Model for the Propagation of Malicious Objects in a Computer Network with Variable Infection Intensity

Antoaneta Popova [1]

[1] *Technical university of Sofia, Kliment Ohridski Bld. 3, 1000 Sofia, Bulgaria*

Abstract **– There are a significant number of scientific publications that use epidemic models to propagate malicious objects in a computer network. These models are based on Markovian models with a constant intensity of transitions, which does not correspond to real conditions, since the intensity of transitions changes due to the fact that after some time antivirus systems start to recognize malware. This paper proposes an original approach based on an epidemic model with variable infection intensity of hosts in a computer network. In the beginning, when the threat is not recognized, the malware spreads rapidly. After a certain period of time, the antivirus system recognizes the malicious code, which leads to a decrease in the infection intensity. Simulations have been done for different infection intensity and threat recognition. Demonstrated models account for the infection time of hosts in the computer network, latency phase, malware detection, and clearance from the system.**

Keywords **– Antivirus agent, epidemic model, malicious objects, Markov model, threats.**

1. Introduction

Global society is becoming increasingly dependent on information technology.

DOI: 10.18421/TEM123-06
https://doi.org/10.18421/TEM123-06

Corresponding author: Antoaneta Popova,
Technical university of Sofia, Kliment Ohridski Bld. 3, 1000 Sofia, Bulgaria
Email: apopova@tu-sofia.bg

Received: 23 March 2023.
Revised: 12 July 2023.
Accepted: 19 July 2023.
Published: 28 August 2023.

The growth of this dependency has also led to the creation of so-called safety-critical systems [1]. In order to ensure technogenic safety, two main problems related to reliability [1], [2] and security [3], [4], [5] must be solved.

The problem of reliability is almost solved with the help of hardware and software redundancy, and various forms of diversity used in cloud technologies.

The main challenge of modern information technology is the security of systems compromised by the proliferation of malicious code spreading as viruses, worms, Trojans, email viruses, botnets, rootkits, etc. The distinction between the different threats lies in the mode of spread, the objectives of the malicious actor and the damage caused by the intrusion.

Cloud-based information systems use various cutting-edge technologies to ensure their security, but on the other hand, contain a large amount of critical information that will be compromised in the event of a breach.

There are many models of the spread of malicious objects in an information environment [6], [7] based on pandemic models [8], [9], [10] for the spread of infections in society.

In [11], [12], dynamic (Markovian) models based on a system of Kolmogorov equations for malware propagation depending on the topology of the computer network are considered.

An improved S-E-I (Susceptible-Exposed-Infected) model for virus propagation is proposed by studying the propagation of infected files in a P2P network [13]. The authors assume that the probability of a peer downloading an infected file is proportional to the spread of infected files in the network.

A relatively large majority of authors [14], [15], [16] use the SEI(Q)R (Susceptible-Exposed-Infected-(Quarantined)-Recovered) model to describe in detail the process of infection propagation in networks.

A threshold Rcq (reproduction number) defining the evolution of contagion spread is proposed in analogy with the spread of epidemics among biological species [17]. If the reproduction number is less than 1, the contagion spread is self-limiting and disappears.

The characteristic of the models considered is that they use a constant infection rate β = const. (some papers use the term contact rate per unit population). In practice, the infection rate β is not constant and depends

The methods considered are essentially semi-Markov models. Based on them, a system of equations is derived and solved on simulators or by numerical methods. Such a solution technique does not allow solving models with variable coefficients.

2. Justification of the Model

The aim of this work is to propose a model with a variable propagation coefficient of computer viruses depending on the activity of antivirus agents.

The classification of antivirus programs according to the detection method of malware is signature analysis, behavior-based methods and machine learning. The first of these assumes that the signature of the virus code is known and this signature is searched for in the exposed information resources. In order to detect such a signature, the other two methods, behavior-based anomaly and/or machine learning, are mainly used. These methods are less effective than signature analysis because they have a latency period - it takes some time to accumulate statistical information to identify an object as infected. If the virus does not have a variable signature, then a search can be performed using signature analysis.

An example architecture for file analysis in the cloud has been presented in study [17]. Most antivirus engines use a similar architecture (Fig. 1), where suspicious objects are sent for analysis and if a malware is detected information about it is sent to the agent.

