Stochastic model of information - computer network blocking in generation and transmission of malicious big data

Sergey Lesko1,*, Dmitry Zhukov2, Anton Aleshkin2

1MIREA - Russian Technological University, Department of management and systems modelling, 119454, 78 Vernadsky Prospect, Moscow, Russia

2 MIREA - Russian Technological University, Department of informational confrontation, 119454, 78 Vernadsky Prospect, Moscow, Russia

Abstract. This work reviews a stochastic model of information - computer networks blocking in generation and transmission of large volumes of malicious data, developed by the authors. Rapid development of the Internet of things (IoT) leads to short time emergence of tens and hundreds of millions of new devices with their IP – addresses in the global network, capable to form a very large -scale botnets in case of infection. Therefore, development of network blocking models when generating and transmitting large amounts of malicious data is a crucial task to solve the problem of traffic filtering and balancing. When creating the model, we considered probability diagrams of transitions between possible states (share of infected devices) of computer networks, which describe the logic of the on-going processes. Based on approach used, the nonlinear second-order differential equation was deduced, allowing to formulate and solve boundary problems for determination of time-dependence of the probability density function for observation of different system states. The resulting differential equation contains the second and first time - derivatives, and the derivatives with respect to a variable, describing the change in the state of the system under consideration. Considering the second time - derivative of the probability density corresponds to the case in which the existing states generate additional new states, causing acceleration of processes and self-organization. The created model allows to assess time to achieve the state, when there is a limit value of share of potentially infected devices or, for example, achievement of its percolation threshold in a computer network. Percolation threshold is the minimum percentage of blocked nodes, at which the entire network loses the information transmission features (there is no free path between any randomly selected nodes). Using the approach adopted in the theory of percolation allowed to link structural and information characteristics of networks, such as dependence of their percolation threshold on the average number of links per node (network density) with dynamic characteristics of their blocking (time to reach the percolation threshold).

* Corresponding author: sergey@testor.ru
1 Introduction

One of the problems, arising when working with big data, is the need to process huge traffic and to load balance information -computer networks in real time. An important task is the necessity to filter useful and harmful traffic. An example of such harmful traffic can be, for example, spamming, viruses, DDoS attacks, etc. In general, all this can be described as computer virus attacks resulting in malfunction or other negative consequences. Given that the Internet of things (IoT) is rapidly developing currently and there are tens and hundreds of millions of new devices with IP addresses, appearing in the global network in a short time, which are capable to form a very large-scale botnet network in the event of infection, the problem of harmful traffic filtering becomes very relevant.

Nowadays studying of threat spreading in computer networks and their protection is mainly the field of computer security, but in our opinion, this problem is interdisciplinary, and the study of analysis methods, as well as the development of new models for describing the spread and filtering of harmful traffic should be done from the point of view of operating with big data, i.e. this is one of the areas of Big Date. One of the tasks to be solved in this area is the development of new theoretical models of generation and distribution of any harmful traffic in computer networks (hereinafter we will use the term viruses to be short, although it may be, for example, spam, as was mentioned above).

Currently, there are many mathematical models [1-11], created to study and simulate propagation of computer viruses based on different approaches, taking into account virus behavior algorithms, network topology, possible node states, effect of protection, diversity of viruses in vulnerability, polymorphism to overcome protection using signature analysis, and more. Comparison of different mathematical models of virus propagation is summarized in table 1.

| Worm propagation models          | Network topology | Topology graphical format | Modelling method | Propagation process | Model type   |
|----------------------------------|------------------|---------------------------|-----------------|---------------------|--------------|
| Classical simple epidemic model  | H                | UG                        | A               | C                   | SI           |
| Worm model with sustainable scanning | H                | UG                        | A               | C                   | SI           |
| RCS model                        | H                | UG                        | A               | C                   | SI           |
| Classical general epidemic model | H                | UG                        | A               | C                   | SIR          |
| Two-factor model                 | H                | UG                        | A               | C                   | SIR          |
| AAWP model                       | H                | UG                        | A               | D                   | SIR          |
| Local preference model           | Non-H            | UG                        | A               | C                   | SI           |
| LAAWP model                      | Non-H            | UG                        | A               | D                   | SIR          |
| Mail worm simulant               | R/SW/PL          | UG                        | S               | D                   | SI           |
| Logic matrix model 0-1           | R/PL             | DG                        | A               | D                   | SIR          |
| OSN worm model                   | PL               | UG                        | S               | D                   | SI           |
| Spatiotemporal model             | H/PL             | DG                        | A               | D                   | SIS          |

