

EPIDEMIC NOISE MODELS AND EPIDEMIC FILTERING – PRELIMINARY PROPOSAL AND ANALYSIS OF CAPABILITIES

Horia-Nicolai TEODORESCU ^{1,2}

¹ “Gheorghe Asachi” Technical University Iasi, Romania

² Institute of Computer Science of the Romanian Academy, Iasi, Romania

E-mail: hteodor@etti.tuiasi.ro

Based on a model for correlated noises that uses analogies with epidemic spreading, I introduce the concept of epidemic filtering and an epidemic filtering algorithm that makes guesses on the contaminated signal samples and cleans only those samples. A preliminary analysis of the performances of the filters is presented and discussed.

Key words: epidemic noise model, burst noise, epidemic filter, nonlinear filter.

1. EPIDEMIC MODEL FOR NOISE

Recently, epidemic algorithms, mimicking the spreading of epidemics, were introduced for modeling communication processes [1], networks of sensors [2, 3], searches in databases for correspondences and disparities in images [4, 5], computer infection transmission [6-8], image search and annotation [9]. In this paper, a model for noise contamination of signals and a preliminary “epidemic filter” for such signals are proposed. The filtering procedure was outlined in [10, 11]. The signal model corresponds to noises that are impulsive and correlated. Such noises occur for example in audio and ultrasound signals, where impulsive bursts of noise occur due to movements of the objects and people on the scene, to wind, to flowing water, to fire, or to other similar causes. The filter presented herein is for one-dimensional (1D) signals; extensions to 2D signals are straightforward.

In the epidemic model of the noisy signals proposed, the samples of the signal compose the population. The epidemic model follows the standard epidemic SIR model [12, 13] that uses three states (S = susceptible, I = Infected, R = recovered). All samples are equally susceptible to infection except those already “cured” by filtering, when the filtering is applied between the epochs of contamination with noise (case not addressed in this paper). The equal susceptibility means that every sample has the same probability of being infected, under the same conditions. In other words, in the models discussed, the probabilities are not related to the samples.

In the model, while all samples are equally susceptible to infection, the noise, which stands for the infection, is a correlated process in the sense that samples close to an infected one have greater probability of being noisy (infected). The contamination process in the vicinity of an infected element in the population creates a correlation between neighboring samples. Examples of natural noises that are highly correlated are the pink noise, which has a $1/f$ spectrum, and the red noise, which has a $1/f^2$ spectrum. Similarly to these natural processes, the probability of contamination in the neighborhood is related to the ratio of low frequency to high frequency content of the noise. When the noise is high frequency, like the salt and pepper noise, the probability of infecting the neighbors tend to zero, while in a noise with high content of low frequency, as in burst noise, the probability is high. Notice that we address noise models where the noise is created by non-independent time events.

Another parameter describing the noise model is the “age /stage” of the infection, measured in epochs (or time since inception, in a continuous model). For a given probability, the age shows the number of potential propagations of the infection from the initially infected samples to the adjacent neighbors and the

other neighbors in determined vicinity. Notice that higher is the low frequency content, higher is the age of infection, for a given contamination probability, because more samples will be infected in a longer time, thus summing up to larger noise impulses. Therefore, the noise seen as an epidemic process has three main information pieces: the initially infected population, the probability of infection between neighbors, and the age of the infection (the last reflecting the infection propagation steps, epochs).

The above model creates only binary noises, in the sense that a sample is either infected or not infected. To account for various levels of noise, we must assume that the infected samples have specific probabilities to show various levels of infection. So, a complete model is characterized by the probabilities to get infected samples with a specified level, $p(j)$, where j is a level of infection. In our models, $p(j)$ represents a uniform distribution, with $p(j)$ distributions are identical for all noise samples. Thus, the noise model is based on non-independent, identically distributed (non-i.i.d) process.

We may assume that the initially infected sub-population is composed by the samples that are somehow distant (outliers) in their vicinity. That property will serve to identify these samples. Once the initially infected samples are detected, further infected samples are determined probabilistically by identifying (selecting) neighbors with a probability p for the first epoch. For the second epoch, we determine again in the vicinity of the infected samples at step 1 what samples are infected, including the samples that are not adjacent to the originally infected samples. The infection process described continues until the age of the infection in the population, that is, a predetermined number of epochs.

