Research of Algorithms for Processing Information in Wireless Networks and Filling the Missing Data

I.Ya. Lvovich[1], A.P. Preobrazhenskiy[1], O.N. Choporov[2]

[1]Voronezh Institute of High Technologies, Lenina str. 73a, Voronezh, Russia, 394043
[2]Voronezh State Technical University, Moscow dist. 14, Voronezh, Russia, 394026

Abstract. In this paper, we consider the problem of transmitting information wirelessly indoors. In this case, data loss due to re-reflections of radio waves or interference is possible. Algorithmization of procedures for eliminating false data is carried out. The results of the study of the influence of the interference environment on the communication range and data transfer time on a Wi-Fi network are presented. Between workstations of the studied area, video, image and text document files are transferred. Evaluation of the effectiveness of the methods of filling in the missing data is given.

1. Introduction

Wi-Fi standard wireless networks are widely used for organizing small-sized local networks in apartments, offices and public areas of access (Hot-Spot), etc. But the optimal design of local service areas of such networks always encounters certain difficulties [1]. They are associated with the need to take into account the peculiarities of the propagation of radio frequency signals of the Wi-Fi range in rooms of complex configuration, which have numerous reflective elements and structures under the influence of a wide variety of interference from numerous electronic devices [2]. Therefore, the real dimensions of the formed service areas are much smaller than expected [3].

2. Formulation of the problem

For the study, the following technical devices were used:

- -Wireless router ZYXEL NBG6604;
- -Two repeaters of the wireless signal TP - LINK TL - WA855RE;
- -Network adapter Wi - Fi TP - LINK TL - WN725N USB 2.0;
- -Smartphone;
- -Laptop.

The ZYXEL NBG6604 access point (access point) is a stand-alone module with a built-in microcomputer and a transmitter / receiver unit.

Through it, radio-frequency channel interaction and information exchange between wireless adapters is carried out, as well as communication with a wired network segment.

Measurements were taken on the premises of the office building.

3. Algorithmization of procedures for eliminating false data

The main heuristic rule to the exclusion of unreliable data is a selection of information from the observations the most probable, i.e. the most typical of this situation, a set of data. All background information can be represented as a set of objects [4]

$$G_{èñõ} = \bigcup_{n=1}^{N_{iskh}} g_n . \tag{1}$$

where N_{iskh} - the volume of the original sample. Each observation is characterized by a set of indicators:

$$\forall n:\ g_n \rightarrow P_n = \left\{P_n^1, P_n^2, ..., P_n^i, ..., P_n^{I_{èñõ}}\right\} \tag{2}$$

where $i = \overline{1, I_{iskh}}$ - indicator index, $n = \overline{1, N_{iskh}}$ - serial number of observations. In the first stage exclusion of false data for each parameter P_i ($i = \overline{1, I_{iskh}}$) Set the upper and lower margins and P_{imin}, P_{imax}, ($i = \overline{1, I_{iskh}}$). The output of which is only possible due to measurement errors or recording indicator. Processing then screening observation that may not be accurate because of output values of a particular parameter (or group of parameters) within range. The result is a set of

$$G_{gr} = \bigcup_{n=1}^{N_{gr}} g_n . \tag{3}$$

which includes only the measurement satisfying the following condition:

$$\forall n, \forall i:\ P_{\min}^i \le P_n^i \le P_{\max}^i . \tag{4}$$

Solution of the second problem, the main stage of exclusion is invalid data in the selection information from the original set of observations with reliability estimate w_n ($n = \overline{1, N_{gr}}$) Above a significant value w_0. Information containing fixed semantic (linguistic) of observations should be based on expert judgment converted into numerical [5].

For information having numerical values of information, the solution of the problem is the elimination of false data mapping a plurality of information observations (2) into a plurality of initial estimates of the reliability of observations

$$W = \bigcup_{n=1}^{N_{gr}} w_n , \tag{5}$$

and forming a plurality of $G \subseteq G_{gr}$ selected observations $G_{\hat{o}} = \bigcup_{f=1}^{N_{\hat{o}}} g_f$ according to the rule

$$\forall n \left(n = \overline{1, N_{ãð}}\right):\ g_n \rightarrow \begin{cases} g_f \mid & w_n \ge w_0 \\ 0 \mid & w_n < w_0 \end{cases} \tag{6}$$

where 0 - a null set.

As a rule, the various indicators are measured in different units. For efficient operation of the algorithm exclude inaccurate data necessary to make the normalization of all parameters. Normalization is a uniform transition to some description for all characteristics, to introduce new units admitting a formal matching objects. one of the following methods normalization (change from baseline values to normalized x z) may be used [6]:

$$z^1 = (x - \bar{x}) / \sigma, \tag{7}$$

$$z^2 = x / \bar{x}, \tag{8}$$

$$z^3 = x / x', \tag{9}$$

$$z^4 = x / x_{\max}, \tag{10}$$

$$z^5 = (x - \bar{x}) / (x_{\max} - x_{\min}), \tag{11}$$

where $\bar{x}, \sigma$ – respectively the mean and standard deviation of x; x'- a reference (normative) value x; $x_{\max}$, $x_{\min}$ - the largest and smallest values of x.