Notes to symbols in table: H – homogenous hashing; R – random network; SW – small world network; PL – power-line network; UG – unoriented graph; DG – oriented graph; C – continuous event; D – discrete-time event; A - analytical; S - simulation; SI – infection-sensitive model; SIR – sensitive – infected – resistant model; SIS – sensitive-infected – susceptible.
We discuss some of the existing models in more detail. The SI – model of virus propagation means that any of computers included in the attacked network can be in one of two states: susceptible (S) and infected (I). According to this model, there is a network consisting of a constant number of (N) computers, with N=S+I, and there can be only one copy of the worm on each infected node, which randomly selects a potential victim in the available address space with some constant average attack rate per unit time. In the SIR model, network nodes exist in three states: susceptible (S), infected (I), and resistant (R). Note that the nodes are unsusceptible recovered of infection only, and N-the total number of nodes is S+I+R. Introducing a constant average rate of immunization and attacks per unit time to describe dynamics of epidemics, you get a system of differential kinetic equations [6, 12,13], describing the process of virus epidemic spread.

The paper [14] reviews a kinetic model, describing viral epidemics in computer networks based on epidemiologic threshold concepts, latency of infection, replication factor (multiplication factor), probability of infection and immunization, time of node immunity, etc. The paper [15] improves mathematical models of computer viruses spread in a heterogeneous computer network based on its topological and architectural features. Generalized structure of computer networks was considered on the basis of the PSIDR model: N = S(t) + I(t) + D(t) + R(t), where N is the total number of objects in the system; S(t) – the number of susceptible objects; I(t) – number of infected objects; R(t) –number of cured objects with immunity; D(t) –number of objects with virus detected. Topological and architectural features of the networks were taken into account by multiplying some terms of kinetic differential equations by empirical correction coefficients. In particular, for the topology of the “star” network, the member, accounting for the loss (infection) of susceptible objects was multiplied by a coefficient equal to 0.6.

General issues of virus epidemics in computer networks were considered in works [16, 17]. In particular, [16] specifies the need to develop protection strategies, invulnerable to changes in network topology and requiring no knowledge of the epidemic evolution mechanisms. For example, development of mechanisms to adjust the number of connections between nodes per time unit and limit them in the event of attacks or development of preventive vaccination methods. The article [17] discusses elaboration of countermeasures to prevent the spread of viruses. The authors argue that release of software updates after detection of vulnerabilities does not provide a reliable security guarantee. To increase the level of protection, the authors propose the idea that it is necessary to allocate a subnet in a computer network, in which the antivirus will be distributed purposefully, targeted at virology.

Paper [18] analyzes four models of virus propagation: the classical SI model, the independent cascade model, the dynamic distribution model and the model considering the topology of networks. Comparison of the simulation results showed that from the point of view of protection mechanisms development the most promising are models based on the network graph description.

The paper [19] discusses the development model of viral epidemic not with random virus spread order but taking into account the error of the attack results due to the impact of viruses on the already infected network nodes. To do this, the authors present the network as a directed probability graph without loops, the nodes of which are described by variables that specify the probabilities of their state (infected, immunized, susceptible) and the arcs specify the interaction between the variables of the graphical model. Viral propagation is determined by network characteristics and is similar to the action of a cellular automaton.

Paper [20] reviews a model to describe the development of virus epidemics based on stochastic models of interactive Markov chains where the state of network nodes at each consequent step of the epidemic depends on its condition and the state of neighbors in the previous step, and the network itself is represented as undirected graph.
Matching techniques can be used to analyze and model virus epidemics in computer networks. Two models are described in the paper [21]: one is based on autoregressive analysis and the other based on Fourier analysis. The results of the analysis show an acceptable correlation between the time of virus spread. Autoregressive and Fourier analysis make possible to predict the increase and decrease in trends in the spread of a certain type of virus using the experience gained from other epidemics. 

The advantage of the existing and described above models is that they describe well the dynamics of viruses such as Code Red I [22], Code Red II, and Nimda [23], the spread of which was observed in the early 2000s. The result of use of these models to describe the dynamics of virus propagation are typical S-curves, the form of which coincides well with the observed data, presented in figure 1.

![Fig. 1. Dynamics of malicious traffic generated by Code Red I [22] and Nimda [23] viruses.](image)

However, if we discuss the dynamics of later incidents, for example, Sapphire/Slammer (January 2003) [24] or Conficker (November 2008) [9], it has a more complex step character (see figure 2) with the presence of significant oscillations, which is significantly different from the data presented in figure 2, and cannot be explained within the existing models.

It should be noted that unfortunately, different studies use non-standardized methods of collecting and reporting data on changes in dynamics (see figures 1 and 2) of harmful traffic generated in information-computer networks as a result of viral activity. The second problem is that despite the large number of publications on viral activity, it is very difficult to find properly measured data. Nevertheless, the available observations allow us to assume the general nature of the occurring processes and give an approximate estimate of their quantitative characteristics.
Fig. 2. Dynamics of malicious traffic generated by Sapphire/Slammer [24] and Conficker [9] viruses.

