Processing information on the state of a computer system using probabilistic automata

Semyonov S.G., Gavrilenko S.Y., Chelak V.V.

National technical university «Kharkiv Polytechnical Institute»

Kharkiv, Ukraine

s_semenov@ukr.net, gavrilenko08@gmail.com, Victor.Chelak@gmail.com.

***Abstract* - The paper deals with the processing of information about the state of a computer system using a probabilistic automaton. A model of an intelligent system for detection and classification of malicious software is proposed, which compares a set of features that are characteristic for different classes of viruses with multiple states of the machine. The analysis process is reduced to modeling the operation of the automaton taking into account the probability of transition from state to state, which at each step is recalculated depending on the reaction of the environment. The received results of research allow to reach a conclusion about the possibility of using the offered system for detection of the harmful software.**

***Keywords*: information processing, heuristic analyzer, probabilistic automaton, malicious software, anti-virus information protection system**

I. INTRODUCTION

Technical progress, as well as ubiquitous computerization and informatization, have become inalienable attributes of modern society. At the same time, one of the most difficult tasks, constantly arising at all stages of the development of information and computer technologies, was the task of providing information and functional security.

According to experts [1-6], a huge threat to the security of computer systems (CS) is malicious software or computer viruses. So, according to Cisco's semi-annual report on information security in the first half of 2016, the growth in the number and quality of attacks with computer viruses has led to huge material losses (more than 100 billion dollars) to the economy and public image losses of various organizations around the world. [1].

It should be noted that anti-virus vendors conduct ongoing monitoring and improvement of specialized software, introduce modern technical and mathematical solutions for processing information on the state of CS. However, more and more frequent cases of successful cyber-attacks, as well as the growth of various kinds of losses in organizations, indicate a lack of measures taken, as well as the imperfection of modern methods of processing and analyzing information on the state of the CS.

Thus, improvement existing, and also development of new methods of information processing about a condition of CS is an urgent scientific task.

The carried-out analysis of literature [2-16] has shown that the problem of processing and the analysis of data on a condition of CS bases generally on heuristic analyzers, intended for recognition of again created malicious applications.

Heuristic analyzers, as a rule, include the intellectual subsystems of data processing which are based on the theory of artificial intelligence. The main complexity in their use consists in the correct choice of a set of signs which would allow to divide classes of abnormal behavior among themselves and to separate them from normal behavior. All of them have also big expenses of time for training and calculation of coefficients, and are also characterized by high probability of false operation.

For elimination of the specified shortcomings authors suggest to use basic provisions of the known and well proved theory of probabilistic automatic machines [17-21] for development of a method of information processing about a condition of CS.

Therefore, the purpose of this work is development of a method of information processing about a condition of CS on the basis of the probabilistic automatic machines.

II. DEVELOPMENT OF AN INFORMATION PROCESSING SYSTEM

From literature, it is known that basic provisions about the probability automata (PA) were for the first time formulated in 1963 in fundamental operation of M. Rabin [17].

The automatic machine is intended for creation of mathematical models of dynamic systems at which there is an

uncertainty described by statistical regularities. This uncertainty is connected:

- with inaccuracy of knowledge of states in which the modelled systems are in process of the functioning;

- with non-determination of rules of change of these statuses.

The conducted researches showed that the probable automatic machine (fig. 1) functions by execution of transitions, after each of which there is an up-dating of values of probable variables of such automatic machine depending on response of the environment.

Generally PA [19] works in some environment in which it gives output signals of y_i and from which he receives input x_i.

If the automata at a point in time t moved from state s_m to state $s_к$ at a point in time $t+1$ received the «fine» signal, then the probability p_{mk} is replaced with αp_{mk}, where the index α is more than 0 and less than 1, and the rest of the probabilities in the line are replaced with $(1-\alpha)p_{mk}/M$. If the signal received is «notfine», then the probability p_{mk} is increased to $(1-\alpha)+\alpha p_{mk}$, and the rest are decreased to $(1-\alpha)(1-p_{mk})/M$, where M is the number if internal automata states.