Once the probably infected samples are determined, they are cured by replacing them with the average of the non-infected samples in a specified window, or with their median etc. We thus obtain median epidemic filters, epidemic average filters (subsequently denoted by EA) etc. These filters are as good as the corresponding epidemic model of the noise is.

Two models of noise are used in this paper for comparing the filtering results. The first type of noise is an impulsive noise, with no correlation. The second one is a correlated noise. Subsequently, t stands for the time variable, which is a discrete variable in the simulations, $r_2(t)$ denotes a random variable with uniform distribution and average \bar{r}_2 , $r_1(t)$, is a random variable unrelated to $r_2(t)$, $s_0(t)$ denotes a deterministic signal, and $s(t) = s_0(t) + n(t)$ is the noisy signal, where $n(t)$ is defined as

$$n(t) = \begin{cases} 0, & \text{if } r_1(t) > 0.4 \\ 0.3 \cdot (r_2(t) - \bar{r}_2), & \text{if } r_1(t) \leq 0.4 \end{cases}$$

In simulations, both r_1 and r_2 are uniformly distributed between 0 and 1. The successive samples of the noise $n(t)$ are not correlated.

In case that the noise in a sample is not zero, the sample is named *infected*. A state variable is assigned to the samples, denoted by $S(s_n)$, where s_n is the sample; the state variable take two values, non-infected and infected, the last one denoted by $S(s_n) = I$.

The model for the correlated noise is generated by a simple correlation procedure. First, the procedure uses a random variable uniformly distributed in [0,1] and compares its value to 0.4. Then, the difference between the average of the samples in a window and the average of the current and precedent samples is compared with the standard deviation in the window. When $|(s_n + s_{n-1})/2 - \bar{s}_n^W| > 0.3 \cdot STDV(s_n^W)$, the previous sample, s_{n-1} , is assumed infected, that is noise is assumed in that sample. Either one of the two conditions is satisfied, a uniform noise with peak-to-peak amplitude of 0.3 and zero average is added to the sample. The corresponding formula is:

$$n(t) = \begin{cases} 0.3 \cdot r_1(t) - 0.15 & \text{if } |(s_n + s_{n-1})/2 - \bar{s}_n^W| > 0.3 \cdot STDV(s_n^W) \text{ or } r_2(t) > 0.4 \\ 0 & \text{else} \end{cases}$$

where the window has five samples and is centered on the current sample.

Thus, the probability that the current sample is contaminated when the neighboring ones are not is 0.4 in the simulations. The values of the noise in the infected samples are 0.3 peak-to-peak and have uniform distribution. Notice that this manner of defining the noise also takes into account the signal, which is a not

typical procedure in noise simulation. However, this procedure of noise generation has the characteristics of processes where noise is both correlated with itself and with the signal. Also note that when the noise is small in the previous sample (compared to the average of the signal in the window), it has lower chance to propagate in the subsequent sample.

2. THE EPIDEMIC FILTERS

The filtering algorithm has three stages. In the first one, the most probable noisy samples are determined and marked as initially infected. The second stage consists in running an epidemic model that determines the samples that probably are noisy and related to the previously detected noisy samples due to the infection propagation (standing for the correlation in the noise). Thus, in the first and the second stages, the algorithm performs a guess of the noisy samples, based on an epidemic model of the noise. Notice that the epidemic model of the noise is assumed to be correct and a priori known. In the third stage of the algorithm, a standard filtering procedure is applied only in windows centered on the samples that were marked as noisy in the previous two stages.

Subsequently, the propagation of the infection during the epidemic epochs is described by the probabilities of infection at various epochs of the current sample, due to infected neighbors. The infection propagation mimics the correlation of the noise.

The algorithm is as follows:

For samples from the first to the last

Determine the probable initial infection, that is, determine the highly probable "infected" samples of the signal.