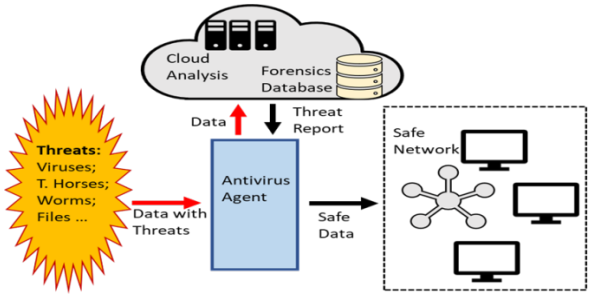

Figure 1. Example architecture of an antivirus agent

The basis for further research is the SEIQRS model Fig. 2 proposed in [18].

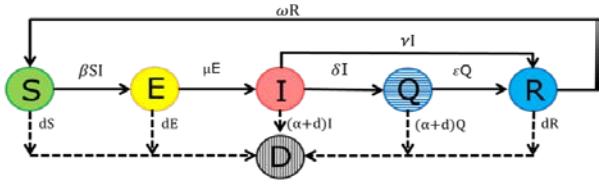

Figure 2. A semi-Markovian graph of the SEIQRS model describing the distribution of malware

State S (Susceptible) contains the size of the population, respectively all the resources that are susceptible to infection. With a spread intensity β and directly proportional to the number of resources in states S and I (Infected), the resources move to state E (Exposed), from where they become infected with intensity and move to state I. The output graphs of state I are two: the object is quarantined Q (Quarantined) with intensity δ or it is cleared from the malware with intensity ν and goes to R (Recovered). The cleaned information resources are returned again from R to S with intensity ω.

In order to more adequately represent the contagion process, the contagion intensity β is represented in the model as the product of two variables, Infection Rate and Iteration Rate.

State D (Dead) contains all infected resources that are destroyed for some reasons e.g., disk corruption, file deletion, etc.

Obviously, this model describes the behavior of the system for a short time, it is not logical that in the presence of quarantined and cleaned resources, (located in states Q and R) the coefficients β and μ are constant.

To analyze the impact of antivirus agents, the model given in Fig. 2 is simplified Fig. 3, yielding a classical SIR model:

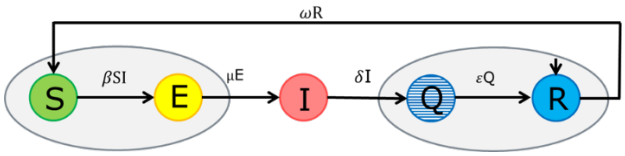

Figure 3. A simplified view of the SEIQRS model from Fig.2

It is assumed that all susceptible are exposed to the infection. Due to the fact that in states Q and R the infected objects are known, they are merged into one state. Ignored is the transition of infected objects to state D, i.e., the destruction of infected files for other reasons.

The model of Fig. 3 is recast so that the transition from the susceptible state to the infected state depends directly proportionally but nonlinearly on the number of infected resources detected.

A model where the coefficient β depends on the number of resources cleared of infection is given in Fig. 4.

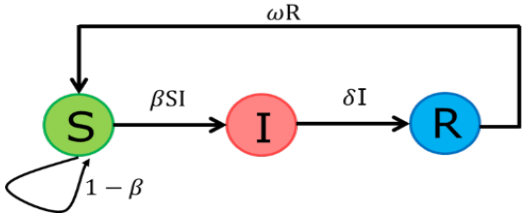

Figure 4. SIRS model with variable infection rate

The Kolmogorov equations [19] describing the system of Fig. 4 are:

$$\begin{aligned} \frac{dS}{dt} &= \omega R - \beta SI \\ \frac{dI}{dt} &= \beta SI - \delta I \\ \frac{dR}{dt} &= \delta I - \omega R \\ S + R + I &= const \\ \beta &= f(R) \end{aligned} \quad (1)$$

The last equation gives the relationship between the vulnerability of the system resulting from the number of resources cured. Logically, after a certain amount of detected and cleaned resources, the infection should be recognized by the antivirus agent and the coefficient β modeling the intensity of infection propagation should start to decrease.

