The Selection Algorithm of Mechanisms for Management of Information Risks

Victor Zavgorodniy, Pavel Lukyanov, Stanislav Nazarov *

Financial University under the Government of the Russian Federation, Moscow, Russia

Abstract

This paper is devoted to research of a selection problem of mechanisms for management of information risks. We propose to use a heuristic algorithm of a selection, which relates to the class "greedy algorithms". This algorithm allows to consider incompatibility of control mechanisms and use of integrated mechanisms. We compared the modified greedy algorithm, the genetic algorithm, and algorithm of full search. Model testing showed the advantage of the modified greedy algorithm. The algorithm can be used for the solution of other tasks, for example, for creation of information systems.

© 2014 The Authors. Published by Elsevier B.V.
Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.

Keywords: information risk; management system of information risks; incompatible mechanisms of control; "greedy algorithm", genetic algorithm.

1. Introduction

When the information system is used, it is necessary periodically to improve management system of information risks (MSIR).
Such need is caused:
• Increase of annual damage from a certain information risk;
• Appearance of new information risks;
• Change of existing risks;
• Upgrade of information system;
• Change of level and difficulties of business processes of the enterprise;
• Changes in adjacent information systems;
• Natural phenomena;
• Appearance of new legal acts;
• Changes in political and economic life of society.

For MSIR system improving we shall change existing mechanisms of control of information risks or create new mechanisms. The mechanism of control of information risks is an algorithm or a tool of control information risks. Possibilities to change different mechanisms of control significantly differ. There are no basic difficulties to modify of legal and organizational algorithms of protection, program and cryptography tools.

At the same time, modification of hardware is almost impossible as electronic circuits have high integration scale. It is possible to change hardware only by changeover of large blocks – the processor, memory, external storage devices, etc.

Anyway, modification of mechanisms is a complex problem. Its decision, as a rule, requires the considerable expenses. Therefore before making the decision on the upgrade of system we shall analyze expediency of upgrade. The ratio of expenses for upgrade and lowering of summary damage from information risks is usually analyzed. The decision on need of MSIR modification an enterprise manual makes based on the analysis of the full expenditures associated with information risks. Information on risks for last years and forecasts about risks next year is analyzed. Then it is necessary to calculate the general damage of the enterprise from information risks next year.

The general damage is the sum of damages in an information system (for example, server recovery) and damage of business processes from violation of information security and use of low-quality information. It is necessary to have also data about control mechanisms of information risks, which can be used in MSIR. This data shall contain the cost of the mechanism or its development, and cost of use of the mechanism within a year.

Then the problem of modification of MSIR consists in a selection of an optimum set of control mechanisms of information risks. The set of mechanisms is optimum if it provides a minimum of the amount of all possible damages from information risks and the general costs of purchase (development) and use of control mechanisms of information risks.

The task of optimization of expenditures on enhancement of MSIR consists in determination of such set of mechanisms of counteraction to information risks, which would provide a minimum of estimated summary general expenditures on control of information risks. This problem can be solved both with use of insurance, and without use of insurance of information risks.

2. The Proposed Algorithm

We will consider the solution of the control task of information risks without insurance. We can formalize the task of a selection of mechanisms as follows [1].

The set of risks \( R = \{r_1, r_2, ..., r_N\} \) is known. For each risk \( r_n \) the damage in the monetary form \( u_{r_n} \) is defined. Then the set of damages can be present as follows \( U = \{u_{r_1}, u_{r_2}, ..., u_{r_N}\} \) in decreasing order of value of damage. Any damage \( u_{r_n} \) is counted on condition that no mechanisms are applied to lowering of value of damage.

The set of control mechanisms \( M=(m_1, m_2, ..., m_K) \), which can be used in MSIR is defined. Each the mechanism of control \( k \) is characterized by sets of parameters \( R_k \) and \( E_k \), and the parameter \( c_k \), too.

The set \( R_k=(r_1, r_2, ..., r_J) \) consist from information risks. The mechanism \( k \) can control each risk from this set.