Noisy samples related to the ones previously found to be noisy are determined by simulation of an epidemic propagation as follows:

For the first epoch of infection, if there are contaminated neighbors of the current sample, then it becomes contaminated with probability p_{e1} . A sample once contaminated remains contaminated.

In the subsequent epochs, if the current sample has neighbors that are contaminated from the previous epochs or initially contaminated, then it becomes contaminated with a probability p_{ek} , where k is the number of the epoch. All infected samples are marked with state I .

For samples from the third to the last -2 (in a window of 5 samples)

If state of current sample = I , apply the average filter to it.

The detection of the samples assumed initially contaminated is performed by definition as follows. In a discrete time process, a sample is chosen to be initially contaminated when

$$|s_n - \bar{s}_n^W| > 0.5 \cdot STDV(s_n^W),$$

where \bar{s}_n^W represents the average of the samples of the noisy signal in a window of W samples centered on the current sample, s_n , and $STDV(s_n^W)$ is the standard deviation of the samples in the same window.

Two slightly different epidemiologic models were tested and the results are compared in this paper.

First model of contamination. From the point of view of the time-dependent signal, the first contamination process is a non-causal one, as a sample can get contaminated by the noise in a future, not-yet-extant sample. The first contamination model is as follows. After the initial epoch, a sample becomes contaminated according to the rule: a non-initially contaminated sample is assumed contaminated with a probability p_{e1} in the first epoch of the contamination process if at least one of the samples in the window is contaminated,

$$p\left((S(s_n^{e1}) = I) \mid \left((S(s_{n-2}^{in}) = I) \vee (S(s_{n-1}^{in}) = I) \vee (S(s_{n+1}^{in}) = I) \vee (S(s_{n+2}^{in}) = I) \right)\right) = p_{e1}.$$

The notations are as follows: $S(s_{n-1}^{in}) = I$ means that the state of the initial sample at time moment $n-1$ is I (infected) etc. The vertical bar $|$ denotes the condition in the conditional probability. Subsequently, logical conditions \vee (OR) are used without parentheses between the clauses.

The contamination in the second epoch is similarly determined by $p(S(s_n^{e2}) = I | C_2) = p_{e2}$, where the condition C_2 is

$$C_2 = (S(s_{n-2}^{in}) = I) \vee (S(s_{n-1}^{in}) = I) \vee (S(s_{n+1}^{in}) = I) \vee (S(s_{n+2}^{in}) = I) \vee (S(s_{n-2}^{e1}) = I) \vee (S(s_{n-1}^{e1}) = I) \vee (S(s_{n+1}^{e1}) = I) \vee (S(s_{n+2}^{e1}) = I).$$

The above condition could be written in a more compact manner using the Polish notation as

$$C_2 = \vee(S(s_{n-2}^{in}), S(s_{n-1}^{in}), S(s_{n+1}^{in}), S(s_{n+2}^{in}), S(s_{n-2}^{e1}), S(s_{n-1}^{e1}), S(s_{n+1}^{e1}), S(s_{n+2}^{e1})) = I.$$

Similarly, the contamination during the third epoch is conditioned by the infected neighbors with a probability $p(S(s_n^{e3}) = I | C_3) = p_{e3}$, where the condition C_3 is

$$C_3 = C_2 \vee (S(s_{n-2}^{e2}) = I) \vee (S(s_{n-1}^{e2}) = I) \vee (S(s_{n+1}^{e2}) = I) \vee (S(s_{n+2}^{e2}) = I).$$

The second epidemic model. The samples after the first epoch become infected with the conditional probability $p((S(s_n^{e1}) = I) | (S(s_{n-1}^{in}) = I) \vee (S(s_{n-2}^{in}) = I)) = p_{e1}$, that is, taking into account only neighboring samples from the past of the signal.