In this case, an exponential dependence of the β coefficient on R was used:

$$\beta(R) = \alpha e^{-k\frac{tA}{R}}, \quad (2)$$

where:

β is the infection intensity variable;

α – The initial intensity of infection;

k – Coefficient determining the effectiveness of infection detection;

A – Volume of the population susceptible to infection;

R – The number of detected infected resources.

Other β dependencies used for activation functions in neural networks [20], [21], such as linear or identity activation function, Nonlinear activation function, Sigmoid or logistic activation function, Tanh or hyperbolic tangent activation function, ReLU (Rectified Linear Unit) activation function, etc. are logical.

3. Model Simulation

Numerous tools for simulating Markov chains, epidemiological simulators, etc. are available on the Internet. Unfortunately, most of them cannot work with the integrity of transitions specified as a function.

Differential equations are analogous to differential and functional-differential equations. They appear to be a convenient tool for solving a similar class of problems in which the coefficients of the intensities of the transitions are set as functions.

MS Excel is a convenient environment for this kind of simulation because it allows on the one hand easy solution to a system of differential equations representing a Markov process and on the other hand their graphical visualization [22].

A fragment of the simulation is shown in Fig. 5: the time step, the variable infection factor β, and the states S, I, and R, are given. The last column performs the check S+R+I=const, significantly showing the correctness of the computation.

Time	β	Susceptible	Infected	Recovered	Total
0	0,39	999	1	0	
1	0,3899961	997,909103	1,99089712	0,1	
2	0,3899767	995,739718	3,96119196	0,29908971	
3	0,3899189	991,433421	7,87137053	0,69520891	1000
4	0,3897695	982,916554	15,6010997	1,48234596	1000
5	0,389409	966,19657	30,7609736	3,04245593	1000
6	0,3885752	933,859645	60,0218013	6,1185533	1000
7	0,386715	873,166487	114,71278	12,1207334	1000
8	0,3827304	765,826928	210,581061	23,5920114	1000
9	0,3746845	596,637328	358,712555	44,6501175	1000
10	0,3599161	380,95411	538,524517	80,5213729	1000
11	0,3365634	187,622411	678,003765	134,373825	1000
12	0,3062121	78,5544505	719,271348	202,174201	1000
13	0,2733908	35,3025202	690,596144	274,101336	1000
14	0,2415768	18,8116686	638,027381	343,16095	1000
15	0,2122074	11,6800993	581,356212	406,963688	1000
16	0,1857234	8,14897066	526,75172	465,09931	1000
17	0,1621653	6,19991344	476,025605	517,774482	1000
18	0,1413973	5,03145044	429,591507	565,377042	1000
19	0,1232032	4,28581107	387,377996	608,336193	1000
20	0,1073342	3,78685274	349,139155	647,073992	1000
21	0,0935356	3,44058504	314,571507	681,987908	1000
22	0,0815615	3,19341561	283,361526	713,445059	1000
23	0,0711831	3,01305925	255,20573	741,781211	1000
24	0,0621929	2,87915411	229,819062	767,301784	1000
25	0,0544055	2,77835605	206,937954	790,28369	1000
26	0,0476575	2,70163456	186,32088	810,977486	1000
27	0,0418058	2,64271181	167,747714	829,609574	1000
28	0,0367266	2,59712435	151,018531	846,384345	1000
29	0,0323127	2,56163859	135,952163	861,486198	1000

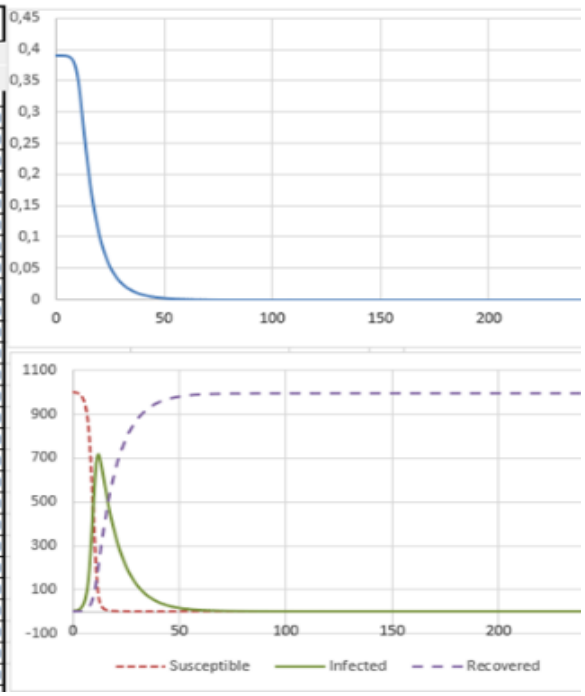

Figure 5. Example solution of the SIRS model with variable Infection rate using MS Excel

Different types of simulations have been performed to investigate the influence of the parameters.