The principle of the offered approach of information processing about a condition of CS consists in check of possible habitats of viruses and detection of commands in them (groups of teams), viruses, characteristic of this type. Each of suspicious teams is compared with a set of conditions of s. The possibility of transition from a condition of s_m to a condition of s_k in timepoint of t is defined by an entrance condition and value of a marker of K of the transition depending on probability from a condition of sm in a condition of s_k, and from reaction of the environment.

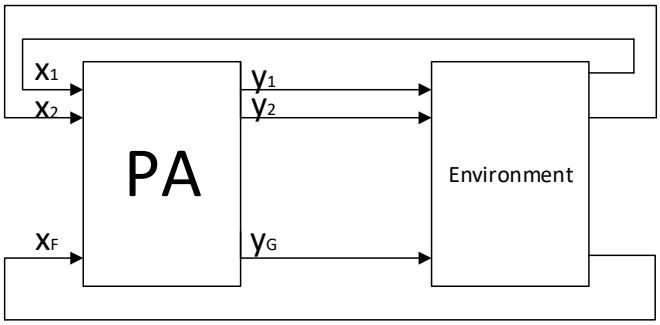


Fig. 1. Probability automata work scheme

If there is a sequence of transitions from start state in finite, then the automatic machine gives the message on possible infection of system with a certain type of a virus or on the safe software. Introduction of additional transitions and obsession with statuses allow to find modification of the known viruses.

At the same time two stages are executed: generation of structure of the automatic machine for certain classes of viruses (fig. 2) and testing of the checked files regarding belonging to the generated classes of viruses or their modifications (rice.3).

