Computing and Informatics, Vol. 29, 2010, 1089–1101

EWMA BASED THRESHOLD ALGORITHM FOR INTRUSION DETECTION

Petar Čisar

Telekom Srbija
Prvomajska 2-4
Subotica, Serbia
e-mail: petarc@telekom.rs

Saša Bošnjak

Faculty of Economics
Segedinski put 9-11
Subotica, Serbia
e-mail: bsale@ef.uns.ac.rs

Sanja Maravić Čisar

Subotica Tech
Marka Oreškovića 16
Subotica, Serbia
e-mail: sanjam@vts.su.ac.rs

Manuscript received 2 April 2009; revised 17 September 2009
Communicated by János Fodor

Abstract. Intrusion detection is used to monitor and capture intrusions into computer and network systems which attempt to compromise their security. Many intrusions manifest in dramatic changes in the intensity of events occuring in computer networks. Because of the ability of exponentially weighted moving average control charts to monitor the rate of occurrences of events based on their intensity, this technique is appropriate for implementation in threshold based algorithms.

Keywords: Intrusion detection, EWMA, threshold algorithm, optimization, network traffic, autocorrelation

Mathematics Subject Classification 2000: 94A13, 94C12, 68M15, 68W99, 62B15

1 INTRODUCTION

The exponentially weighted moving average (EWMA) is a statistic for monitoring the process that averages the data in a way that gives less and less weight to data as they are further removed in time. For the EWMA control technique, the decision regarding the state of control of the process depends on the EWMA statistic, which is an exponentially weighted average of all prior data, including the most recent measurements.

By the choice of weighting factor λ, the EWMA control procedure can be made sensitive to a small or gradual drift in the process. The statistic that is calculated is:

$$EWMA_t = \lambda Y_t + (1-\lambda) EWMA_{t-1} \qquad t = 1, 2, \ldots, n \tag{1}$$

where

- $EWMA_0$ is the mean of historical data (target)
- Y_t is the observation at time t
- n is the number of observations to be monitored including $EWMA_0$
- $0 < \lambda \leq 1$ is a constant that determines the depth of memory of the EWMA.

This equation is due to Roberts [8].

The parameter λ determines the rate at which "older" data enter into the calculation of the EWMA statistics. The value of $\lambda = 1$ implies that only the most recent mesurement influences the EWMA. Thus, a large value of $\lambda = 1$ gives more weight to recent data and less weight to older data – a small value of λ gives more weight to older data. The value of λ is usually set between 0.2 and 0.3 [15] although this choice is somewhat arbitrary.

In real situations, the exact value of the shift size is often unknown and can only be reasonably assumed to vary within a certain range. Such a range of shifts deteriorates the performance of existing control charts. The most usual quality control procedures used in intrusion detection systems, such as the cumulative sum (CUSUM) and EWMA charts are based on a mean shift with a given size. This shift can be caused by intrusion or attack, for example. Lucas and Saccucci [16] have shown that although the smoothing factor λ used in an EWMA chart is usually recommended to be in the interval 0.05 to 0.25, in practice the optimally designed smoothing factor depends not only on the given size of the mean shift δ, but also on a given in-control Average Run Length (ARL). With shift $= \delta/\sigma_Q$ and ARL $= 370$,

optimal choices are given by the following figure (an average line has been drawn for a lot of ARLs instead of one line for each ARL) [10].

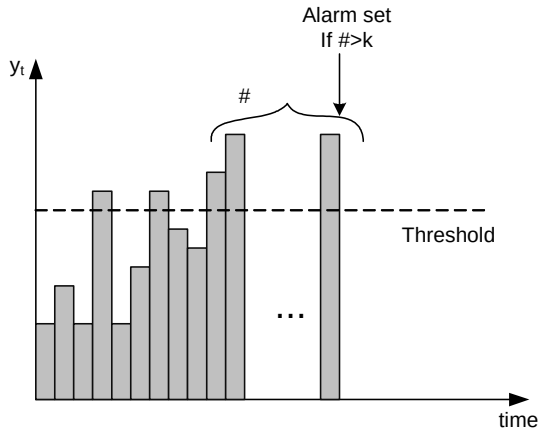


Fig. 1. Weighting factor

The estimated variance of the EWMA statistic is approximately:

$$\sigma^2_{EWMA} = \frac{\lambda}{2-\lambda} \cdot \sigma^2 \tag{2}$$

when t is not small, where σ is the standard deviation calculated from the historical data. The center line for the control chart is the target value or $EWMA_0$. The upper and lower control limits are:

$$\begin{aligned} UCL &= EWMA_0 + m\sigma_{EWMA} \\ LCL &= EWMA_0 - m\sigma_{EWMA} \end{aligned} \tag{3}$$

where the factor m is either set equal 3 (the 3-sigma control limits) or chosen using the Lucas and Saccucci tables (ARL = 370):

λ	0.05	0.1	0.2	0.3	0.4	0.5	0.75	1
m	2.49	2.70	2.86	2.93	2.96	2.98	3.00	3.00

Table 1. Choice of m

In addition to the aforementioned authors the theme of EWMA statistics and statistical anomaly detection in computer networks has also been addressed in [1, 3–7, 9–11, 13, 18].

2 ADAPTIVE THRESHOLD ALGORITHM

This relatively simple algorithm relies on testing whether the traffic, i.e. the number of packets, exceeds a particular threshold over a given interval. In order to account for seasonal (daily and weekly) traffic variations, the value of the threshold is set adaptively, based on an estimate of the mean number of packets, which is computed from recent traffic measurement.

If x_n is the number of packets in the n^{th} time interval and μ_{n-1} is the mean rate calculated from measurements prior to n, then the alarm is active if [2]:

$$x_n \geq (\lambda + 1) \cdot \overline{\mu}_{n-1}. \tag{4}$$

Then alarm is signalized at the moment n. Here $\alpha > 0$ is the parameter that indicates the percentage above the mean value that we consider to be an indication of anomalous behaviour. The mean μ_n can be computed over some past time interval or using the EWMA of previous measurements [2]:

$$\overline{\mu}_n = \lambda \cdot \overline{\mu}_{n-1} + (1 - \lambda) \cdot x_n, \tag{5}$$

where λ is the EWMA factor. Direct application of the above algorithm would yield a relatively high number of false alarms (false positives). A simple modification that can improve its performance is to signal an alarm after a certain number of consecutive violations of the threshold # (Figure 2). In this case, the alarm is active if [2]:

$$\sum_{i=n-k+1}^{n} 1_{\{x_i \geq (\alpha+1)\overline{\mu}_{i-1}\}} \geq k \tag{6}$$

where $k > 1$ is the parameter that indicates the number of consecutive intervals the threshold must be exceeded for generating an alarm.

The configurable parameters of this algorithm are the threshold factor α, the number of successive threshold violations k before signalling an alarm, the EWMA factor λ and the time interval T over which the number of packets are taken.

3 EXPONENTIAL SMOOTHING

Calculating the optimal value of parameter λ is based on the study of authentic samples of network traffic. Random variations of network traffic are normal phenomena in the observed sample. In order to decrease or eliminate the influence of individual random variations of network traffic on occurence of false alarms, the procedure of exponential smoothing is applied, as an aspect of data preprocessing.

For any time period t, the smoothed value S_t is determined by computing:

$$S_t = \lambda y_{t-1} + (1 - \lambda) S_{t-1} \text{ where } 0 < \lambda \leq 1 \text{ and } t \geq 3. \tag{7}$$

This is the basic equation of exponential smoothing. The formulation here is given by Hunter [15].