3.1. No-Feedback Model ω=0 With Constant Infection Intensity β=Const

In this case (Fig. 6), the model is transformed into a typical epidemiological model of SIR with constant infection intensity β = const = 0.19.

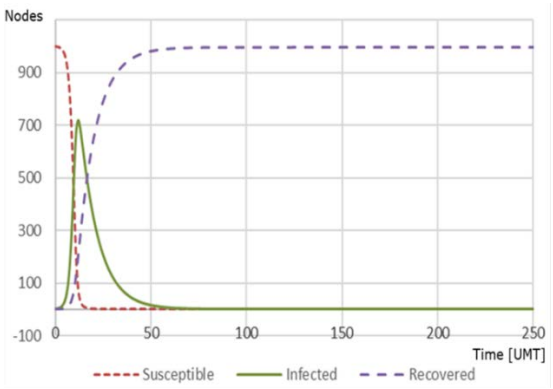

Figure 6. Graphical representation of the spread of infection states S, I, and R. At the end of the simulation, all hosts are infected.

It can be seen that all hosts in the network become infected after a certain time due to the constant intensity β.

3.2. No-Feedback Model ω=0, with Variable Infection Intensity β and α=0.1

The above graph of Fig. 7 is visualized the variation of the infection intensity β= f(R), calculated by the formula (2), with an initial value of β=0.19, α=0.1 and k=1.

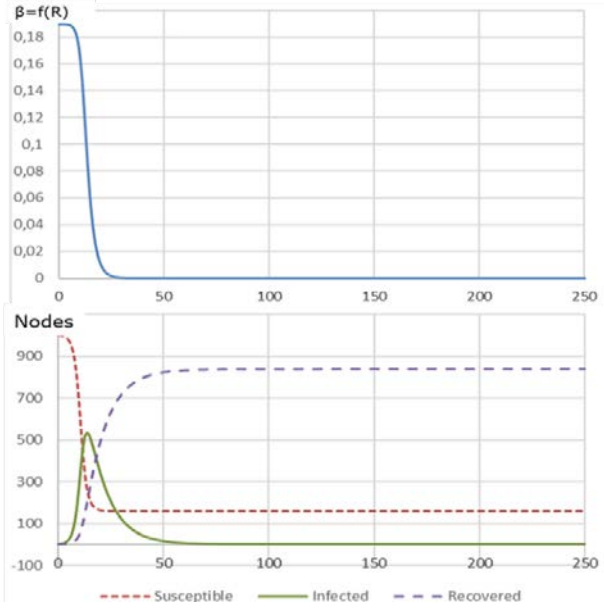

Figure 7. Graphical representation of the infestation intensity variable β=f(R) and the infestation spread states S, I, and R. At the end of the simulation, all hosts are infected

The above graph in Fig. 7 shows that after 15 units of model time (UMT), the contagion propagation intensity starts to decrease and tends to zero after 50 UMT.

Information about the infected hosts is given in the bottom graph. It can be seen that all hosts in the network become infected after a certain amount of time. The explanation for this lies in the fact that the intensity of infection decreases once the infection of the hosts is done.

3.3. No-Feedback Model ω=0, With Variable Infection Intensity β and α=0.5

The results are given in Fig. 8.