With no detailed discussion of the virus behavior mechanisms (vulnerabilities they scan and attack methods, as this is not the topic of the presented work) of Code Red I, Code Red II, Nimda, Slammer and Conficker we give only some necessary explanations to figure 2. Solid lines in figure 2 show trends in the dynamics of computer virus attacks. A characteristic feature of the observed data is that during the spread of Conficker, as well as, for example, the Nimda virus (see figure 1), there are significant fluctuations relative to the solid line showing the general trend. Harmful traffic of network worms Slammer and Conficker was measured using network telescopes. In particular, for tracking of scan infected device TCP/445 we used network telescope UCDS covering a large sample of monitored devices (curve I in the right part of figure 2) and sampling of lower dimension (curve II on the right side of figure 2). Both studies showed similar results in observed dynamics. The data gaps in figure 2 are due to several technical failures in the process of data collection. However, figure 2 suggests that the dynamics of incidents, related to Sapphire/Slammer and Conficker is complex in nature (see figure 2) with significant oscillations compared to Code Red and Nimda (see figure 1).

Describing topology of network nodes blocking while spread of viruses there is currently a prevailing approach, according to which the development of the epidemic is represented as a process, reminiscent of the structure of the Cayley tree with a random number of links [25]. Attention should be paid to work [26] in which the problem of determination of node infection probability is reviewed, depending on remoteness of a node from a source of infection in networks with various scale and number of nodes. The topological parameters here were the scale and number of nodes, but the diversity of the network structures was not studied in these works.

Obviously, in case there are not many blocked nodes, at least one “open” path (a path consisting of non-blocked nodes) will exist between two randomly selected non-nearby nodes. We call as the percolation threshold the share of blocked nodes, at which the network comes out of action, below its value the network is healthy, even though it has some nodes or their groups (clusters) blocked by viruses. Above the percolation threshold, the entire network shuts down and loses data throughput. Note that saturation of the traffic curve during viral activity can be considered as a percolation threshold as well, when virus activity on scanning the network and sending packets reaches the limit of the horizontal section. This may be due to reaching the maximum possible number of devices on the network that have this vulnerability or a set of them to be infected.
It should be noted that there are many papers, describing the study of network structures percolation properties [27-35]. However, few people have studied the relationship between the structural properties of networks and the dynamics of their blocking.

The study of harmful traffic dynamics in networks with different (including random) topology is one of the directions in the field of processing and transmission of big data, and is of great scientific and practical interest for development of methods of its filtering in order to balance the computer networks loads and ensure their performance.

2 Percolation properties of network structures

Percolation theory (probability theory in graphs) studies the solution to the problem of nodes and communication tasks for different network of regular (2D structures – triangular, hexagonal, Cayley tree, etc. and 3D – hexagonal, cubic, etc.) and random structure. When solving the communication task, they determine the share of links that need to be broken, so that the network is divided into at least two unrelated parts. The problem of nodes determines the part of blocked nodes at which the network breaks into unconnected clusters, storing connection inside themselves (or vice versa, the proportion of transmitting nodes, when conductivity occurs). The share of blocked nodes (in the node problem) or broken links (in the link problem), in which the conductivity between two randomly selected nodes of the network is lost, is called the percolation threshold (flow threshold).

Use of the term of the blocked nodes or links share is equivalent to the notion of the probability of finding a randomly selected node (or link) to be in a blocked (broken) state. Therefore, the value of the percolation threshold determines the probability of information transmission through the whole network if some of its nodes (or links) are blocked (excluded), i.e. the average probability of blocking a node (disconnection) is set.

In works [34, 35] we performed numerical simulation of dependence of the random networks percolation thresholds on the average number of connections per node (density) of the network. The results obtained in these papers for the problem of node blocking, for small network densities, show that for random structures, the dependence of their natural logarithm \( \ln P(x) \) on the inverse value of the network density \((1/x)\) can be described by a linear equation:

\[
\ln P(x) = \frac{4,02}{x} - 2,26
\]

with the value of the correlation coefficient of numerical data and the equation of linear dependence, equal to 0,97.

This dependence can be used to calculate the percolation thresholds of networks by their density values. Further, using dynamic models, we can determine the time of its achievement, and the network exit from a healthy state.

3 Stochastic model of host blocking and time to reach percolation or saturation threshold

Suppose that at some point of time \( t \) the proportion (or number) of blocked (due to congestion or virus infection) nodes of the data network equals some value \( x_i \) (which we call the state of the network).
The state observed at time $t$ can be described as $x_i$ ($x_i \in X$). In addition, we introduce the time interval $\tau_0$, during which the state $x_i$ can change. In this case, any value of the current time $t=h \cdot \tau_0$, where $h$ is the step number of the state transition (transition between states becomes quasi-continuous with infinitely small time interval $\tau_0$, $h=0,1,2,3,...,N$. Current state $x_i$ at step $h$, after transition to step $h+1$ may increase by a certain amount $\varepsilon$ (infection), or decrease by some amount $\xi$ (cure), and respectively, to be equal to $x_i+\varepsilon$, or $x_i-\xi$. In the simplest case, $\varepsilon$ and $\xi$ are some constant values for any step $h$.