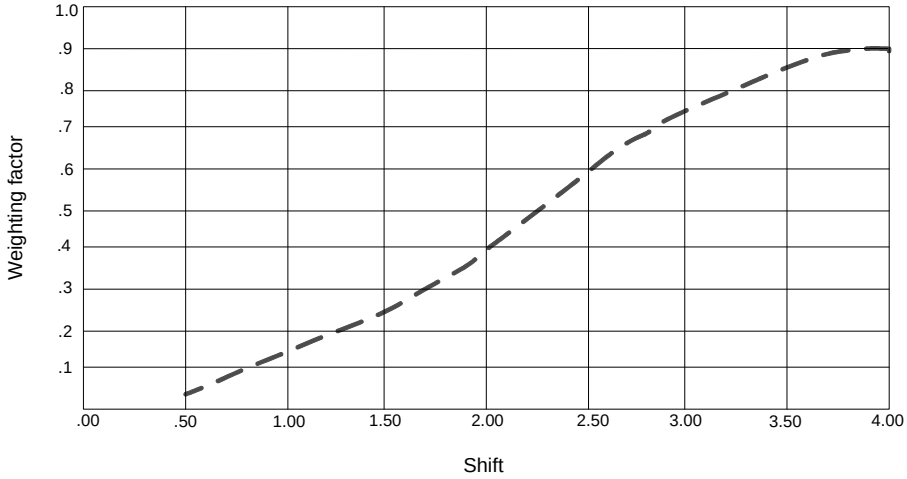


Fig. 2. Adaptive Threshold – k

This smoothing scheme begins by setting S_2 to y_1, where S_i stands for smoothed observation or EWMA, and y_i stands for the original observation. The subscripts refer to the time periods 1, 2, ..., n. For example, the third period is $S_3 = \lambda y_2 + (1 - \lambda)S_2$ and so on. There is no S_1. The optimal value for λ is the value which results in the smallest mean of the squared errors (MSE). Here is an illustration of this principle through an example. Consider the following data set consisting of n observations of data flow over time-for starting $\lambda = 0.1$:

Time	**Flow (y_t)**	**S_t**	**Error ($y_t - S_t$)**	**Error squared**
1	y_1			
2	y_2	y_1	E_2	E_{22}
3	y_3	S_3	E_3	E_{32}
...	...	...	...	...
n	y_n	S_n	E_n	E_{n2}
				SSE_n

Table 2. Smoothing scheme

The sum of the squared errors (SSE) is $SSE_{0.1}$. The mean of the squared errors is $MSE_{0.1} = SSE_{0.1}/(n-1)$. After that, the MSE is calculated for $\lambda = 0.2$. If $MSE_{0.2} < MSE_{0.1}$ then $MSE_{0.2}$ is better value for λ. This iterative procedure is related to the range of λ between 0.1 and 0.9. In this way, the best initial choice for λ is determined and then, for getting more precise value, search optionally continues between $\lambda - \Delta\lambda$ and $\lambda + \Delta\lambda$, where $\Delta\lambda$ is arbitrarily small interval around λ (for instance, in practical applications, $\pm 10\,\%$ around optimal λ).

The initial EWMA plays an important role in computing all the subsequent EWMA's. There are several approaches in defining this value:

1. setting S_2 to y_1
2. setting S_2 to the target of the process
3. setting S_2 to average of the first four or five observations.

It can also be shown that the smaller the value of λ, the more important is the selection of the initial EWMA. The user would be wise to try a few methods, before finalizing the settings.

Due to the lack of exactness in available publications about the determination of initial S_2 in the procedure of exponential smoothing, the authors of this paper have dealt with researching the link between selection of $S_2 = y_1$ and λ_{opt}, i.e. $S_2(\lambda_{opt})$. In that sense, the range of S_2 is determined from the lowest to the highest sample value during the period of observation. This research was conducted on an authentic sample of network traffic of an Internet service provider and observed were the values of local maximums (in this example from 8 to 34 Mb/s), with great enough number of values $n = 33 > 30$ (great sample), taking into account the generality of conclusions. The period of observation was one month. The next figure shows the numerical and graphical dependence $S_2(\lambda_{opt})$. Calculation of partial optimal values is realized using the application in Excel.

Since a set of different results is obtained for partial values of λ_{opt}, the authors of this paper suggest for the overall optimal parameter Λopt to accept the average of all the partial results (in this particular case it is 0.3879).