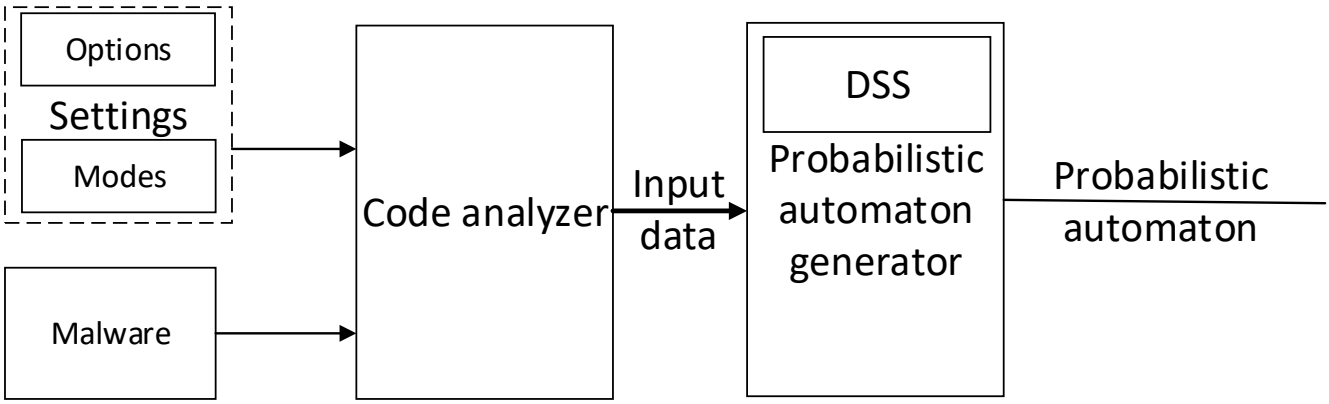


Fig.2. Generation of automata structure

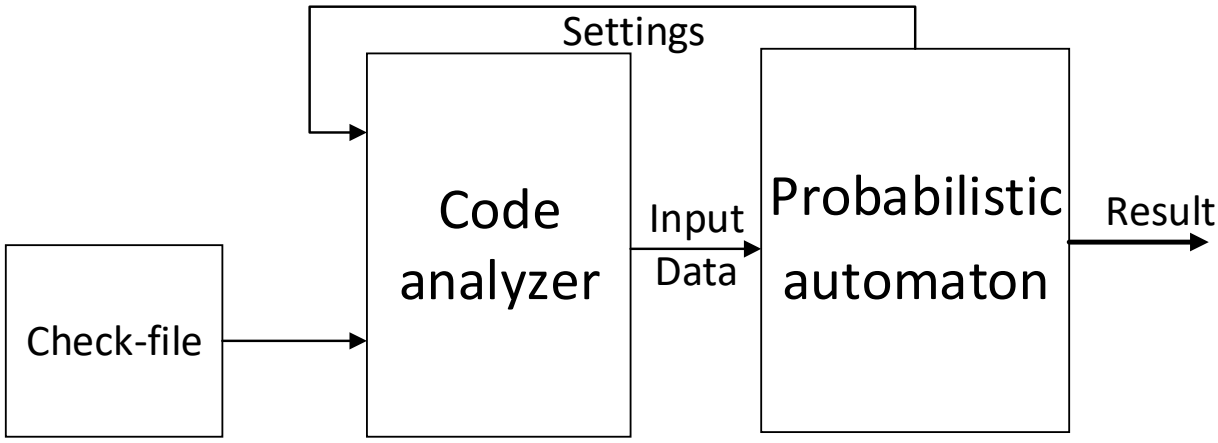


Fig. 3 – Testing of the checked files regarding belonging to the generated classes of viruses or their modifications

At a stage of generation of structure of the automatic machine on an entrance of system harmful files enter. The analyzer of a code forms a file signature, compares each component of a signature with a condition of the automatic machine, forms structure of the probabilistic automatic machine, carries out its optimization, forms a class of types of viruses and generates the table of probabilities of transitions.

The conducted researches have shown that performance of an objective of information processing is impossible without formation of the structured knowledge of probability of transitions from a state to a state. For the solution of this task it is offered to use the created table of probabilities of transitions on the basis of decision support system (fig. 4). The System of decision support system (DSS) consists of two parts – the memory block storing information on ancestors flowing states and the block of rules of transitions of system.

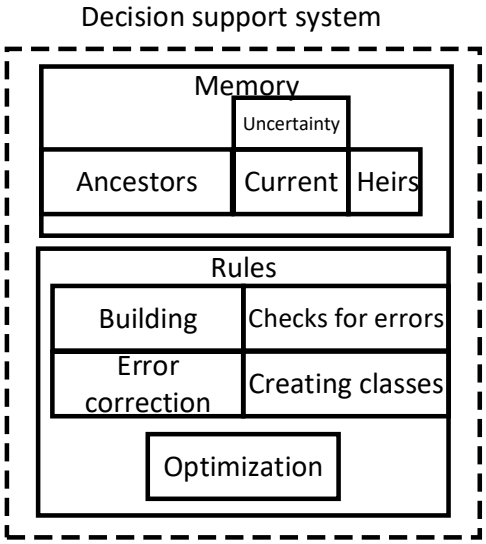
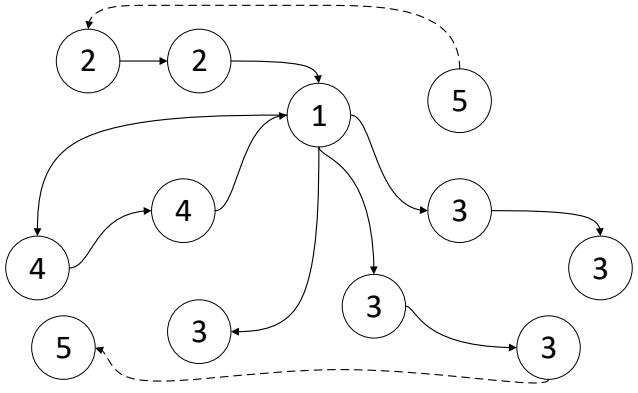


Fig. 4. Decision support system

For making decision of a problem of information processing on a condition of DSS the CS recursion depth is set. Fig. 5 shows an example of the DSS contents for a recursion depth of R_k=2.

The marker value K depends on N – are long the input sequence of the automatic machine and from coefficient of Perfect_Key which allows to consider as the input sequence only a part of a code. Perfect_Key is initially set in settings of system, and then every time is enumerated in case of detection of the repeating code locations.

