UDC 004.49: 004.056.57: 004.942 DOI: 10.30748/soi.2017.151.19

V. Shevchenko[1], Ju. Shcheblanin[2], A. Shevchenko[2]

[1] *Taras Shevchenko National University of Kyiv , Kyiv*
[2] *State University of Telecommunication, Kyiv*

THE EPIDEMIOLOGICAL APPROACH TO PROGNOSIS AND MANAGEMENT OF INFORMATION INCIDENTS

The subject of the article is prognosis and management of information incidents. The purpose of the article – use the experience of mathematical modeling of biological epidemics, to predict the results of large-scale information and cyber-attacks. Numerical simulation methods are used. Existing approaches to modeling epidemics in biological and the computer world are analyzed. Analogies and differences of epidemics in the biological and computer worlds were established. The obvious form of mathematical models for prediction of information and cyber-attacks dynamics was proposed. The latent period features of infections were analyzed. Opportunities of not destructive existing of alien objects in biological and computer systems were discussed. Simulation features are discussed. The conditions of deterministic chaos occurrence in the simulation are analyzed. Dependencies of epidemic peaks from specific ratios ware determined. Using of ratios for guidance of information and cyber incidents epidemic process was proposed.

***Keywords**: prediction model, the logistic function, epidemic, information, attack, incident.*

Introduction

Action algorithms of biological viruses are similar to algorithms of computer viruses. For prevention of biological virus intrusion, usually use vaccination. At same way for prevention of computer viruses intrusion, usually use antivirus. In both of cases we have so called "zero day threats" when type of virus is unknown. Therefore prevention means are unknown too. In this case we defend our body (computer of biological) from "Diseases of dirty hands". By another words, common sanitary rules may protect us from infection.

Information security incidents mainly associated with virus attacks because it often using computer virus tools. But the goal of information attacks – not just to bring the computer down. The main objectives are economic losses; hit the image, undermining confidence, promoting the desired information content.

Number of information and cyber incidents increased on average by 66% annually. This is at least twice the growth of the mobile market and the growth of global GDP (fig. 1). The most dangerous is that at least 71% realized unauthorized access (attacks) remain undiagnosed [1–2]. The intensity of the attacks and the consequences need to be predicted for the effective decision making support on counter attacks. So study of forecasting models of information and cyber-attacks is relevant. Information and cyber-attacks have much in common with biological epidemics. The results of the study of biological epidemics regularities accumulated centuries and can be useful for predicting the consequences of information and cyber-attacks.

The epidemics scale led to high levels of systemic actions of doctors. Therefore epidemics mathematical research methods are also used in the study of patterns of noninfectious diseases [3]. Modern epidemiology is based on a systems approach. Much attention is paid prediction of epidemics options for the timely adoption of adequate preventive measures. To predict is possible to use previous experience [4–6], or predictive modeling [7–8]. Previous experience does not cover all possible situations, so it provides not enough information to make appropriate decisions. Modeling requires adequate mathematical models. On the one hand the model should reflect the main features of the modeled process. On the other hand, the model should be simple enough to ensure their input data, for timely adjustment of the structure and model parameters in according to change of the situation or problem statement.

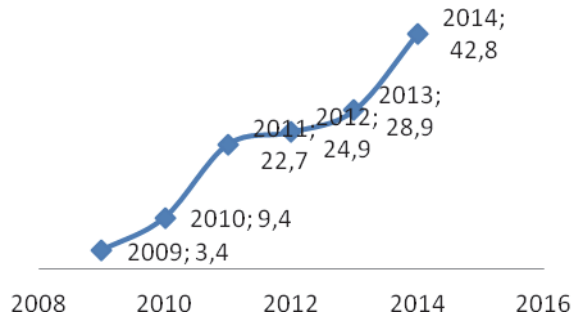

Fig. 1. Growth of information security incidents

The practice raises for the models contradictory conditions [7]: speed, accuracy, clarity, completeness consideration of influencing factors and others. The complexity of the model should meet the complexity of the process and on the other hand, opportunities to provide input data. The more complex the model, the

more difficult to provide its inputs, the higher the degree of uncertainty in which it operates.

