# Data Fusion Algorithm in Intelligent Distributed Sensor Networks

A. Sachenko, V. Kochan, V. Turchenko, V. Koval

Institute of Computer Information Technologies
Ternopil Academy of National Economy
3 Peremoga Square, 282004 Ternopil UKRAINE
Phone: +380 (352) 33-0830 Fax: +380 (352) 33-0024
E-Mail: sachenko@cit.tane.ternopil.ua

***Abstract***

*The offered data fusion algorithm is based on the averaging information received from reasonable sensors, whose readings always contain errors. The decision can be applied for receiving the most correct result about the object condition that is determined by set of various physical measures (coordinate, height, speed, temperature, etc.) in a space, defensive industry and measuring engineering.*

## 1. Introduction

At the modern stage of engineering development, the surrounded world can be presented as hierarchical totality of automated systems, which elements are linked in networks. The latter contains, as a rule, heterogeneous sensors, which are distributed by means of various criteria: logically, spatially, geographically [1]. Therefore rather important and actual problem of sensor's processing appears in space, defensive industry, electric power industry, measuring engineering, medical area, robotics, chemical industry and in other spheres. Intelligent distributed sensor systems and networks are widely used at the present moment [2, 3]. In most cases the role of intelligent systems functions can be reduced to the fact that they are capable of doing sensor perception of environment and possessing sufficient knowledge (intelligence) to give an adequate reaction to the researched situation or environment. That is why the task of creation data fusion algorithms for reasonable sensors (RS) in intelligent distributed sensor networks is very actual.

As the existing sensor fusion algorithms [4, 5, 6] have some negative sides, the comparative estimation of these algorithms is conducted below. Also the algorithm of mean square weight factors (MSWF) is offered.

## 2. Sensor fusion algorithms

Let us consider a network of reasonable sensors. Reasonable sensors are sensors, which fulfill some primary processing of an input signal, and the result of processing is represented in digital form. The simple distributed network (Fig. 1) contains set of RS, which are linked among them by the full graph scheme.

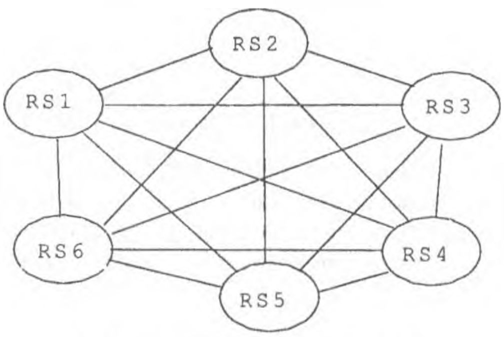


**Fig.1. Distributed network of RS**

With the reference to radar engineering [1], receiving and processing of information from RS is a process of getting the possible data about object (coordinate, height, speed, course corner, location time etc.), which is in a zone of RS visibility. As RS areas action are crossed the information about one object can arrive from several RS. The data about object that received from RS should be imposed in ideal case. Nevertheless, the coincidences are not observed in practice because of systematic and casual errors. By virtue of these reasons there are difficulties in data fusion. When sensor data are fused the accuracy and reliability of the received result is decisive criteria of data fusion algorithms functioning.

D. Dolev has presented Byzantine agreement algorithm for the solving of the Byzantine generals problem, which was researched by L. Lamport and his colleagues [7, 5, 8]. The main idea of the Byzantine generals problem solving is the following: if there are $N$ of independent RS, then is supposed such number $t$ of faulty elements (sensors) at which the inequality should be carried out

$$N \geq 3 * t + 1 . \qquad (1)$$

That is each RS should be linked with not less than $2 * t + 1$ other RS.

The main principle of data fusion algorithm offered by D. Dolev is firstly checking of Byzantine agreement condition (1). The highest ($x_{max}$) and lowest ($x_{min}$) RS values are rejected, then calculated an average value of remained elements (reference value of element)

$$Z_j = \frac{\sum_{i=1}^{N} x_{ij} - (x_{\max j} + x_{\min j})}{N - 2} ,$$

where $Z_j$ - the reference value of element, $j$ - number of RS that sends the data, $i$ - number of RS that receives the data, $N$ – RS numbers, $x_{ij}$ – value which receives $i$-RS from $j$-RS.