The set of indexes \( E_k = (e_{k1}, e_{k2}, ..., e_{KN}) \) estimate efficiency of the control mechanism number \( k \). A set member shows what part of damage from information risk number \( n \) will be prevented when using the mechanism of number \( k \). Value \( e_{kn} \) changes in the range \( 0 \leq e_{kn} < 1 \). Efficiency of all control mechanisms can be characterized means of
matrix \( E \).

\[
E = \begin{bmatrix}
e_{11} & e_{12} & \ldots & e_{1N} \\
e_{21} & e_{22} & \ldots & e_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
e_{K1} & e_{K2} & \ldots & e_{KN}
\end{bmatrix}
\]

Often in practice, some mechanisms of control influence on one risk. It means that in columns of a matrix \( E \) can be some elements, which are not equal 0. The effect from impact of several mechanisms on risk \( n \) cannot be defined as an additive index \( \sum_{k=1}^{K} e_{kn} \). In this case the summary index can equal and even to exceed 1. We use a next multiplicative index for the characteristic of damage lowering from risk \( n \), on condition that all \( K \) mechanisms are included in MSIR:

\[
\prod_{k=1}^{K} (1 - e_{kn}) = (1 - e_{1n})(1 - e_{2n})\ldots(1 - e_{Kn})
\]

(1)

This index characterizes a common part of damage from risk of \( n \), which will remain in case of application of all \( K \) control mechanisms.

The parameter \( c_k \) represents costs of the enterprise of buying or of upgrade, development, creation, and on implementation and maintenance of the mechanism number \( k \). Often the manual of the enterprise cannot direct funds on enhancement of MSIR, which exceed a certain amount of \( C_{\text{max}} \).

In practice, some mechanisms are incompatible. They cannot be used in the same system. The reason can consist, for example, in technological incompatibility. Therefore, application can be incompatible with a software platform. Mechanisms can duplicate each other and existence these mechanisms in the same system are inexpedient. Compatibility of mechanisms is set by a compatibility matrix:

\[
D = \begin{bmatrix}
d_{11} & d_{12} & \ldots & d_{1K} \\
d_{21} & d_{22} & \ldots & d_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
d_{K1} & d_{K2} & \ldots & d_{KK}
\end{bmatrix}
\]

Value of an array element \( d_{ij} \) is defined by the following condition:

\[
d_{ij} = \begin{cases} 
1, & \text{if mechanism number } i \text{ and mechanism number } j \text{ are compatibility;} \\
0, & \text{otherwise}
\end{cases}
\]

The set of the mechanisms entering MSIR, is defined as a binary vector of a configuration:

\[ X = (x_1, x_2, \ldots, x_K). \]

Components of a vector accept the following values:

\[
x_i = \begin{cases} 
1, & \text{if mechanism number } k \text{ enter MSIR;} \\
0, & \text{otherwise}
\end{cases}
\]

Mechanisms of control \( x_i, x_j \in X \) are compatible if the condition is satisfied:

\[ x_i x_j \leq d_{ij}, \quad i = 1, K, \quad j = 1, K. \]
The general damage of $U_o$, which is expected after introduction in MSIR of mechanisms of control, we will designate as residual damage. A binary vector of a configuration defines the residual damage. Taking into account the entered designations, expression for computation of residual damage can be report in terms of:

$$U^o(x_1, x_2, \ldots, x_K) = \sum_{n=1}^{N} u_{n} \prod_{k=1}^{K} (1 - e_{kn} x_k).$$

Then we can represent the problem of an optimum selection of mechanisms of control by information risks: to define a binary vector $X^* = (x_1^*, x_2^*, \ldots, x_K^*)$, which provides a minimum of the sum of expenditures on use of mechanisms of control and residual damage from all significant risks

$$\sum_{k=1}^{K} (c_k x_k^*) + \sum_{n=1}^{N} u_{n} \prod_{k=1}^{K} (1 - e_{kn} x_k^*)$$

In case of execution of conditions:

$$x_i x_j \leq d_{ij}, \quad i = 1, K, \quad j = 1, K;$$

$$\sum_{k=1}^{K} (c_k x_k^*) \leq C_{max}.$$}

The task of determination of an optimum set of mechanisms of MSIR is the non-linear discrete binary task of enumeration type. Solution of such tasks possibly algorithms of full enumeration, branch-and-bound procedures and dynamic programming algorithms. Heuristic algorithms can receive the suboptimal solutions, which satisfying to the practical purposes.

