Acta Technica 62 No. 1B/2017, 427–436

Leakage monitoring algorithm of water supply pipe network based on information entropy difference[1]

Zhixia Han[2]

Abstract. In order to improve the performance of leakage monitoring and reduce the constraint of single discrimination parameter in unit process, this paper proposes a leakage monitoring algorithm for water supply pipe network based on information entropy difference. Firstly, the Kalman set is used to predict the fault quickly. Secondly, the parameter fault coverage is introduced. The information entropy difference is used to filter the fault. Finally, the entropy difference of the parameter fault information is defined to complete the source leakage monitoring. The simulation results show that the fault set predicted by this algorithm has compressibility, and the set of faults after screening retains the real fault, and has higher fault detection rate and lower false alarm rate.

Key words. Information entropy difference, leakage monitoring, Kalman set, information entropy.

1. Introduction

As the amount of data carried in the network increases, the survivability of the network is increased and the loss caused by the network fault is reduced. This is of great significance. In order to realize the timely recovery of network fault-missing service, fast and accurate leakage monitoring and monitoring mechanism is needed. Therefore, as an essential part of survivability research, high performance monitoring algorithm has been a hot research topic both at home and abroad [1, 2].

The alarm obtained by the monitoring module in the network can be regarded as an external symptom of the fault. According to the collected symptom set, the most likely water supply network leakage collection can be predicted, and the root

[1]This work was support by the Industrial Research Project of Shaanxi Science and Technology Agency (2014K05-47). The Research of Leakage Monitoring Application Technology of Water Supply Pipe Network Based on Complex System Theory.

[2]Electronics & Electric Engineering College, Baoji University of Arts and Sciences, Baoji, Shaanxi, China, 721016

cause fault can be detected. In order to improve the monitoring performance, a fault propagation model must be introduced to represent the causal relationship between the fault and the symptom, because one symptom may correspond to multiple faults.

Based on the traditional monitoring methods such as monitoring loop, monitoring trace and monitoring tree [3–7], good monitoring performance can be achieved. However, the need to use additional detection wavelength to achieve leakage monitoring, monitoring costs are high. Therefore, in order to reduce the consumption of network resources for leakage monitoring, the leakage monitoring algorithm using traffic capture symptom in the network has been extensively studied. In this kind of algorithm, the leak detection algorithm based on Kalman set can realize the accurate leakage monitoring in the network even if the symptom is not complete, which has great advantage. The fault propagation model based on Kalman set is used in literature [8, 9]. The leakage detection is realized by the approximate reasoning algorithm. The monitoring performance is better, but the computation complexity is higher. In the paper [10], a simplified Kalman set is used as the fault propagation model. A discriminating parameter is defined as the criterion of fault diagnosis, which reduces the computational complexity and achieves better monitoring performance. Because of the limited parameters, it is difficult to achieve more accurate leakage monitoring using a single discriminating parameter.

In order to improve the performance of the monitoring algorithm in the bipartite fault propagation model, an information entropy difference based on water supply pipe network leakage detection (IWLD) is proposed in this paper. The algorithm divides the monitoring process into three modules: fault prediction, screening and monitoring. First, the prediction module quickly finds the largest possible fault set corresponding to the symptom. Then, the filter module introduces parameter fault coverage to convert the maximum possible fault set into the signal, and uses the information entropy difference (CS) method to eliminate the redundancy of the signal. Finally, the entropy difference (ED) of the parameter information is defined to identify the root cause fault. The fault detection and monitoring are completed by several parameters respectively, and the monitoring performance is improved.

2. Fault information entropy difference monitoring algorithm

2.1. Fault prediction module

Probabilistic Weighted Bipartite Graph (PWBG) is chosen as the fault propagation model in order to achieve fast fault prediction and accurate leakage monitoring, as shown in Fig. 1. According to PWBG, it is possible to quickly and accurately identify all possible faults associated with the sign in the symptom set, get the maximum possible fault set.

The maximum possible failure set is the set of all possible faults associated with the symptom set.