Samples after the second epoch are contaminated with the probability conditioned by the state of the samples that appear in time before the current sample, that is, by samples with indices s_{n-1}, s_{n-2} , $p(S(s_n^{e2}) = I | ((S(s_{n-1}^{in}) = I) \vee (S(s_{n-2}^{in}) = I) \vee (S(s_{n-1}^{e1}) = I) \vee (S(s_{n-2}^{e1}) = I))) = p_{e2}$. Samples after the third epoch are contaminated with the probability

$$p(S(s_n^{e3}) = I | \vee(S(s_{n-1}^{in}), S(s_{n-2}^{in}), S(s_{n-1}^{e1}), S(s_{n-2}^{e1}), S(s_{n-1}^{e2}), S(s_{n-2}^{e2})) = I) = p_{e3}.$$

Summarizing, in all epochs, the current sample becomes infected with the corresponding probability when at least one of its neighbors in the past of the signal is infected. In the simulations reported in the next section, the conditional probabilities are all equal, $p_{e1} = p_{e2} = p_{e3} = 0.3$.

The filter applies the average in an asymmetric window where the past dominates the future, when the current sample is infected, that is when the condition

$$C = (S(s_n^{in}) = I) \vee (S(s_n^{e1}) = I) \vee (S(s_{n-1}^{e2}) = I) \vee (S(s_{n-1}^{e3}) = I)$$

is true. The filter action is expressed by

$$s_n^* = \begin{cases} s_n & C = \text{false} \\ (s_{n-2} + s_{n-1} + s_n + s_{n+1}) / 4 & C = \text{true} \end{cases}.$$

The result of the epidemic average (EA) filter described above is compared to the average filter with symmetrical window centered on the current sample in the next section.

3. RESULTS

The two epidemic filters described above were run for several signals with several noise types, several contents, and with several parameters of the algorithm. Examples of results are provided subsequently. In the first part of this section I exemplify results obtained with non-correlated noise, while in the second part of this section the case of filtering correlated noise is shown. The signal used in tests is the arbitrary signal with non-harmonic components, $s_0(t) = 3 + \sin(t/5) + 0.4 \cdot \sin(t/17)$.

In all the results exemplified, the filtering was performed with the following parameters: number of epochs $N_e = 3$; the probability of contamination of the neighbors was between 0.1 and 0.4; initially contaminated samples are detected by the condition that the difference of the current sample and the average in a window of 5 samples centered on the current sample is larger than $0.3\sigma \dots 0.5\sigma$. In all figures showing filtering results, the upper trace is the filtered signal with epidemic (average-type) filter, while the lower

trace is unfiltered signal. Noise in the signal has the peak amplitude 0.15 or 0.2 (peak to peak 0.4). The filtered version of the signal is shifted in all figures for convenience. Figures 1, 2, and 3 exemplify the results obtained with the epidemic filter for non-correlated noise.

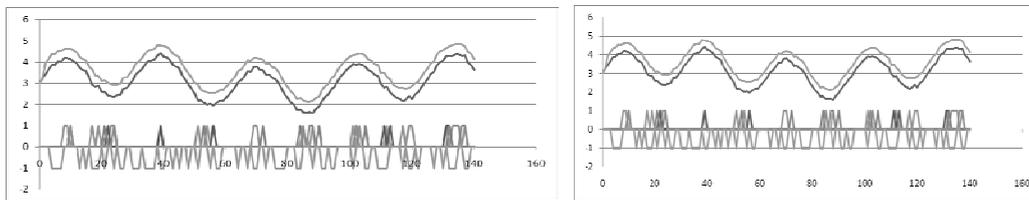


Fig. 1 – First epidemic filter, non-correlated noise, $N_e = 3$; contamination $p = 0.1$; initially 0.5 sigma, peak 0.2, average filter; – two examples of noise contamination. Negative impulses (0, -1) mark the samples actually contaminated by noise; positive impulses mark the samples predicted by the epidemic algorithm. The lower analog trace is the noisy signal. The upper trace is the filtered signal (shifted up for visibility).

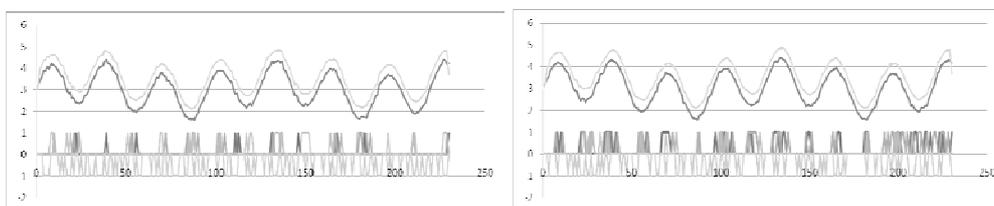


Fig. 2 – Same, but with probability of contamination $p = 0.2$ (see legend to Fig. 1) – two examples of noise contamination (shown case not in among those reported in Table 2).

