Hindawi Publishing Corporation
Journal of Computer Networks and Communications
Volume 2012, Article ID 192913, 5 pages
doi:10.1155/2012/192913

Research Article

Usage of Modified Holt-Winters Method in the Anomaly Detection of Network Traffic: Case Studies

Maciej Szmit[1,2] and Anna Szmit[3]

[1] *Computer Engineering Department, Technical University of Lodz, 18/22 Stefanowskiego Street, 90-924 Lodz, Poland*
[2] *Corporate IT Security Agency, Orange Labs Poland, 7 Obrzezna Street, 02-691 Warsaw, Poland*
[3] *Department of Management, Technical University of Lodz, 266 Piotrkowska Street, 90-924 Lodz, Poland*

Correspondence should be addressed to Maciej Szmit, maciej.szmit@gmail.com

Received 25 November 2011; Revised 15 March 2012; Accepted 29 March 2012

Academic Editor: Yueh M. Huang

The traditional Holt-Winters method is used, among others, in behavioural analysis of network traffic for development of adaptive models for various types of traffic in sample computer networks. This paper is devoted to the application of extended versions of these models for development of predicted templates and intruder detection.

1. Intruder Detection Systems

Intruder Detection Systems (IDSs) are software or hardware solutions aimed at detection of intrusion attempts to a protected network or a host. This is done by monitoring network traffic, usage of the resources of a protected computer system or by the analysis of system logs in order to detect suspicious actions and then take appropriate actions, which in the majority of cases is the generation of an alert informing about the detected danger. In the literature, the following are usually distinguished: Intruder Detection Systems, Active Response Systems, and Intruder Protection Systems (IPSs).

The next generation of security devices is the so-called Unified Threat Management (UTMs), which integrate, apart from the traditional IPS, also mechanisms such as Gateway Antivirus, Gateway Antispam, Content Filtering, Parental Control, Load Balancing, Bandwidth Management, and On-Appliance reporting, while obviously not every UTM system must have all of the above mechanisms implemented.

Another type of specialized security solutions, which can be implemented in UTM systems or constitute standalone solutions, is Information Leak Prevention systems, also known as Data Loss Prevention, Data Leak Prevention (DLP), or Information Loss Prevention (ILP),

Book [1, page 179] presents a listing of Intruder Detection Systems and Intruder Protection Systems, which includes more than 60 systems. Issues relating to IDS are also presented in many other research works (see e.g., [2–4]).

Anomaly detection is one of the three groups of methods, including misuse detection systems and integrity verification, used in Intruder Detection Systems.

Misuse detection is the detection of specific behaviours which confirm that an attack occurred, whereas anomaly detection involves predictive pattern of behaviours, deviations from which instances of an attack on a protected system are considered. Misuse detection has, in the majority of cases, deterministic character (the rules matching the observed phenomena or action is found or not), and it is easier to algorithmize, whereas anomaly detection necessarily refers to uncertain observations and has to use statistical methods (statistical methods have been used in IDS systems since 1987, and the first IDS in which they were implemented was the "Haystack" project conducted in Los Alamos National Laboratory (see e.g., [5, page 432]).

Paper mentioned in [6] describes the application of the traditional Holt-Winters method in behavioral analysis of network traffic for development of adaptive models for various types of traffic in four sample computer networks. The next obvious step, after evaluation of the model, is the development of and predicted pattern and alert generation algorithm (see e.g., [5, 7, 8, page 419]).

2. Holt-Winters Model: Brutlag's Anomaly Detection Algorithm

The Holt-Winters model, called also the triple exponential smoothing model, is a well-known adaptive model used to modeling time series characterized by trend and seasonality (The Holt model was formulated in 1957 and the Winters model in 1960. See [9, 10, page 248], a comprehensive review of the literature about this and other models based on exponential smoothing is given in [11]). In its additive version, it presents the smoothed variant of the y_t time series as the sum of three constituents

$$\hat{y}_t = L_t + T_t + S_{t-r}, \tag{1}$$

where $\hat{y}_t$ is the value estimated by the model of the variable in moment t, r is the length of the seasonal periodicity,

$$L_t = \alpha(y_t - S_{t-r}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{2}$$

is the constituent smoothing out the level of the time series,

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{3}$$

represents the increase of the time series resulting from the trend,

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-r} \tag{4}$$

is the seasonal component of the time series, α, β, and γ are smoothing parameters, estimated for the particular time series, while y_t is the real value of the variable in moment t, and the parameters α, β, and γ belong to $[0; 1]$ interval.