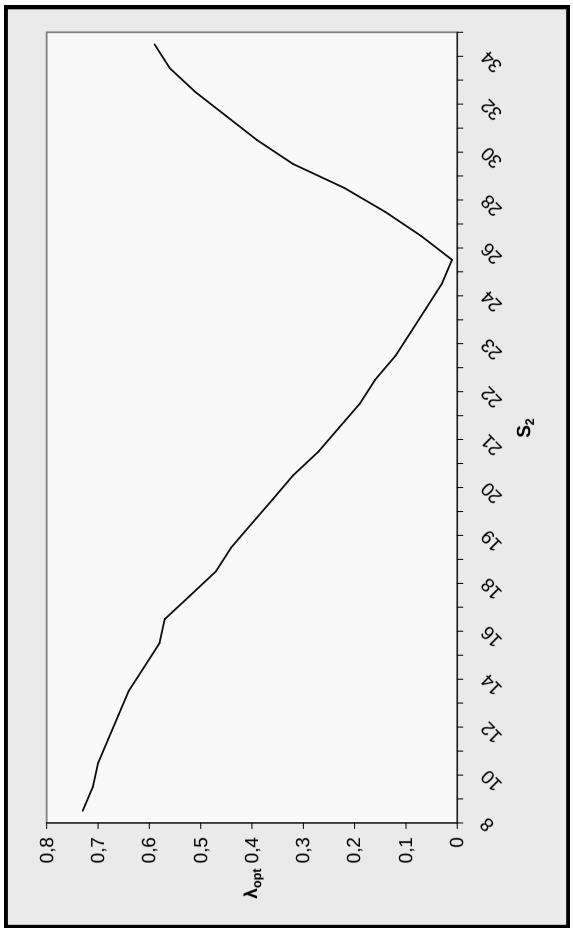
4 THE FACTOR OF TIME IN EWMA INTRUSION DETECTION

Having in mind the great diversity of attacks on computer networks, it is quite difficult to define the start of the timeline of attacks. Some types of attacks realized in form of one or more packages can be recognized very quickly – for less than one second [14]. Other types of attacks are active during a much longer period and can not be identified as attacks until watching a lot of results as a whole. Although each attack has a specific beginning, the initial point of attack is not always possible to determine at the time of occurrence. Attack detection in real time means recognition of the attack in a certain period of time zero (the moment when the attack actually started) plus arbitrarily small interval after that – in less than a few seconds [14]. In term of industry definitions, the real-time interval is 5 s–5 min.

Reaction time of a security system on the attacks is changing over time. So in the nineties of the past century reaction time amounted to twenty days, in 2000–2002 to about two hours, while from 2003 and later this time was less than 10 seconds.

In the context of the above data, the EWMA algorithm for intrusion detection can be made more efficient if an alarm signal appears only if both conditions are fulfilled simultaneously:

- EWMA value of the current traffic exceeds the value of UCL
- time duration of the exceeding is greater than 10 s.



	S_2	λ_{opt}
1	8	0,73
2	9	0,71
3	10	0,7
4	11	0,68
5	12	0,66
6	13	0,64
7	14	0,61
8	15	0,58
9	16	0,57
10	17	0,52
11	18	0,47
12	18,5	0,44
13	19	0,4
14	19,5	0,36
15	20	0,32
16	20,5	0,27
17	21	0,23
18	21,5	0,19
19	22	0,16
20	22,5	0,12
21	23	0,09
22	23,5	0,06
23	24	0,03
24	25	0,01
25	26	0,07
26	27	0,14
27	28	0,22
28	29	0,32
29	30	0,39
30	31	0,45
31	32	0,51
32	33	0,56
33	34	0,59

λ_{opt}:	0,3879

Fig. 3. Calculation of $S_2(\lambda_{opt})$

By implementation of time factor in EWMA algorithm, filtering of situations is performed with exceedings of the upper control limits, which further decreases the number of false alarms. It is necessary to note that the introduction of a considerably shorter time dimension implies the use of such network software, which will be able to measure and calculate the average value of traffic in intervals that match the accepted time limit of 10 s.