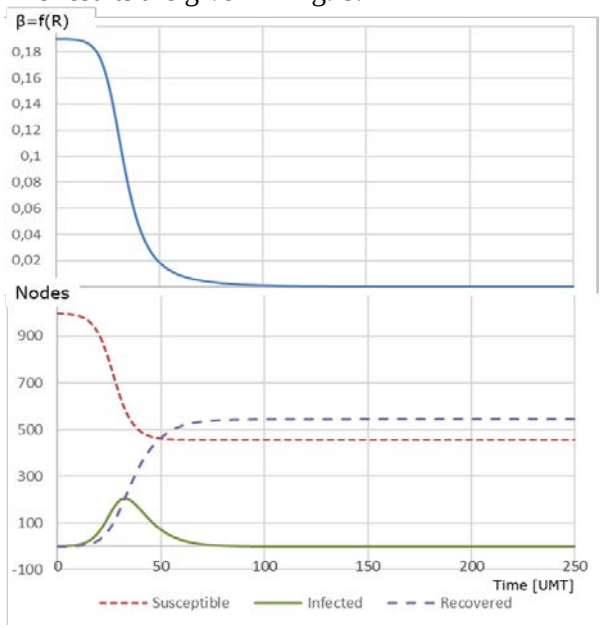

Figure 8. Graphical representation of the contagion intensity variable β= f(R) and the contagion spread states S, I, and R. At the end of the simulation, not all hosts are infected

The difference with respect to the previous case is the higher sensitivity of the functor β= f(R), determined by the higher value of α=0.5. The other coefficients are the same: initial value of β=0.19 and k=1.

Fig. 8 shows that the coefficient β tends to zero after 60 UMT. At this point, the infection of all hosts has not completed, which is shown in the bottom graph.

3.4. No-Feedback Model ω=0, with the Variable Infection Intensity β and α=0.5

The results are given in Fig. 9. Here the sensitivity of the functor β=f(R) is increased again. The values of the coefficients are the initial values of β=0.09, α=1, and k=1.

Fig. 9 confirms the claim that the effectiveness of antivirus protection depends inversely and nonlinearly on the intrusion detection time.

The graph shows that most hosts were not infected, since the computer virus was detected in a timely manner and its signature was passed to the antivirus agent, which in turn stopped the infection from spreading.

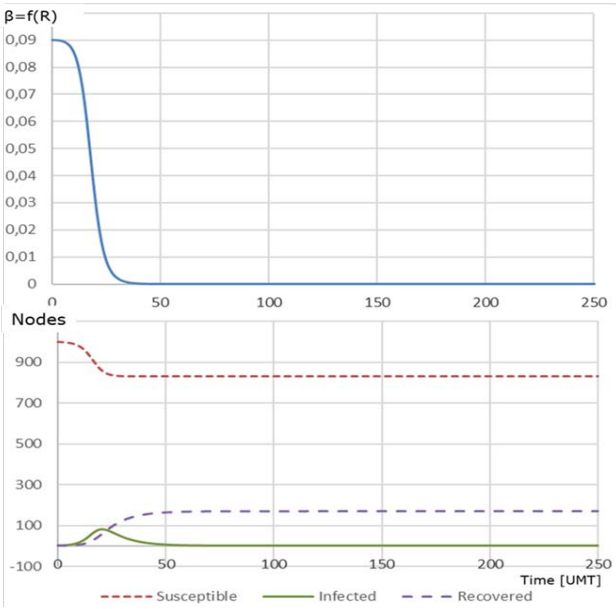

Figure 9. Graphical representation of the contagion intensity variable β=f(R) at α=1 and the contagion spread states S, I and R. At the end of the simulation, about 12.5% of the hosts are infected

3.5. SIRS Model with Feedback ω=0.01, Variable Infection Intensity β and α=1

In Fig. 10, a SIRS model is given in which the infestation-cleaned hosts are tied into service with intensity ω=0.01.

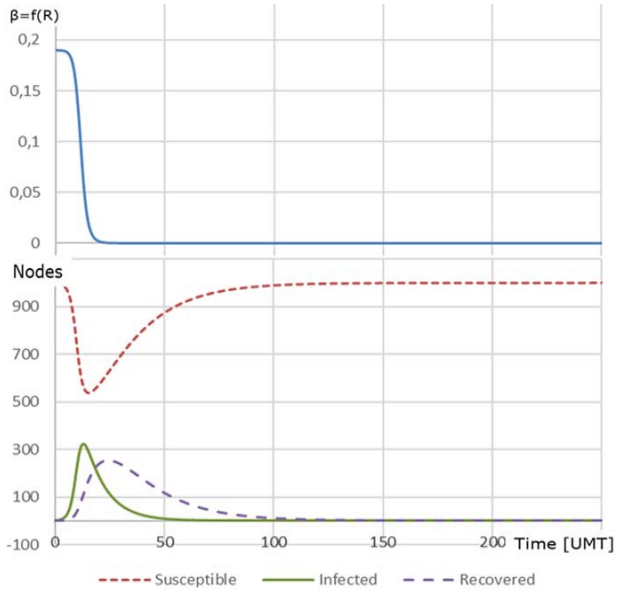

Figure 10. SIRS model with feedback. Above is given the contagion intensity variable β=f(R) at α=1 and the contagion spread states S, I, and R. There are no infected hosts at the end of the simulation