On the other hand, the higher the degree of uncertainty of modeled object, the easier it should be a model. To resolve the contradiction appointed using rough model that is not only simple, but also reflects the most significant features of the modeled processes. Modeling of epidemics should also start with rough models. Rough models benefits: speed training to use (speed of structural and parametric synthesis), clarity and simplicity of timely parameters correction in according to changing internal and external conditions of the modeling object.

The purpose of the article – use the experience of mathematical modeling of biological epidemics [9], to predict the results of large-scale information and cyber-attacks.

Initial Model

The most famous rough model: linear, exponential and logistic [7]. Linear and exponential model used for the well-studied processes in a limited range of input values that are at the same stage of the life cycle. Exponential growth with saturation models used in processes that have reached the limit of its development. Moore's Law is most famous among the models of unlimited exponential growth. Moore's Law is based on the assumption of no limits growth. If the resource provision varies randomly or seen several stages of the life cycle, it is more adequate the S-shaped logistic model [7; 10] in the form of ordinary differential equations

$$\frac{dy}{dt} = m \cdot (y - Y_{min}) \cdot (Y_{max} - y), \qquad (1)$$

or as a function that is its decision

$$y(t) = Y_{min} + \frac{Y_{max} - Y_{min}}{1 + e^{-m \cdot (Y_{max} - Y_{min}) \cdot (t - \Delta t)}}, \qquad (2)$$

here y – dynamic development variable (eg, infected); t – time; Y_{min}, Y_{max} – lower and upper limits of y values; m – a permanent factor; Δt – abscissa of symmetry point (shift along the abscissa axis).

Examples of S-shaped patterns in biology and medicine: the dynamics of disease risk after the Chernobyl accident, population dynamics and population [7; 10].

Similar models are used to simulate the dynamics of growth of infection by computer viruses. Dynamics models of virus spread (SI, SIS, SIR, AAWP, PSIDR) take into account specific of distributed environment (computer network topology) and specifics of combating against the viruses [11–15].

Integral-differential equations are most appropriate for epidemics simulate [16–18]. The result of solving these equations is the family of S-shaped curves and curves resulting additive convolution the latter. Integral-differential equation is mathematically rigorous, but not quite comfortable in dealing with. In addition, the interim decision results are not obvious for epidemiologist without special mathematical training. Sometimes this limitation is relevant for information security managers (or system administrators). Transition to the logistics ordinary differential equations in finite increment and replacement of integration by finite sums was used in [9] for model simplifying.

Modified Model

Modifies the known structural Boyev-model (fig. 2) [16]. Let: P – the total number of infected sites; S, N – susceptible and resistant to infection; E – in incubation (infected themselves, but have not infect others and not identified); I – contaminated sites that are actively infect others; R – objects that are treated and received immunity (antivirus); F – items that had to be completely removed from work after infection; K_S, K_E, K_F – coefficients susceptibility to infection, transmission of infection, withdrawal from work (total disability); $f(I, S, K_E)$ – logistical dependence of infection among susceptible. Consider in more detail the mathematical dependent of transitions between states of objects.

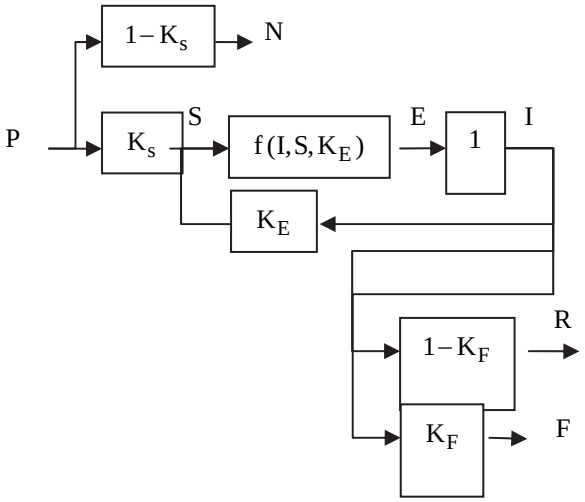

Fig. 2. The modified structural epidemic model by Boyev

First, find the initial number of objects resistant and susceptible to infection:

$$N_0 = P \cdot (1 - K_S), \qquad (3)$$

$$S_0 = P \cdot K_S. \qquad (4)$$

Found values are taken as the initial conditions at the initial time t_0. Subsequently, the number of objects resistant to the infection rate may be adjusted by K_S, reflecting natural immunity (operating systems features) immunity, which was formed through antivirus tools, complete isolation share objects using quarantine measures and so on.