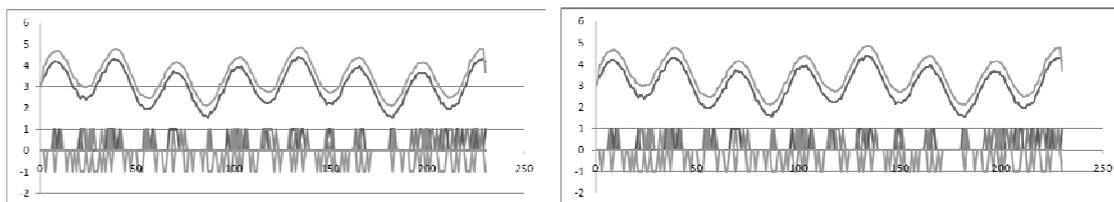


Fig. 3 – Same, but with probability of contamination $p = 0.3$ (see legend to Fig. 1) – two examples of noise contamination.

Notice the differences in the samples that are truly affected by noise and those “guessed” by the described epidemic filtering algorithm. Many samples that the algorithm guesses are wrongly guessed, while many truly noisy samples are not detected. These errors are specific to the algorithm and constitute the most serious limitation of the proposed filters.

Table 1

The error produced by the averaging epidemic filter on signals with non-correlated noises (19 runs)

Total error in signal	Epidemic average filter – total error	Standard average filter– total error
0.6976	0.6649	0.3193

The MSE of the filters, compared to the MSE for a standard average filter with the same window, are given in Table 1. Notice that the results produced by the EA filter are significantly weaker on average than those for the average filter. Moreover, in some cases, the results are poorer than the initial signal, as the MSE error after the EA filter is higher than the MSE in the input noisy signal.

Epidemic predictive filtering for correlated noise. Samples of the simulation results are subsequently presented with details on the simulation parameters, for correlated noises. Figure 4 shows that for correlated noise, the results of the EA filtering can be poorer than the results obtained with the simple average filter. However, in this case, sometimes better results can be produced by the EA filter as shown in Fig. 5.

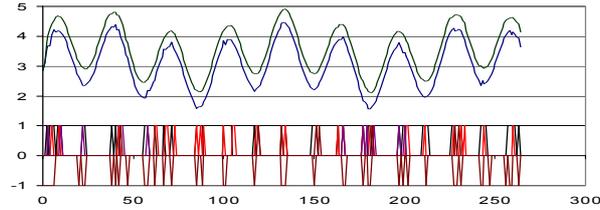


Fig. 4 – A case when the EA filter behaves significantly poorer than the average filter. Results obtained: MSE = 0.268 for the EA filter, MSE = 0.220 for the standard averaging filter; MSE in the original signal was 0.4219.

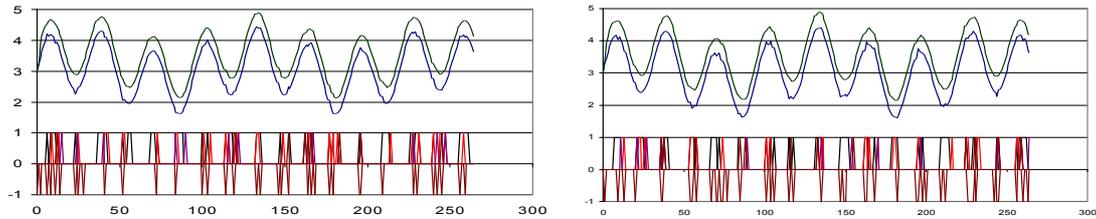


Fig. 5 – Results obtained in two favorable cases. Left: MSE = 0.2538 for the EA filter, MSE = 0.2846 for the standard averaging filter; MSE in the original signal was 0.3077. Right: MSE = 0.3166 for the EA filter, MSE = 0.3195 for the standard averaging filter; MSE in the original signal was 0.3996

