KAUNAS UNIVERSITY OF TECHNOLOGY

Remigijus Laurutis

THE INVESTIGATION OF VIRUS PROCESSES IN TELECOMMUNICATION NETWORKS

Summary of Doctoral Dissertation

Technological Sciences, Electronics and Electrical Engineering (01T)

Kaunas, 2004

The research was accomplished during the period of 2000–2004 at Kaunas University of Technology.

Academic supervisor:

Prof. Dr. Habil. Romualdas Gudonavičius (Kaunas University of Technology, Technological Sciences, Electronics and Electrical Engineering–01T), 2000–2002.

Prof. Dr. Habil. Danielius Eidukas (Kaunas University of Technology, Technological Sciences, Electronics and Electrical Engineering–01T), 2002–2004.

Council of Electronics and Electrical Engineering trend:

Prof. Dr. Habil. Danielius Eidukas (Kaunas University of Technology, Technological Sciences, Electronics and Electrical Engineering–01T)**;**

Prof. Dr. Brunonas Dekeris (Kaunas University of Technology, Technological Sciences, Electronics and Electrical Engineering–01T)**– chairman**;

Prof. Dr. Habil. Vincas Laurutis (Šiauliai University, Technological Sciences, Electronics and Electrical Engineering–01T);

Assoc. Prof. Dr. Rimantas Plėštys (Kaunas University of Technology, Technological Sciences, Electronics and Electrical Engineering–01T);

Dr. Rimantas Kalnius (JSC „Telebaltikos konsultacijos“, Technological Sciences, Electronics and Electrical Engineering–01T).

Official opponents:

Prof. Dr. Habil. Pranciškus Balaišis (Kaunas University of Technology, Technological Sciences, Electronics and Electrical Engineering–01T);

Assoc. Prof. Dr. Gintautas Daunys (Šiauliai University, Technological Sciences, Electronics and Electrical Engineering–01T).

The official defense of the dissertation will be held on 13.00, December 16, 2004 at the Council of Electronics and Electrical Engineering trend public session in the Dissertation Defense Hall at the Central Building of Kaunas University of Technology (K.Donelaičio g. 73–403a, Kaunas).

Address: K.Donelaičio g. 43, LT–44029 Kaunas, Lithuania.
Tel.: (370) 7 300 042, email: mok.grupe@adm.ktu.lt

The send–out date of summary of the Dissertation is on November 16, 2004.

The dissertation is available at the library of Kaunas University of Technology.

KAUNO TECHNOLOGIJOS UNIVERSITETAS

Remigijus Laurutis

VIRUSINIŲ PROCESŲ ANALIZĖ TELEKOMUNIKACIJŲ TINKLUOSE

Daktaro disertacija

Technologijos mokslai, elektros ir elektronikos inžinerija (01T)

Kaunas, 2004

Disertacija rengta 2000–2004 metais Kauno technologijos universitete

Mokslinis vadovas:

Prof. habil.dr. Romualdas Gudonavičius (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T), 2000 m.–2002 m.

Prof. habil.dr. Danielius Eidukas (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T), 2002 m.–2004 m.

Elektros ir elektronikos inžinerijos mokslo krypties taryba:

Prof. habil.dr. Danielius Eidukas (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T);

Prof. dr. Brunonas Dekeris (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T)**– pirmininkas**;

Prof. habil.dr. Vincas Laurutis (Šiaulių universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T);

Doc. dr. Rimantas Plėštys (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T);

Dr. Rimantas Kalnius (UAB „Telebaltikos konsultacijos“, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Oficialieji oponentai:

Prof. habil.dr. Pranciškus Balaišis (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T);

Doc. dr. Gintautas Daunys (Šiaulių universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties tarybos posėdyje, kuris įvyks 2004 m. gruodžio 16 d.13 val. Kauno technologijos universiteto Centrinių rūmų disertacijų gynimo salėje (K.Donelaičio g. 73–403a, Kaunas).

Adresas: K.Donelaičio g. 73, LT–44029 Kaunas, Lietuva.
Tel.: (8–37) 300 042, el.paštas: mok.grupe@adm.ktu.lt

Disertacijos santrauka išsiųsta 2004 m. lapkričio 16 d.

Su disertacija galima susipažinti Kauno technologijos universiteto bibliotekoje.

INTRODUCTION

One of the most definitive problems of Internet consumers of telecommunication services is the email viruses ("viruses"), amount of which increases yearly. These viruses – harmful programs created by humans are sent by e-mail protocol. The programs on the net broadcasted execute programmed functions, and are capable of self-copying without the intervention of on-line consumers. Computer viruses are acknowledged as the first artificial intellect representatives because they are able to spread and reproduce as biological viruses.

The problems are caused not only by the increasing number of new virus programs but by the expeditious spread of virus on the net as well. Virus is spreading rapidly on the net and requires a defensive reaction while technologies evolve. It is possible that, the majority of problems will be caused not by slowly circulating viruses but by the speed ones -"zero days" i.e. the programs which overload the net, cause epidemics and decrease the QoS of services.

The abundance of scientific projects indicates the importance of service quality i.e. the problem of QoS analysis, and the security of information to be an important parameter of the quality of the telecommunication services. The problems of QoS and security are being ventilated by universities, scientific institutions, and telecommunication companies: ITU, IEEE, EICAR, SANS Institute, Usenix, IBM Research Centre, SNORT, and Silicon Defence.

Such scientific researches are being prosecuted in Lithuania as well. The scientific program "The quality of telecommunication nets and services" is being held in KTU (prof. R.Gudonavičius, prof.B.Dekeris, dr.L.Narbutaitė, dr.R.Jankūnienė, dr.G.Činčikas etc.). The academe of prof.D.Eidukas, prof.P.Balaišis and others scholars, which researches the topics of quality and security of the electronic equipment, analyses the topics of the quality of telecommunication nets and services in the doctoral thesis. Investigating those topics, a group of scholars in KTU which analyses the software security problems, the problem of evaluation of net connection quality is being researched widely. The research of telecommunication nets – an object of QoS investigation- is being investigated by the scholars of KTU.

The author has been carrying out scientific researches for several years and then presents publications about the security of telecommunication nets, the detection of virus epidemics, the anomalies and their influence to QoS. The main attention is paid to the viruses, as one of the most important components, which reduces the QoS quality of services. The author's research about the artificial neuron network methods for the operation of the telecommunication nets showed the advantages of these technologies. The main attention is paid to the problems of software security and anomalies in the telecommunication nets. On purpose to estimate the efficiency of neuron networks, the experiments were carried out and their results were published in scientific magazines.

Aim of the work

1. To suggest the techniques, which allow detecting already known virus epidemics of the telecommunication nets and unknown ones effectively.
2. To calibrate the efficiency of the technique behaviour.

Goals of the work

1. To accomplish the comparative analysis of virus identification systems which are suggested by the international software security organizations; to estimate which kind of the viruses' epidemics causes the greatest damage.
2. To acquaint with the subsistent models which depict the biological epidemics, and their parameters. To adjust the chosen model of the viruses' epidemics to the modelling of the most dangerous telecommunication net viruses' epidemics.