We introduce the concept of probability to find a system in exact state. Assume that after some number of steps $h$, we can say about the described system that:

$P(x-\varepsilon,h)$ – probability that the system is in state $(x-\varepsilon)$;

$P(x,h)$ – probability that the system is in state $x$;

$P(x+\xi,h)$ – probability that the system is in state $(x+\xi)$.

After each step the state $x_i$ (hereinafter index $i$ can be disregarded to be short), can change for the value of $\varepsilon$ or $\xi$.

Probability $P(x, h+1)$ – that in the next ($h+1$) step the system (or the process) will be in state $x$, will equal to (see Figure 3):

$$P(x, h+1) = P(x-\varepsilon, h) + P(x+\xi, h) - P(x, h)$$

(2)

Fig. 3. Diagram of possible transitions between system (process) states at $h+1$ step.

Here is explanation of equation (2) and the diagram presented in figure 3. Probability of transfer to state $x$ from step $h$, is $P(x,h+1)$, which is defined by sum of probabilities of transfers to this state from states $(x-\varepsilon)$, that is $P(x-\varepsilon,h)$ and $(x+\xi)$, that is $P(x+\xi,h)$ in which the system was at step $h$ minus probability of transfer ($P(x,h)$) of the system from the state $x$ (in which it was at step $h$) to any other state at step $h+1$. We consider that transfers themselves are done with probability equal to 1.

Considering, that $t=h \cdot \tau$, where $t$ – time of process, $h$ – step number, $\tau$ – continuity of one step, we move from $h$ to $t$. Take the equation (2) in Tailor series nearby point $x$. Passing from probabilities to probability density ($\rho(x,t)=dP(x,t)/dx$) and taking into account maximum the second derivative by $x$ and the first by time $t$, we get (3):

$$\frac{d\rho(x,t)}{dt} = a \frac{d^2 \rho(x,t)}{dx^2} - b \frac{d\rho(x,t)}{dx}$$

where $a = \frac{\varepsilon^2 + \xi^2}{2\tau}$ and $b = \frac{\varepsilon - \xi}{\tau}$

(3)

Equation term of type $-d\rho(x,t)/dt$ – determines the total change of system state or the process in the course of time;
Equation term of type \( \frac{d\rho(x,t)}{dx} \) – describes ordered transition either to the state, when it increases \((E > \xi)\), or decreases \((E < \xi)\);

Equation term of type \( \frac{d^2\rho(x,t)}{dx^2} \) – describes stochastic change of state.

We formulate and solve a boundary value problem to describe the operation of the network, considering its percolation properties or the limiting proportion of nodes that can be infected. When the number of blocked nodes in the network is equal to the number of all available devices \( x=L=1 \) (100%), it stops operation fully. As the share cannot exceed 1, the condition should be met: \( \rho (x, t)_{x=L}=0 \).

State \( x = 0 \) means, that the network has no blocked nodes, thus, taking into account, that the number of blocked nodes cannot outcome to the sphere of negative meanings, with \( x=0 \) we should use reflection condition of type:

\[
\rho(x, t)_{x=0} = 0.
\]

As at the start moment of time \( t=0 \) there can be some number \( x_0 \) of blocked nodes in the system already, the initial condition is set in form:

\[
\rho(x, t = 0) = \delta(x - x_0) = \begin{cases} 
\int \delta(x - x_0) dx = 1, & x = x_0 \\
0, & x \neq x_0
\end{cases}
\]

As the initial condition contains delta function, solution for \( \rho(x,t) \) is divided into 2 areas with \( x>x_0 \) and \( x\leq x_0 \). Using methods of operator calculus for probability density \( \rho_1(x, t) \) and \( \rho_2(x, t) \) of system state detection in one of the values in the interval from 0 to \( L \) we get the below equation system:

with \( x > x_0 \)

\[
\rho_1(x, t) = 2 \frac{e}{e^{(\lambda-t)/\xi} + \xi^2} \sum_{n=1}^{M} \frac{\sin(\pi n x_0/L) \sin(\pi n L-x/L)}{\cos(\pi n)} e^{-\pi^2 n^2 \pi^2 \xi^2 / 2}.
\]

with \( x \leq x_0 \)

\[
\rho_2(x, t) = 2 \frac{e}{e^{(\lambda-t)/\xi} + \xi^2} \sum_{n=1}^{M} \frac{\sin(\pi n x_0/L) \sin(\pi n L-x_0/L)}{\cos(\pi n)} e^{-\pi^2 n^2 \pi^2 \xi^2 / 2}.
\]

If we calculate the primitive function \( P(L, t) \):

\[
P(\lambda, t) = \int_{0}^{\lambda_0} \rho_2(x, t) dx + \int_{\lambda_0}^{\lambda} \rho_1(x, t) dx
\]

function \( P(\lambda, t) \) will define probability that system state by the time moment \( t \) will be in the interval from 0 to \( \lambda \), that is the percolation threshold value, equal to \( \lambda \), will not be achieved.