The resulting value is

$$Z_{rez} = \frac{\sum_{j=1}^{N-1} Z_j}{N} .$$

As it is visible from Fig. 2 the algorithm has essential lack. Initial signal (resulting value) value is less correct owing to rejection of the highest and lowest element, than average value of sample (on Fig.2 the average value of sample comes nearer to true value). Therefore such algorithm is correct only in that case, when the value of the highest and (or) lowest element is by rough errors, which deviate the average value from real signal.

In sensor fusion algorithm by S.Mahaney and F.Schneider [4, 6], as against from the previous algorithm, the concepts of accuracy and precision of RS are introduced. If to designate an error through $U_e(t)$ and true signal through $U_t(t)$ then RS readings can be in boundaries

$$U_e(t) - U_t(t) \leq U_t(t) \leq U_e(t) + U_t(t) .$$

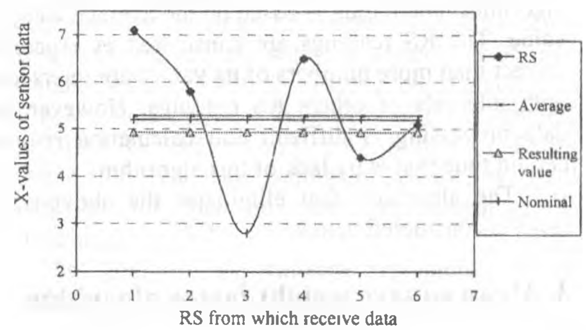


**Fig. 2. Functioning of D.Dolev algorithm on example of 6 RS**

Therefore, it is possible to set the readings of each RS by an interval of allowable values. S. Mahaney's and F. Schneider's algorithm uses groups of the allowable data. The value is admitted if it in common boundaries for all RS area (this area is formed on the basis of the maximal value among the lowest boundaries of RS accuracy and minimal value among the highest boundaries). If the interval of the RS readings is in common boundaries for all RS of values area, it is considered as admitted, differently its readings are rejected (Fig. 3). It is obvious, that any value that is not admitted can not be correct. The algorithm result is average arithmetic value of the middle of allowable value intervals of the RS readings. This algorithm as against previous will not carry out elimination in the RS readings at the large rejections from average value.

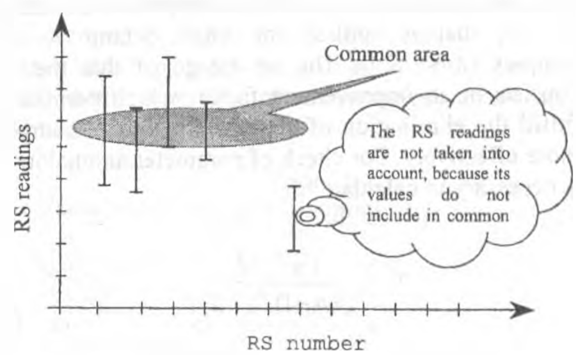


**Fig. 3. Common area of sensor readings accuracy**

In the Brooks-Iyengar hybrid algorithm [4], there are groups with $N$ real values and accuracy is determined by distance from unique correct value. As all RS have limited accuracy then its values consist of the highest and lowest boundaries. This algorithm uses the intervals of allowable values of RS readings and takes into account Byzantine agreement problem. The algorithm functioning is based on the average weighed value. The RS readings are considered as especially correct than more numbers of its values are intersected with intervals of others RS readings. However, the data processing is difficult and calculation requires certain time that is by lack of this algorithm.

The algorithm that eliminates the above-stated lacks is considered below.

## 3. Mean square weight factor algorithm

### 3.1. Background

The offered algorithm of mean square weight factor is based on assumption that all sensor's values of RS output signals in intelligent distributed sensor network represent by data sample. Conditions described in Byzantine generals problem by D.Dolev [5] is checked firstly. After this the preliminary data processing is executed that applying of statistical methods will be more correct.

The preliminary processing basically consists of elimination of rough errors. It is possible to explain essence of such errors (abnormal or strongly detailed values) by example [9]. Let us to allow that 10% of measurement results represented by abnormal values, differ from average more than *3l* (*l* – line segment on axis *ÎÕ*). If the rest of readings are placed within the *l* limits then these 10% redouble this estimation at least.