3. To perform the experimental modelling, to estimate which parameters of epidemic is the most influential for reducing the damage caused by epidemics, and to suggest the most efficient technique of preventing from the viruses' epidemics.
4. To suggest the method which allows controlling the epidemic parameters, reducing the amount and damage of viruses' epidemic, efficiently.
5. To calibrate the effectiveness of the designed techniques experimentally.

Novelty of the Work

1. The author defends the techniques, detecting the email virus epidemics in the principles of the biological epidemiology, characterized to have the blocking feature of new and unknown viruses' epidemics.
2. 2. The results of the imitated modelling, which were obtained using the suggested techniques, confirm the antiviral email security system, using this technique, to be more efficient than traditional ones.

Practical value of the work

The method, designed to detect the viruses' epidemics in the telecommunication nets, allows reducing the damage of the most popular viruses' epidemics. This software security system should help to avoid the main disadvantage of current software security systems - i.e. incapacity to identify the new viruses. Installing the protection systems, which use that method, would help to reduce the overloaded nets during the period of the viruses' epidemics, the on-line consumers would be protected, the reliability of the services would increase, and the QoS quality of the net services would be improved. The most important - the damage caused by the viruses' epidemics would be reduced. After the suggested technique to detect the epidemics, proved out, the establishment of the new type of antiviral equipment could be initiated.

Approbation of the work

Four publications based on the topic of the dissertation are acclaimed in the criticized magazines such as "Electronics and Electrotechnics" (KTU), "The Sciences of Information" (VU), two essays are published in the conference editions.

The main results of scientific and experimental investigations are approbated in the international conferences "Electronics" in Kaunas (2000-2004 yr.), in the conference "The days of programmers 2001" KTU, in the conference "IT 2002"KTU, in the conference of the science academy XIX of the Lithuanian Catholics (2003) ŠU.

THE CONTENT OF THE DISSERTATION

The analyzed questions are defined, the relevance and scientific novelty of the research are indicated, the aims are formulated and the practical value of the research is formed in the **introduction**.

In the first section the tendencies of the viruses' epidemics are surveyed, the methods which allow depicting the viruses are analyzed. It is pointed out that the antiviral software security systems according to the type of the acceptance of solutions are divided into 2 kinds:

1. Knowledge-based systems;

2. Behaviour-based systems.

The knowledge-based systems use the cumulative knowledge about the previous anomalies and attacks for the detection of anomalies. The more the software security systems have the information about the known attacks and harmful actions, the more perfectly those systems function. The advantage of such systems - is a small number of inaccurate solutions. The systems compare the information with the known rules in the net and if some similarities are traced, they generate an alarm signal. Those systems require the permanent renewal of rules because the viruses are becoming more inventive and their brand new attacks often differ from the previous. This is the reason of why those systems can identify only already existing attacks but cannot identify the new ones. The second disadvantage – those systems hinge on the environment of an operation: operating systems and technical equipment.

The behaviour-based software security systems function observing the variations of net condition and the information flux. The situation of regular situations or actions is simulated by collecting various parameters of the regular state of net. The system compares this information with the current net activity and if any variation is detected, the system generates the alarm signal. In that case, the system reacts to all what is new and unknown.

Conclusions of the section

1. After accomplishing the review of literature it is defined that the telecommunication viruses circulate faster and the detection system of the virus epidemics is becoming an important component of net operations.
2. It is defined, that the knowledge-based systems used for the antiviral security, protect from known virus attacks effectively but cannot protect from unknown ones.
3. The behaviour-based security systems are capable to protect from unknown viruses, but their disadvantage would be a huge number of false alarms.
4. The frequentative Hollinger's and the methodology of accumulative circulation are slow, i.e. the epidemic is diagnosed after sending more than several e-letters.
5. The register method is used to prevent from viruses based only on the TCP/IP protocol. This method allows to expand the time for equipment protection from the viruses on the net,

preserving the regular quality QoS of services. This methodology is suggested to adapt in blocking the virus epidemics.

6. The behaviour-based methodology is suggested to use in detecting the unknown virus

epidemics. To avoid the slow human factors, it is suggested to adapt such techniques which would allow to diagnose epidemics without the human intervention and to close the circuit of security system.

In the second section modeling of popular e-mail viruses and virus epidemics is executed. To value what damage could be caused by virus epidemics and what parameters let to control virus spread one should choose models of viruses and to analyse them. The methods used in biology are also used in the modeling of telecommunication viruses as virology and epidemiology is already explored:

SIS method (susceptible-infectious-susceptible) is used in modeling biological epidemics during which infected and cured individuals could be infected again, i.e. get ill with the same epidemic.

SIR method (susceptible-infectious-resistant) is opposite to SIS method. Infected and cured individuals become resistant to the same virus attack in the future.

It is known that telecommunication epidemics have two periods and to their modelling SIDR method is adjusted. Periods of SIDR method:

1. Period when epidemic spreads unstoppable. During the first period a virus infects devices which are connected by telecommunication net. The virus spreads freely and does not raise any suspicions.

2. Period when viruses which rise epidemic are detected and started to be blocked. The virus spread during the second period cause QoS or other problems and is noticed. It infects many network devices. Security specialists take virus examples, explore them and create antiviral programs which stop viruses of that type. Antiviral programs are placed in to the Internet and users can swap them in. Not infected devices to which the antiviral program was installed become virusproof. Infected units must be detected, fixed and the antiviral program must be installed.

The spreading virus which is analysed by SIDR method has two stages reviewed above: before and after detecting it in the net.

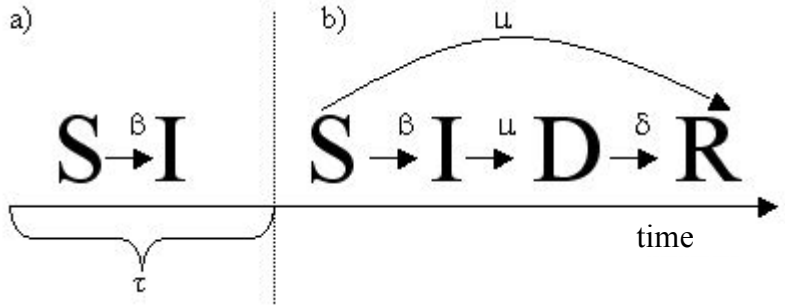


Fig.1. Model of PSIDR epidemic adapted to e-mail networks. a – epidemic is not detected, b – epidemic is detected.

Till the virus is detected in the e-mail network ($t<\tau$), devices in the net are infected and their state fluctuates from S to I with intensity (β). When ($t>\tau$), model changes because time needed for detecting virus in the net (μ) and time needed for virus neutralization (δ) must be evaluated.