Define Redundancy rate to represent percentage of the number of redundant faults in the largest fault collection, which is defined as shown in expression

$$R\left(H_{\mathrm{MaX}}\right)=\frac{|H_{\mathrm{R}}|}{|H_{\mathrm{MaX}}|}=1-\frac{|F_{\mathrm{C}}|}{|H_{\mathrm{MaX}}|}. \tag{1}$$

Here, H_{MaX} represents the largest possible set of failures, H_{R} represents the set composed of redundant faults in H_{MaX}, and F_{C} represent the set of real failures in the network.

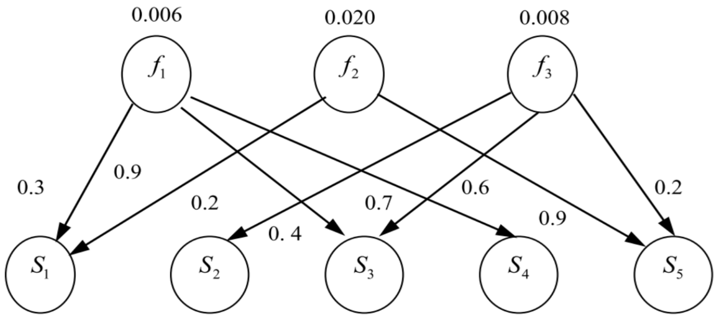

Fig. 1. Probabilistic Weighted Bipartite Graph (PWBG)

The maximum number of possible fault sets is always greater than the number of faults actually occurred on the network, i.e., most of the failures are not actually occurring. Therefore, it is necessary to filter the failures in the set, filter out the possibility of less likely failure, get fewer possible number of elements of the fault set, the maximum possible elimination of redundant faults on the monitoring algorithm to achieve more accurate leakage damage monitoring.

2.2. Information entropy difference fault screening algorithm

In order to reduce the redundancy and reduce the influence of redundancy fault on fault judgment, we must eliminate redundant faults in the largest fault collection as much as possible, and select the maximum fault set, screening out the fault of larger possibility, taking the failure as a signal, the introduction of coverage as the signal strength of the fault, the fault screening problem will be converted into signal processing problems. Information entropy difference method can be used as a signal processing method to achieve important information to retain the signal and remove part of the signal or all the redundant information purposes, to achieve fault collection screening.

Workflow of information entropy difference is shown in Fig. 2. Define the signal strength of the corresponding fault signal as the important information, and the signal strength of the redundant fault corresponding signal as redundant information.

(1) The first step of the entropy difference is to verify the compressibility of the signal, i.e. to show that the processed signal contains redundant information. The maximum fault set contains redundant faults with high redundancy, and the

corresponding signals contain more redundant information and are compressible. Therefore, it is reasonable to use the information entropy difference method to deal with the fault signal.

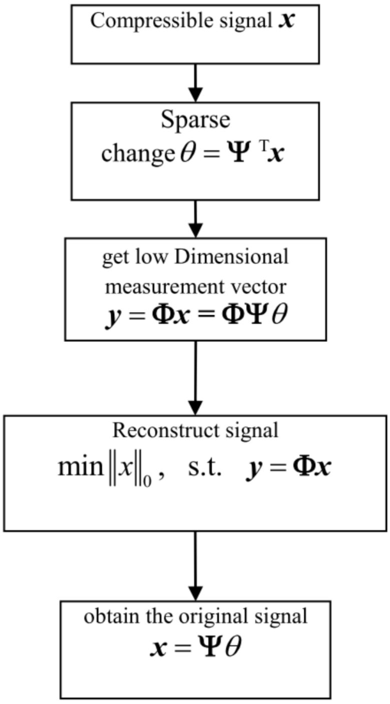

Fig. 2. Information entropy difference workflow

(2) The second step of the information entropy difference is to obtain the observation vector value of the signal according to the observation model. The observation model mainly uses the perceptual matrix $\boldsymbol{\Phi}$ to project the signal and obtain the observed vector value of the signal. It is the key to improve the filtering performance of the perceptual screening method by designing the appropriate sensing matrix so that the observed vector values contain both the important information of the original signal and the redundant information as much as possible.