Estimation of model parameters is iterative, usually though minimization of arbitrarily selected measures of error (e.g., the Mean Squared Error of expired estimations or the sum of absolute values of the residuals of the model see e.g., [12, page 187], [13, page 226], [14, page 77], and [6, 15], [16, page 223]).

Holt-Winters method was used to detect network traffic anomalies as described in [17]. In the paper concept of "confidence bands" was introduced. As described in the paper, confidence bands measure deviation for each time point in the seasonal cycle, and this mechanism bases on expected seasonal variability.

The estimated deviation of the real value of the dependent variable is

$$d_t = \gamma\left|y_t - \hat{y}_t\right| + (1-\gamma)d_{t-r}, \tag{5}$$

where d_t is the estimated deviation of the real value of the dependent variable y in moment t from the estimated value $\hat{y}_t$, where the value of parameter γ is the estimated value in the model described above in (4). The event when the real value of the dependent variable y_t differs from the estimated value $\hat{y}_t$ by more than d_t multiplied by the scaling factor m is considered an anomaly (an alert is triggered in the IDS system). In [17] the extension of the RRDtool is presented, covering real-time determination and marking of values $\hat{y}_t + md_t$ and $\hat{y}_t - md_t$ on the chart and generating information on occurring anomalies. The author assumed an arbitrary method of determining the initial values of parameters α, β, and γ as well as the iterative method of adapting only parameter α (see: [17]), which, from a statistical point of view, may provoke doubts, as it leads to development of suboptimal models from the perspective of minimization of the value of any measure of error. Additionally, in the Brutlag method, the calculated value of the parameter d_t for the purposes of determining the value above or below which anomalies will be reported is multiplied by intuitively selected scaling factor m of value between 2 and 3, which makes the model even more arbitrary (see e.g., [18]).

Thirdly, an important feature of the Holt-Winters model is the assumption on single seasonality (periodicity) of the given series, while in the case of network traffic one could expect double seasonality: daily and weekly. Anyone intending to use the Holt-Winters model to develop an anomaly detection system needs to select which periodicity should be used in the model.

3. Adaptative Models with Double and Triple Seasonality (Taylor Models)

In, [19] a suggestion is made to extend the Holt-Winters method to cover series with double, while in [20] with triple seasonality. In [21] the Taylor model with double seasonality was used to modelling internet traffic. Obviously it is theoretically possible to develop analogous models for time series with multiple periodicity; however, issues are raised in the literature (see [22]) concerning the unstable behaviour of such models, as well as the doubtful impact of third and further seasonalities on the calculated value of the predicted variable. Similar reservations also apply to double-seasonal Taylor models, in which the duration of the first period is considerably longer than that of the second one.

Double-seasonal Holt-Winters-Taylor model (referred to as HWT2 in subsequent sections) is determined by the following equations:

$$\hat{y}_t = L_{t-1} + T_{t-1} + D_{t-r1} + W_{t-r2}, \tag{6}$$

where r_1 is the length of the seasonal 1 (day) periodicity, r_2 is the length of the seasonal 2 (week) periodicity,

$$L_t = \alpha(y_t - D_{t-r1} - W_{t-r2}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{7}$$

is the constituent smoothing out the level of the series,

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{8}$$

corresponds to the increase of the series resulting from the trend,

$$D_t = \gamma(y_t - L_t - W_{t-r2}) + (1-\gamma)D_{t-r1} \tag{9}$$

is a seasonal component of the series for seasonality 1 (day), and

$$W_t = \delta(y_t - L_t - D_{t-r1}) + (1-\delta)W_{t-r2} \tag{10}$$

is a seasonal component of the series for seasonality 2 (week).