For the solution of set a problem, we suggest to use the following algorithm, which can be carried to a class of greedy algorithms [2]. The entity of an algorithm consists in a selection on each step of one of the possible mechanisms, providing maximum effect. The effect is defined by a difference of value of lowering of expenses as a result of use of the next mechanism and the missed opportunity.

The missed opportunity is defined as impossibility of use on the subsequent steps of the mechanisms incompatible with the next mechanism, which will include in system. The missed opportunity is value on which the general costs of risk management would decrease when using on the subsequent steps of the mechanisms incompatible with the mechanism, which is included in system.

Costs of control of information risks are equal to the amount of expenditures on development (purchase) and use of mechanisms of risk management and the general damage from information risks.

Thus, on each step, unlike classical greedy algorithm, are analyzed not only local effect from included on in mechanism system, but also consequences of this step in further operation of algorithm are considered.

Restrictions on the expenditures connected using mechanisms of control of information risks are considered also. We entered the following designations for the formal representation of algorithm:

- $h$ – number of the executed step of algorithm;
- $X_h(x_{h1}, x_{h2}, \ldots, x_{hk})$ – condition of a vector of a configuration after a step number $h$ of algorithm;
- $W^i(h)$ – set of the mechanisms which are already used in system on a step number $h$ of algorithm;
- $S^i(h)$ – a set of mechanisms previously not included on system in $h$ steps of algorithm, but compatible to mechanisms of a set of $W^i(h)$;
- $\Omega^i(h)$ – set of the mechanisms incompatible with a set of $W^i(h)$, i.e. excluded from further reviewing;
- $U^o_n(h)$ – residual value of damage from risk number $n$ after a selection of mechanisms on the first $h$ steps.

Thus, values $x_{hk}=1$ correspond to the mechanisms, which have been already selected on the first $h$ steps of
algorithm, i.e. mechanisms which were entered into a set of $W(h)$.

Let $m_{h+1} \in S^l(h)$ – the mechanism selected on a step $h+1$ from a set of $S^l(h)$. The selection of the $m_{h+1}$ mechanism means that the corresponding component in $X_h(x_{h1}, x_{h2}, \ldots, x_{hk})$ becomes equal to 1.

We will assume that the selected mechanism $m_{h+1}$ in a vector $X_h$ corresponds a component with number $k$. Then value by which the damage from risk $n$ will decrease in case of a selection on a step of $h+1$ of the mechanism $m_{h+1}$ with number $k$, is equal $\Delta U_n(h+1, k)$ and is defined as follows:

$$\Delta U_n(h+1, k) = U_n(h) e_{kn}.$$  \hspace{1cm} (6)

Residual value of damage from risk number $n$ will be equal:

$$U_n(h+1, k) = U_n(h)(1 - e_{kn}).$$  \hspace{1cm} (7)

Summary reduction $\Delta U(h+1, k)$ of damages from risks of all types in case of a selection on a step number $h+1$ of the mechanism number $k$ is equal:

$$\Delta U(h+1, k) = \sum_{n=1}^{N} \Delta U_n(h+1, k) = \sum_{n=1}^{N} U_n(h)e_{kn}.$$  \hspace{1cm} (8)

We denote missed opportunity of lowering of damages on the subsequent steps of algorithm as $\Delta U^-_\tau(h+1, k)$.

The missed opportunity is defined as impossibility of use on the subsequent steps of the mechanism $\tau$, which incompatible with the mechanism $k$ ($\tau \in \Omega^1(h)$).