When $t<\tau$, than:

$$S(t)+I(t)=N \ ; \tag{1}$$

And:

$$\frac{dS}{dt}=-\beta SI \ ; \tag{2}$$

$$\frac{dI}{dt}=\beta SI \ . \tag{3}$$

When $t>\tau$, than:

$$S(t)+I(t)+D(t)+R(t)=N \ ; \tag{4}$$

And:

$$\frac{dS}{dt}=-\beta SI-\mu S \ ; \tag{5}$$

$$\frac{dI}{dt}=\beta SI-\mu I \ ; \tag{6}$$

$$\frac{dS}{dt} = -\beta SI - \mu S \quad ; \tag{5}$$

$$\frac{dD}{dt} = \mu I - \delta D \quad ; \tag{7}$$

$$\frac{dR}{dt} = \delta D + \mu S \quad ; \tag{8}$$

Here: N – the number of devices in the network; S – the number of devices frail to virus; β - virus spreading speed; I – the number of infected devices; μ - virus detecting speed; D – the number of infected devices which were detected; δ - virus neutralization speed (neutralization, re storage of information, etc.); R – the number of fixed devices.

One of the advantages of S-I-D-R model is that it lets to calculate the loss caused by a email virus.

1. *Expenditures of devices restoration.* This parameter is calculated by evaluating for how long devices were in a state of D (detected). If the state of device changed while being in D, it means that the device was repaired:

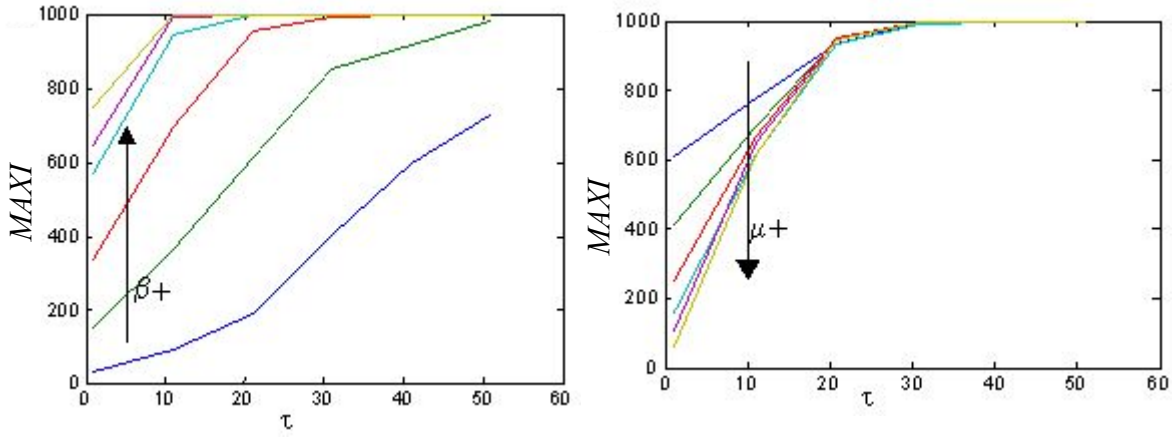
$$IAS = \sum_{\pi}^{T} D(t)$$

2. *Damage of epidemic.* It is calculated by evaluating how many net devices were infected and for how long were not working:

$$EZ = \sum_{t_0}^{T} I(t)$$

3. *Maximum number of infected devices.* One of the most important parameters is the number of infected devices during the period of epidemic:

$$MAXI = max(I(t)), t \in [t_0 \ldots T].$$



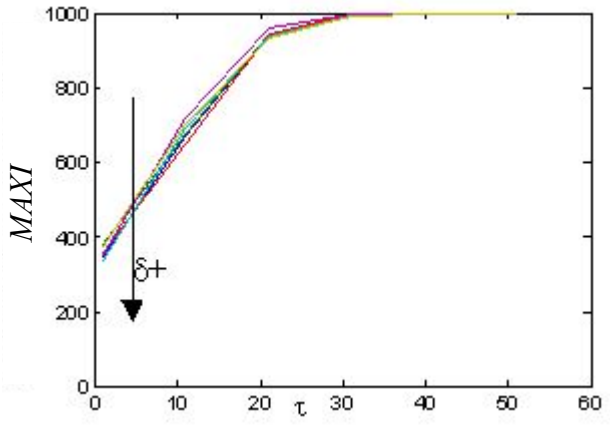


Fig.2. Epidemic's behaviour in e-mail network by changing parameters β,δ, μ and τ [0 - 50]

In more detail analysis of virus spreading in e-mail networks we rated what influence other parameters have to the damage caused by epidemic.

Conclusions of the Section

1. When the number of individuals is large, the value of incidental events approach their average and then dynamics of population is characterized by average values which coincides in determinable and stochastic models. This is the reason why determinable methods also can be applied to model big technological epidemics. Model of matrixal epidemics is not beneficial for modelling of epidemics of technological viruses, because it does not use too much redundant information: age of the virus, number of times. These parameters are not beneficial for modelling of epidemics of technological viruses.
2. Parameter τ – *the speed of viruses' epidemics identification,* mostly influences the maximum of epidemics.
3. The second important parameter is β – *the speed of viruses' reproduction.* If its meaning becomes very high, even minimal lateness time of identification of viruses would not be able to reduce the damage caused by epidemics for devices infected by viruses.
4. Parameter μ – *speed of viruses' identification*, influences the maximum of epidemics just to a certain meaning of identification lateness of epidemics. Still when the viruses' identification speed increases, the damage caused by epidemics is reduced, irrespectively to the meaning of identification lateness of epidemic. Wanting to reduce the damage caused by epidemics, the speed of viruses' identification must be high as possible.
5. Parameter δ – time for the repair of device damaged by virus, also does not have influence for the maximum of epidemics. The expenses for the repair of devices reduces with shortening of the time.
6. When the lateness of the start of epidemics' identification of epidemics modelling evaluation was detected, the best e-mail antiviral security strategy was formed: it needed the installation to e-mail servers of systems identifying epidemics, operating e-mail sending. The identification system of letters infected with viruses would immediately react to new viruses and it would shorten the viruses' identification time, and control system of letters' sending would reduce the speed of viruses' spread by e-mails. These tasks would be effectively performed by automatic systems of epidemics identification installed in e-mail servers.

The suggested method of epidemics identification using artificial neural networks (NN) is described in the third section.

NN system most frequently used as independent manager – source of knowledge. It operates as separate module and it performs the work of decision making in the operational system, it transmits the result to other elements of the system which do not have direct connection with NN. In these kinds of systems NN is trained to systemize receivable data and is characterized as having features of artificial intelligence. Technological viruses are also named ones of the first representatives of artificial intelligence, it is quite logical against to undesirable phenomenon to use the system characteristic with similar features.

For the training of NN network it is necessary to prepare training data set. This data consists of sets of inlet and preferable outlets necessary for NN. It is very important to single out such variables of the system that mostly affect the result. NN is trained to find connection between data of inlet and outlet. Usually the training set consists of experimentally collected data. Unseen data set for NN is chosen for its testing. If the network performs similarly with the testing data and with the training set and we obtain the satisfactory bias, NN is rated as properly trained and is able to work with practical tasks.

NN used for identification of e-mail viruses epidemics; the suggested working algorithm of software security system is presented in the image 3.

Algorithm disables 2 means of detection with viruses' epidemics: NN controller and virus throttling. Virus throttling security was proposed at HP Labs laboratory which aim is to block viruses spreading in TCP/IP topical networks. Virus throttling security mean was applied only to TCP/IP package management up till now. Our suggested algorithm with NN controller allows using this method also for blocking of e-mail viruses' epidemics.