Note that traffic saturation during viral activity can be also considered as percolation threshold, when viral activity on scanning the network and sending packets reaches the limit of stationary value. This may be due to reaching the maximum possible number of devices on the network for infection, that have this vulnerability or a set of them.
Respectively, probability $Q(t)$ of the percolation threshold $\lambda$ to be reached by the time moment $t$ can be defined as follows:

$$Q(\lambda, t) = 1 - P(\lambda, t)$$

Let us analyze the result. Take the arbitrary value $x_0$, $\varepsilon$ and $\xi$ ($\varepsilon > \xi$). To model the process, we consider the initial share of infected devices $x_0$, for example, to be 2% at the start of activity observation (refer to Figure 4) from all near susceptible, the $\tau$ value equal to one conditional unit of time ($\tau = 1$), $\varepsilon = 0,03$ (3%) and $\xi = 0,01$ (1%). Curve 1 in figure 4 is drawn for dynamics of percolation threshold achievement (or possible boundary share of infected devices) equal to 0,10; curve 2 for 0,12; curve 3 for 0,14; curve 4 for 0,16; curve 5 for 0,18 values. Note that the share of potentially susceptible nodes is not large and can hardly exceed some percent.

Percolation threshold can be determined by equation (1) starting from network topology and density of its connections (number of links per one node).

As the share of potentially susceptible devices increases, the dynamics slows down and the time to reach a stationary state (in which all potentially susceptible devices are infected) increases (see curves 2 – 5 in figure 4). A specific feature of the dynamics with a large proportion of potentially susceptible devices is the possibility of slowing down the process after the initial growth and even a slight decrease in the share of infected devices, and then a new rapid growth (see curve 5 in figure 4). Such dynamics are not represented on the observed virus propagation curves (see Fig. figure 1), however, several researchers said that after the initial growth, there is some slowdown, and then a new increase in the activity of viruses.

Theoretical calculations received from analysis of the model (see figure 4) are well within the observed data presented in figure 1 for the dynamics of malicious traffic generated by Code Red I and Nimda viruses (curve 1 in figure 4). However, in this case, for small values of share of potentially vulnerable network devices, there is no double transition, as for malicious traffic generated by the Sapphire/Slammer and Conficker viruses (see figure 2).
Let us add the model we suggest. For this take equation (2) and after decomposition of its terms in Tailor series we account for all second and first derivatives, arriving at equation (6).

$$\frac{d^2 P(x,t)}{dt^2} = a \frac{d^2 P(x,t)}{dx^2} - b \frac{dP(x,t)}{dx} - c \frac{d^2 P(x,t)}{dx^2}, \quad \text{where} \quad a = \frac{\tau^2 + \xi^2}{\gamma \tau}, \quad b = \frac{\xi}{\tau}, \quad c = \tau$$  \hspace{1cm} (6)

Consideration in differential model of equation term of $\frac{d^2 P(x,t)}{dt^2}$ type allows for description of situation with increase of virus spread due to self-organization, resulting from virus polymorphism, that is their ability to modify the initial code.

Since the differential equation (6) contains the second time derivative, the second initial condition is necessary. It is not as obvious as the first one, but in this case, we can use the continuity of the function for any point in time. Available $\delta$ – function for any point in time. Available $\delta$ – function gives rise to the solution, being steady at point $x=x_0$, experiences the derivative discontinuity herein, but the second initial condition can be set in form $\frac{\partial G(x,t)}{\partial t} \bigg|_{t=0} = 0$. As the initial condition contains delta function, the solution for $\rho(x, t)$ is broken into two areas with $x>x_0$ and with $x\leq x_0$. Using methods of operator calculus for probability density $\rho(x, t)$ and $\rho_0(x, t)$ of system state detection in one of the values in the interval from 0 to $L$ we get the below equation system:

With $x \geq x_0$
Let us analyze the obtained model. To model the process, we consider the initial share of infected devices \( x_0 \), equal to 2% (refer to Figure 5) from all potentially subsessile, the \( \tau \) value equal to one conditional unit of time (\( \tau = 1 \)), \( \varepsilon = 0.03 \) (3%) and \( \xi = 0.01 \) (1%). Curve 1 in figure 5 is drawn for percolation threshold (or possible boundary share of infected devices) equal to 0.10; curve 2 for 0.12; curve 3 for 0.14; curve 4 for 0.16; curve 5 for 0.18 value. Note that the share of potentially susceptible nodes is not large and can hardly exceed some percent.