We will review an example of generation of structure of the probable automatic machine for two classes of the viruses and their modifications presented in the form of the table of Tab. 1. A, B, C, D is a single-byte feature set, characteristic of a certain class of a virus, for example And – is compared with existence in a virus body of API of the GetProcAddress function which derives the address of the exported function or a variable from the given dynamic link library (DLL).



The fig. 5 Contents of Memory of Decision support system for Rk =2 recursion depth

where:

1 – the index on current state of the automatic machine (rather this state other addresses are written down);

2 - conditions ancestors of current state;

3 - conditions successors of current state;

4 – uncertain states (states which the system with the set R_k, sees both as the ancestor, and as the successor);

5 – conditions of the automatic machine which aren't stored in memory of SPR.

We will enter designations of a status of the automatic machine: $S_0(t)$ – start state of the automatic machine, $S_i(t)$ – a current status of the automatic machine, S_{k0} – finite secure state of the automatic machine, S_{k1} – finite value consisting automatically with identification of a S_{K1}, S_{k2} virus finite consisting automatically with identification of a S_{K2} virus. For this example of $N = 8$ bits. As it is long a code small, we will set initially value Perfect_Key = 0.25, P = 0.

After submission on an input of the generator of structure of the automatic machine given on fig. 1, a S_{K1}, system virus will generate structure of the automatic machine given on fig. 7.

DSS is engaged in regulation of value of the table of probabilities on each step of operation of the automatic machine. In fig. 6 it is presented an example of one of rules of optimization of SPR.

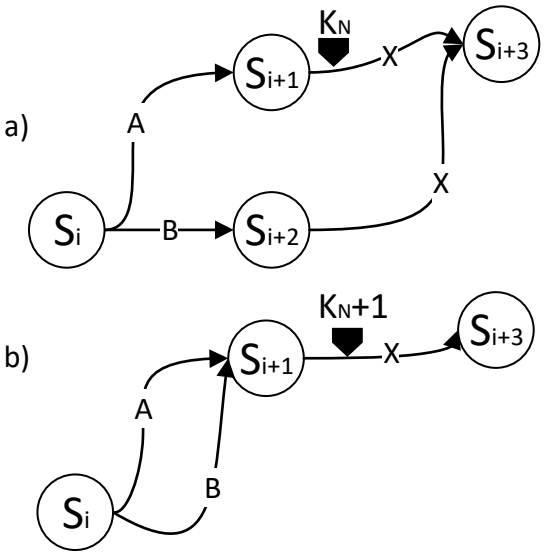


Fig. 6. An example of one of rules of optimization of SPR (a – the Diagram before optimization, b – the Diagram after optimization).

TABLE 1. TABLE OF VIRUS CODE EXAMPLES

Virus type	**Virus code**								
	Code 1			**Code 2**			**Code 3**		
SK_1	AA	BB	CC	DD	BB	CC	EE	BB	CC
SK_2	DD	DD	DD	DD	DD	BA	DD	DD	CB