The trained NN analyses the flow of e-mails generated by consumers and decides whether the behavior of the consumer is similar to the behavior during the epidemic or not. If NN decides that the consumer's sent/received letter accessed system inlet in general behavior manner is not similar to epidemic, the letter is immediately sent. If the letter is under suspicion that it contains viruses, its data is stored to global data base (GDB), and the letter itself is transferred to sending mechanism of throttling. The suspicious letter is stored to the queue of throttle and is kept there a fixed period of time *E_tmax*. If *E_tmax* time has run out and other identical to NN suspicious letters have not queued, the letter is sent to the recipient. If other new identical NN suspicious letters have queued, this is treated as the beginning of virus epidemics and letters one by one are deleted.

Such algorithm of NN operations allows early identification of virus epidemics and decrease time τ to minimum, and throttling mechanism decreases speed of spread β of viruses, that did not seem suspicious to NN.

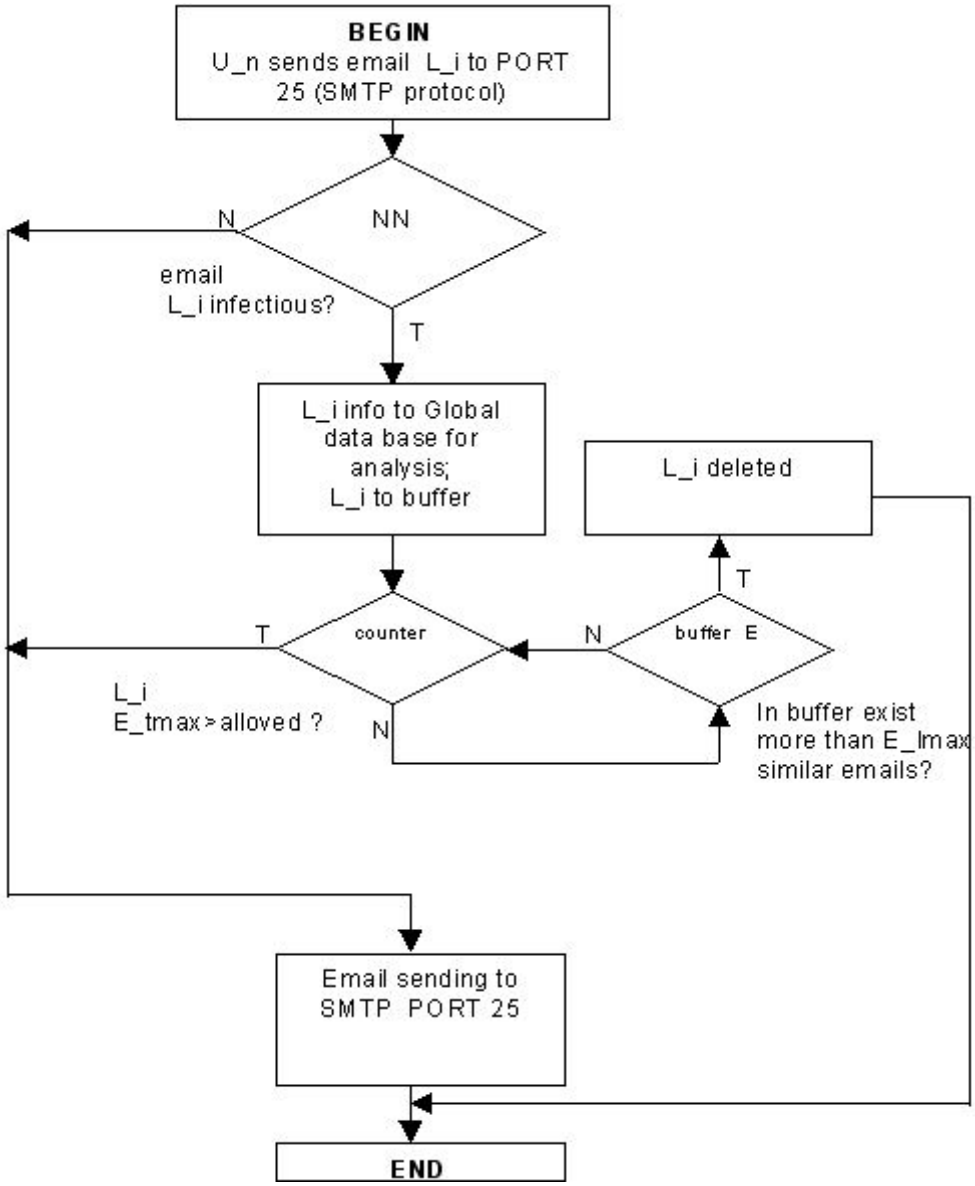


Fig.3. Operational algorithm of identification and blocking system of e-mail viruses' epidemics operated by NN.

If we want that artificial neuron network would classify infected and non-infected letters as precise as possible, it needs proper network training, which has its own peculiarities. It is advisable to transform expressions of time parameters into variation and avoid connection with concrete dates during the training of NN. The column of operation "Time" was transformed into the subtraction of current and past times Δt (measurement: minutes).

After insertion of the flow of letters generated by viruses into the flow of "normal" consumers, we would get data sets for NN training.

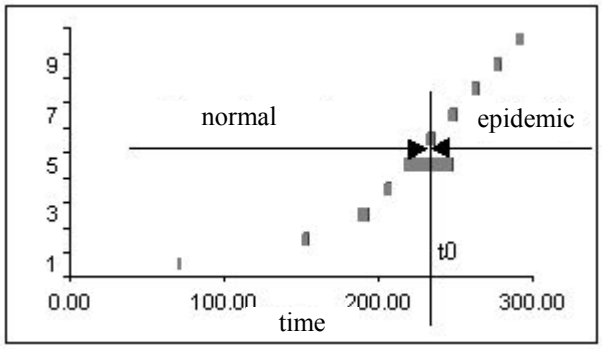


Fig.4. Graphical image of normal and of the beginning of the epidemic flows.

It seen from the image 4, that to the time moment t0, the consumer has sent letters in different time moments, the meanings of letters' sizes were varied. At the time moment t0the epidemic started, because the speed of letters increased, the creation time and sizes of letters got fixed meanings. This data set is proper for the NN training, using this principle of transformation of time parameters, we are able to prepare NN training set from experimentally collected data.

Software security system – it is a system with all attributes characteristic to a system, one of them is operating outline. That is why its effectiveness will be conditioned by working effectiveness of operating outline. There are open and closed software security systems. In open software security systems a part of is left to perform for a human, humans are not a part of closed software security systems. The application of artificial neuron networks for automatic identification of viruses' epidemics allow to work for the system automatically, without operator – human, in this way this type of software security systems become closed. This is the reason why the viruses' epidemics identification time would be shortened to minimum; in this way the damage caused by viruses would be reduced.

Before choosing of evaluation methods of software security systems, we have to evaluate the effectiveness of each mean separately and to define its influence to the effectiveness of all system.

