10	24.3906
Median	4445.0	8.3995	3.55E+10	24.2936
Mode	N/A	N/A	N/A	N/A
Sample variance	4,802,193	0.1759	6.67E+20	0.3099
Stand. deviation	2191.4	0.4195	2.58E+10	0.5567

Skewness	0.7745	-0.0110	0.7270	0.2036
Kurtosis	2.7757	2.4321	2.0864	1.7674
Minimum	1529	7.3324	1.32E+10	23.3068
Maximum	10,671	9.2753	1.01E+11	25.3371

In addition to the basic traffic data, Table 2 also shows descriptive statistics of the so-called log-volumes:

$$y_t^{(j)} := \ln\left(T_t^{(j)}\right), t = 0, 1, \ldots, T_j, j = 1,2. \qquad (20) \tag{20}$$

where $(T_t^{(j)})$ are aggregate traffics (in bits) of observed Series A and B. Log-volumes also represent an aggregate indicator, obtained as a natural logarithm of the traffic data volumes. As is pointed from several authors, e.g. (So, Chen, Chiang, & Lin, 2007), their usage changes the interpretation of activity shocks, because unexpected values are not affected by the growth trend in their dynamics. Also, the variance of log-volatility shocks is more uniform across the sample, that is, the timeline of the observed data, which can be seen from Table 2, too. In addition, the corresponding Split-MA(1) processes for both series are read as follows:

$$X_t^{(j)} := y_t^{(j)} - y_{t-1}^{(j)} = \ln\left(T_t^{(j)}/T_{t-1}^{(j)}\right), t = 1, \ldots, T, j = 1, 2,$$

i.e., they represent the so-called log-returns of aggregate traffic volumes.

We further consider the possibility of using the GSB process as a suitable stochastic model of logarithmic volume dynamics. According to them, as well as using Equations (1) and (3), the martingale means $(m_t^{(j)})$ and innovations $(\varepsilon_t^{(j)})$ can be obtained by iterative procedure:

$$\begin{cases} \varepsilon_t^{(j)} = y_t^{(j)} - m_t^{(j)}, \\ m_t^{(j)} = m_{t-1}^{(j)} + \varepsilon_{t-1}^{(j)} I\left\{\left(\varepsilon_{t-2}^{(j)}\right)^2 \geq \tilde{c}\right\}, \end{cases} \qquad (21) \tag{21}$$

where $j = 1,2$ and $\tilde{c}$ is the estimated critical value, obtained by using Equation (11). As initial values in (21), values $\varepsilon_0^{(j)} = \varepsilon_{-1}^{(j)} = 0$, as well as $m_0^{(j)} = y_0^{(j)} = \hat{\mu}$, $j = 1, 2$ were taken. The estimated values of basic statistical indicators of the increments $(X_t^{(j)})$, $j = 1, 2$, as well as two modelled series, martingale means $(m_t^{(j)})$, $j = 1, 2$ and innovation series $(\varepsilon_t^{(j)})$, $j = 1, 2$, are shown in the following Table 3.

Note first that both Series A and B have similar statistical indicators, which indicates a certain similarity in their dynamics and other stochastic characteristics. This can be seen by comparing statistical indicators of the increments $(X_t^{(j)})$, $j = 1, 2$, and the innovation series $(\varepsilon_t^{(j)})$, $j = 1, 2$. It is noticeable that the sample means of both series are close to zero, that is, they have the property of symmetry of their empirical distributions.

Table 3 *Statistical indicators of increments, martingale means and innovations of the GSB processes*

Statistics	Series A			Series B		
	$\left(X_t^{(1)}\right)$	$\left(m_t^{(1)}\right)$	$\left(\varepsilon_t^{(1)}\right)$	$\left(X_t^{(2)}\right)$	$\left(m_t^{(2)}\right)$	$\left(\varepsilon_t^{(2)}\right)$
Mean	2.25E-04	8.4747	-0.0074	5.93E-04	24.410	-0.0274
Median	-8.26E-04	8.4031	-0.0102	-2.80E-02	24.359	-0.0279
Mode	N/A	8.6716	N/A	N/A	24.267	N/A
Sample variance	8.70E-04	0.1779	0.0034	0.0283	0.3127	0.0766

Stand. deviation	0.0295	0.4218	0.0582	0.1681	0.5592	0.2767
Skewness	-2.2574	-0.0900	-0.0436	1.3059	0.3127	1.0009
Kurtosis	2.3966	2.4052	2.1456	1.9767	1.5896	2.8703
Minimum	-0.2992	7.3700	-0.3951	-0.4726	23.429	-0.8103
Maximum	0.1580	9.2753	0.2082	0.6074	25.238	1.1539

As an illustration, Figure 3 shows the empirical distribution of both innovation series, as well as their fitting with the three distributions mentioned earlier. It is noticeable that fitting with non-Gaussian distributions (Laplace and Cauchy) is somewhat more adequate than the standard Gaussian distribution.