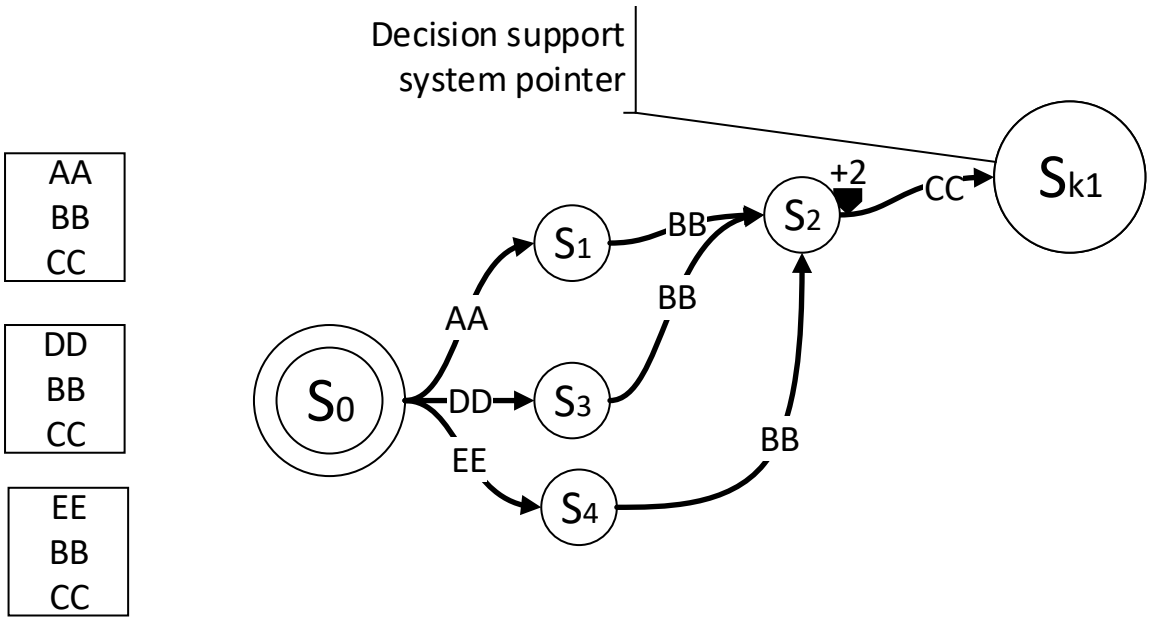


Fig. 7. The flowgraph of transitions of the probabilistic automatic machine without optimization for a SK1 virus

At generation of structure of the automatic machine there was a change of Perfect_Key upon transition from a condition of S2 to a condition of Sk1 due to identity of the subsequent fragments of a code. Increase in Perfect_Key will allow to raise requirements to the accuracy of coincidence of the checked site of a code to a transition condition. (i.e. the lower threshold will be increased)

System of decision-making having analyzed the received automatic machine, will execute its optimization due to reduction of quantity of states, recalculation of a marker of M and the table of probabilities of transitions.

Optimization happens as follows. If in visible area for the set depth of Rk=2 transitions to one and too a state with identical conditions of transition are observed, then it is necessary to count the number of such repetitions, to increase marker value M by number of repetitions, to unite the repeating ways in one edge. Thus, we have received into 2 fortune and 5 edges instead of 4 states and 7 edges (fig. 8).

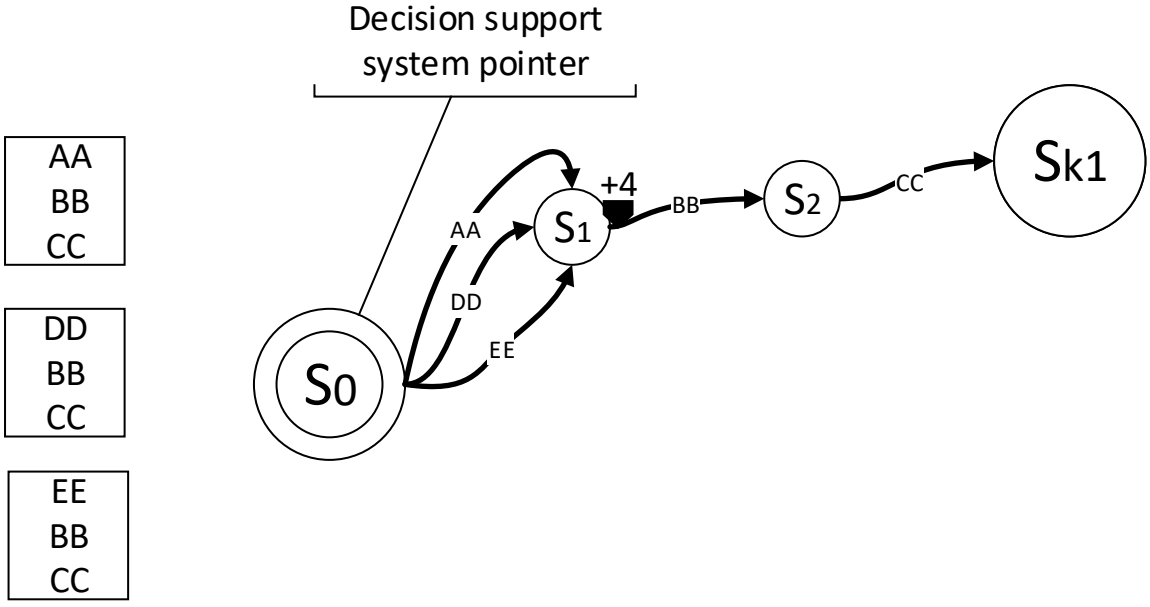


Figure 8. The flow graph of transitions of the probabilistic automatic machine after optimization