Set the signal intensity threshold α_{SI}, the design of the sense matrix, to retain fault signal of the original signal $(x = (x_1, x_2, ..., x_n))$, whose signal strength is higher than the threshold. Here, α_{SI} is obtained from (2), in which, $\mu\,(0 \leq \mu \leq 1)$ is the scale factor, the size of α_{SI} being under flexible control.

$$\alpha_{\mathrm{SI}} = \mu \cdot \mathrm{Max}\,\{x_1, x_2, ..., x_n\}\ . \tag{2}$$

Since the diagonal matrix is multiplied by the target matrix, the size of the elements within the target matrix can be scaled. Thus, a diagonal matrix $\boldsymbol{A}$ ($\boldsymbol{A} =$ diag $(a_1, a_2, ..., a_n)$) may be introduced as the perceptual matrix $\boldsymbol{\Phi}$ of the original signal $\boldsymbol{x} = (x_1, x_2, ..., x_n)$. The diagonal matrix element values are given by equation

(3) below. According to the designed perceptual matrix $\boldsymbol{\Phi} = \boldsymbol{A}$, the observation vector $\boldsymbol{y}$ of the original signal $\boldsymbol{x}$ can be obtained by equation (4).

$$a_i = \left\{ \begin{array}{l} 1, x_i \geq \alpha_{\mathrm{SI}} \\ 0, x_i \geq \alpha_{\mathrm{SI}} \end{array} \right\} . \tag{3}$$

(3) The third step of the information entropy difference is to reconstruct the original signal value according to the observation vector $\boldsymbol{y}$. As shown in equation (2), in all cases which meet $\boldsymbol{y} = \boldsymbol{\Phi x}$ to find the sparsest characteristics of the signal $\boldsymbol{x}'$ is the demand, which meet the requirements of the collection, $\boldsymbol{x}'$ has the minimum number of non-zero elements.

The set of fault components corresponding to non-zero elements in the reconstructed signal $\boldsymbol{x}'$ is called the filtered fault set H_{S}. Use the filtered fault set H_{S}.

The true fault coverage $\boldsymbol{\eta}\left(H_{\mathrm{S}}\right)$ and redundancy of the set $R\left(H_{\mathrm{S}}\right)$ is used to observe the performance of the screening algorithm. The calculation of $\boldsymbol{\eta}\left(H_{\mathrm{S}}\right)$ and $R\left(H_{\mathrm{S}}\right)$ is shown in equation (4) and equation (5), H_{RS} being a set of redundancy fault components in H_{S} and $F_{\mathrm{C_S}}$ being a set of real faults in H_{S}.

$$\boldsymbol{\eta}\left(H_{\mathrm{S}}\right) = \frac{|H_{\mathrm{S}}| - |H_{\mathrm{RS}}|}{|F_{\mathrm{C}}|} = \frac{|F_{\mathrm{C_S}}|}{|F_{\mathrm{C}}|} , \tag{4}$$

$$R\left(H_{\mathrm{S}}\right) = \frac{|H_{\mathrm{R_S}}|}{|H_{\mathrm{S}}|} = 1 - \frac{|F_{\mathrm{C_S}}|}{|H_{\mathrm{S}}|} . \tag{5}$$

The larger the number $\boldsymbol{\eta}\left(H_{\mathrm{S}}\right)$, the more important information that H_{S} represents the original fault collection H_{Max} will contain. Expression $\boldsymbol{\eta}\left(H_{\mathrm{S}}\right) = 1$ means that all important information of H_{Max} is retained in H_{S}. In contrast to $R\left(H_{\mathrm{Max}}\right)$, the smaller $R\left(H_{\mathrm{S}}\right)$ is, the less number of redundant faults, the better the performance of the filtering algorithm, and when $R\left(H_{\mathrm{S}}\right) = 0$, there is no redundant fault in H_{S}.

2.3. Leakage monitoring module based on fault information entropy difference monitoring algorithm

Because there may be multiple faults with maximum coverage, the possible fault set after filtering contains multiple possible failures. At this point, the use of coverage as a discriminating parameter for leakage monitoring is highly likely to lead to misjudgment of the situation. Therefore, it is necessary to introduce new reasonable parameters for leakage monitoring. This module defines the information entropy difference $\Delta H\left(f\right)$ of the fault f and uses it as the discriminating parameter. A leakage monitoring algorithm based on information entropy difference is proposed. The following theoretical analysis shows that the use of information entropy difference $\Delta H\left(f\right)$ as a discriminating parameter is reasonable.