The initial values of components were arbitrarily set as

$$\begin{aligned} L_1 &= y_1, \\ T_1 &= 0, \\ D_1 &= D_2 = \cdots = D_{r1} = 0, \\ W_1 &= W_2 = \cdots = W_{r2} = 0. \end{aligned} \tag{11}$$

4. Application of Brutlag's Anomaly Detection Algorithm in the HWT2 Model

In order to identify indications of anomalies in the modelled system, an analogous solution to the one presented in [17] can be used. In view of the double seasonality in the Taylor model, one might imagine two types of scatter permitted for the value of the predicted variable—one based on the parameter γ, and the other on δ

$$\begin{aligned} d_t &= \gamma \left| y_t - \hat{y}_t \right| + (1-\gamma) d_{t-r1}, \\ w_t &= \delta \left| y_t - \hat{y}_t \right| + (1-\delta) d_{t-r2}. \end{aligned} \tag{12}$$

The initial values of components were arbitrarily set as

$$w_{r2+1} = d_{r2+1} = \left| y_{r2+1} - \hat{y}_{r2+1} \right|. \tag{13}$$

One needs to remember that the parameters of the exponential smoothing models may be interpreted as a measure of the impact of the last measurement (parameters α, β, γ, and δ) or earlier measurements (values $1-\alpha$, $1-\beta$, $1-\gamma$, and $1-\delta$) on predicted values. Contrary to descriptive models, where the estimated values of parameters given the appropriate dependent variable has an intuitive meaning (in the case of single-equation additive model, the impact of the explanatory variable on the value of the dependent variable), and the criterion of minimizing the adopted measure of adjustment is decisive, the parameters of adaptive models of time series with exponential smoothing may be interpreted as a measure of smoothing—the greater the values of γ and δ, the greater the impact of values of the last measurements (i.e., measured correspondingly one day and one week earlier), the lower the values of the parameters, the better the model "remembers" the previous values (whose impact is weighted with $1-\gamma$ and $1-\delta$ coefficients). One might, at least to a certain extent, that is if does not have too great an influence on the adopted measure of adjustment, decide to arbitrarily change the values of smoothing parameters, especially if the given series displays periodicity.

The existence of two types of permitted scatter results in the necessity of distinguishing between two types of alerts: the first one related to exceeding the thresholds determined by the parameters of daily seasonality and the second one—of weekly seasonality. As the thresholds may intertwine it is necessary to distinguish in the alerts (see Figure 1) all three possible events ("daily" threshold exceeded, "weekly" threshold exceeded, both thresholds exceeded).

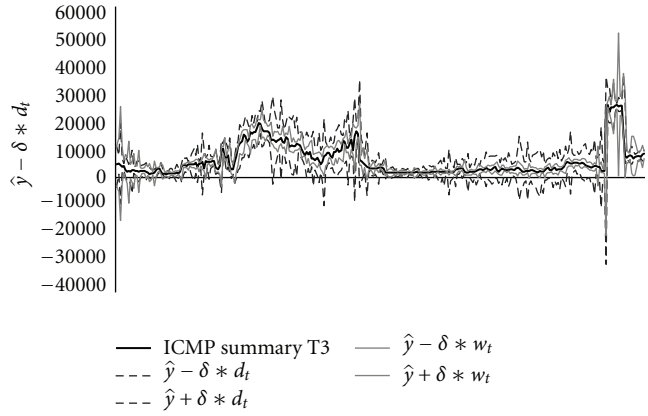

Figure 1: Sample fragment of series (308 observations, that is, 2.14 hours) ICMP in T3 network (from [6]) with daily thresholds and weekly thresholds for the scaling factor = 2.5.

5. Results and Conclusions

For the networks described in [6], the HWT2 model analysis was carried out for which the parameters were estimated through minimization of the expression

$$\frac{\text{Mean\_Absolute\_Error}}{\text{Mean}}, \tag{14}$$

where

$$\text{Mean} = \frac{\sum_{t=1}^{n} y_t}{n}, \tag{15}$$

where n is the length of the data series.

We decided to use MAE rather than mean squared error (MSE)-based measure because there were a lot of so-called outliers noted in the analysed samples and MSE-based measure that can be oversensitive in those cases (see, e.g., [21]).

In the referenced article network traffic for aggregated series was modelled (hourly data) and in the case of smaller-scale research series resulting from measurements of traffic every 10 minutes were used.