After giving on an entrance of the generator of structure of the automatic machine, a S_{K2}, system virus will generate the structure given in fig. 9 and the following table of probabilities (Tab. 2) .

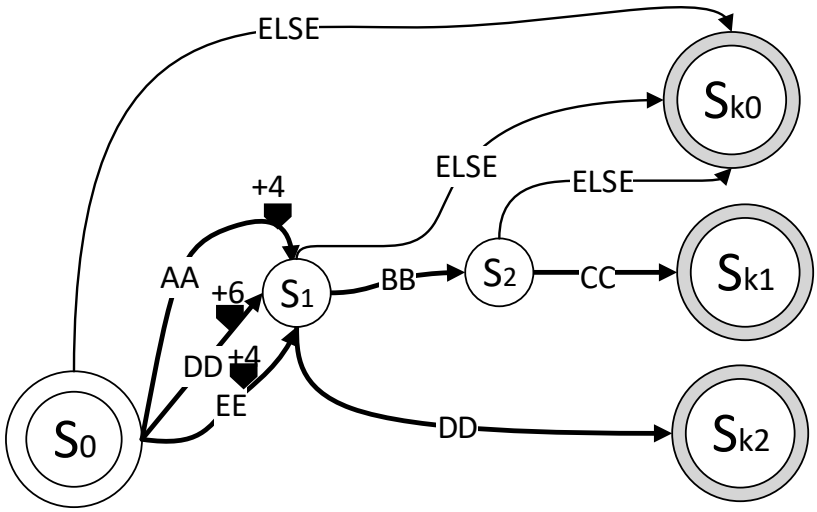


Fig. 9. Flow graph of Transitions of the Probabilistic Automatic Machine

TABLE 2. TABLE OF PROBABILITIES OF TRANSITIONS OF THE AUTOMATIC MACHINE

S(t)/X(t+1)	AA	BB	CC	DD	EE	PERFECT_KEY
$S_0(t)$	1	0.25	0.25	1	1	0.25
$S_1(t)$	0.75	1	0.75	1	0.75	0.75
$S_2(t)$	0.75	0.75	1	0.75	0.75	0.75

CONCLUSION

In operation the question of information processing of KS status with use of the developed probable automatic machine is considered. The model of the intelligent detection system and classification of malicious software is offered.

The conducted research in the conditions of use of ten test programs from which 5 – it were the modifications of programs recognized by automatic machine and the second 5 – the programs which aren't belonging to the classes of the recognized programs showed, rather high frequency of the correct assessment of a status of KS (received the correct answers in 9 of 10 cases). These results give the grounds to claim about a possibility of use of the developed approach in practice of an assessment of statuses of KS as additional tool for detection of the virus attacks in the general system of detection of the malicious software.

Further enhancement of the offered model of processing can lie in the plane of adaptation of the probable automatic machine to large volumes of processed data and carrying out researches with use of bigger volume and different modifications of the malicious software.

LITERATURE

[1] The semi-annual report on IB from Cisco. [Electronic resource]. – Access mode: http://www.securitylab.ru/blog/personal/Informacionnaya_ bezopasnost _v_detalyah/316275.php.

[2] Shelukhin O. I. Detection of invasions into computer networks / O. I. Shelukhin, D. Zh Sakalema, A. S. Filinova. – M.: Garyachy liniya-Telecom, 2013. – 220 pages.

[3] Goshko S. V. Technologies of fight against computer viruses / S. V. Goshko. – M.: Solon Press, 2009. – 352 pages.