But the effect of $1 - K_S$ factor regarding transfer objects to unfavorable group will apply only to the amount of susceptible objects S that are not transferred

to other groups (E , I , R , F). If at any time t_i (as a result of the undertaken preventive measures) the number of resistant facilities increased by an amount ΔN, the total number of objects impervious to time t_i will be calculated as the sum of the previous value and the corresponding increment

$$N(t_i) = N(t_{i-1}) + \Delta N . \quad (5)$$

Set the equations time integration step equal Δt and since the initial time at each step sequentially calculate the change of state facilities. Differential equations of increasing number of the objects that are at the first time interval - the incubation period, written in finite differences

$$\frac{\Delta E_1}{\Delta t} = K_E \cdot I \cdot S . \quad (6)$$

Last dependence is similar to one of the Lotka-Volterra equations. Decision of equation (including complex interactions with other variables) is a dependence that is qualitatively similar to the logistics. Find the value increment of the number of objects that are in a state of the first period of the incubation period

$$\Delta E_1 = K_E \cdot I \cdot S \cdot \Delta t . \quad (7)$$

Then reduce the number of susceptible objects on the founded amount

$$S = S - \Delta E_1 . \quad (8)$$

Because the incubation period T_E and the period for the infection T_I is more than integration step Δt, then for the variables E, I, we will write a subscript indexes that will mark the interval in the incubation period of infection state (eg, if Δt is minute, therefore the index is the number of minutes in the same period):

$$E_i ,\ i = \overline{1, i_{end}^E} ,\quad I_i ,\ i = \overline{1, i_{end}^I} , \quad (9)$$

here $i_{end}^E = \frac{T_E}{\Delta t}$, $i_{end}^I = \frac{T_I}{\Delta t}$ – numbers of last periods in the respective periods.

Then execute the shift of state of contaminated sites. All those who were in the state $(i-1)$-th time interval transit into a state (i)-th time interval

$$E_i = E_{i-1},\ i = \overline{1, i_{end}^E - 1} . \quad (10)$$

Those who were in the last period of time incubation period i_{end}^E will transit to the first time interval the of infected status $I_1 = E_{end}$.

Further procedure is performed similarly for shift of states of infected objects

$$I_i = I_{i-1} ,\ i = \overline{1, i_{end}^I - 1} . \quad (11)$$

Note that the total number of infected sites and sites in incubation period are calculated as relevant amounts of contaminated sites at all time intervals of corresponding periods

$$E = \sum_{i=1}^{i_{end}^E} E_i ,\ I = \sum_{i=1}^{i_{end}^I} I_i . \quad (12)$$

Increment of the number of objects that are cured and objects had to withdraw from work are calculated by the appropriate coefficients of the number of contaminated sites that are at the last infection time interval

$$R = K_R \cdot I_{end} , \quad (13)$$

$$F = (1 - K_R) \cdot I_{end} . \quad (14)$$

Implementation of the model in the software environment MatLab proved its efficiency and adequacy (fig. 3). Timeline is different for different types of information and cyber-attacks.

The main attention is given for multi-layered attacks. Therefore, in the model were saved purely biological characteristic – the incubation period.

The incubation period in the computer world correspond the latent period during which the malicious code executes additional adjustment, additional penetration in complete secrecy of their actions. In multi-level attack malicious code type 1 initially weakened defense, prepares virtual channels guaranteed access to information resources and resource management in future. Then, by well-prepared channels malicious code type 2 enters the system (or in another more secure or more controlled part of system) and perform basic tasks malware. Such attacks levels may be several. These levels can combine different attacks ways from highly technical to social engineering.