5 AUTOCORRELATION

Autocorrelation or serial correlation of time series means that the value of the observed variable in a time unit depends on values which appear sooner or later in series. In practical situations, autocorrelation of the first order is usually examined, which may be shown by a simple correlation coefficient or so-called autocorrelation coefficient. Let R_t be the time series data, where $t = 1, 2, \ldots, T$, then the autocorrelation coefficient of the first order is given by:

$$\rho(R) = \frac{\sum_{t=2}^{T} R_t R_{t-1}}{\sqrt{\sum_{t=2}^{T} R_t^2 \sum_{t=2}^{T} R_{t-1}^2}} \qquad -1 \leq \rho \leq 1. \tag{8}$$

One of the standard features of traffic time series is that increasing rates of traffic R_t are not mutually significantly autocorrelated, i.e. the value of autocorrelation coefficient is near to zero. At the same time, this means that the distribution of positive and negative values of increasing rates is random and does not follow a specific systematic regularity. Positive autocorrelation implicates that the positive values are followed by mainly positive values and negative values by negative ones and then $\rho \approx +1$. In case of negative autocorrelation, the change of sign appears very often, i.e. in most cases the positive rate leads to a negative rate and vice versa and then $\rho \approx -1$. Since there is no typical scheme, on the basis of positive rate in one particular time period it can not be concluded with a significant probability that in the next period the growth or decline will appear. The same situation is with the importance for negative rate. Examples of positive and negative correlation, as well as for random distribution of characters are given in the following figures.

Researchers in [12] dealt with the influence of autocorrelated and uncorrelated data on the behaviour of intrusion detection algorithm. In their work they came to the conclusion that EWMA algorithm for autocorrelated and uncorrelated data works well in the sense of intrusion detection in some information system. The advantage of EWMA technique for uncorrelated data is that this technique (as opposed to the case of autocorrelated data) can detect not only rapid changes in the intesity of events, but also small changes in mean value realized through the gradual increase or decrease of the intensity of events. However, in EWMA for uncorrelated data, initial value of smoothed intensity events is to be reset after

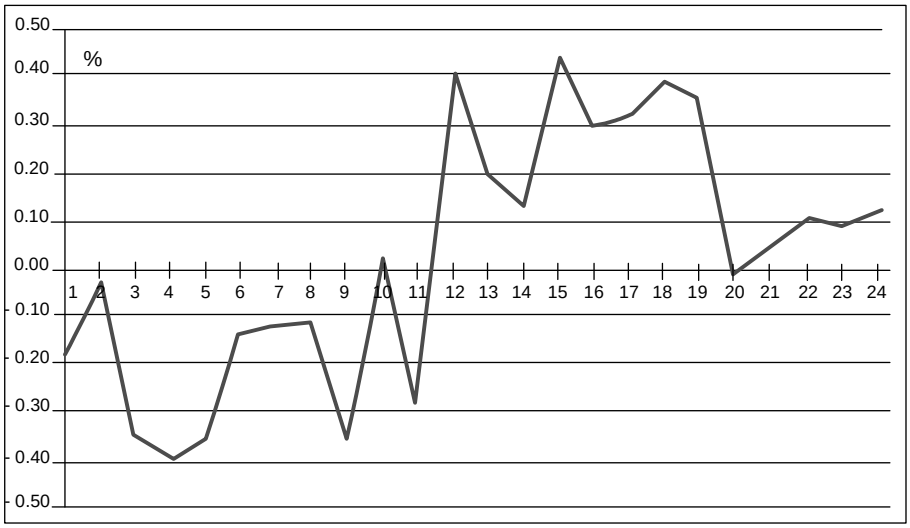


Fig. 4. Positive autocorrelation, $\rho = 0.61$

intrusion detection, in order to avoid the impact of current values of parameters on future results (*carry-over effect*). In the case of EWMA for autocorrelated data this reset is not necessary, because EWMA automatically adjusts the upper and lower control limits. Generally, the smoothing constant should not be too small, so that short-term trend in the intensity of events in the recent past could be detected.