The obtained results and the number of alerts generated on exceeding the data threshold ("gamma" alert and "delta" alert) is presented in Table 1. For comparison, the table also contains the number of alerts obtained in modelling traffic using the classic Holt-Winters model (analogous to [6], this is with an iterative estimation of the value of parameters with minimization of MAE/M error, however, for 10-minute interval measurement, which is a departure from the referenced article).

As compared to the traditional Holt-Winters model, the magnitude of error is virtually unchanged (the table presents results with accuracy down to one per cent. In reality, there were differences between models in adjustment measures, at no more than one thousandth percentage point, which in practice is insignificant) whereas in all cases the number of alerts generated for the same scaling factor has lowered.

Table 1: Parameters of the HWT2 model for particular types of traffic.

Network	Protocol	Alpha	Beta	Gamma	Delta	MAE/M	Number of gamma alerts	Number of delta alerts	Number of gamma and delta alerts	Number of Brutlag's Holt Winters alerts
T3	TCP	0,999497	0,00224	0,487092	0,492391	4,11%	0	0	0	1052
T3	UDP	0,887382	0,007949	0,006263	0	15,45%	1413	2056	717	1377
T3	ICMP	0,999501	0,004789	0,487093	0,492397	8,69%	0	0	0	944
W1	TCP	0,968618	0,005005	0,002687	0,00513	45,92%	918	2193	714	1070
W1	UDP	0,997541	0,001291	0	0,15	30,19%	2743	1358	1047	3293
W1	ICMP	0,8433	0,002486	0	0,2	31,27%	2743	2504	1687	3247
T2	TCP	0,999385	0,00449	0,616195	0,499587	4, 20%	0	0	0	1014
T2	UDP	0,884039	0,000771	$3{,}54E{-}05$	0	15,87%	4562	4685	2806	4562
T2	ICMP	1	0,006488	0,5	0,5	8,53%	0	0	0	1174

Source: own research.

The potential application of the HWT2 model with the so-determined two types of thresholds may prove useful for reducing the number of false positives.

The application of the parameters γ and δ as weights determining the permissible scale makes the model have a relatively "short memory." Therefore, if the present value of the periodical constituent is closely related to its previous value, the extent of permitted thresholds will be relatively high. In the case of models with "better memory," they will be more sensitive to unitary changes in traffic. In the analyzed network traffic series, together with the high value of parameter α, the usually estimated value of parameter γ was relatively low (10^{-2} or less), the consequence of which was a relatively high number of alerts, the vast majority of which, as it would seem, would be deemed false positives. In the examined series, noted was greater sensitivity of the model to changes in the α level parameter than the seasonality impact (γ and δ), that had relatively low values. One might thus consider, instead of increasing the scaling factor, to arbitrarily increase the value of these parameters to the maximum values which do not cause significant deterioration of the model-matching score (e.g., below 1 percentage point). As it would seem, this is an approach which is better substantiated from a statistical point of view than manipulating the value of the scaling factor (an interesting challenge of the method adopted in [17] for determination of the values not provoking alerts contained [18]).

6. Further Actions

Among the methods used in anomaly detection, the following may be mentioned:

(i) entropy measurement—see, for example, [23, 24],

(ii) the so-called correlation of packets—see [25, 26]—where algorithms of simulated annealing were used,

(iii) principal components analysis—see, for example, [27, 28],

(iv) support vector machines—see, for example, [4],

(v) adaptive threshold algorithm and the cumulative sum algorithm—see, for example, [29–31],

(vi) data clustering—see also [32],

(vii) k-nearest neighbors method—see [33], decision trees—see, for example, [34],

(viii) artificial neural networks (ANNs)—see, for example, [35].

(ix) distributed ANN—see, for example, [35, 36],

(x) decision rule induction—see, for example, [37],

(xi) immune algorithms—see, for example [38],

(xii) genetic algorithms—see, for example, [39],

(xiii) fuzzy logic—see, for example [40, 41],

(xiv) zero-one models—see [42] and so forth.

As presented in [6] the characteristics of various networks or even the various types of network traffic in the same network are very different. Therefore, even if one of the widely used models of traffic or methods for their creation finds even the slightest application in test trials, its research work is practically useful.