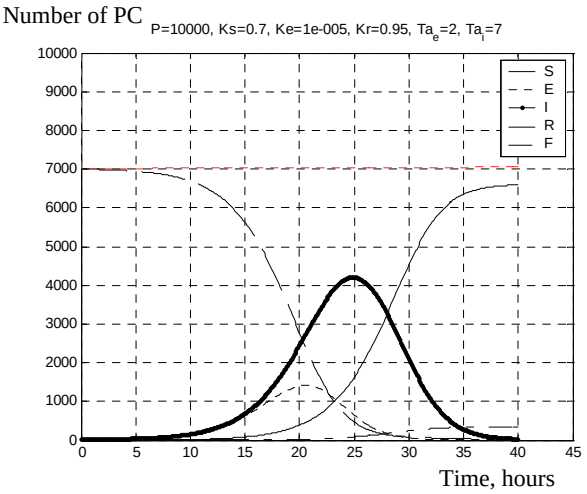

Fig. 3. The results of numerical simulation of a cyber-attack

The simulation showed that integration step increasing leading to significant damage to the quality of the epidemic model (observed phenomenon of deterministic chaos). Reducing almost does not change either to quality or quantity of modeling of process development, but proportionally increases the simulation time.

The graphs show the logistics nature of reducing the number of favorable sites and the increasing number

of cured objects and objects removed from work. General view of the number of infected objects and objects that are in the incubation period corresponds to existing statistical data on the development of biological epidemics, allowing the use of known biological laws for the computer world.

The main practical result of simulation is a useful "bell"- dependence of the number of infected objects.

Amplitude of this dependence determines the level of epidemic danger. This is fundamental epidemics difference in biological and computing world. In the computer world is dangerous, any infection. But hold analogy with the biological world. As is well known in the biological world is not exists organisms completely free from viruses, bacteria, parasites or other objects that use resources without permission of donor body. We call it alien biological objects. In most cases, the organism keeps a balance with alien biological objects or even enters them in cooperation - symbiosis.

If exterminate all alien biological objects, then their place will come other, which can be more harmful. Therefore, in the biological world organism fights against not all alien object. A similar situation is possible in the computer world. But some mechanisms are different. For example, the availability of useful (or rather harmless) alien objects not ensures absence of other (more malicious) sites. Although, sometimes it works in biological world too. On the other hand, unlike the biological world, freedom from the presence of alien object is not necessarily leads to other (malicious) objects incomes. The overall conclusion about the biological and computer alien object: not mandatory to fight against all alien objects.

Return to the simulation results. The simulation showed that the first prerequisite of the epidemic is the emergence some (non-zero) number of infected objects or objects in incubation period state.

Based on simulation results, the second prerequisite of the beginning of the epidemic is a certain proportion of resistant objects and conditions of transmission from infected to susceptible objects. Mathematically, it is defined by a certain ratio K_s and ratios K_E.

The sufficient condition of the epidemic is the simultaneous occurrence of the first and the second required conditions. Thus under epidemic understand the condition where the percentage incapacitated object exceeds a certain value. In a technical sense is the amount at which loses exceeded normal performance information infrastructure of a business area (enterprise, organization, industry).

Therefore investigation of dependencies of epidemic peaks from ratios K_s (fig. 4) and ratios K_E (fig. 5) is interesting from appropriate decision making point of view.

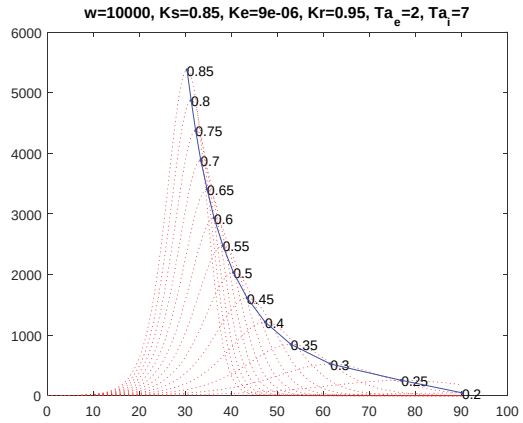

Fig. 4. Dependencies of epidemic peaks from K_s

Appropriate Matlab listing for ratios variation:

```
Ksb= 0.2;
    dKs= 0.05;    j3end = 14;
    KKs = Ksb:dKs:Ksb + dKs*(j3end-1);
Keb= 0.000003;
    dKe= 0.000001;    jend = 7;
    KKe = Keb:dKe:Keb + dKe*(jend-1);
```