Among set of methods of elimination rougherrors [9], it is necessary to note method of the maximal relative deviation with improvement factor [9, 10] that is applied for small behind volume samples (n <= 25). The advantage of this method consists in an improvement factor, which permits to fulfill the elimination of abnormal values of samples more effectively. For check of parameter anomalous it is necessary to calculate

$$\tau'_i = \frac{|x_{ij} - \bar{x}_i|}{\sqrt{(n-1)/n} * S_i} \qquad (2)$$

where $x_{ij}$ - value receiving by $i$-RS from $j$-RS that is checked on anomalous, $x_i$ - average value $i$-sample, $n$- number of sample elements, $\tau'$ - quantile of statistics distribution.

$$S_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2}$$

- mean square deviation of the unbiased estimation variance.

Calculated according to (2) values $\tau'_i$ is compared to tabulared value $\tau_{1-p}$ [9] that is calculated with confidence probability $q=1-p$. If the calculated value $\tau'_i \leq \tau_{1-p}$ then it is possible to approve that value not abnormal with probability $P$. In other case value reject from sample and check anomalous of value carry out repeatedly on rest elements. For large on the volume samples n>25 it is expedient to use the tables of Student distribution [9, 10].

As it is visible from Fig. 4 the abnormal error value significantly hold away the average sample value from true value therefore roughvalue needs to rejected from sample. It is necessary to note that each result of separate measurement is equal to the sum of true value and casual error. Therefore law of measurement result distribution will coincide with the law of casual error distribution.

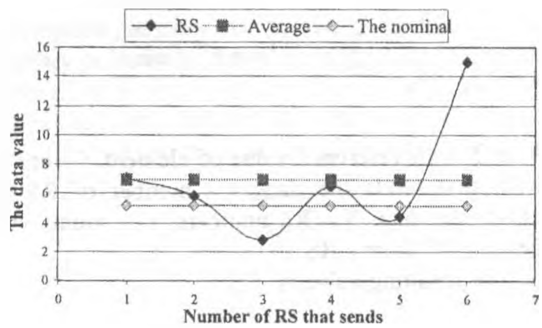


**Fig. 4. The schedule of rough errors influence on the result**

The accuracy of measurements is estimated by average square law deviation from average arithmetic sample. After elimination of rougherrors the average square deviation of measurement result from average [9] is

$$\sigma = \sqrt{\frac{\sum_{j=1}^{n}(x_{ij} - \overline{x_i})^2}{n-1}} \qquad (3)$$

From expression (3) it is possible to find estimation of average quadratic error of measurement result [11].

$$\bar{\sigma}_i = \frac{\sigma_i}{\sqrt{n}} \tag{4}$$

Having average square estimation (4), it is possible to note that the true value is in limits [11, 12]

$$\bar{x}_i - \bar{\sigma}_i \le value \le \bar{x}_i + \bar{\sigma}_i .$$

It is obvious that the greatest value of average square law deviation, then measuring device bring large error into measurement result. Therefore, as against the previous methods it is necessary to use not arithmetic mean, but average weighed value (taking into account factors of each value importance) for averaging of sensors data

$$\bar{X}_i = \frac{\sum_{j=1}^{n} k_{ij} x_{ij}}{\sum_{j=1}^{n} k_{ij}}$$

where $k_{ij}$ - factor of importance of $j$ – parameter that is accepted by $i$-RS. As last uses back proportional to square of measurement error

$$k_{ij} = \frac{1}{\left(x_{ij} - \bar{x}_i\right)^2}$$

When all average weighed values $^{i}$-RS are found one uniform resulting value is calculated as arithmetic mean of weighed average

$$\bar{X}_{rez} = \frac{\sum_{i=1}^{n-t} \bar{X}_i}{n-t} \tag{5}$$

where $n$ – number of RS, $t$ - quantity of fault elements.

Thus, from averaging result founded average estimation has smaller casual error than separate values (by that it is founded).

### 3.2. Experimental results

For estimation of data fusion algorithm quality we shall fulfill experiments with the help of imitating model. In experiments is used 6 RS which receive data about measurement object. Thus there is one fault RS, which sends to another RS various readings about measurement signal and with the greatest error. The true signal value is 5.230. For estimation of algorithms quality we shall consider their work at measurements errors: first RS ±2.5, second ±0.41, third ±1.7, fourth ±2.9 and fifth ±1.8. The readings of sixth RS we shall consider as fault, therefore measurement error of this RS true signal is ±5.4.