As effectiveness' rates of the software security system depend on its purpose, protecting systems from their fulfillment, damages can be defined with economical rate: $C_{\sum AS}=C_1+C_2$, here $C_{\sum AS}$ - average damage connected with the software security system in time unit; C_1 - the average costs for software security system in time unit; C_2 - the average damages because of onsets in the operation of this system in time unit.

It is seen from (9) that average damages $C_{\sum AS}$ depend on the price of the software security system, its durability, installation expenses, exploitation processes and other. That is why the effectiveness will be defined by this rate:

$$E=1-\frac{C_{\sum AS}}{C_{ML}}\cdot K_1 \quad ; \tag{9}$$

Here C_{ML} - maximal allowed damages connected with exploitation of software security system in time unit; K_1 - coefficient regulating the speed of effectiveness variations ($K_1<1$).

In the selection of optimal version of software security system, we will use this operator:

$$\min_i (C_{1i}+C_{2i}) \Big| C_{\sum AS} \leq C_{ML} \quad ; \tag{10}$$

Here $i=\overline{1,N}$; N – number of possible variants of security systems.

The results of training and capability of artificial neuron network can be expressed in several main parameters, they are – average square bias, correlation ratio, network training bias and Akaike informational criteria of artificial neuron network. These parameters, their calculation methods and their meanings will be reviewed in this section.

The mean squared error expresses the average square inequality between desirable and obtained network meanings of outlets. MSE calculations are as follows:

$$MSE = \frac{\sum_{j=o}^{P} \sum_{i=o}^{N} (d_{ij} - y_{ij})^2}{N\,P} \quad ; \tag{11}$$

Here: P – the number of outlet elements of neural network; N – the number of inlet data examples; the net outlet of y-j element for i example; d – the desirable outlet of element j for i example. During the training of NN it is observed whether the meaning of MSE increases or decreases. The increase of MSE presents that network has either overworked or is not capable to complete its task.

MSE may be calculated for different data sets: training, testing, cross validation sets. During the NN training process it is necessary to observe the meaning of MSE with testing or cross validation sets and to stop the training when MSE starts increasing, and not to pay attention to the decreasing MSE of training set. This kind of situation shows that the network started to memorize the training set and works with other data ineffectively.

MSE is used for measuring of how exactly match the meanings of the network outlet and meanings of desirable outlets, though it does not show the changes of shifts of these data arrays towards one direction. For example, after changing the meanings of network outlet in proportion, MSE would change, and working of NN would not change. Correlation ratio r solves the problem. Correlation ratio changes in the range [–1...1], and it is excellent direct correlation between x and d, when r=-1, it means that excellent reverse correlation, in this way x and d, i.e. when x increases, decreases linearly. When r=-0, there is no correlation between x and d. Transitional meanings indicate the transitional correlation rate, for example, training NN and having r=0,9, it is considered that correlation rate is quite good.

Akaike informational criterion is used for the measurement of proportion of training abilities and the frequency of the network. During the training of the network it is necessary to minimize this parameter:

$$AIC(k) = N \ln(MSE) + 2k \quad ; \tag{12}$$

Here: k – number of weights neuron network; N – the number of examples in training data; MSE - mean squared error.

Conclusions of the Section

1. The mean of identification of viruses' epidemics using artificial neural networks was suggested. During the analysis of changes of consumers' e-mail flow, artificial neural networks allow the identification of new and unknown e-mail viruses' epidemics. It shortens the time of epidemics identification to minimum, also the damage caused by the epidemic

diminishes.

2. For the blocking of undesirable e-mails, the method of blocker of TCP/IP protocol. This method allows blocking the virus infected e-mail flow more effectively, without reducing of the quality of services Qos provided to the consumers.
3. The proposed method allows transforming data of e-mail flow into format appropriate to artificial neural networks. Also the method of transformation of time e-mail flow characteristics into the format appropriate to the training of artificial neural networks.
4. The method was analysed, which is possible to rely on for the calculations of parameters of software security system effectiveness. He allows optimizing the selection of software security system means. The proposed parameters, which allow rating most effectively artificial neuron network training and functioning quality. It gives the opportunity to use the resources of the devise, operating the analysis of e-mail flow.

The fourth section is devoted to the verification using experimental method, weather the neural network is capable to identify the flow created by viruses and the flow created by consumers effectively.

For the accomplishment of the research the following tasks have to be fulfilled:

1. The collecting of experimental training data set.
2. The transformation of the collected data to the format appropriate to artificial neural networks.
3. The selection of appropriate types of neural networks for this task.
4. The experimental verification of the type of NN (artificial neural network) performing this type of task best.
5. The defining of efficiency of this software security system, the comparison of this software security system with the efficiency of already known software security systems.
6. The presentation of conclusions.

Data for NN training were collected during one month at public institution "Šiauliai Business Incubator". The premises of this institution are occupied by approximately 40 companies, while there are 100 consumers of topical and Internet networks. Server, which collected the statistics of the e-mail sending, was integrated in topical network. Overall 4000 e-mails were collected. Since we have been interested in active operations of consumers, operations, left in the training data array were performed during working hours. The example of data fragment is listed in the table (Table 1), all senders and recipients are modified for confidentiality.

Table 1. Fragment of training data set

The Time of The Operation	Sender	Recipient	Subject	Size (Bytes)
2004.03.12 16:37	jena@jsss.lt	kodavmas@vbb.lt	Re: NETO	4254
2004.03.12 16:50	h.damks@lc.lt	k.gecas@lc.lt	Offer for web site creation	5374
2004.03.12 16:50	h.damks@lc.lt	k.gecas@lc.lt	infomation	1512
2004.03.12 16:51	h.damks@lc.lt	k.gecas@lc.lt	Emailing: Bank sheet	1523
2004.03.12 16:51	h.damks@lc.lt	k.gecas@lc.lt	Sheet	18435
2004.03.12 17:04	hujgs@sv.lt	lna.date@splus.lt	Read: GPS Update	59875
2004.03.12 17:15	hujgs@sv.lt	mlda.v@splus.lt	Read: JSC Gama	1094362
2004.03.12 17:15	hujgs@sv.lt	mlda.v@splus.lt	Message	2319237
2004.03.12 17:16	sajnas@nfoeja.lt	egdhus@nfoeja.lt	?	189899
2004.03.12 17:41	jjena@takas.lt	zvle@flame.lt	About presents	6829

2004.03.12 17:57	saunas@nfoeja.lt	levs@nfoeja.lt	About extra package	920
2004.03.12 18:32	e.dlnsks@saula.net	denkej@wdant.com	Re: ?	131499
2004.03.12 18:36	levs@nfoeja.lt	sajunas@nfoeja.lt	Information	1275
2004.03.12 18:37	palba@nfoeja.lt	zene@skdas.net	hall carpet	324805

Data generated by consumers was also collected from primary data array and listed in separate tables. 25 consumers were chosen at random and their data was used in further experiments.

NN is not capable to operate with dates directly; consequently the received data set is not suitable for NN training. Data of time and date has to be transformed to appropriate format for NN according to the methods stated in the third section of the thesis. Expressions of time parameters were transformed into variations on purpose of avoiding connections with concrete data during NN training. The column of operation "Time" was transformed into the subtraction of current and past times Δt (measurement: minutes): After this kind of transformation the time Δt presents the time period from the last sent letter and the letter before last.