When a symptom occurs in the network, it will provide a certain amount of information about the fault associated with it. Information Entropy is the average amount of information provided by a number of indications associated with a failure. The greater the amount of information is, indicating that the greater the uncertainty of the variable is the failure to determine the probability of failure that will be less likely.

In order to analyze the reasonability of fault information entropy difference as the parameter of fault judgment, the following two parameters are defined: ideal information entropy and actual information entropy.

The ideal information entropy of the fault f is the information entropy of the fault f in the ideal case, that is, all the symptoms associated with the fault f occur. The calculation of the ideal information entropy $H_1(f)$ is shown in the equatio.

$$H_1(f) = -\sum_{s_i \in S(f) \bigcap S_{\mathrm{O}}} p(f|s_i) \log p(f|s_i) \,. \tag{6}$$

Here, $S(f)$ is a corresponding set of signs to f, the network can all be shown outside the signs of the set S_O, that $p(f|s_i)$ can be obtained through the Bayes formula

$$p(f|s_i) = \frac{p(f)\, p(s_i|f)}{\sum_{f_j \in F} p(f_j)\, p(s_i|f_j)} \,. \tag{7}$$

The actual information entropy of the fault f is the information entropy gathered by f the symptom set S_{N} in the real situation. The calculation of the actual information entropy $H_2(f)$ is shown in the expression

$$H_2(f) - \sum_{s_I \in S(f) \bigcap S} p(f|s_i) \log p(f|s_i) \,. \tag{8}$$

The smaller the information entropy difference $\Delta H(f) = H_1(f) - H_2(f)$ of the fault f is, the closer the greater the probability of occurrence of the fault f to the ideal condition is, therefore, the information entropy difference $\Delta H(f)$ of the fault f can be used as a parameter in the leakage monitoring judgment.

2.4. Simulation and result analysis

In this paper, a fault prediction module and a fault screening module are added to the proposed algorithm. Among them, the failure prediction module quickly predicts the maximum possible fault set H_{Max}, and calculates the redundancy. Fault filtering module outputs possible filtered fault sets H_S and filters algorithm performance. When the scale factor $\mu = 0$, that H_{Max} is not actually screening for treatment; at that time $0 < \mu \leq 1$, H_{Max} was the corresponding screening to be filtered after the possible failure of the collection H_{S}. The algorithm is IWLD, and the comparison algorithm is MCA [10] and BSD [11].

In order to observe the performance of the algorithm in different random networks, 10 random networks are generated, and 50 valid single fault cases are generated in each network. The input of each case algorithm is S_i, the output is the fault hypothesis collection H, the detection rate $\mathrm{DR}(S_i)$, $\mathrm{DR}(S_i) = |H \bigcap F_{\mathrm{C}}| / |F_{\mathrm{C}}|$, and the false positive rate $\mathrm{FPR}(S_i)$, $\mathrm{FPR}(S_i) = |H - F_{\mathrm{C}}| / |H|$. In addition, in order to observe the performance of the filter module, this algorithm outputs the maximum possible fault set redundancy $R_{\mathrm{Si}}\left(H_{\mathrm{MaX}}\right)$, possible fault set redundancy $R_{\mathrm{Si}}\left(H_{\mathrm{S}}\right)$ and fault coverage $\eta_{\mathrm{Si}}\left(H_{\mathrm{S}}\right)$ under different scale factors.