The goal of experimental researches is to consider functioning of data fusion algorithms with use of measurement errors distributed on the various distributivelaws. Thus for elimination of the received results chance and partial cases it is necessary to fulfill not less than 10 experiments of imitating model.

Let us consider algorithms functioning at the equal distributive law of measurement errors. The table with true signal values (Table 1.1), table with deviation from true signal values (Table 1.2) and the table with relative deviations of all algorithm values from MSWF algorithm (Table 1.2) are showed below. From these tables it is necessary to note that the MSWF algorithm has shown the best results (greatest quantity of the smallest deviations from true signal; there is no one worse result; best results of relative deviations) in comparison with other algorithms (Fig. 5). Concerning cases, when the results of MSWF algorithm worse from the results of other algorithms, it is necessary to node that their size is insignificant in comparison with the best results. The quantity of the worse results is insignificant in comparison with the best results also (one result from ten for algorithm "D.Dolev" and three results from ten for "R.Brooks and S.Iyengar" and "S.Mahaney" algorithm).

The results of functioning of all algorithms at the normal distributive law of measurement errors are presented in the Tables 2.1 - 2.3.

**Table 1.1.**
**The calculated values of true signal**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| True signal | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 |
| Dolev | 5,1592 | 5,1485 | 5,3249 | 5,5192 | 5,6599 | 5,3535 | 4,9532 | 5,9629 | 5,5192 | 5,1485 |
| Mahaney | 5,1110 | 5,1736 | 5,4982 | 5,5395 | 5,7466 | 5,2241 | 5,0750 | 5,7461 | 5,5396 | 5,1736 |
| Brooks | 5,3309 | 5,0693 | 5,3748 | 5,1509 | 5,4673 | 5,0979 | 5,4787 | 5,3187 | 5,1509 | 5,0693 |
| MSWF | 5,1704 | 5,1855 | 5,3918 | 5,2714 | 5,5225 | 5,2619 | 5,0690 | 5,8880 | 5,2714 | 5,1855 |

**Table 1.2.**
**Deviation from true signal values**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Dolev | 0,0708 | 0,0815 | 0,0950 | 0,2892 | 0,4299 | 0,1235 | 0,2768 | 0,7329 | 0,2892 | 0,0815 |
| Mahaney | 0,1190 | 0,0564 | 0,2682 | 0,3095 | 0,5166 | 0.0059 | 0,1550 | 0,5161 | 0,3096 | 0,0564 |
| Brooks | 0,1010 | 0,1607 | 0,1448 | 0,0791 | 0,2373 | 0,1321 | 0,2487 | 0,0887 | 0,0791 | 0,1607 |
| MSWF | 0,0596 | 0,0445 | 0,1618 | 0,0414 | 0,2925 | 0,0319 | 0,1610 | 0,6580 | 0,0414 | 0,0445 |

**Table 1.3.**
**Relative deviations of all algorithms values from MSWF algorithm**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Dolev | 1,1879 | 1,8315 | 0,5869 | 6,9855 | 1,4697 | 3,8715 | 1,7193 | 1,1138 | 6,9855 | 1,8315 |
| Mahaney | 1,9966 | 1,2674 | 1,6576 | 7,4758 | 1,7662 | 0,1850 | 0,9627 | 0,7843 | 7,4783 | 1,2674 |
| Brooks | 1,6941 | 3,6112 | 0,8949 | 1,9106 | 0,8113 | 4,1411 | 1,5447 | 0,1348 | 1,9106 | 3,6112 |

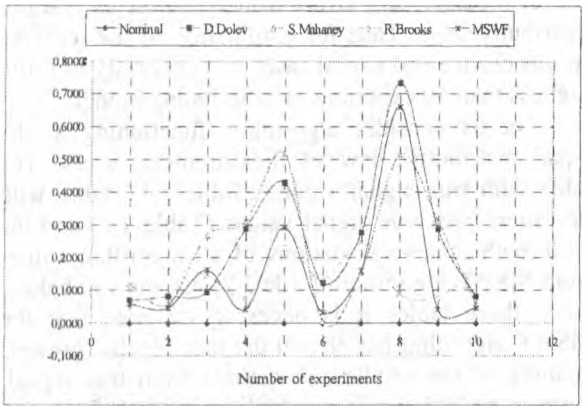


Fig 5 Diagram of deviation of true signal from its real value (equal distribution law)

The analysis of these tables has shown that the MSWF algorithm is more effective. The number of the best results (minimal deviations from true signal) is greatest (Fig. 6). Also the MSWF algorithm is better taking into account the worst cases, when the deviations of the resulting readings from true signal is greatest (numbers of the worse results at MSWF algorithm is equal 2 and in "D.Dolev" algorithm is 4).