We can draw the conclusion from the collected data, that during epidemics and non - epidemic period the MGS and averages of sizes of letters interlay. The flow of MGS (the speed of e-mail creation) of interlay range is presented in the picture 5.

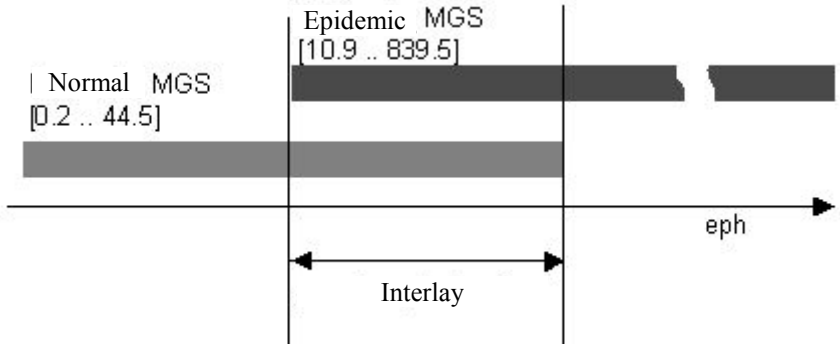


Fig.5. The MGS interlay during normal and epidemic flow.

The interlay appears due to the fact that slow viruses can spread in similar speed as e-mails are sent by active consumers – their MGS varies from 10.9 to 44.5. This leads to a conclusion that virus - epidemics cannot be identified only through MGS. It leads to the usage of more parameters and analysis of dynamics of subtotals variation of e-mail flow parameters: variations of e-mail size and MGS. According to this statement, data must be transformed according to the methods stated in the thesis (section 3.2, 3.1-3.6 formulas). The table 2 presents the fragment of data prepared for the training, when the length of buffer 1=3

Table 2. Fragment of e-mail operations prepared for NN training

Time	u(i)		y(i)	a(i)						
	MGS	Size	Virus	MGS (t)	Size (t)	MGS (t+1)	Size (t+1)	MGS (t+2)	Size (t+2)	Virus (t)
t = 1	6,7	2,7	1	6,7	2,7	3	1,9	0,3	2,4	1,0
t = 2	3	1,9	1	3	1,9	0,3	2,4	0,2	487,7	1,0
t = 3	0,3	2,4	1	0,3	2,4	0,2	487,7	22,4	598,8	1,0
t = 4	0,2	487,7	1	0,2	487,7	22,4	598,8	20,7	832,7	1,0
t = 5	22,4	598,8	1	22,4	598,8	20,7	832,7	7,8	365	1,0
t = 6	20,7	832,7	1	20,7	832,7	7,8	365	87,8	365	1,0
t = 7	7,8	365	1	7,8	365	87,8	365	180	350,8	1,0

t = 8	87,8	365	1	87,8	365	180	350,8	7,9	367,9	1,0
t = 9	180	350,8	1	180	350,8	7,9	367,9	0,2	383,1	1,0
t = 10	7,9	367,9	1	7,9	367,9	0,2	383,1	0,2	4,7	1,0
t = 11	0,2	383,1	1	0,2	383,1	0,2	4,7	0,2	172	1,0
t = 12	0,2	4,7	1	0,2	4,7	0,2	172	24	53,7	1,0
t = 13	0,2	172	1	0,2	172	24	53,7	128,6	69,4	1,0
t = 14	24	53,7	1	24	53,7	128,6	69,4	22,6	3,8	1,0
t = 15	128,6	69,4	1	128,6	69,4	22,6	3,8	1,5	1,6	1,0
t = 16	22,6	3,8	1	22,6	3,8	1,5	1,6	0,9	4,7	1,0
t = 17	1,5	1,6	1	1,5	1,6	0,9	4,7	0,4	9	1,0
t = 18	0,9	4,7	1	0,9	4,7	0,4	9	85,7	210,3	1,0

The condition as a base point was, that software security operated by NN should identify virus epidemics maximum in an hour. It leads to this kind of queue length measurement. This table gives the maximal, medium and minimal speed of spread of viruses. The most difficult for identification is the minimum speed epidemics, because its MGS=10.9 and is similar to MGS of normal active consumers. MGS of this size means, that within an hour virus sends approximately 10 infected e-mails. And we can draw the conclusion that on purpose of identification of within an hour minimally spreading epidemics, the analysis needs to keep 10 e-mail parameters in a queue. And the maximum queue length in NN operating system cannot be more than 10. But wanting to ascertain whether it is impossible to identify the infection from the data from shorter queues, we need to make experiments with shorter length of queues. This allows evaluating the identification period of time of spreading epidemics at medium and maximum speed. The data is presented in table 3.

Table 3. The subjection of epidemics identification time to MGS and length of the queue.

MGS	T_max	Length of the queue	T_apt
10,00	60,00	10,00	60,000
10,00	60,00	5,00	30,000
10,00	60,00	3,00	18,000
200,00	60,00	10,00	3,000
200,00	60,00	5,00	1,500
200,00	60,00	3,00	0,900
839,00	60,00	10,00	0,715
839,00	60,00	5,00	0,358
839,00	60,00	3,00	0,215

Explanation of the table: MGS – the speed of viruses' reproduction, T_max – the minimal time of possible identification of the epidemics (min.); Length of the queue –length of at the same time NN analysed queue; T_apt - time of identification of the epidemics (min.).

The third section of this thesis was devoted to the methods of proper evaluation of quality of NN training. It needs data unseen during the NN training – cross validation set. During the training of the network the bias is measured, which is made by network itself with cross validation set, and when this bias reaches the minimum and starts increasing, it is recommended to stop the NN training process, because this fact shows that NN does not start learning the training set data, but it starts to memorize them. As this kind of effect exists, testing of already trained NN has to be run with unknown data for NN. It needs to be observed, because data sets of training, cross validation and testing have to be formed from one data set. Three different buffer length testing data sets A, B and C were further subdivided into three parts: training set (comprised 50% of all collected data), cross validation set (comprised 25% of all collected data), and testing set (comprised 25% of all

collected data). Using this method we obtained nine data arrays with which training and testing of NN networks were experimented. Neurosolutions programming package was used for training and testing of NN networks.

The first stage of experimental modelling – to test which NN network is most effectively trained with training data set. As identification task of epidemic and normal flow reduces to the task of classification of parameters in time, Elman and TLRN network were chosen for experimental modelling. In literary sources these networks are recommended for work with tasks of dynamic classification.

- Elman network. The variety of multi-layer perceptron extended in elements, which remember past states. Memory elements allow sourcing the information transforming in time.
- TLRN network (Time Lagged Recurrent Network). It is the variety of multi-layer perceptron extended in structures of flash memory. Structures of the memory can remember passed states only of input or inner layers that is why this type of network is also recommended for work with tasks of management of dynamic systems.

As it is unknown what structure of NN to choose, we need to search for best training structure using the experimental method. The network is structured using the principle that is offered in literary sources: the network of minimal structure is enlarged to maximal recommended limits. It is recommended that the inner weight number of NN would be smaller than quantity of examples of training data variety, if this condition will not be put in operation, the network can only memorize the data, but will not be trained.