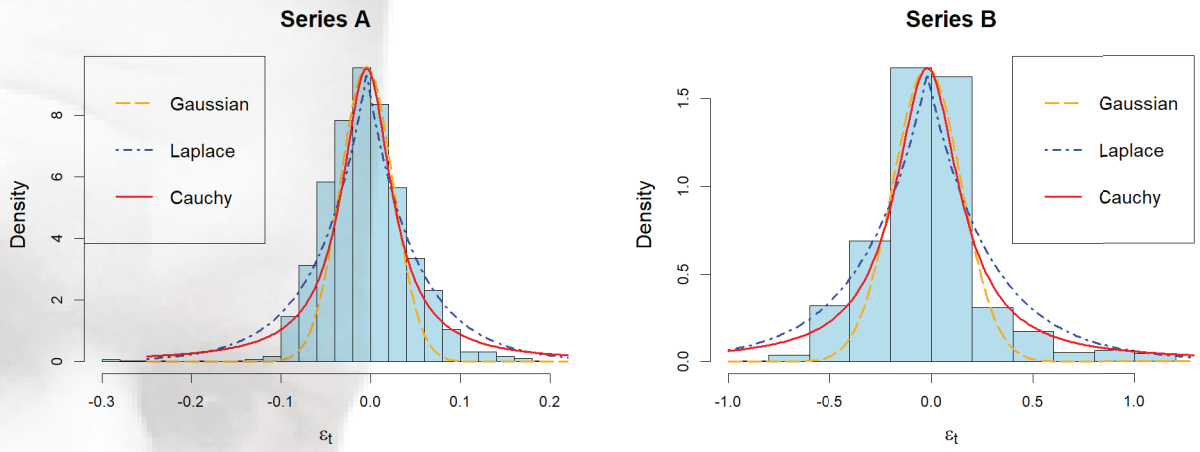


Figure 3 *Empirical distribution of GSB process innovations (given by histograms) according to theoretical distributions (given by lines).*

In the following, using the previously described estimation procedure, the parameters of both time series are estimated. Table 4 shows the estimates obtained by applying the above-mentioned procedures, that is, two kinds of parameters estimates. Additionally, some other estimates, such as the first-order sample correlation $\hat{\rho}_X(1)$, and the estimates of the threshold parameter b_c, are also shown. We can notice that the condition $-0.5 < \hat{\rho}_X(1) < 0$ is satisfied in both series cases, which enables the estimation of the parameter b_c.

Table 4 *Estimated parameters values of the log-volumes series*

Parameters estimates		Series A	Series B
Mean value	$\tilde{\mu}$	16.832	12.672
	$\hat{\mu}$	17.004	12.587
Sample correlation	$\hat{\rho}_X(1)$	-0.1911	-0.3948
Threshold parameter	$\tilde{b}_c$	0.2362	0.6523
	$\hat{b}_c$	0.3766	0.5546
Critical value	$\tilde{c}$	0.0196	0.2868
	$\hat{c}$	0.0459	0.1487
Scale parameter	$\tilde{\lambda}$	0.5201	0.5069
	$\hat{\lambda}$	0.4520	0.5113