The results of analysis of the model presented in figure 5 show the possibility of two stages of growth of viral activity growth and the presence of oscillations.

\[
\rho_1(x, t) = -\frac{2}{L} e^{-\frac{t}{\tau}} e^{k(x-x_0)} \sum_{n=1}^{\infty} \frac{\sin\left(\frac{\pi n x_0}{L}\right) \sin\left(\frac{\pi n L-x}{L}\right)}{\cos(\pi n)} \left(1 + \frac{\varepsilon \frac{\sin^2(\frac{\pi n x_0}{L})}{\cos(\pi n)}}{\tau} \right)^{\frac{2}{\alpha^2} \mathcal{N}^2}
\]

With \( x < x_0 \)

\[
\rho_2(x, t) = -\frac{2}{L} e^{-\frac{t}{\tau}} e^{k(x-x_0)} \sum_{n=1}^{\infty} \frac{\sin\left(\frac{\pi n L-x_0}{L}\right) \sin\left(\frac{\pi n L}{L}\right)}{\cos(\pi n)} \left(1 + \frac{\varepsilon \frac{\sin^2(\frac{\pi n x_0}{L})}{\cos(\pi n)}}{\tau} \right)^{\frac{2}{\alpha^2} \mathcal{N}^2}
\]

Where \( \frac{k}{\alpha} = \frac{(\varepsilon - \xi)}{(\varepsilon^2 + \xi^2)} \)

Let us analyze the obtained model. To model the process, we consider the initial share of infected devices \( x_0 \), equal to 2% (refer to Figure 5) from all potentially subsessile, the \( \tau \) value equal to one conditional unit of time (\( \tau = 1 \)), \( \varepsilon = 0.03 \) (3%) and \( \xi = 0.01 \) (1%). Curve 1 in figure 5 is drawn for percolation threshold (or possible boundary share of infected devices) equal to 0.10; curve 2 for 0.12; curve 3 for 0.14; curve 4 for 0.16; curve 5 for 0.18 value. Note that the share of potentially susceptible nodes is not large and can hardly exceed some percent.

The results of analysis of the model presented in figure 5 show the possibility of two stages of growth of viral activity growth and the presence of oscillations.

![Fig. 5. Time dependence of probability value to reach the set threshold of device susceptibility to infection with regard of self-organization.](image)

It should be noted that in a detailed analysis of the Conficker worm activity, oscillations were also observed in the horizontal areas represented on the right side of figure 2 (see figure 6).
Fig. 6. Conficker scanning of unique IP-addresses by TCP/445 (daily scanning pattern with peaks at 14:00 is seen at increase).

For tracking of infected device scanning TCP/445 ports, curve I in figure 6 was obtained, using a network telescope UCDS covering a large sample of monitored devices, and curve II in figure 6, when scanning the sample with smaller dimension. Both studies showed similar results on the observed dynamics and qualitatively coincide with the theoretical calculations in figure 5 (on the horizontal sections of curves 1-5, oscillations are also observed, but on a smaller scale than in figure 6).

4 Summary and conclusions

Rapid development of the Internet of things (IoT) leads to short time emergence of tens and hundreds of millions of new devices with their IP – addresses in the global network, capable to form a very large -scale botnets in case of infection. Therefore, development of network blocking models when generating and transmitting large amounts of malicious data is a crucial task to solve the problem of traffic filtering and balancing.

There are many models that describe well the dynamics of such viruses as Code Red I, Code Red II and Nimda, the spread of which was observed in the early 2000s. As a result of using these models to describe the dynamics of virus propagation, we obtain the typical S – curves, the form of which coincides well with the observed data.

However, emergence of new generations of viruses, for example, such as Sapphire/Slammer and Conficker has led to the fact that dynamics of their spread is already of a two-step character with oscillations on horizontal sections. This is due to significant difference in behavior of Sapphire/Slammer and Conficker viruses unlike Code Red I, Code Red II and Nimda, not only in use of multi-vector for vulnerabilities to infiltrate onto the infected device, but the use of polymorphism to hide from signature-based analysis by antivirus programs. Thus, Sapphire/Slammer and Conficker are already evolutionary viruses, having elements of self-organization in their behavior.
We developed a model consider to the possibility of self-organization in virus spread, which considers the probability diagrams of transitions between possible states of computer networks (the share of infected devices), which describe the logic of the ongoing processes. This allows us to derive a nonlinear differential equation of the second order, as well as to formulate and solve on its basis boundary value problems to determine the time dependence of probability density functions for observation of certain system states.

The resulting differential equation contains the second and the first -time derivatives, and variable derivatives, describing change in the state of the system under consideration. Considering the second derivative of the probability density in time corresponds to the case in which existing states generate additional new states, which leads to acceleration of processes and self-organization.