Suppose the number of cases in each network is n, the output fault detection rate $\mathrm{DR}\left(\mathrm{Net}_i\right) = \sum_{j=1}^{n} DR\left(S_j\right)/n$, the error detection rate $\mathrm{FPR}\left(\mathrm{Net}_i\right) = \sum_{j=1}^{n} \mathrm{FPR}\left(S_j\right)/n$, the algorithm outputs the maximum possible fault set redundancy $R_{\mathrm{Net}_i}\left(H_{\mathrm{MaX}}\right) = \sum_{j=1}^{n} R_{S_j}\left(H_{\mathrm{MaX}}\right)/n$, the probability of failure aggregation under different scale factors $R_{\mathrm{Net}_i}\left(H_S\right) = \sum_{j=1}^{n} R_{S_j}\left(H_S\right)/n$, fault coverage $\boldsymbol{\eta}_{Net_i}\left(H_S\right) = \sum_{j=1}^{n} \boldsymbol{\eta}_{S_j}\left(H_{\mathrm{S}}\right)\Big/n$.

Suppose that the number of random networks is m, the final output is: fault detection rate $\mathrm{DR} = \sum_{i=1}^{m} DR\left(\mathrm{Net}_i\right)/m$, fault detection rate variance $\mathrm{VDR} = \sum_{i=1}^{m} \left\{\mathrm{DR}\left(\mathrm{Net}_i\right) - \mathrm{DR}\right\}^2 \Big/ m$, fault false detection rate $\mathrm{FPR} = \sum_{i=1}^{m} \mathrm{FPR}\left(\mathrm{Net}_i\right)/m$. In addition, the proposed algorithm can output the maximum possible fault set redundancy $\mathrm{R}\left(H_{\mathrm{MaX}}\right) = \sum_{i=1}^{m} R_{\mathrm{Net}_i}\left(H_{\mathrm{MaX}}\right)/m$, the possible failure set redundancy under different scale factors $\mathrm{R}\left(H_{\mathrm{MaX}}\right) = \sum_{i=1}^{m} R_{\mathrm{Net}_i}\left(H_{\mathrm{MaX}}\right)/m$ and fault coverage $\boldsymbol{\eta}\left(H_{\mathrm{S}}\right) = \sum_{i=1}^{m} \boldsymbol{\eta}_{\mathrm{Net}_i}\left(H_{\mathrm{S}}\right)/m$.

2.5. Simulation results and analysis

Figure 3 shows the filtered fault set redundancy degree under different scale factors. For $\mu = 0$, the value of $R\left(H_{\mathrm{MaX}}\right)$ is 83.34 % $\sim$ 94.67 %, the mean 90.84 %. For $\mu = 0.2$, the value of $R\left(H_{\mathrm{MaX}}\right)$ is 80.79 % $\sim$ 88.98 %, with an average of 86.70 %. For $\mu = 0.4$, the value of $R\left(H_{\mathrm{MaX}}\right)$ is between 70.10 % and 79.49 %, the mean value is 76.77 %. For $\mu = 0.6$, the value of $R\left(H_{\mathrm{MaX}}\right)$ is 55.01 % $\sim$ 67.53 %, the mean is 63.50 %. For $\mu = 0.8$, the value $R\left(H_{\mathrm{MaX}}\right)$ was between 42.22 % and 56.26 % with the mean value of 50.56 %. Finally, for $\mu = 1$, the value of $R\left(H_{\mathrm{MaX}}\right)$ is in the range 34.41 % $\sim$ 41.35 %, the mean is 37.97 %.

It can be seen that H_{MaX} has a high redundant fault, the corresponding signal contains more redundant information is with compressibility. In the screening algorithm, with the increase of the scale factor, the redundancy of the possible fault set is reduced.

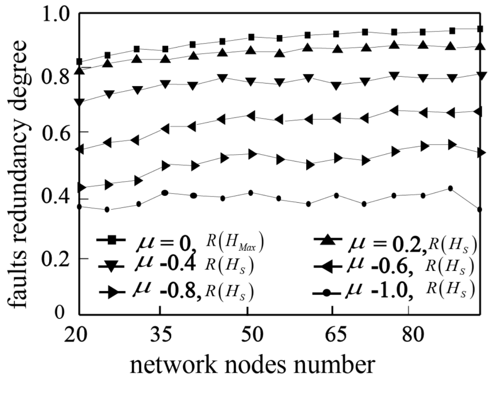

Fig. 3. Faults set redundancy degree after filtering at any scale factor

Table 1 shows the true fault coverage $\boldsymbol{\eta}\,(H_{\mathrm{S}})$ for the possible fault sets H_{S} after filtering. It can be seen from Table 1, in different network sizes, the value of $\boldsymbol{\eta}\,(H_{\mathrm{S}})$ is 1 and H_S can keep true failure. This is because the real fault always has the largest coverage, and at any scale factor, the real fault can always be preserved.