The experiments were fulfilled with use of the exponential and Simpson distributive laws of measurement errors. At using of the exponential distributive law the most stable value receives "R.Brooks and S.Iyengar" algorithm. Its results have little change in comparison with true signal, though it shows the worst results on occasion. At using of the Simpson distributive law the results of MSWF algorithm are better (greatest quantity of the least deviations from true signal and rather not plenty of the greatest deviations). If consider relative deviations then the results of MSWF algorithm in 4.5 times is best in comparison with other algorithms.

The experimental results of data fusion algorithms functioning have confirmed, that at using of equal, normal, exponential and Simpson distributive laws the most preferable is MSWF algorithm. The results of the given algorithm are best at equal, normal and Simpson distributive laws. At exponentially distributed measurement errors it has worse results where high stability has shown "R.Brooks and S.Iyengar" algorithm. Therefore, the developed MSWF algorithm more effective calculates the correct value at the existence of the fault data. Using of this algorithm can improve accuracy and precision in many distributed applications. The MSWF algorithm will be used in intelligent distributed sensor network [3] where sensor errors are mainly distributed on normal distributive law.

**Table 2.1.**
**The calculated values of true signal**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Real signal | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 | 5,2300 |
| Dolev | 5,2083 | 5,2547 | 5,3558 | 5,7200 | 5,3289 | 4,9702 | 5,2167 | 5,3159 | 5,3327 | 4,9898 | 4,9368 | 5,2815 | 5,2696 | 5,4345 |
| Mahaney | 5,2419 | 5,1941 | 5,2930 | 5,6596 | 5,2819 | 5,0843 | 5,2923 | 5,4470 | 5,2692 | 4,8184 | 4,9481 | 5,3165 | 5,2670 | 5,3986 |
| Brooks | 5,0828 | 5,1229 | 5,1841 | 5,5279 | 4,9373 | 5,2438 | 5,1169 | 5,1225 | 5,3586 | 4,9534 | 5,0784 | 5,3080 | 5,0144 | 5,4218 |
| MSWF | 5,2497 | 5,1400 | 5,2185 | 5,6824 | 5,2516 | 5,0761 | 5,1559 | 5,2672 | 5,2391 | 4,9727 | 5,0632 | 5,3796 | 5,2763 | 5,5818 |
| | | | | | | | | | | | | | | |

Table 2.2.
**Deviation from true signal values**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dolev | 0,0216 | 0,0247 | 0,1258 | 0,4900 | 0,0989 | 0,2597 | 0,0132 | 0,0859 | 0,1027 | 0,2401 | 0,293194 | 0,051528 | 0,039608 | 0,204585 |
| Mahaney | 0,0119 | 0,0358 | 0,0630 | 0,4296 | 0,0519 | 0,1456 | 0,0623 | 0,2170 | 0,0392 | 0,4115 | 0,281898 | 0,086592 | 0,037052 | 0,168679 |
| Brooks | 0,1471 | 0,1070 | 0,0458 | 0,2979 | 0,2926 | 0,0138 | 0,1130 | 0,1074 | 0,1286 | 0,2765 | 0,15153 | 0,07801 | 0,215508 | 0.191833 |
| MSWF | 0,0197 | 0,0899 | 0,0114 | 0,4524 | 0,0216 | 0,1538 | 0,0740 | 0,0372 | 0,0091 | 0,2572 | 0,166736 | 0,149691 | 0,046336 | 0,351878 |

Table 2.3.
**Relative deviations of all algorithms values from MSWF algorithm**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dolev | 1,0979 | 0,2750 | 11,035 | 1,0831 | 4,5610 | 1,6888 | 0,1786 | 2,3049 | 11,190 | 0,9334 | 1,758438 | 0,34423 | 0,854811 | 0,581408 |
| Mahaney | 0,6068 | 0,3982 | 5,5247 | 0,9496 | 2,3950 | 0,9466 | 0,8421 | 5,8206 | 4,2772 | 1,5998 | 1,690689 | 0,578468 | 0,799645 | 0,479368 |
| Brooks | 7,4522 | 1,1901 | 4,0226 | 0,6586 | 13,496 | 0,0901 | 1,5265 | 2,8812 | 14,004 | 1,0750 | 0,908801 | 0,521138 | 4,651006 | 0,545169 |