Table 2

The error produced by the averaging epidemic filter on signals with correlated noises

Run	MSE for the epidemic filter	MSE for the average filter	MSE of original noisy signal
1	0.3425	0.2849	0.3021
2	0.2812	0.2731	0.3172
3	0.3077	0.2739	0.3164
4	0.2053	0.1878	0.3372
5	0.3103	0.2528	0.2811
6	0.2232	0.2211	0.2347
7	0.3879	0.3443	0.3757
8	0.3166	0.3196	0.3996
9	0.2728	0.2819	0.3175
10	0.4009	0.3462	0.4416
11	0.3750	0.2353	0.3828
12	0.3276	0.2250	0.3525
13	0.3284	0.3086	0.3239
14	0.3379	0.2694	0.4537
15	0.1729	0.2299	0.3276
16	0.3924	0.3558	0.3344
17	0.2834	0.2385	0.2591
18	0.1752	0.2210	0.2942
19	0.2707	0.2680	0.4154
Average	0.3006	0.2704	0.3404

The results in Table 2 show that, on average, the epidemic average filter accomplishes weaker results than the average filter, as expected; yet, the ratio between the MSE reduction and the number of processed samples is superior to the EA filter. Moreover, in some cases, for example 8, 9, 15, and 18, the results are better than the ones obtained on the same noisy signals with the average filter. More results, as well as Excel files with the implementations of the noise models and of the filters are posted at the Web address that can be obtained from the author.

4. DISCUSSION AND CONCLUSIONS

Epidemic filters with probabilities of infection tending to 1 and with the number of epochs tending to infinity would “guess” that all samples are contaminated and would apply to all samples the standard average filtering procedure. This shows that the presented epidemic filters and other similar ones based on other standard filtering procedure represent the generalization of the corresponding standard filters.

The epidemic filter can be seen as a filter with a decision generator that determines to what samples (windows) the related standard filtering should be applied. The decision generator is in fact a statistical predictor, the prediction referring to the state (noise contamination) of the current sample. The prediction is performed based on the a priori knowledge about the type of noise (that the noise is correlated), on the degree of correlation (how long is the lag of correlation) and on the probabilities related to the noise process. The epidemic model uses this knowledge to make a guess on which samples are noisy.

While on average the epidemic filters describe here can not be better than the average filter because they only limit the averaging operation to a selected set of samples, they may sometimes produce better results.

The first noise model presented is not correlated. Epidemic filters as introduced in this paper are conceived for correlated noises, being a noise model based type of filter. That explains why the epidemic filters produce poor results on signals contaminated with that noise. When a correlated noise is used, the epidemic filter performs considerably better, as expected, occasionally outperforming the average filter, while performing significantly fewer averaging operations.

The proposed class of noise models is able to describe temporally or spatially correlated noise. Based on the model, simple filters aiming at the cleaning of signals contaminated with “epidemic-like” noise are presented. The filters represent versions of the typical average, median, and spreading-based filters. For the class of addressed noises, the proposed filters produce results comparable, yet poorer than the classical filters, but with fewer operations performed. Thus, the main advantage of the epidemic filters is the reduced number of filtering operations; the price paid is the overload due to the epidemic model and decision block. When the overload is smaller than the cost of the omitted filtering operations, the epidemic filters may be of interest in applications.

Further work, to be presented in other papers, includes analysis of filtering applied between the epochs of contamination with noise, use of weights in the average filter dependent on the probability of noise contamination of the samples, and use of other filtering procedures, for example median filtering. An obvious further improvement to the presented algorithm would be to use a window with the width dependent to the number of infected samples around the central sample in the window, as guessed by the algorithm. It seems natural to increase the width of the window when more than one infected sample is guessed in the window with the basic width. Extensions to specialized filters, like the ones presented in [14] are also envisaged.

ACKNOWLEDGMENTS

I thank the partial support of ADBIOSONAR grant (ANCS, Romania) in 2011. I thank colleagues M. Hulea, M. Zbancioc, S. Bejinariu and Cristina Nita for patiently reading preliminary versions of the paper and proposing corrections and improvements.