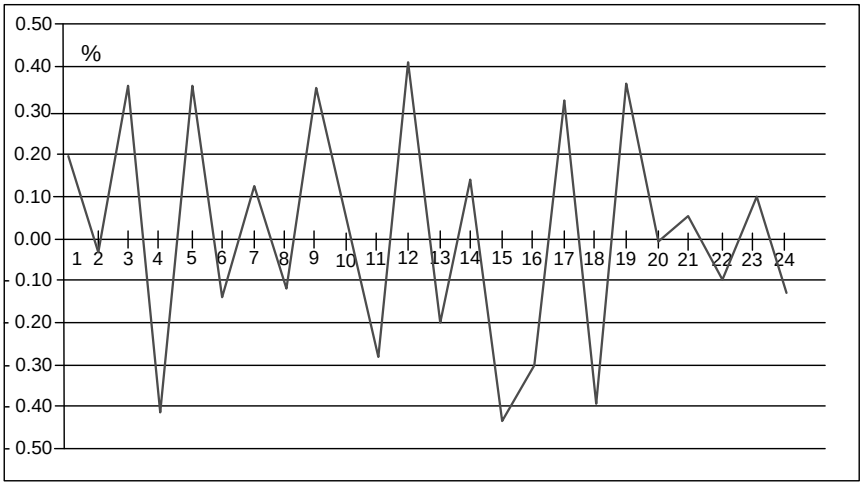


Fig. 5. Negative autocorrelation, $\rho = -0.6$

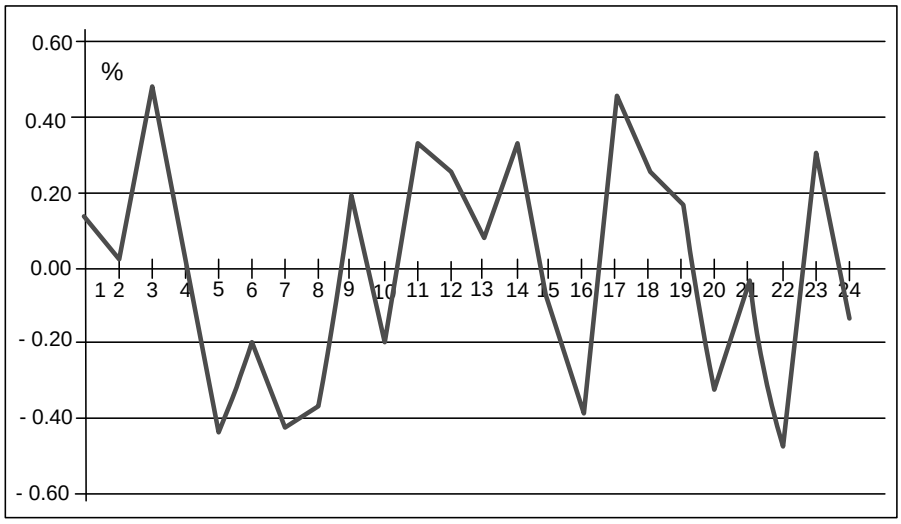


Fig. 6. Random distribution of characters, $\rho = 0.05$

In accordance with what has been said in this chapter, as justification for application of EWMA statistics, it is important to determine the statistical independence of samples, which will be examined by checking the existence of correlation between the used data. In this order, Pearson's correlation coefficient will be used, which is supplied as a ratio of covariances of two variables and the product of their standard deviations:

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y} \qquad -1 \leq \rho_{xy} \leq 1. \tag{9}$$

The value of correlation coefficient ρ_{xy} can be calculated using the statistical function CORREL (array1, array2) in MS Excel.

More authors proposed different interpretation ways of correlation coefficient. Cohen [17] noted that all the criteria are based on the greater or lesser extent of arbitrariness and should not be kept too strictly. Yet, an often used interpretation of these coefficients is [13]:

- ρ between 0 i 0.2 – no correlation or is insignificant
- ρ between 0.2 i 0.4 – low correlation
- ρ between 0.4 i 0.6 – moderate correlation
- ρ between 0.6 i 0.8 – significant correlation
- ρ between 0.8 i 1 – high correlation

6 CONCLUSIONS

This paper gives an overview of the EWMA based threshold algorithm suitable for detecting intrusions in network systems. By the method of exponential smoothing it is practically shown how to compute the optimal value of parameter λ. This algorithm can be made more effective if the component of time is also included. For the application of this method, it is necessary just to examine the extent of the correlation between samples. The paper also describes one of the ways of this testing, such as the interpretation of results.