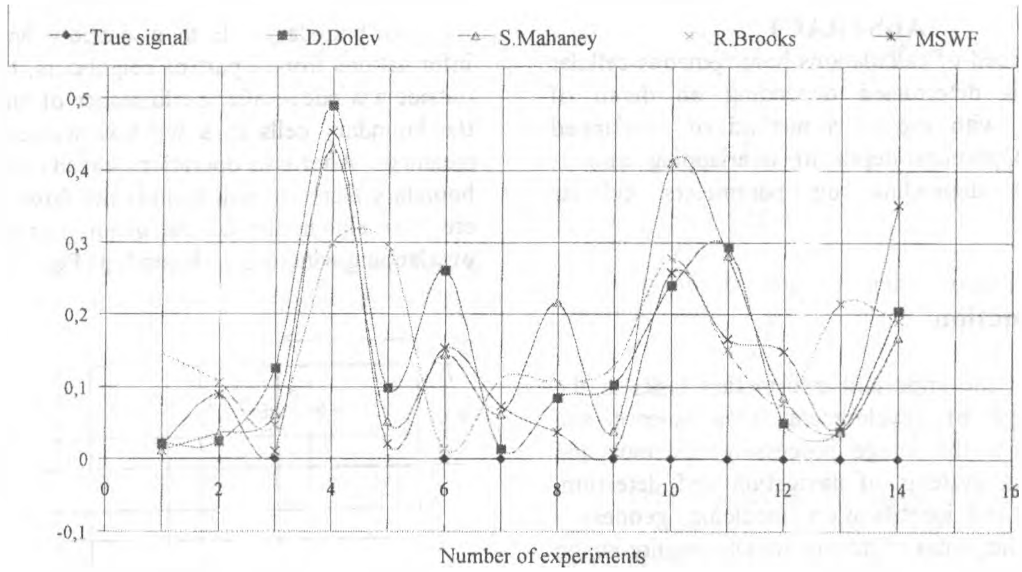


**Fig. 5 Diagram of deviation of true signal from its real value (normal distribution law)**

## 4. Conclusion

Algorithm of mean-square weight factor is offered and experimentally checked. It has shown the best results at calculation of values from sensors with different data disorder.

## References


[1] Dryshinin V., "Handbook of radar-tracking engin ", Moscow, *Voennoe izdatelstvo*, 1967, 751p.

[2] Iyengar S., Member S., "Distributed Sensor Network – Introduction to the Special Section", *Transaction on Systems, Man and Cybernetics*, vol. 21, No. 5, 1991.

[3] Sachenko A., Kochan V., Turchenko V., "Intelligent Distributed Sensor Network", *Proc. of IMTC/98*, St.Paul, USA, vol.1, 1998, pp.60-66.

[4] Brooks R., Iyengar S., "Robust Distributed Computing and Sensing Algorithm", *Computer*, June 1996, pp.53-60.

[5] Dolev D., "The Byzantine Generals Strike Again", *J.Algorithms*, 1982, pp.14-30.

[6] Mahaney S. and Schneider F., "Reaching Approximate Agreement in the Presence of Faults", *J.ACM*, July 1986, 516p.

[7] Barborak M.. Malek M., Dahbura A., "The consensus problem in fault tolerant computing", *ACM Computing Surveys*, Apr. 1986, 349p.

[8] Lamport L., Shostak R., and Pease M, "The Byzantine Generals Problem", *ACM Trans. Program. Lang. Syst.*, July 1982, pp.382-401.

[9] Lvovsky E., "Statistical methods of the formula construction", Moscow, *Vyshaja shkola*, 1988, 239p.

[10] Kuzin O., "The basic of cybernetics", vol. 2, Moscow, *Energia*, 1979, 586p.

[11] Kasandrova O., Lebedev V., "Processing of supervision results", Moscow, *Nauka*, 1970.

[12] Kozlov T., Ovsienko V., Smirnsky V., "Course of common statistic theory", Moscow, *Statistica*, 1965.