Normalization z^1 and z^2 are defined for scales and intervals of relationships, and the rest - only for a ratio scale. The most convenient normalization with respect to the allowable range changes of parameters values. For the problem (3) is determined using the following formula [7]:

$$\forall n, \forall i : P\_norm_n^i = \frac{P_n^i - P_{\min}^i}{P_{\max}^i - P_{\min}^i} \quad (12)$$

Determining the degree of reliability of the information observation to solve the problem (5) - (6) is based on the concept of typicality, i.e. w_n the accuracy of the information of observation shall be the higher than is typical for a given situation (for the entire series of observations). Since details of the observations $\{P_n^i\}$ represented by numerical values legitimate geometric approach allowing observation information considered as a "constellation" in the *i*-dimensional hyperspace characteristics. The method and the adequacy of the solutions depend on the additional a priori information on the degree of "contamination" of the original sample G_{gr} observations.

If a priori known that the sample G_{gr} "clogged" is small, legitimately assume that g_n observations are grouped in some way symmetrical with respect to an imaginary center of gravity and are more likely to the most reliable observations are located at the shortest distance from a hypothetical generalized observation g_0 with the numerical data set $P_0 = \{ P_0^1, P_0^2, ..., P_0^i, ..., P_0^{I_{gr}} \}$ where

$$\forall i : P_0^i = \sum_{n=1}^{N_{gr}} P\_norm_n^i / N . \quad (13)$$

The solution is the distance vector calculation values $S = \{S_1, S_2, ..., S_n, ..., S_{N_{гр}}\}$ by observing $g_n \in G_{gr}$ to a generalized observation g_0 using varying adequate in relation to the metric data, such as the Euclidean:

$$\forall n : S_n = \left[\sum_{\forall i} (P\_norm_n^i - P\_norm_0^i)^2 \right]^{1/2} \quad (14)$$

The degree of reliability of observations g_n

$$wn = S_{min} / S_n, \quad (15)$$

where $S_{min} = \min_{\forall n} S_n$

If the sample G_{gr} "clogged" more significantly, for example half, the more correct to assume asymmetric distribution of the facts and then the notion of a generalized surveillance g_0 can not adequately submit a sample G_{gr}. In this case include a procedure for determining the degree of reliability of observations: using varying adequate in relation to the metric data, such as Euclidean vector values computed total distances $S = \{ S_1, S_2, ..., S_n, ..., S_{N_{gr}} \}$ Information from each observation to the other

$$s_n = \left[\sum_{m=1}^{N_{gr}} \sum_{i=1}^{I_{iskh}} \left(P_n^i - P_m^i \right)^2 \right]^{1/2}, \quad n = \overline{1, N_{gr}} \quad (15)$$

and similarly to (14) is determined by observing the degree of confidence.

If the sample G_{gr} strongly "blocked", but there is a significant probability that a group of reliable information sufficient observations expressed in the hypothesis of compactness sense in relation to other possible groupings, the legitimate approach, based on the classification (cluster) analysis. The following procedure for determining the degree of reliability of the information. If the sample volume is G_{gr} then organized $M = N_{gr} - 2$ iterative cycle index $C = \overline{2, (N_{gr} - 1)}$.

To study the characteristics of high-speed network equipment Wi-Fi using the known client - server Iperf program, which provides for the generation of TCP-traffic to test bandwidth under conditions of structurally similar interference. I_{perf} program has the ability to work as a client request and server scanning mode. One system acts as I_{perf} server on a port, and the other - as a client system on which the client is running Iperf. Both systems use the same executable file, providing the opportunity to select one of the roles - the client or server. After you enter the basic data for testing, run a quick check of the quality of the connection between the two systems [8].

For the experiment within the office premises were organized two wireless service areas based on the ZYXEL NBG6604 access points and repeaters TP-LINK TL-WA855RE wireless signal. For each service area are connected by two workstations via network adapters Wi - Fi TP-LINK TL-WN725N USB 2.0. Between two wireless workstations within each zone, runs a client-server Iperf program with which the fixed data transfer rate between the client and server at various combinations of communication channels, which employed the access point [9].

To create an interference situation in the service area of TP-LINK TL-WA855RE Network A with a wireless workstation to another through a shared folder run the video.

The investigated area - ZYXEL NBG6604 Network W. In this service area at different distances from the access point (ZYXEL NBG6604) portable computer (Wireless Station D) receives fixed-size files from the wireless workstation Wireless Station S. Secured time transmission of files and levels signal Wi -Fi towards and away from the laptop computer of the access point.

4. Results

Between workstations investigated zone transfers video files, image and text document volume of 15 MB. The levels of signal Wi - Fi at the points of measurement are recorded by a special program on a laptop computer. Measurements were taken at 6 and 6 and the channel 1 and 6 channels.