If we ignore the second time derivative in the equation, we get a model that is more complex than the existing models and allows us to describe not only the S - shaped dynamics of virus propagation in computer networks. As the proportion of potentially vulnerable devices increases, the dynamics slows down and the time to reach a stationary state (in which all potentially vulnerable devices are infected) increases. A characteristic feature of the dynamics described by this model (with a large proportion of potentially vulnerable devices) is the possibility of slowing down the process after the initial growth and even a slight decrease in the proportion of infected devices, and then a new rapid growth.

When the second time derivative is taken into account in the initial differential equation, a model is obtained that shows the possibility of two stages of viral activity growth and the presence of oscillations in the horizontal section, which coincides fully with the observed dynamics of the Conficker computer virus spread.

The created models allow to estimate time to reach the condition, at which the limit value of a share of potentially infected devices or, for example, a threshold of its percolation is achieved in a computer network. Percolation threshold is the minimum percentage of blocked nodes, at which the entire network as a whole loses the properties of information transmission (there is no free path between any randomly selected nodes). Using the approach adopted in the theory of percolation allowed to link structural and information characteristics of the networks, such as dependence of their percolation threshold on the average number of links per node (network density) with the dynamic characteristics of their blocking (time to reach the percolation threshold).

Comparison of theoretical calculations and analysis of models with real observed dynamics of such computer viruses as Code Red, Nimda, Sapphire/Slammer and Conficker showed that the created models describe them very well.

The results presented in this paper were obtained within the programme of choosing scientific projects performed by scientific laboratories teams, according to the letter of the Ministry of Education and Science of Russia, 3 July 2019 No. МН-1037/АМ. Project 2019-1393 "Developing new models for analyzing and forecasting the dynamics of stochastic processes in sophisticated systems considering self-organisation and the presence of memory ".

References
1. S. Staniford et al. How to Own the Internet in Your Spare Time. *USENIX security symposium*. Vol. 2, C. 14-15 (2002)

2. Z. Chen, L. Gao, K. Kwiat. Modeling the spread of active worms. *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies* (IEEE Cat. No. 03CH37428), IEEE, Vol. 3, C. 1890-1900 (2003)

3. G. Yan, S. Eidenbenz. Modeling propagation dynamics of bluetooth worms (extended version). *IEEE transactions on mobile computing*. Vol. 8, №. 3, C. 353-368 (2008)

4. X. Fan, Y. Xiang Modeling the propagation of Peer-to-Peer worms. *Future generation computer systems*, Vol. 26, №. 8, C. 1433-1443 (2010)

5. C. C. Zou, D. Towsley, W. Gong Modeling and simulation study of the propagation and defense of internet e-mail worms. *IEEE Transactions on dependable and secure computing*, Vol. 4, №. 2, C. 105-118 (2007)

6. H. Andersson, T. Britton Stochastic epidemic models and their statistical analysis. *Springer Science & Business Media*, Vol. 151 (2012)

7. Z. Chen, C. Ji Importance-scanning worm using vulnerable-host distribution. *GLOBECOM'05. IEEE Global Telecommunications Conference*, Vol. 3, pp. 6 (2005)

8. Z. Chen, C. Ji Optimal worm-scanning method using vulnerable-host distributions. *IJJSN*, Vol. 2, №. 1/2, pp. 71-80 (2007)

9. Z. Chen, C. Chen A closed-form expression for static worm-scanning strategies. 2008 *IEEE International Conference on Communications*, pp. 1573-1577 (2008)

10. Z. Chen, C. Chen, C. Ji Understanding localized-scanning worms. 2007 *IEEE International Performance, Computing, and Communications Conference*, pp. 186-193 (2007)

11. Q. Wang, Z. Chen, C. Chen, N. Pissinou. On the robustness of the botnet topology formed by worm infection. *IEEE Global Telecommunications Conference GLOBECOM*, pp. 1-6 (2010)

12. Benjamin M. Bolker David J. D. Earn, Pejman Rohani, Bryan T. Grenfell. A simple model for complex dynamical transitions in epidemics. *Science*, 287:667–670 (2000)

13. C. Wang, J.C. Knight, M.C. Elder. On Viral Propagation and the Effect of Immunization. *Proc. of 16th ACM Annual Computer Applications Conference*, New Orleans, LA (2000)

14. Bimal Kumar Mishra, Gholam Mursalin Ansari. Dierential Epidemic Model of Virus and Worms in Computer Network, *International Journal of Network Security*, Vol. 14, No.3, pp. 149-155 (2012)

15. S.G. Semenov, V.V. Davydov. Matematicheskaya model rasprostraneniya kompyuternyh virusov v heterogennyh kompyuternyh setyah avtomatizirovannyh sistem upravleniya tekhnologicheskim processom (Mathematical model of computer virus propagation in heterogeneous computer networks of automated process control systems). *Vestnik NTU "HPI", № 38*, pp. 163-171 (2012)