Expression for computation $\Delta U^-_\tau(h+1, k)$ takes the form:

$$\Delta U^-_\tau(h+1, k) = \sum_{n=1}^{N} U_n(h)(1 - e_{kn}) d_{kr} \overline{s}^{1}_{h\tau}.$$  \hspace{1cm} (9)

Where $\overline{d}_{kr}$ – inverse value $d_{kr}$ from a matrix of compatibility $D$ (if $d_{kr} = 1$, then $\overline{d}_{kr} = 0$ and otherwise);

Multiplier $S^{1}_{h\tau} = 1$, if $\tau \in S^l(h)$ and $S^{1}_{h\tau} = 0$ otherwise.

Multiplier $S^{1}_{h\tau}$ allows to consider on step $h+1$ the mechanism $\tau$, which became incompatible only on step $h+1$, as a result of including of mechanism $k$.

Value of $U_n(h)(1 - e_{kn})$ is residual damage from risk $n$ after use of the mechanism $k$ on $h+1$ step.

$$\Delta U^-(h+1, k) = \sum_{\tau=1}^{K} \sum_{n=1}^{N} U_n(h)(1 - e_{kn}) d_{kr} \overline{s}^{1}_{h\tau}. \hspace{1cm} (10)$$

Integrated missed opportunity of lowering of damages, in case of a selection on $h+1$ step the mechanism $k$ is equal:

$$\Delta U^-(h+1, k) = \sum_{\tau=1}^{K} \sum_{n=1}^{N} U_n(h)(1 - e_{kn}) d_{kr} \overline{s}^{1}_{h\tau}. \hspace{1cm} (11)$$

We will enter value $\mathcal{E}(h+1,k)$ for an assessment of effect from including on a step $h+1$ of the mechanism $k$:

$$\mathcal{E}(h+1,k) = \Delta U(h+1,k) - (\Delta U^-(h+1,k) + c_k). \hspace{1cm} (12)$$

Instead of effect, it is more convenient to use the dimensionless value – specific effect:

$$\mathcal{E}_s(h+1,k) = \mathcal{E}(h+1,k)/C_{max}. \hspace{1cm} (13)$$

According to the entered designations, the algorithm of a selection of mechanisms represents sequence of steps.
On each step $h$ for $m \in S^j(h)$ is calculated $\mathcal{E}_j(h+1,k)$ also mechanism $m^*$ with number $k^*$ is selected, for which $\mathcal{E}_j(h+1,k^*)=\max$ and a condition (5) is satisfied. If such mechanism is not present, then the algorithm stops. We consider suboptimal vector is $X^*(x_1^*,x_2^*,\ldots,x_K^*)$.

Testing of model showed that with increase in quantity of variables the accuracy of an algorithm decreases. It is explained by the following reason. Value $\Delta U^-(h+1,k)$ is calculated for all mechanisms, which are not included in a suboptimal subset on h step. Those mechanisms, which won't include to a final suboptimal subset of mechanisms, are considered also.

For increase of accuracy of algorithm, the process of calculation $\Delta U^-(h+1,k)$ is changed. For computation $\Delta U^-(h+1,k)$ we use value $dv$, which received the name "viewing depth". Value of $dv$ defines the maximum number of mechanisms, which participate in calculation $\Delta U^-(h+1,k)$.

On each step, we define mechanisms in quantity $dv$, which can become incompatible after selection of the mechanism $k$. Values $\Delta U^-(h+1,k)$ for mechanisms with the greatest values $\Delta U(h+1,k) - c_k$ are calculated.

Depth of viewing place upper limits of analyzable mechanisms number. In case of algorithm execution, the number of mechanisms incompatible with $k$ can be less $dv$.

The variable of $dv$ depends on number of mechanisms of control $K$. Experimentally it was set that the highest accuracy of an algorithm is reached if value of $dv$ is in an interval $\frac{1}{4}K < dv < \frac{1}{3}K$.