REFERENCES

1. P. TENNENT, M. HALL, B. BROWN, M. CHALMERS, S. SHERWOOD, *Three applications for mobile epidemic algorithms*, Proc. Mobile HCI 2005, Salzberg, Germany, ACM Press, pp. 223–236.
2. M. AKDERE, C. C. BILGIN, O. GERDANERI, I. KORPEOGLU, O. ULUSOY, U. CETINTEMEL, *A comparison of epidemic algorithms in wireless sensor networks*, Computer Communications, **29**, pp. 2450–2457 (2006).
3. D. GANESAN, B. KRISHNAMACHARI, A. WOO, D. CULLER, D. ESTRIN, S. WICKER. *An empirical study of epidemic algorithms in large scale multihop wireless networks*, IRB-TR-02-003, March, 2002.
4. G. OLAGUE, C.B. PÉREZ, F. FERNÁNDEZ, E. LUTTON, *An artificial life approach to dense stereo disparity*, Artif. Life Robotics (2009), **13**, pp. 585–596 © ISAROB 2009 (DOI 10.1007/s10015-008-0621-6).
5. C.B. PEREZ, G. OLAGUE, F. FERNANDEZ, E. LUTTON, *An evolutionary infection algorithm for dense stereo correspondence*, In F. Rothlauf et al. (Eds.): EvoWorkshops 2005, LNCS 3449, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 294–303.
6. R. THOMMES, M. COATES, *Epidemiological modelling of peer-to-peer viruses and pollution*. 25 Annual IEEE Conf. Computer Communications (INFOCOM'06), IEEE Press, Barcelona, Spain, 2006, pp. 15–26 (available as a report at <http://target0.be/madchat/vxdevl/library/Epidemiological%20Models%20of%20Peer-to-Peer%20Viruses%20and%20Pollution.pdf>).
7. J. SCHAFER, K. MALINKA, *Security in peer-to-peer networks: empiric model of file diffusion in BitTorrent*, Fourth Int. Conf. on Internet Monitoring and Protection, 2009 (ICIMP'09), 24–28 May 2009, Venice, Mestre, Italy, pp. 39–44.
8. K.K. RAMACHANDRAN, B. SIKDAR, *Dynamics of malware spread in decentralized peer-to-peer networks, dependable and secure computing*, IEEE Trans. on, July–Aug. 2011, **8**, 4, pp. 617–623, 2010.

9. BOHLOOL, M., MENEZES, R., RIBEIRO, E., *A network-centric epidemic approach for automated image label annotation*, 2nd International Workshop on Complex Networks, Rio de Janeiro, 13–15 Oct. 2010, in *Complex Networks (Series)*, Communications in Computer and Information Science, **116**, pp. 138–145, 2011.
10. H.N. TEODORESCU, *Modeling epidemics by neuro-fuzzy coupled map lattices – natural and artificial populations*, EHB 2011, Nov. 24–26, 2011, Iasi, Romania (unpublished).
11. H.N. TEODORESCU, *Modeling epidemics with population behavioral factors included*, Epidemics'3 Conference, Boston, US, 29 Nov.–2 Dec., 2011 (unpublished, abstract in CD Proceedings).
12. A. KALAFUT, A. ACHARYA, M. GUPTA, *A study of malware in peer to peer networks*, Proceedings 6th ACM SIGCOMM on Internet Measurement IMC '06, 2006; <http://conferences.sigcomm.org/imc/2006/papers/p33-kalafut.pdf>.
13. S. EDLUNDA et al., *Comparing three basic models for seasonal influenza*, Epidemics' 3, 2011, pp. 135–142.
14. H.N. TEODORESCU, *Adaptive filter and estimator for echolocation in air and robotic vision*, in: Stoica A., Arslan T., Huntsberger T. et al. (Eds.), Proc. ECSIS Symposium on Advanced Technologies for Enhanced Quality of Life (AT-EQUAL 2009), Iasi, Romania, Jul 22–26, 2009, pp. 163–168.

Received January 4, 2012