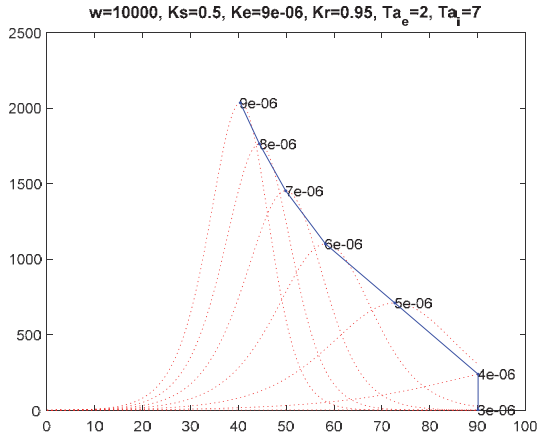

Fig. 5. Dependencies of epidemic peaks from K_E

But more useful is determining of dependencies of epidemics peaks from ratios K_s and ratios K_E simultaneously (fig. 6).

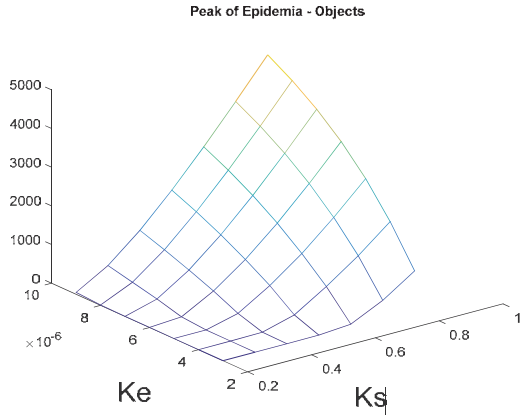

Fig. 6. Dependencies of epidemic peaks from ratios K_s and K_E (3D view)

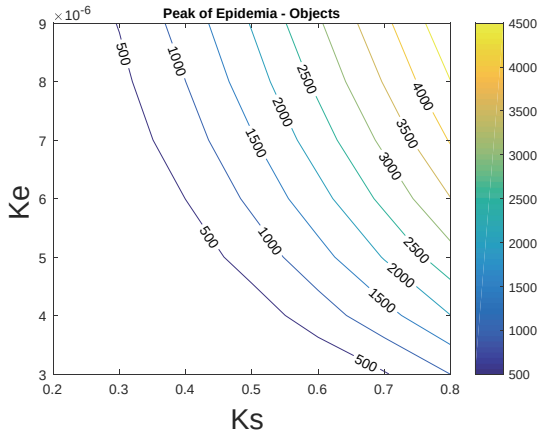

Fig. 7. Dependencies of epidemic peaks from ratios K_S and K_E (Topo-view)

Now if we know dangerous level of epidemic peak, then we can provide some limitation (control) for ratios K_S and K_E. Appropriate values of K_S and K_E will provide non-dangerous epidemic peak level in incidents case. In this meaning K_S and K_E is guidance for information and cyber incidents epidemic process.

Conclusion

The logistic model adequately predicts the epidemic and allows you to schedule regular proactive measures. A great feature of the model is complete visibility of the physical variables and all mathematical transformations. This allows you to accurately monitor the adequacy of the model and make the necessary adjustments in time. New result is separation of data on the number of infected objects and objects that are in the incubation period for the temporal stages corresponding states. This provides additional opportunities for disease control measures.

Mathematical model allowed to separate different types of preventive measures by means the ratios K_S, K_E.

As the results of further simulations, the level of the epidemic depends from second required conditions and the start time depends from the first. This allows to decompose required conditions in the numerical experiment for forecasting of possible epidemics scenarios.

Areas for further research.

1. Increasing the adequacy of the model by a more detailed definition of factors of susceptibility to infection, transmission, extraction of work.

2. Detailing of inner structure of K_S and K_E ratios for more precise guidance and control of information and cyber incidents epidemics.

References

1. *The Global State of Information Security® Survey 2016. Turnaround and transformation in cybersecurity*, (2017), https://www.pwc.com/gx/en/issues/cyber-security/information-security-survey.html (accessed 23 March 2017).

2. *Healthcare cybersecurity challenges in an interconnected world. Key finding from The Global State of Information Security. Survey 2015*, (2017), http://www.pwc.ru/en/ riskassurance/publications/assets/healthcare.pdf (accessed 23 March 2017).