16. Justin Balthrop, Stephanie Forrest, M. E. J. Newman, Matthew M. Williamson. Technological Networks and the Spread of Computer Viruses. *SCIENCE*, vol. 304, pp. 527-529 (2004)

17. Li-Chiou Chen, Kathleen M. Carley. The Impact of Countermeasure Propagation on the Prevalence of Computer Viruses. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, Vol. 34, No. 2, pp. 823-833 (2004)

18. Arnold Adimabua Ojugo, Fidelis Obukowho Aghware, Rume Elizabeth Yoro, Mary Oluwatoyin Yerokun, Andrew Okonji Eboka, Christiana Nneamaka Anujeonye, Fidelia Ngozi Efozia. Evolutionary Model for Virus Propagation on Networks, *Automation, Control and Intelligent Systems*, 3(4), pp. 56-62 (2015)
19. Honoriu Vălean, Adina Pop, Camelia Avram. Intelligent model for virus spreading. *The International Symposium on System Theory, Automation, Robotics, Computers, Informatics, Electronics and Instrumentation*, Craiova, Romania (2007)

20. YA.M. Dalinger, D.V. Babanin, S.M. Burkov. Matematicheskie modeli rasprostraneniya virusov v kompyuternyh setyah razlichnoj struktury (Mathematical models of virus propagation in computer networks of different structures). Modelirovanie sistem (Systems modelling), 2011, №4(30), pp.3-11 (2011)

21. Jos’e R. C. Piqueira, Felipe Barbosa Cesar. Dynamical Models for Computer Viruses Propagation. *Mathematical Problems in Engineering*, Volume 2008, Article ID 940526, 11 pages, doi:10.1155/2008/940526 (2008)

22. C.C. Zou, W. Gong, D. Towsley. Code Red Worm Propagation Modeling and Analysis. *In 9th ACM Symposium on Computer and Communication Security*, Washington DC, USA, P. 138 – 147 (2002)

23. S. Staniford, V. Paxson, N. Weaver. How to Own the Internet in Your Spare Time. *11th Usenix Security Symposium*, San Francisco, USA, August 2002, P.149-167, ISBN 1–931971–00–5 (2002)

24. D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, N. Weaver. Inside the Slammer Worm. *IEEE Security and Privacy*, 1(4):33–39, July 2003 (2003)

25. J. Nazario, "Defense and Detection Strategies against Internet Worms"; Artech House, ISBN: 1580533572 (2004)

26. R. Pastor-Satorras and A. Vespignani, "Epidemics and Immunization in Scale-Free Networks", Wiley-VCH, S. Bornholdt and H. G. Schuster (eds.) Handbook of Graphs and Networks: From the Genome to the Internet, 2005, DOI:10.1002/3527602755.ch5 (2005)

27. S. Mizutaka, T. Tanizawa Robustness analysis of bimodal networks in the whole range of degree correlation. *Physical Review E*, Volume 94, Issue 2, 17, article number 022308 (2016)

28. De Brito, J.B., Sampaio Filho, C.I.N., Moreira, A.A., Andrade, J.S. Characterizing the intrinsic correlations of scale-free networks. *International Journal of Modern Physics C*, Volume 27, Issue 3, article number 1650024 (2016)

29. P.N. Timonin Statistical mechanics of high-density bond percolation. *Physical Review E*, Volume 97, Issue 5, article number 052119 (2018)

30. A. Zhou, S. Maletić, Y. Zhao Robustness and percolation of holes in complex networks. *Physica A: Statistical Mechanics and its Applications*, Volume 502, pp. 459-468 (2018)

31. E. Katzav, O. Biham, A. K. Hartmann Distribution of shortest path lengths in subcritical Erdos-Rényi networks. *Physical Review E*, Volume 98, Issue 1, , article number 012301 (2018)

32. A.G. Hunt, F. Yu, The fractals of percolation theory in the geosciences (Book Chapter). *Fractals: Concepts and Applications in Geosciences*, pp. 114-152 (2017)

33. D. C. Rubie, S. A. Jacobson, Mechanisms and Geochemical Models of Core Formation (Book Chapter), *Deep Earth: Physics and Chemistry of the Lower Mantle and Core*, pp. 181-190 (2015)

34. D. Zhukov, T. Khvatova, S. Lesko, A. Zaltsman. Managing social networks: applying the Percolation theory methodology to understand individuals’ attitudes and moods. *Technological Forecasting and Social Change*, 2018, Volumes 123, pp. 234-245 (2018)

35. D.O. Zhukov, T.Yu. Khvatova, S.A. Lesko, A.D. Zaltsman. The influence of the connections density on clusterisation and percolation threshold during information distribution in social networks. *Informatica i ee Primeneniya, (Informatics and its applications)*, volume 12, issue 2, pp. 90 – 97 (2018)