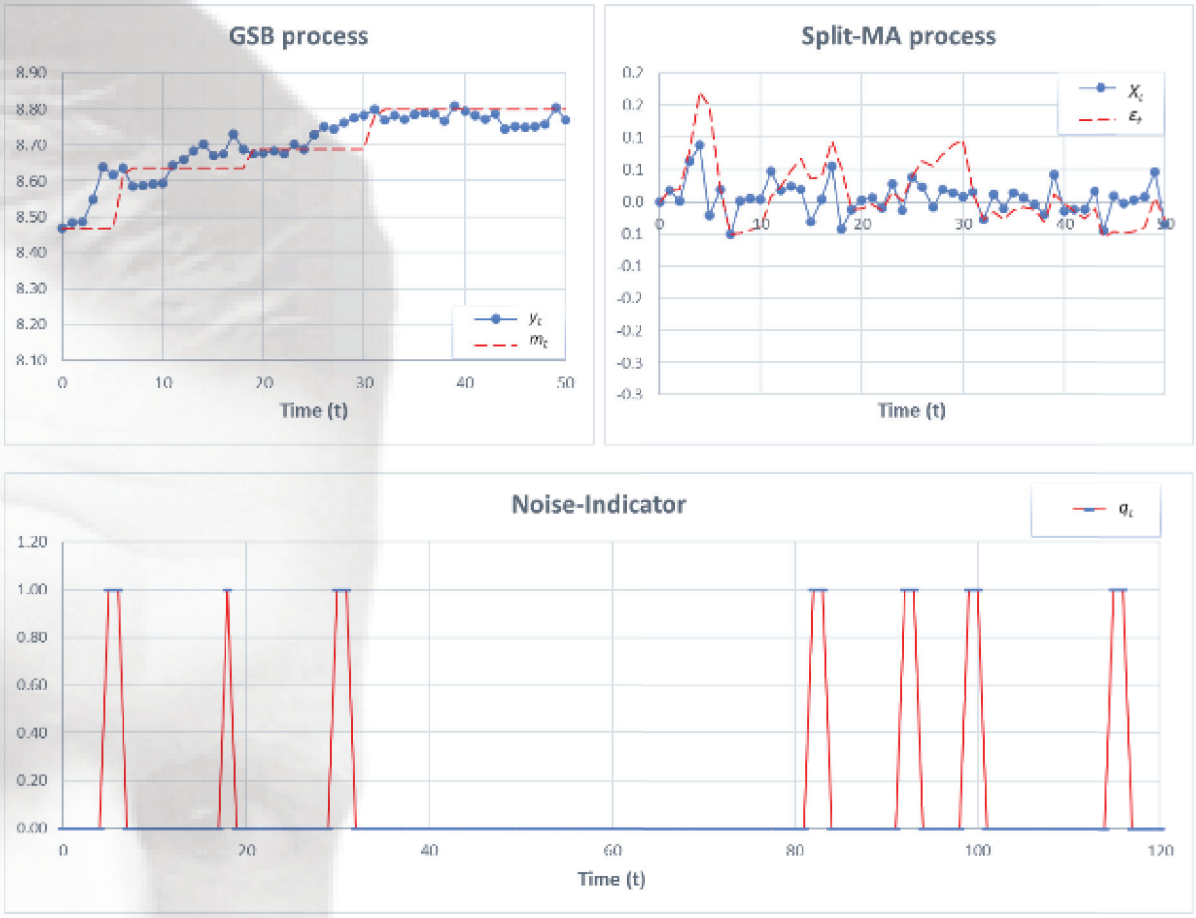


Figure 4 *Plots above: Dynamics of empirical and modeled data for Series A; Graph below: Realization of the Noise-Indicator of Series A.*

The agreement between the modelled and actual data can be seen in above plots of Figures 4 and 5, where in addition to the observed log-volume values ($y_t^{(j)}$), the modelled martingale mean values ($m_t^{(j)}$), increments ($X_t^{(j)}$) and innovations ($\varepsilon_t^{(j)}$) are also shown. It is noticeable high agreement between these series, which can be explained by the theoretical findings presented in Section 2. Namely, the martingale means ($m_t^{(j)}$) are constant in the case when there were no pronounced fluctuations of the series ($y_t^{(j)}$) in the previous time period. Conversely, when fluctuations are pronounced, there is a change in the value of the martingale means, in which case the values of the series ($X_t^{(j)}$) and ($\varepsilon_t^{(j)}$) are then equal.

Finally, the simplest way to detect pronounced fluctuations is based on the value of the Noise-Indicator, that is, the 0-1 series (q_t), which in this case take unit values. As its name suggests, the indicator (q_t) "reacts" to pronounced (and unexpected) changes in network traffic values. Therefore, the case $q_t = 1$ can be considered as "situation" when there is an unusual change in the fluctuations of the observed data series. As an illustration of the aforementioned facts, the realizations of the Noise-Indicators ($q_t^{(j)}$), $j = 1,2$, for both series are shown in the below graphs at Figures 4 and 5. Moreover, as already mentioned earlier, in the case of Series B, there are periodic changes in its values, which can also be observed in the realizations of the indicator ($q_t^{(2)}$).