3. Artyukhov, I.P., Polikarpov, L.S. and Hamnahadaev, I.I. (2010), "*Metody epidemiologicheskogo izuchenija neinfekcionnych zabolevanij*" [*Methods of epidemiological study of noninfectious diseases*], Medical. Univ. by the spec. 060101 – Medicine, Typography KrasHMU, Krasnoyarsk, 145 p.

4. Romanenko, T.A. (2009), "Diagnostichno-prognostichni kriterii prognozuvannja tendencii rozvitku epidemichnogo procesu kashljuku" [The diagnostic and prognostic criteria for predicting trends of the epidemic process of pertussis], *Prophylactics medicine*, No. 4(8), pp. 17-23.

5. Britton, T., Janson, S. and Martin-Lf, A. (2007), Graphs with specified degree distribution, simple epidemics and local vaccination strategies, *Adv. Appl. Prob.*, No. 39, pp. 922-948.

6. Van Doorn, E.A. (1991), Quasi-stationary distribution and convergence to quasi-stationary of birth-death processes. *Adv. Appl. Prob*., No. 23, pp. 683-700.

7. Shevchenko, A.V. and Shevchenko, V.L (2010), "Grubi modeli rozvitku v medycyni" [Rough development models in medicine], *Medical informatics and engineering*, No. 4, pp. 52-55.

8. Kermack, W.O. and McKendrick, A.G. (1927), A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond. A.*, No. 115, pp. 700-721.

9. Shevchenko, A.V. and Gepko, A.L. (2011), "Matematychna model prognozuvannja dynamyky epidemij" [Mathematical model of epidemics dynamics prognosis], *Prophylactics medicine*, No. 3(15), pp. 3-6.

10. Shevchenko, V.L. (2011), "*Optymizacijne modeljuvannja v strategichnomu planuvanni*" [*Optimizing modeling in strategy planning*], CVSD NUOU, Kiev, 283 p.

11. Garetto, M., Gong, W. and Towsley, D. (2003), Modeling Malware Spreading Dynamics, *IEEE INFOCOM 2003*, www.ieee-infocom.org/2003/papers/46_01.PDF (accessed 23 March 2017).

12. Monahov, Yu.M., Hruzdeva, L.M. and Monahov, M.Yu. (2010), "Vredonosnye programy v kompjuternych setjach" [Malicious software in computer networks], Publishing House Vladym. state univ. Press, Vladymyrskyy state.univ., Vladimir. 72 p.

13. Klymentiev, K.E. (2013), "*Kompjuternye virusy s antivirusy: vzgljad programmista*" [*Computer viruses and antiviruses: programmers view*], DMK Press, Moskow, 656 p.

14. Leveille, Jasmin (2002), *Epidemic Spreading in Technological Networks*, 100 p., www.hpl.hp.com/techreports /2002/HPL-2002-287.pdf (accessed 23 March 2017).

15. Bolot, J. and Lelarge, M. (2008), Cyber Insurance as an Incentive for Internet Security, *Workshop in Economics of Information Security (WEIS) Seventh Workshop on Economics of Invormation Security*, June, pp. 25-28.

16. Boev, B.V. and Makarov, V.V. (2004), "Geo-informacionnye sistemy i epidemii grippa" [Geo-information systems and flu epidemics], *Veterinary pathology*, No. 3(10), pp.51-59, http://elibrary.ru/item.asp?id=9165685 (accessed 23 March 2017).

17. Boev, B.V. (2017), "Kompjuternoe modelirovanie v ocenke posledstvij akta biologicheskogo terrorizma" [Computer modeling in evaluation of the aftermath of the biologically terrorism act], *Proceedings. I Russian Symposium on biological security*, Research institute of Epidemiology and Microbiology named by Gamaleia N.F. RAMS, Moscow, www.bio.su. (accessed 23 March 2017).

18. Shevchenko, V. and Shevchenko, A. (2017), The Epidemiological Approach to Information Security Incidents Forecasting for Decision Making Systems. *13-th International Conference Perspective Technologies and Methods in MEMS Design (MEMSTECH). Proceeding*, Polyana, April 20-23, pp. 174-177, http://ieeexplore.ieee.org/document/7937561/, DOI: 10.1109/MEMSTECH.2017.7937561/.