Table 1. H_S set true fault coverage ($\boldsymbol{\eta}\,(H_{\mathrm{S}})$

Network nodes	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
$\boldsymbol{\eta}\,(H_{\mathrm{S}})$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

3. Leakage monitoring performance analysis

Because when $\mu = 1.0$, H_S not only has the smallest redundancy, but also can keep the true breakdown, has the best screening performance. Therefore, the algorithm in this paper under $\mu = 1.0$, conduct fault screening and monitoring, and get monitor the results.

Figure 4 comparisons for the three kinds of algorithms fault detection rate. As shown in the figure, the proposed IWLD algorithm fault detection rate from 94.2 % $\sim$ 97.4 %, mean is 96.17 %. BSD algorithm fault detection rate from 91.6 % to 97.4 %, mean is 94.6 %. The failure detection rate of MCA algorithm is between 73.2 % and 89.8 %, mean value is 79.39 %. It can be seen that the IWLD algorithm has the highest fault detection rate, the BSD algorithm has a lower fault detection rate and the MCA algorithm has the lowest detection rate.

Figure 5 shows the comparison of the three algorithms 'fault detection rate variance in different networks, and shows the stability of the three algorithms' fault detection in different random networks. As shown in the figure, IWLD algorithm fault detection rate variance is between 0.00040 $\sim$ 0.00276, mean 0.00095. BSD algorithm fault detection rate variance is between 0.00038 $\sim$ 0.00264, mean 0.00138. MCA algorithm fault detection rate variance is between 0.00008 $\sim$ 0.00546, mean 0.00299. IWLD algorithm and BSD algorithm are almost equal to zero. Overall, IWLD algorithm is more stable than BSD algorithm, that is, the stability of fault detection rate is higher in different random networks. In contrast, MCA algorithm stability is relatively low.

Figure 6 presents the three kinds of algorithm error rate comparison. As shown in the figure, IWLD algorithm fault error detection rate of 2.6 % to 5.8 %, mean value 3.83 %. BSD algorithm error rate is between 16.16 % $\sim$ 23.25 %, mean value is 20.65 %. The error rate of MCA algorithm is between 10.2 % $\sim$ 27.2 %, and the mean is 20.61 %. IWLD algorithm has the lowest false alarm rate, BSD and MCA algorithms have higher false alarm rate. This is because IWLD algorithm has taken the initial screening of faults, to a large extent reduced the redundant fault on the impact of leakage monitoring.