REFERENCES

[1] Seibold, D.: Enterprise Campus Security-Addressing the Imploding Perimeter. Available on: http://www.itsa.ufl.edu/2003/presentations/IntSec.ppt.

[2] Vasilios, A.—Papagalou, F.: Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks. Availabe on: http://www.ist-scampi.org/publications/papers/siris-globecom2004.pdf.

[3] Sorensen, S.: Competitive Overview of Statistical Anomaly Detection.White Paper, Juniper Networks 2004.

[4] Engineering Statistics Handbook-EWMA Control Charts. Availabe on: http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc324.htm.

[5] Zhao, Y.—Tsung, F.—Wang, Z.: Dual CUSUM Control Schemes for Detecting a Range of Mean Shifts. IEEE Transactions 2005, availabe on: http://qlab.ieem.ust.hk/qlab/download/papers/paper\%2035.pdf.

[6] Fengmin, G.: Deciphering Detection Techniques: Part II, Anomaly-Based Intrusion Detection. White Paper, McAfee Security 2003.

[7] Mahadik, V. A.—Wu, X.—Reeves, D. S.: Detection of Denial-of-QoS Attacks Based on A2 Statistic and EWMA Control Charts. Available on: http://arqos.csc.ncsu.edu/papers/2002-02-usenixsec-diffservattack.pdf.

[8] Roberts, S. W.: Control Chart Tests Based on Geometric Moving Averages. Technometrics 1959.

[9] Viinikka, J.—Debar, H.: Monitoring IDS Background Noise Using EWMA Control Charts and Alert Information. Available on: http://viinikka.info/ViiDeb2004.pdf.

[10] Neubauer, A. S.: The EWMA Control Chart: Properties and Comparison with other Quality-Control Procedures by Computer Simulation. Clinical Chemistry, available on: http://www.clinchem.org/cgi/content/full/43/4/594.

[11] Engineering Statistics Handbook-Single Exponential Smoothing. Available on: http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm.

[12] Ye et al.: Computer Intrusion Detection Through EWMA for Autocorrelated and Uncorrelated Data. IEEE Transactions on Reliability, Vol. 52, 2003, No. 1.

[13] Savannah State University, Office of Institutional Research & Planning. Available on: http://irp.savstate.edu/irp/glossary/correlation.html.

[14] Roesch, M.: Next-Generation Intrusion Prevention: Time Zero. Available on: http://searchsecurity.techtarget.com/tip/.

[15] Hunter, J. S.: The Exponentially Weighted Moving Average. Journal of Quality Technology, Vol. 18, 1986, pp. 203–210.

[16] Lucas, J. M.—Saccucci, M. S.: Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. Technometrics 32, 1991, pp. 1–29.

[17] Cohen, J.: Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates, Hillsdale (New Jersey), 1998.

[18] Dulanović, N.—Hinić, D.—Simić, D.: An Intrusion Prevention System as a Proactive Security Mechanism in Network Infrastructure. YUJOR – Yugoslav Journal of Operations Research, Vol. 18, 2008, No. 1, pp. 109–122.

Petar Čisar graduated at the Faculty of Electrical Engineering in Belgrade. Master's study completed in information engineering at the Faculty of Economics in Subotica. He currently works on his Ph. D. thesis. The spheres of his interest are mobile technologies, as well as the development of security methods in network environments.

Saša Bošnjak is an Associate Professor of computer science at the Faculty of Economics in Subotica. He holds a range of courses in information engineering. His research interests include databases, software development, computer networks, reuse methodology, e-business and internet technology. He received his Ph. D. degree in information systems from Faculty of Economics in Subotica in 1995.

Sanja Maravić Čisar graduated at the Faculty of Electrical Engineering in Belgrade. Master's study completed at the Technical Faculty in Zrenjanin. She works as lecturer at Subotica Tech in the following courses: visual programming, object-oriented based programming, JAVA and multimedia systems. Currently she works on her Ph. D. thesis.