3.6. Model with Feedback ω=0.01, and Variable Intensity β and α=0.5

Fig. 11 shows a graphical representation of the model in Fig. 10 with a smaller value of at α=0.5. At the initial moment, many hosts are infected, but after some time the antivirus system detects the threat. The end result is that the number of infected hosts decreases and approaches zero.

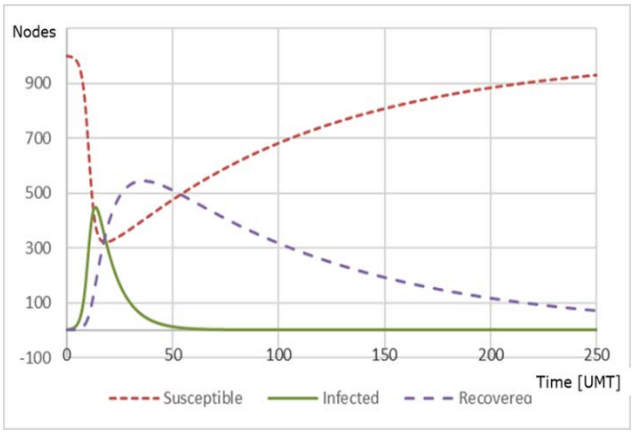

Figure 11. SIRS model with feedback ω=0.01, the variable infection rate β=f(R) at α=1. Although more than half of the hosts get infected, there are no infected hosts after a while

3.7. SIRS Models with Fluctuations

When the infection detection is weak (low values of α<0.1) and the intensity of transitions from state R to state S is high, ω>0.1, temporal fluctuations in the model are observed (Fig. 12), which are explained by the fact that some hosts become reinfected.

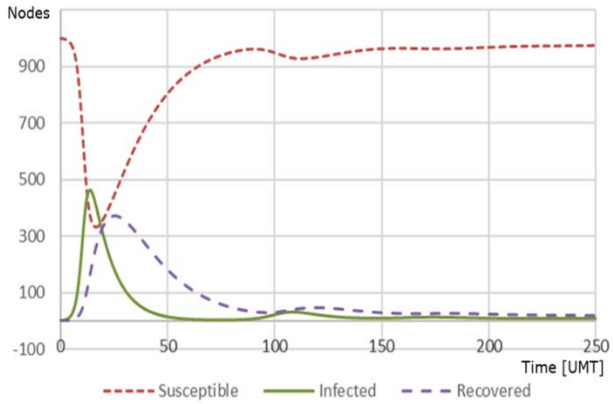

Figure 12. Temporal fluctuations of SIRS at high values of ω>0.1 and low values of α<0.1. The explanation is that some of the hosts reinfect again

4. Conclusion

The introduction of epidemic models of malware spread with variable infection intensity ratios makes it possible to model not only the spread of infection but also the performance of the antivirus system.

In contrast to most models based on constant coefficients, here the dynamics of the infection process, the latent phase of the malware, its detection by the antivirus system and the subsequent eradication can be seen.

The simulations performed show the impact of the detection time of a threat and the percentage of infected hosts susceptible to it. In this way, one can qualitatively and quantitatively assess the spread of malware in a computer network and the effectiveness of the antivirus system.

In the real conditions, a computer network is subjected to a continuous stream of different threats. To model such a case, a model of each threat is made separately, and the superposition (overlay) of the partial states of the individual models gives the overall picture of the system infection. Of course, the arrival time of the individual threats must be taken into account, which is most often modeled as a Poisson flow.

The resulting analytical dependencies and the associated graphical results are confirmed by practice, indicating the applicability of the models.

In the scientific literature, there are numerous models for malware spread using constant infection intensity (SEI, SIR, SEIRS, SEIQRS, etc.), which in turn can be easily recast into models with variable infection intensity and thus yield more accurate results for malware spread and eradication.