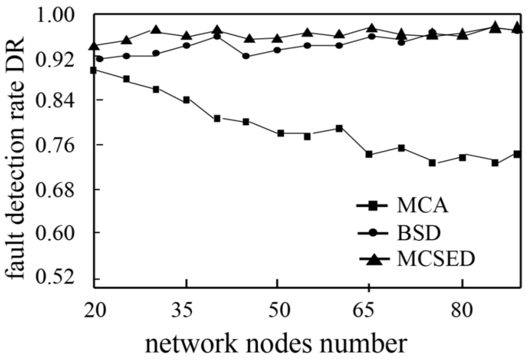

Fig. 4. Fault detection rate

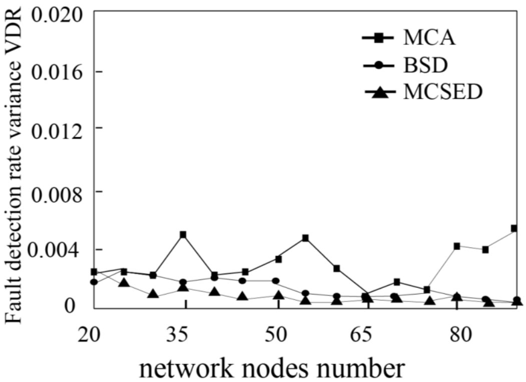

Fig. 5. Fault detection rate variance

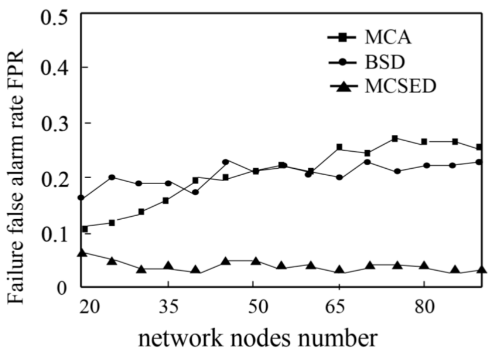

Fig. 6. Failure false alarm rate

4. Conclusions

In this paper, a water supply pipe network leakage monitoring algorithm based on information entropy difference is proposed. Firstly, the Kalman set of the algorithm is used in the prediction module to get the prediction result. Secondly, the parameter fault coverage is introduced into the filtering module. The fault set is transformed into the signal, and the redundancy of the signal is eliminated by the information entropy difference. Lower redundancy fault set is got; finally, the parameter information entropy difference is defined as the criterion of fault detection, and the root fault collection is monitored. The simulation results show that the proposed algorithm can stably show high fault detection rate in different random networks and greatly reduce the false detection rate. In order to obtain more accurate monitoring performance, it is necessary to find a better combination of parameters in the monitoring algorithm of information entropy, which will be a further step of future research work.

References

[1] J. Tapolcai, P. Ho, L./,Ronyai, P./,Babarczi, B./,Wu: *Failure localization for shared risk link groups in all-optical mesh networks using monitoring trails.* Journal of Lightwave Technology *29* (2011), No. 10, 1597–1606.

[2] Y. Xiong, Z. Y. Xiong, D. P. Wu, Y. Li: *Multi-fault aware parallel localization protocol for backbone network with many constraints.* Photonic Network Communications *24* (2012), No. 3, 210–218.

[3] R. Y. Wang, J. F. Chang, K. P. Long, X. L. Yang, W. L. Zhu: *A fault detection and location mechanism for optical burst switching network.* Journal of Optoelectronics Laser *17* (2006), No. 12, 1477–1481.

[4] R. Y. Wang, J. F. Chang, K. P. Long, H. Chang, P. Z. Zhang: *Performance and fault monitoring scheme of data channel for optical burst switching networks.* Optical Communication Technology *30* (2006), No. 08, 8–11.

[5] A. Haddad, E. E. Doumith, M. Gagnaire: *A fast and accurate meta-heuristic for failure localization based on the monitoring trail concept.* Telecommunication Systems *52* (2013), No. 2, 813–824.

[6] E. Hosszu, J. Tapolcai, L. Ronyai, P. Soproni, P. Babarczi, P. H. Ho: *Fast failure localization in all-optical networks with length-constrained monitoring trails.* International Congress on Ultra Modern Telecommunications and Control Systems, 3–5 Oct. 2012, St. Petersburg, Russia, IEEE Conference Publications (2005) 677–683.

[7] J. Tapolcai, L. Ronyai, P. Ho: *Link fault localization using Bi-directional M-trails in All-optical mesh networks.* IEEE Transactions on Communications *61* (2013), No. 1, 291–300.

[8] R. Barco, P. Lázaro, V. Wille, L. Diez, S. Patel: *Knowledge acquisition for diagnosis model in wireless networks.* Expert Systems with Applications, Part: 1 *36* (2009), No. 3, 4745–4752.

[9] M. Steinder, A. S. Sethi: *Probabilistic fault diagnosis in communication systems through incremental hypothesis updating.* Computer Networks *45* (2004), No. 4, 537 to 562.

[10] X. H. Huang, S. H. Zou, L. W. Chu: *Internet Services Fault Management:Layering Model and Algorithm.* Journal of Software *18*, (2007), No. 10, 2584–2594.

[11] C. Zhang, J. X. Liao, X. M. Zhu: *Heuristic fault localization algorithm based on Bayesian suspected degree.* Journal of Software *21* (2010), No. 10, 2610–2621.

Received July 12, 2017