Since the training results during the training period of the network frequently depend on primary and totally random network weight subtotals, the network of every structure must be trained five times. The averages of these training results with obtained testing data are presented below.

Table 4. The averages of NN training results with buffer queue length of l=3

Eksp. No.	Type	Network Structure	MSE	r	%Error	AIC	00	01	10	11	Notes
1	elman	3:5:1	0.01	0.98	0.8	–2164	99.80	0.20	1.16	98.84	
2	elman	3:10:1	0.008	0.99	0.84	–2632	99.90	0.10	1.16	98.84	
3	elman	3:20:1	0.001	0.99	0.23	–3588	99.90	0.10	0.10	99.00	
4	tlrn	3:5:1	0.01	0.98	0.63	–2239	99.80	0.20	2.30	97.70	
5	tlrn	3:10:1	0.01	0.98	0.5	–2207	99.90	0.10	3.40	96.60	unstable
6	tlrn	3:20:1	0.01	0.97	0.77	–2060	99.88	0.12	3.40	96.60	Very unstable

Table 5. The averages of NN training results with buffer queue length l=5

Eksp. No.	Type	Network Structure	MSE	r	%Error	AIC	00	01	10	11	Notes
1	elman	5:5:1	0.01	0.97	1.2	–2241	99.90	0.10	2.10	97.90	
2	elman	5:10:1	0.01	0.98	1.1	–2210	99.80	0.20	2.10	97.90	
3	elman	5:20:1	0.03	0.96	1.89	–1699	99.40	0.60	4.30	95.70	
4	tlrn	5:5:1	0.03	0.96	1	–1566	99.22	0.78	3.50	96.50	
5	tlrn	5:10:1	0.03	0.95	0.97	–1779	99.20	0.80	5.30	94.70	
6	tlrn	5:20:1	0.03	0.95	1.4	–1670	98.80	1.20	4.40	95.60	

Table 6. The averages of NN training results with buffer queue length l=10

Eksp. No.	Type	Network Structure	MSE	r	%Error	AIC	00	01	10	11	Notes
1	elman	10:5:1	0.03	0.97	1.1	–1698	99.50	0.50	2.60	97.40	
2	elman	10:10:1	0.03	0.96	1.5	–1558	99.30	0.70	4.40	95.60	
3	elman	10:20:1	0.03	0.96	1.4	–1803	99.30	0.70	4.40	95.60	
4	tlrn	10:5:1	0.04	0.95	1.8	–1331	98.90	1.10	4.30	95.70	
5	tlrn	10:10:1	0.03	0.96	1.8	–1389	100.0	0.00	5.40	94.60	
6	tlrn	10:20:1	0.03	0.96	1.6	–1333	99.10	0.90	5.30	94.70	

Contractions of fields: MSE – mean squared error; r – correlation ratio; % error – percentage of network training error; AIC – Akaike informational criterion; 00 – identification process of normal flow (no detected viruses); 01 – non-identification percentage of normal flow (false alarm); 10non-identification (error) of infected flow; 11- correct identification of flow of viruses (true alarm).

We can state from data the presented in tables 4 – 6 that in all cases the training percentage bias is not higher than 1,9%. We can also state that better training characteristics belong to Elman networks.

For the visual presentation of obtained data, results are illustrated graphically. Obtained results during the training with the continuing data set of Elman network are presented in all images of the left and on the right – TLRN network.

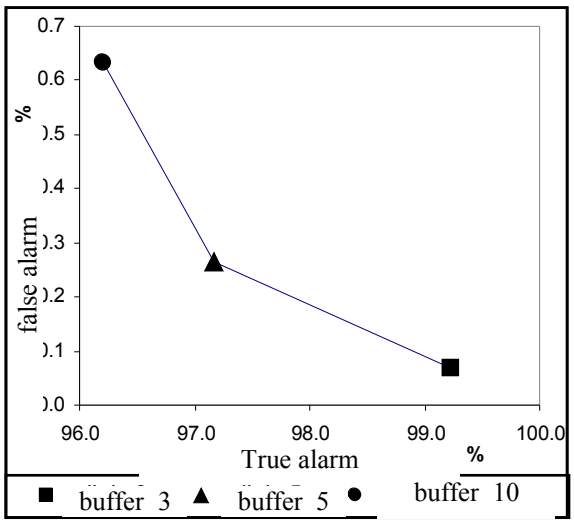


Fig.6. The Efficiency of Elman network Classification

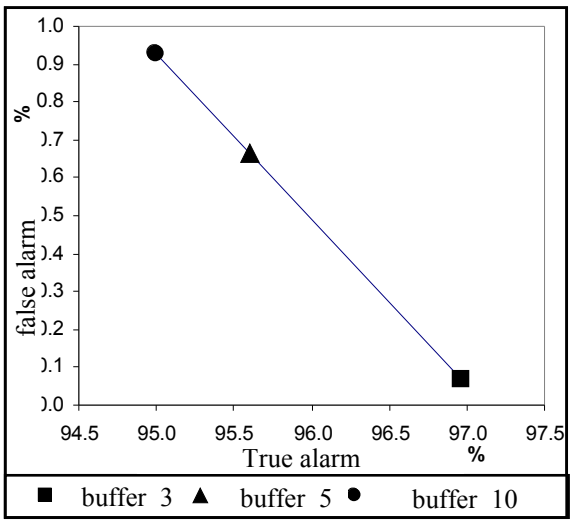


Fig.7. The Efficiency of TLRN network Classification

We can state images 6 and 7 that highest percent of identification of epidemics and lowest percent of false alarms are reached when three e-mails are queued. In the analysis of the data which is in the buffer queue length of 10 e-mails, the worst result is achieved – the lowest percent of epidemics identification and highest percent of false alarms.

From MSE analysis we can decide which NN network may be trained more effectively I it can be stated that MSE averages of Elman networks are slightly lower than correlation ratio of TLRN networks. The quality of network training is higher when the correlation ratio is due to 1.

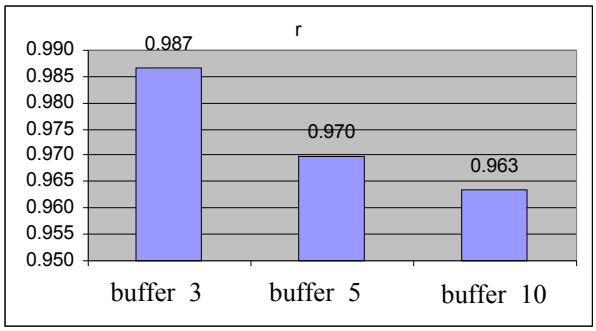


Fig.8. Correlation coefficient in Elman network

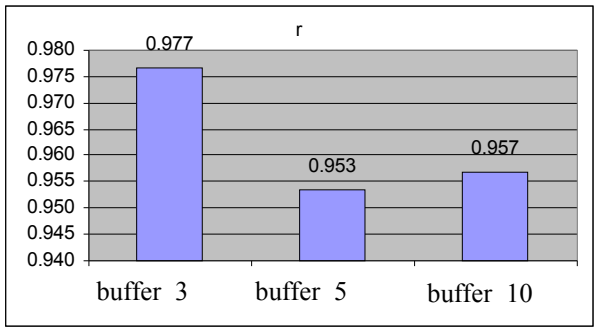


Fig.9. Correlation coefficient in TLRN network

AIC – Akaike informational criterion depicts the ratio of training abilities and network complexity. During the process of NN training it is essential to get the AIC value as low as possible. Images 10 and 11 illustrate the change of AIC averages in the process of the experiment.