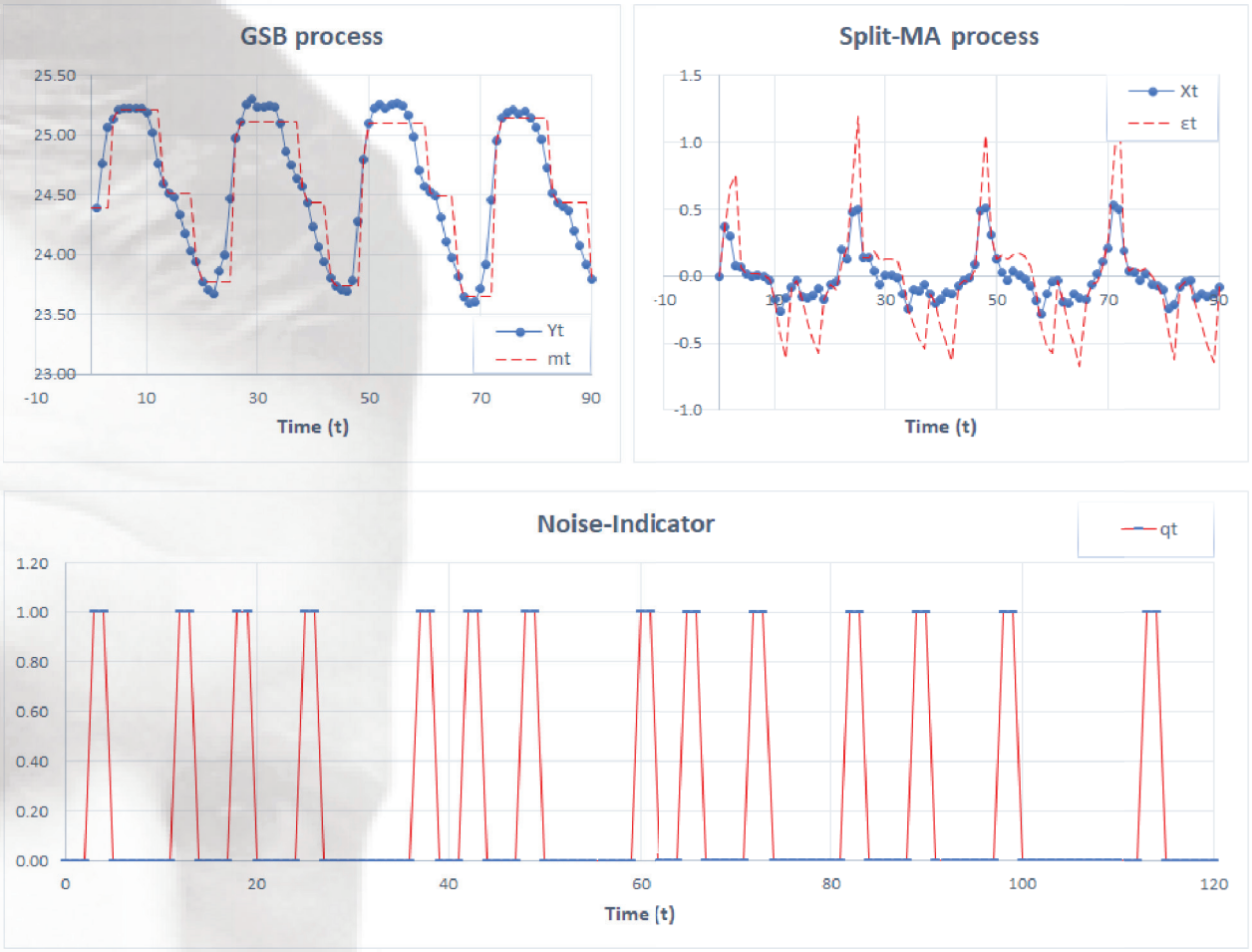


Figure 5 *Plots above: Dynamics of empirical and modeled data for Series B; Graph below: Realization of the Noise-Indicator of Series B.*

Conclusion

The application of the GSB process presented here confirms its possibility in modelling actual the network intrusion detection. It is worth to notice that one of the advantages of this kind of stochastic modelling is that it allows the simultaneous use of both stationary and non-stationary components. Thereby, the asymptotic behavior of these series and the corresponding estimates thus obtained are of particular importance. It should also be noted that the proposed parameter estimation procedure can be implemented algorithmically in a relatively simple way. Finally, let us notice that for threshold parameter estimation some other methods can be used, such as the Empirical Characteristic Function (ECF) method described in (Stojanović, Milovanović, & Jelić, 2016). Certain modifications of this approach are certainly desirable in some future research, in order to more successfully detect various types of attacks on computer networks and other important information systems.