Acknowledgements

This research is conducted and funded in relation to the execution of a scientific-research project No KII-06-H35/12 „An Innovative Approach in Developing an Intelligent Information System for Detection and Prevention of Financial and Customs Fraud“ under the contract with National Science Fund in Bulgaria.

References:

[1]. Knight, J. C. (2002). Safety critical systems: challenges and directions. In *Proceedings of the 24th international conference on software engineering*, 547-550.

[2]. Sommerville, I. (2011). *Software engineering* (9th ed.). America: Pearson Education Inc.

[3]. Maurya, A., & Kumar, D. (2020). Reliability of safety-critical systems: A state-of-the-art review. *Quality and Reliability Engineering International*, *36*(7), 2547-2568.

[4]. Winther, R., Johnsen, O. A., & Gran, B. A. (2001). Security assessments of safety critical systems using HAZOPs. In *International Conference on Computer Safety, Reliability, and Security*, 14-24. Berlin, Heidelberg: Springer Berlin Heidelberg.

[5]. Troubitsyna, E., Laibinis, L., Pereverzeva, I., Kuismin, T., Ilic, D., & Latvala, T. (2016). Towards security-explicit formal modelling of safety-critical systems. In *Computer Safety, Reliability, and Security: 35th International Conference, SAFECOMP 2016, Trondheim, Norway, September 21-23, 2016, Proceedings 35*, 213-225. Springer International Publishing.

[6]. Piqueira, J. R. C., & Araujo, V. O. (2009). A modified epidemiological model for computer viruses. *Applied Mathematics and Computation*, *213*(2), 355-360.

[7]. Mishra, B. K., & Jha, N. (2010). SEIQRS model for the transmission of malicious objects in computer network. *Applied Mathematical Modelling*, *34*(3), 710-715.

[8]. Kermack, W. O., & McKendrick, A. G. (1932). Contributions to the mathematical theory of epidemics. II.—The problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, *138*(834), 55-83.

[9]. Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability* , *4*, 149-165. Berkeley: University of California Press.

[10]. Vynnycky, E., & White, R. (2010). *An introduction to infectious disease modelling*. OUP oxford.

[11]. Zou, C. C., Gong, W., Towsley, D., & Gao, L. (2005). The monitoring and early detection of internet worms. *IEEE/ACM Transactions on networking*, *13*(5), 961-974.

[12]. Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. *Journal of the royal society interface*, *2*(4), 295-307.

[13]. Thommes, R. W., & Coates, M. J. (2005). Modeling virus propagation in peer-to-peer networks. In *2005 5th International Conference on Information Communications & Signal Processing* , 981-985. IEEE.

[14]. Yang, F., & Zhang, Z. (2021). Hopf bifurcation analysis of SEIR-KS computer virus spreading model with two-delay. *Results in Physics*, *24*, 104090.

[15]. Özdemir, N., Uçar, S., & Billur Eroğlu, B. (2020). Dynamical analysis of fractional order model for computer virus propagation with kill signals. *International Journal of Nonlinear Sciences and Numerical Simulation*, *21*, 239-247.

[16]. Yuan, H., & Chen, G. (2008). Network virus-epidemic model with the point-to-group information propagation. *Applied Mathematics and Computation*, *206*(1), 357-367.

[17]. Oberheide, J., Cooke, E., & Jahanian, F. (2008). CloudAV: N-Version Antivirus in the Network Cloud. In *USENIX Security Symposium*, 91-106.

[18]. Mishra, B. K., & Saini, D. (2007). Mathematical models on computer viruses. *Applied Mathematics and Computation*, *187*(2), 929-936.

[19]. Anceschi, F., & Polidoro, S. (2019). A survey on the classical theory for Kolmogorov equation. *arXiv preprint arXiv:1907.05155*.

[20]. Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. *Towards Data Sci*, *6*(12), 310-316.

[21]. Wang, Y., Li, Y., Song, Y., & Rong, X. (2020). The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences*, *10*(5), 1897.

[22]. Popov, G., & Nakov, O. (2021). An epidemic model of COVID-19 disease with variable spreading. In *AIP Conference Proceedings*, *2333*. AIP Publishing.