The offered algorithm provides time of implementation of model less than a minute in case of quantity of mechanisms equal 100.

In case of simulation in the field of applicability of an algorithm of full search (to 30 mechanisms) the maximum relative error didn't exceed 7%, and the mean relative error equaled 0.84%. The maximum relative error of an algorithm does not exceed 15% in an interval of basic data from 10 to 100 algorithms.

In the market of control mechanisms of information risks we can meet subsystems consisting of several mechanisms. We will name such subsystem the integrated mechanism. We can optimize MSIR using normal and integrated mechanisms of control. We can solve such problem also by means of modified "greedy algorithm".

3. Algorithm of a selection of mechanisms for information risk management by means of the genetic algorithm

The solution of the selection task of mechanisms for MSIR can be executed with use of genetic algorithms [3]. The selection of genetic algorithms as an algorithm of the solution of this task is defined by properties of genetic algorithms:

- To target function it is not imposed such restrictive requirements as a monotonicity, a continuity, differentiability, etc.;
- The solution of the task with a large number of parameters is possible;
- Probably simultaneous change of many parameters;
- Decisions are near absolute, instead of to a local extremum;
- The method is simple.

There is a set of modifications of genetic algorithms. We offer to use one of possible genetic algorithm for determination of mechanisms composition for MSIR.

The purpose of development of algorithm we considered need of comparability of results of operation of
algorithm with the results received by other algorithms, and possibility of test of algorithm in broad range of basic data. We provided use of deterministic basic data (real data), and the generated arrays of random data. There is a possibility of test of algorithm by a selection of settings for adaptation to basic data.

Under the terms of the task (9-11) it is necessary to determine a vector \( \mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_K^*) \) which provides a minimum of target function:

\[
K \sum_{k=1}^{K} (c_k x_k) + \sum_{n=1}^{N} u_{r_n} \prod_{k=1}^{K} (1 - e_{kn} x_k).
\]

It is necessary to consider compatibility of mechanisms and restriction at cost. In the notation of genetic algorithms, the vector \( \mathbf{X}=(x_1, x_2, \ldots, x_k) \) appears as an individual, and target function – as fitness function. The vector \( \mathbf{X} \) has simple structure. The parameter of the target function \( x_k \) is binary and accepts values 1 or 0. In terminology of genetic algorithms, the chromosome of an individual consists of one gene. On the other hand the number of binary parameters is great.

On the first step of algorithm are entered or generated number of mechanisms and their cost, quantity and values of damages, matrices of compatibility of mechanisms and coefficients of lowering of damages from use of mechanisms.

On the second step, the initial population consisting of binary vectors (individuals) is generated. Number of mechanisms defines the quantity of discharges in vectors. Appearance 1 or 0 in each discharge is equally possible. In case of generation of initial population, compatibility of mechanisms in population is not checked and restriction at summary cost is not checked.

On the third step, it is necessary to define couples of individuals to which the operator of a crossover is applied. Pair for the selected individual is found by dint of the random number generator with uniform distribution. The random number generator selects number of an individual from not separated earlier individuals. The selected individual is integrated in pair with the next individual. The operator of a crossover with one discontinuity point is applied. The discontinuity point is selected in a random way.

On the fourth step, the operator of a mutation is executed. In model, the probability of change of each bit of an individual on inverse value is set.

The fifth step of algorithm is the most variable. On this step, the new generation of individuals is created. We suggest using one of modifications of an elite algorithm. We range individuals in the previous and current population in the order of increasing of fitness function, taking into account of restrictions at the summary cost and compatibility of mechanisms. We include in new generation on a half of quantity of individuals from each population, with the best function values of fitness.

On the sixth step, we check a condition of completion of operation of algorithm. If the condition of completion of algorithm is not satisfied yet, the new generation becomes previous and transition to a step 3 is carried out. On a step 3, we make the same operations, which were executed, for the initial population. Operation of algorithm comes to the end in case of achievement of limit value of number of repetitions of steps from the third to the fifth.