Received by Editorial Board 10.10.2017
Signed for printing 7.12.2017

Відомості про авторів:
Шевченко Віктор Леонідович
доктор технічних наук професор
професор кафедри Київського національного
університету ім. Тараса Шевченка,
Київ, Україна
https://orcid.org/0000-0002-9457-7454
e-mail: gii2014@ukr.net

Щебланін Юрій Миколайович
кандидат технічних наук старший науковий співробітник
доцент кафедри
Державного університету телекомунікацій,
Київ, Україна
https://orcid.org/0000-0002-3231-6750
e-mail: gii2014@ukr.net

Шевченко Аліна Віталіївна
аспірант Державного університету телекомунікацій,
Київ, Україна
https://orcid.org/0000-0003-3793-9364
e-mail: gii2014@ukr.net

Information about authors:
Shevchenko Viktor
Doctor of Technical Sciences Professor
Professor of Department of Taras Shevchenko
National University of Kyiv,
Kyiv, Ukraine,
https://orcid.org/0000-0002-9457-7454
e-mail: gii2014@ukr.net

Jury Shcheblanin
Candidate of Technical Science Senior Researcher
Assistant professor of Department
of State University of Telecommunication,
Kyiv, Ukraine,
https://orcid.org/0000-0002-3231-6750
e-mail: gii2014@ukr.net

Alina Shevchenko
Ph.D student of State University of Telecommunication,
Kyiv, Ukraine
https://orcid.org/0000-0003-3793-9364
e-mail: gii2014@ukr.net

ЕПІДЕМІОЛОГІЧНИЙ ПІДХІД ЩОДО ПРОГНОЗУВАННЯ ТА УПРАВЛІННЯ ІНФОРМАЦІЙНИМИ ІНЦИДЕНТАМИ

В.Л. Шевченко, Ю.М. Щебланін, А.В. Шевченко

Предметом статті є прогнозування та управління інцидентами інформаційної безпеки. Мета статті: використання досвіду моделювання біологічних епідемій для прогнозування результатів великомасштабних інформаційних та кібернетичних атак. В роботі використані методи чисельного моделювання. Проаналізовані існуючі підходи щодо моделювання епідемій в біологічному та комп'ютерному світах. Встановлені аналогії та відмінності епідемій в біологічному та комп'ютерному світах. Запропоновані наочні форми математичних моделей щодо прогнозування динаміки розвитку інформаційних та кібернетичних атак. Проаналізовані властивості латентного періоду інфекцій. Обговорені можливості не руйнівного існування чужинних організмів в біологічних та комп'ютерних системах. Обговорені особливості моделювання. Обговорені умови можливого виникнення детермінованого хаосу. Встановлені залежності впливу певних параметрів на величину піку епідемії. Запропоноване використання визначених параметрів для керування епідеміологічним процесом.

Ключові слова: *прогнозуюча модель, логістична функція, епідемія, інформація, атака, інцидент.*

ЕПИДЕМИОЛОГИЧЕСКИЙ ПОДХОД К ПРОГНОЗИРОВАНИЮ И УПРАВЛЕНИЮ ИНФОРМАЦИОННЫМИ ИНЦИДЕНТАМИ

В.Л. Шевченко, Ю.Н. Щебланін, А.В. Шевченко

Предметом статьи является прогнозирование и управление инцидентами информационной безопасности. Цель статьи: использование опыта моделирования биологических эпидемий для прогнозирования результатов крупномасштабных информационных и кибернетических атак. В работе использованы методы численного моделирования. Проанализированы существующие подходы к моделированию эпидемий в биологическом и компьютерном мирах. Установлены аналогии и отличия эпидемий в биологическом и компьютерном мирах. Предложены наглядные формы математических моделей прогнозирования динамики развития информационных и кибернетических атак. Проанализированы свойства латентного периода инфекций. Обговорены возможности неразрушающего существования чуждых организмов в биологических и компьютерных системах. Обсуждены особенности моделирования. Обсуждены условия возможного возникновения детерминированного хаоса. Установлены зависимости влияния определенных параметров на величину пика эпидемии. Предложено использование определенных параметров для управления эпидемиологическим процессом.

Ключевые слова*: прогнозирующая модель, логистическая функция, эпидемия, информация, атака, инцидент.*