References

Čisar, P., & Maravić-Čisar, S. (2010). EWMA-based threshold algorithm for intrusion detection. *Computing and Informatics*, 1089-1101.

Čisar, P., & Maravić-Čisar, S. (2010). Network Statistics in Function of Statistical Intrusion Detection. *Studies in Computational Intelligence*, 27-35.

Čisar, P., & Maravić-Čisar, S. (2010). Skewness and Kurtosis in Function of Selection of Network Traffic Distribution,. *Acta Polytechnica Hungarica*, 95-106.

Čisar, P., & Maravić-Čisar, S. (2012). Network Statistical Anomaly Detection Based on Traffic Model. *Annals of Faculty Engineering Hunedoara – International Journal Of Engineering*, 89-96.

Čisar, P., Maravić-Čisar, S., Popović, B., Kuk, K., & Vuković, I. (2022). Application of Artificial Immune Networks in Continuous Function Optimization. *Acta Polytechnica Hungarica, 19*(7), 153-164.

Čisar, P., Popović, B., Kuk, K. Č., & Vuković, I. (2022). Machine Learning Aspects of Internet Firewall Data. In *Security-Related Advanced Technologies in Critical Infrastructure Protection - Theoretical and Practical Approach.* NATO Science for Peace and Security Series C: Environmental Security: Springer Dordrecht.

Eskin, E. (2000). Anomaly Detection over Noisy Data using Learned Probability Distributions. *Proceedings of the 17th International Conference on Machine Learning* (pp. 255–262). Stanford University.

Fengmin, G. (2003). *Deciphering Detection Techniques: Part II Anomaly–Based Intrusion Detection.* White Paper: McAfee Security.

Jovanović et al. (2018). Soserbia: Android-based software platform for sending emergency messages. *Complexity*, Article ID: 8283919.

Jovanović, M., Stojanović, V., Kuk, K., Popović, B., & Čisar, P. (2022). Asymptotic Properties and Application of GSB Process: A Case Study of the COVID-19 Dynamics in Serbia. *Mathematics*, Article No. 3849.

Ljajko, E., Stojanović, V., Tošić, M., & Božović, I. (2023). Cauchy Split-Break Process: Asymptotic Properties and Application in Securities Market Analysis. *Sci. Bulletin, Series A: Aplied Mathematics & Physics*, In press (accepted manuscript).

So, M. K., Chen, C. W., Chiang, T. C., & Lin, D. S. (2007). Modelling Financial Time Series with Threshold Nonlinearity in Returns and Trading Volume. *Applied Stochastic Models in Business and Industry, 23*(4), 319-338.

Sorensen, S. (2004). *Competitive Overview of Statistical Anomaly Detection.* White Paper: Juniper Networks.

Spathoulas, G., & S. Katsikas, S. (2010). Reducing false positives in intrusion detection systems. *Computers & Security*, 35-44.

Stojanović, V., Bakouch, H., Ljajko, E., & Božović, I. (2023). Laplacian Split-BREAK Process with Application in Dynamic Analysis of the World Oil and Gas Market. *Axioms*, Article No. 622.

Stojanović, V., Milovanović, G., & Jelić, G. (2016). Distributional properties and parameters estimation of GSB Process: An approach based on characteristic functions. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 835–861.

Stojanović, V., Popović, B., & Popović, P. (2011). The Split-BREAK Model. *Brazilian Journal of Probability and Statistics*, 44-63.

Stojanović, V., Popović, B., & Popović, P. (2014). Stochastic Analysis of GSB Process. *Publ. Inst. Math.*, 149-159.

Stojanović, V., Popović, B., & Popović, P. (2015). Model of General Split-BREAK Process. *REVSTAT–Statistical Journal*, 145-168.