After the carried-out tests of model, we receive the results characterizing accuracy, time of simulation and boundary of applicability of model. Results of simulation allow defining the directions of enhancement of algorithm also.

Accuracy of algorithm depends on quantity of individuals of initial population and quantity of cycles. Such result is expected. It is explained by exceptions of algorithm (fig. 1).

Other result received during simulation, is not obvious. Accuracy and stability of algorithm significantly depends on a level of incompatibility of mechanisms. So in case of number of mechanisms of equal 30 and 10% of incompatibility of mechanisms (10% of total number of mechanisms), and also in case of 50 individuals in initial population and 50 repetitions of a cycle, in 30% of cases of implementations the result of simulation is not received. Any vector (individual) is not admissible.
Algorithm accuracy sharply decreases on boundary of steady operation of algorithm. The reason of it is that the considerable part of actions of algorithm is the operation with incompatible combinations of individuals.

If to set the following parameters of simulation of 40 mechanisms, 30 types of damages, quantity of individuals in initial population 200, the quantity of cycles 1000, incompatibility of mechanisms of 10%, a mutation of 5% of bits in each individual, algorithm works with high stability and provides the relative error not quitting for 5%.

However, in case of 80 mechanisms and 50 types of damages the settings of algorithm given above do not ensure its steady functioning.

Other direction of adaptation of the genetic algorithm to the solution of tasks with a high level of incompatibility and the significant amount of mechanisms is transition to generation of initial population in which only compatible individuals would be located. In this case, process of formation of initial population could combine an accidental selection with use of special templates. Templates can be created based on a matrix of compatibility of mechanisms.

4. Conclusion

To the considered algorithms it is necessary to add an algorithm of full enumeration. It is necessary for an assessment of accuracy of the algorithms based on the genetic algorithm and modified greedy algorithm, and for establishment of boundaries of practical application of an algorithm of full enumeration.

The exact algorithm of full enumeration works in real time if the number of mechanisms is in the range $1<k<20$. In case of $k=25$ value the operating time is equal about 10 minutes, if $k=27$ the operating time is equal 45 minutes (all measurements are made in implementation process of algorithms on the personal computer with clock rate of 2 GHz) (fig. 2).

Thus, use of an algorithm of full enumeration for obtaining the exact solution of practical tasks of a selection of mechanisms of control by information risks is impossible on personal computers. The algorithm is important for obtaining data on the accuracy of other algorithms in case of small values of basic data.

Tests of the modified greedy algorithm showed its practical significance for the solution of the selection task of control mechanisms. The advantage this algorithm is its minor time of implementation. Modeling time not large depends on dimensionality of the task. In case of number of mechanisms the equal 100, operating time of algorithm does not exceed 1 minute.
Algorithm accuracy quite meets requirements of practical tasks. The mean relative error does not exceed 5% on all operating range of basic data. The worst implementations on accuracy taking into account adequate selection of depth of the analysis of incompatibility do not exceed 15%.

The genetic algorithm on time of implementation takes a line item between the modified greedy algorithm and full enumeration. On all operating ranges of basic data, a time of implementation of algorithm does not exceed several tens of minutes.

Accuracy of algorithm depends both up dimensionality of the task, and up a level of incompatibility of mechanisms. The best results on accuracy and time of implementation are received in case of an insignificant level of incompatibility of mechanisms (less than 1%). In case of quantity of individuals in population equal 200 and quantity of cycles the equal 1000, relative error of results does not exceed 3%. In case of an average and high level of incompatibility of mechanisms (>10%) the algorithm shows results on accuracy the worst results of the modified greedy algorithm.

Thus, it is expedient to apply the modified greedy algorithm to practical tasks of a selection of mechanisms. For tasks with a small level of incompatibility of mechanisms, both algorithms can be applied to obtaining the most exact decisions.

The algorithm can also be used for the solution of other tasks, for example, for creation of information systems.

References