Presently, our works are carried out on implementing both models (traditional Winters and HWT2) in the Anomaly Detection preprocessor, referred to in articl [6].

References

[1] A. Fadia and M. Zacharia, "Network intrusion alert. An ethical hacking guide to intrusion detection," in *Proceedings of the Thomson Source Technology*, Boston, Mass, USA, 2008.

[2] S. Sooyeon, K. Taekyoung, J. Gil-Yong, P. Youngman, and H. Rhy, "An experimental study of hierarchical intrusion detection for wireless industrial sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 4, pp. 744–757, 2010.

[3] E. A. Patkowski, "Mechanizmy wykrywania anomalii jako element bezpieczeństwa," Biuletyn Instytutu Automatyki i Robotyki nr 26/2009, Wydawnictwo Wojskowej Akademii Technicznej, Warsaw, Poland, 2009.

[4] F. Palmieri and U. Fiore, "Network anomaly detection through nonlinear analysis," *Computers and Security*, vol. 29, no. 7, pp. 737–755, 2010.

[5] J. Pieprzyk, T. Hardjono, and J. Seberry, *Teoria Bezpieczeństwa Systemów Komputerowych*, Helion, 2005.

[6] M. Szmit and A. Szmit, "Use of holt-winters method in the analysis of network traffic: case study," *Communications in Computer and Information Science*, vol. 160, pp. 224–231, 2011.

[7] L. Fillatre, D. Marakov, and S. Vaton, "Forecasting seasonal traffic flows," in *Proceedings of the Workshop on QoS and Traffic Control*, Paris, France, December 2005.

[8] I. Klevecka, "Forecasting network traffic: a comparison of neural networks and linear models," in *Proceedings of the 9th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat '09)*, Riga, Latvia, October 2009.

[9] P. Goodwin, "The holt-winters approach to exponential smoothing: 50 years old and going strong," in *Proceedings of the FORESIGHT Fall*, pp. 30–34, 2010, http://www.forecasters.org/pdfs/foresight/free/Issue19_goodwin.pdf.

[10] B. Guzik, D. Appenzeller, and W. Jurek, *Prognozowanie i Symulacje. Wybrane Zagadnienia*, Wydawnictwo AE w Poznaniu, Poznań, Poland, 2004.

[11] E. S. Gardner, "Exponential smoothing: the state of the art-Part II," *International Journal of Forecasting*, vol. 22, no. 4, pp. 637–666, 2006.

[12] J. Gajda, *Prognozowanie i Symulacja a Decyzje Gospodarcze*, C. H. Beck, Warsaw, Poland, 2001.

[13] A. Zeliaś, B. Pawełek, S. Wanat et al., *Prognozowanie Ekonomiczne. Teoria, Przykłady, Zadania*, Wydawnictwo Naukowe PWN, Warszawa, Poland, 2004.

[14] M. Cieślak, Ed., *Prognozowanie Gospodarcze*, Wydawnictwo AE Wrocław, 1998.

[15] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, New York, NY, USA, 2nd edition, 2002.

[16] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*, Springer, Berlin, Germany, 2008.

[17] J. D. Brutlag, "Aberrant behavior detection in time series for network monitoring," in *Proceedings of the 14th System Administration Conference*, pp. 139–146, New Orleans, Fla, USA, 2000.

[18] E. Miller, "Holt-Winters Forecasting Applied to Poisson Processes in Real-Time," August, 2010, http://www.scribd.com/doc/35521051/Miller-Automated-Error-Detection-in-Web-Production-Environment.

[19] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of Operational Research Society*, vol. 54, pp. 799–805, 2003.

[20] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *European Journal of Operational Research*, vol. 204, pp. 139–152, 2010.

[21] S. Gelper, R. Fried, and C. Croux, "Robust forecasting with exponential and holt-winters smoothing," *Journal of Forecasting*, vol. 29, no. 3, pp. 285–300, 2010.

[22] R. Lawton, "On the Stability of the Double Seasonal Holt-Winters Method," http://forecasters.org/submissions09/LawtonRichardISF2009.pdf.