Table 1. Results of the study on the influence of the interference environment, communication range and data transfer time in the Wi-Fi network (channel 6 and 6).

measurement parameter	The distance from the transmitter to the receiver, meters					
	one	four	8	12	16	twenty	24

Table 2. Results of the study of influence of noise conditions on the operating distance and time data in the Wi-Fi network.

Level signal dBm	-41	-55	-61	-67	-72	-78	-81
time video transmission, with	65	79	115	128	152	no transmission	no transmission
image transfer time, with	62	81	117	135	156	no transmission	no transmission
the transmission time of a text file with	68	79	121	152	163	no transmission	no transmission

The received data are entered in an Excel table. To evaluate the effectiveness of the modified filling a comparative analysis of methods for filling gaps: maximum likelihood and approximation «Deductor 2.0» statistical package and the modified method «ZET». The results are shown in Table 3. Modified method «ZET» has the minimum error fill gaps, which justifies its use in pre-processing of information.

5. Conclusion

1. For a comprehensive statistical data processing and modeling in wireless systems there is a need in the transformation information comprising fixed meaning (language) to the numerical values mean. This transformation serves to implement an algorithm based on expert using linguistic variables.

2. The accuracy of the model any complex wireless systems is largely determined by the quality of the source data, so that the necessary pre-processing of statistical information, namely the exclusion of unreliable data, providing the possibility of selection of reliable data and exclude outlying observations; filling the gaps in the data matrix, as most models provide high accuracy in a completely filled matrix.

3. Improve the efficiency of the method used to fill gaps can be by using regression equations of higher order data recording on homogeneous groups of unreliable data exclusion method and sequential method is applied two to four times and using the results of the previous pass as inputs. Extend the range of application of the method on the data from an unknown distribution function can be replaced with correlation coefficient estimate formation matrix an amount tips inverse distance.

Table 3 Evaluating the effectiveness of filling the missing data methods.

real meaning	Reconstructed			Mistake		
	The extrapolation of the second order	maximum likelihood		The extrapolation of the second order	maximum likelihood	«ZET» modified method
0.846	0,401	0.214	0.409	-0.445	-0.632	-0.437
0,450	-0.020	0.357	0,421	-0.470	-0.093	-0.029
0,333	0,666	0.357	0.335	0,333	0,024	0,002
0.765	0.997	0,643	0.391	0.232	-0.122	-0.374
0	0.427	0.357	0.337	0.427	0.357	0.337
0.559	0.490	0,500	0,508	-0.069	-0.059	-0.051
0.427	0.138	0.357	0.343	-0.289	-0.070	-0.084
0.514	0.557	0,500	0.494	0,043	-0.014	-0.020
0,266	0.426	0.426	0,380	0,160	0,160	0,114
The total absolute error				2,468	1,521	1,448
Mean absolute error				0,274	0.170	0.161

6. References

[1] Lutakamale, A.S. Wildfire Monitoring and Detection System Using Wireless Sensor Network: A Case Study of Tanzania / A.S. Lutakamale, S. Kaijage // Wireless Sensor Network. – 2017. – Vol. 9. – P. 274-289.

[2] Talluri, S. The relationship between operating efficiency and service quality: are they compatible? / S. Talluri, M.K. Kim, T. Schoenherr // Intern J Prod Res. – 2013. – Vol. 51. – P. 2548-2567.

[3] Minerva, R. Towards a definition of the Internet of Things (IoT) / R. Minerva, A. Biru, D. Rotondi // IEEE Internet Initiative – Torino, Italy, 2015.

[4] Odu, G.O. Review of Multi-criteria Optimization Methods / G.O. Odu, O.E. Charles-Owaba // Theory and Applications. – 2013. – Vol. 3. – P. 01-14.

[5] Shah, A. Parallel predictive entropy search for batch global optimization of expensive objective functions / A. Shah, Z. Ghahramani // Advances in Neural Information Processing Systems, 2015. – P. 3330-3338.

[6] Neittaanmki, P. Mathematical Modeling and Optimization of Complex Structures; Series: Computational Methods in Applied Sciences / P. Neittaanmki, S. Repin, T. Tuovinen // Springer International Publishing AG, Switzerland, 2016.

[7] Rios, L.M. Derivative-free optimization: a review of algorithms and comparison of software implementations / L.M. Rios, N.V. Sahinidis // Journal of Global Optimization. – 2013. – Vol. – 54. – P. 1247-1293.

[8] Morais, H. Optimal Scheduling of a Renewable Micro-Grid in an Isolated Load Area Using Mixed-Integer Linear Programming / H. Morais, P. Kádár, P. Faria, Z.A. Vale, H.M. Khodr // Editorial System for Renewable Energy Magazine. – 2010. – Vol. 35(1). – P. 151-156.

[9] Orlova, D.E. Stability of solutions in ensuring the functioning of organizational and technical systems // Modeling, optimization and information technologies. – 2018. – Vol. 6(1). – P. 325-336.