[4] Rabin, M.O. Probabilistic automata. Information and Control 6(3), 230–245 (1963) (Russian translation: Rabin M. O. Probabilistic automatic machines / Cybernetic collection, Issue 9. – M.: Foreign literature, 1964. - Page 123-141.

[5] Semenov. S.G. Data protection in the computerized operating systems (monograph) / S. G. Semenov, V. V. Davydov, S.Yu. Gavrilenko. – "LAP LAMBERT ACADEMIC PUBLISHING", Gemany, 2014. – 236 pages.

[6] Semenov S. G. Control methods and identifications of a condition of computer systems on the basis of BDS testing / S. G. Semenov, S.Yu. Gavrilenko//Proceedings of the symposium "Metrology and metrology assurance" – Sozopol, Bulgaria, 2015. – page 400-405.

[7] S. Gavrilenko, V. Chelak, Hornostal O. Intrusion detection in computer systems. Proceedings of the symposium "Metrology and metrology assurance" – Sozopol, Bulgaria, 2016, pp. 342-347.

[8] Eskin E. Anomaly detection over noisy data using learned probability distributions. In Proc. 17th International Conf. on Machine Learning, pages 255-262. Morgan Kaufmann, San Francisco, CA, 2010.

[9] Gavrilenko of Page Yu. Development of templates of identification of a condition of computer systems on the basis of BDS testing/Semenov S. G., Gavrilenko of Page Yu., Bangs V. V//V_snik NTU "HP _". _nformatika that моделювання. – Харків, 2016. – No. 21. – page 118-125.

[10] [10] Gavrilenko S. Developing parametrical criterion for registering abnormal behavior in computer and telecommunication systems on the basis of economic test//G. Semenov, S. Gavrilenko, V. Chelak//Actual problems of economics. – Kiev, 2016, Vol 4(178), ss. 451-459.Semenov S. Approximating computer system operation technologies under external action through the brusselator model with perturbation in the form of dynamic chaos / S. Semenov, S. Gavrilenko // Revista RECENT – Industrial Engineering Journal – Transilvania University of Brasov – Romania, Vol. 16 (2015), No. 1 (44).

[11] Semenov S. Development of antivirus security system// S. Semenov, I. Sheverdin S. Gavrilenko//Proceedings of the seventh world congress “Aviation in the XXI-st century” Safety in Aviation and Space Technologies.–2016, pp 1.10.39 – 1.10.41

[12] Semenov S. G. Gert-model прогнозування параметрів funkts_onalno ї bezpek tekhn_chnikh of systems / S. G. Semenov, S.Yu. Gavrilenko//Sistemi of an obrobka інформац і ї: ZB. sciences. the ave. – X: HU PS, 2016. VIP. 2 (139). From 50-52.

[13] Yazov Yu. K. Fundamentals of methodology of a quantitative assessment of efficiency of information security in computer systems. Rostov - on - Don: SKNTs VSh publishing house, 2006. – 274 pages.

[14] Nesterenko V. A. Statistical methods of detection of violations of safety in network//Information processes. 2006. T. 6, No. 3. Page 208-217.

[15] Bezobrazov S. V., Golovko V. A. Use of neural network detectors in artificial immune systems for detection and classification of computer viruses//Neyrokompyyutera. 2010. No. 5. Page 17-31.

[16] Rabin, M.O., Probabilistic automata. Information and Control 6(3), 230-245 (1963). (Russian translation: Rabin M. O. Probabilistic automatic machines / the Cybernetic collection, - the Issue 9. - M.: Foreign literature, 1964. - Page 123-141.)

[17] Pospelov D. A. Probabilistic automatic machines. M.: Energy, 1970. – 88 pages.

[18] Хопкрофт D., Motvani R., Ullman J. Introduction to the theory of automatic machines, languages and calculations 2nd prod.: The lane with English - M.: Publishing house "Williams 2002. - 528 pages [

[19] Mateus, P., Qiu, D., Li, L.: On the complexity of minimizing probabilistic and quantum automata. Information and Computation 218, 36–53 (2012)

[20] Bukharayev R. G. Bases of the theory of probabilistic automatic machines. — M.: Science, 1985. - 287 pages.