[23] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, "An empirical evaluation of entropy-based traffic anomaly detection," in *Proceedings of the Association for Computing Machinery (ACM '08)*, 2008.

[24] Y. Gu, A. McCallum, and D. Towsley, "Detecting anomalies in network traffic using maximum entropy estimation," in *Proceedings of the IMC Conference*, http://conferences.sigcomm.org/imc/2005/papers/imc05efiles/gu/gu.pdf.

[25] SPADE 092200, http://rpmfind.net/linux/RPM/mandriva/9.2/i586/Mandrake/RPMS/snort-2.0.1-3mdk.i586.html.

[26] T. J. Kruk and J. Wrzesień, "Korelacja w wykrywaniu anomalii," in *Proceedings of the Materiały Konferencji CERT Secure*, Warsaw, Poland, 2003.

[27] H. Ringberg, A. Soule, J. Rexford, and C. Diot, "Sensitivity of PCA for Traffic Anomaly Detection," San Diego, Calif, USA, 2007, http://www.haakonringberg.com/work/papers/pca_tuning.pdf.

[28] A. Lakhina, M. Cronvella, and C. Diot, "Diagnosis network-wide traffic anomalies," in *Proceedings of the ACC SIGCOMM*, February 2004, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.7011&rep=rep1&type=pdf.

[29] V. A. Siris and F. Papaglou, "Application of anomaly detection algorithms for detecting syn floodinfg attacks," in *Proceedings of the IEEE Global Telecommunications Conference*, vol. 4, pp. 2050–2054, 2004.

[30] R. Mbabazi, *Victim-based defense against ip packet flooding denial of service attacks*, M.S. thesis, Makerere University, 2009.

[31] R. Blazek, H. Kim, B. Rozovskii, and A. Tartakovsky, "A novel approach to detection of "Denial-of-Service" attacks via adaptive sequential and batch-sequential change-point detection methods," in *Proceedings of the IEEE Systems, Man, and Cybernetics Information Assurance Workshop (West Point '01)*, June 2001.

[32] O. Siriporn and S. Benjawan, "Anomaly detection and characterization to classify traffic anomalies case study: TOT public company limited network," *Proceedings of World Academy of Science, Engineering and Technology*, vol. 37, pp. 706–714, 2009.

[33] A. Sharma, A. K. Pujari, and K. K. Paliwal, "Intrusion detection using text processing techniques with a kernel based similarity measure," *Computers and Security*, vol. 26, no. 7-8, pp. 488–495, 2007.

[34] S. O. Al-Mamory and H. Zhang, "New data mining technique to enhance IDS alarms quality," *Journal in Computer Virology*, vol. 6, no. 1, pp. 43–55, 2010.

[35] D. Tian, Y. Liu, and Y. Xiang, "Large-scale network intrusion detection based on distributed learning algorithm," *International Journal of Information Security*, vol. 8, no. 1, pp. 25–35, 2009.

[36] Snort+AI, http://snort-ai.sourceforge.net/.

[37] R. Cichocki, "Algorytmy indukcji reguł decyzyjnych w Systemach Wykrywania Intruzów," in *Proceedings of the XII Konferencja Sieci Komputerowe*, Zakopane, Poland, 2005.

[38] D. Dasgupta, "Immunity-based intrusion detection system: a general framework," in *Proceedings of the 22nd National Information Systems Security Conference (NISSC '99)*, 1999.

[39] W. Li, "Using genetic algorithm for network intrusion detection," in *Proceedings of the United States Department of Energy Cyber Security Group 2004 Training Conference*, Kansas City, Mo, USA, 2004.

[40] J. Luo, S. Bridges, and R. Vaughn, "Fuzzy frequent episodes for real time intrusion detection," *International Journal of Intelligent Systems*, vol. 15, no. 8, pp. 687–704, 2000.

[41] S. Bridges and R. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in *Proceedings of the National Information Systems Security Conference (NISSC '00)*, Baltimore, Md, USA, October 2000.

[42] M. Szmit, Využití nula-jedničkových modelů pro behaviorální analýzu síť ového provozu, [w:] Internet, competitiveness and organizational security, Tomas Bata University Zlín, pp. 266–299, 2011.