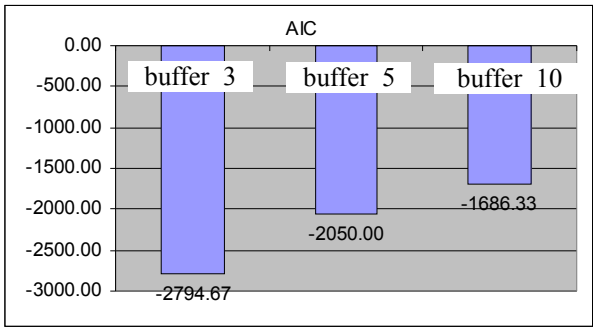


Fig.10. AIC averages of Elman

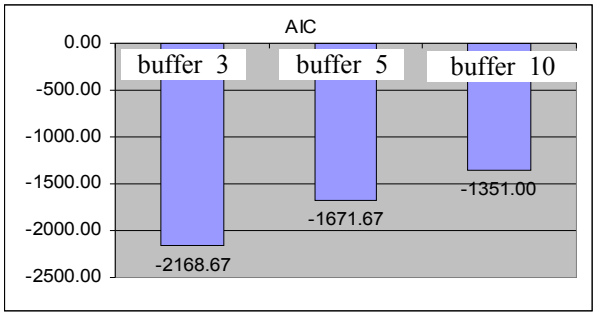


Fig.11. AIC averages of TLRN

And again the Elman network according to the AIC parameter is superordinate to TLRN network. The experiment is outstanding, when three letters are queued; Elman network reaches the lowest AIC value.

It can be stated that the increasing the complexity of the network structure, increases the AIC value. When three e-mails are queued – the network structure – 3:xx:1, and when ten e-mails are queued – the network structure is 10:xx:1. NN which have 3:xx:1 structure, are trained faster and more effectively, their AIC is to 40%lower. This can be explained that for the training of more complex structures of NN require more recourses.

On the tasks of system resources economy is to maintain the AIC as low as possible. NN having lowered AIC, would save resources of service operators when used practically and this would allow supplying service for more clients.

After the review of all results of the experiment we can state, that Elman network is trained most effectively and produce the lowest bias, when parameters of three queued e-mails are analysed. Correlation ration of this type of network is due to 1, the value of Akaike criterion is lowest during the training of this type network. It can be firmly stated that Elman network secures best results using lowest outlay of resources. And for the further experiments the Elman structure network will be used. The number from inner layers of NN must be chosen in every case, because it is dependant from the amount of collected training data.

The data set, which reflects the most ultimate epidemic parameters, was created for the purpose of evaluation of what kind of queue bias is to be expected identifying virus infections in the nets. For the trained NN (the most effective Elman structure – 3:20:1 was chosen) unseen data set with inserted various epidemic data set of normal consumers was introduced.

The set which consists of four different epidemics were presented to the trained NN network. The NN network chosen according to the best obtained criterion during the training process; the most effective identification of epidemics is performed by Elman type networks, when three queued e-mails are analysed. Graphical results of the test are presented in image 12.

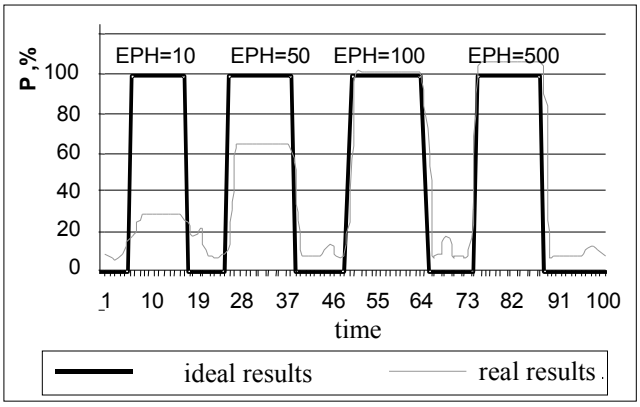


Fig.12. Graphical results of the epidemic identification test

The desirable NN network outlet was marked by wide black line. The high level of the signal presents big possibility of the epidemics, low level presents that the flow is normal. The thin grey line presents the results of the trained NN network. It can be stated that when epidemics MGS=10, the NN with ~30% possibility diagnoses that the epidemics is operating. When MGS=50, the NN presents the epidemics with ~60% possibility, and when MGS>100, the NN identifies the infected e-mail flow which causes epidemics with the precision of all 100%. But the diagram does not present the flow of epidemics, when NN has to identify the flow of normal e-mails, it can be stated that NN does not perform as we would like and returns the possibility of 10-20% that the epidemics is operating. This is the action causing false alarm, which is not desirable, the same as non-identification of epidemics. For its avoidance we offer to use the method of changeable sensibility threshold.

Conclusions of the section

1. The task of classification is well performed by the artificial Elman neural networks (NN). TLRN (Time Lagged Recurrent Network) NN having more complicated structures, in the course of training is often unstable; very often the meaning of the training errors ranges, while it doesn't happen so when using the same data containing, similar sized Elman networks in training. In most cases Elman networks are less likely to produce the bias, than TLRN networks respectively.
2. The criterion importance of Elman networks in Akaike is not as high as of TLRN networks. It shows that the Elman network training requires fewer resources of equipments and with less equipment better results may be achieved.
3. The maximum epidemics identification and the minimal false alarms percentage is achieved by analysing the parameters of the three queued e-mails. When analysing the three letters in the NN queue by means of cross validation set, it also correctly identifies 99 percent of epidemics and 0, 1 percent of normal letters it ascribes to epidemic flow of letters. It guarantees that epidemics will be found at its beginning.
4. The results of the experiment illustrate, that training artificial neural networks to classify virus infected and normal flow of e-mails the value of correlation ratio ranges from 0,95 to 0,99. It leads to the conclusion that data arrays of the beginner lever training are transformed correctly by means of the suggested method. The artificial neural networks acquire

knowledge how to qualitatively classify the data arrays.

5. The suggested method of software security systems effectiveness in valuing rates allows us to compare effectiveness rates of NN software security system with the effectiveness rates of the standard antiviral systems. The results of the comparison indicate that under the same circumstances NN software security effectiveness rate is higher than the standard antivirus system's effectiveness rate.

CONCLUSIONS OF THE THESIS

1. The analysis of virus identification methods shows that the systems which analyse changes of consumers' behaviour are more effective in identifying new viruses. Their software configuration profile can be easily closed in order to avoid people's interference, which is usually relatively slow, in their management.

2. It was suggested that virus infections in telecommunication should be modelled according to the models of epidemics types used in biology. On the base of these models the e-mail virus epidemic model is formed. With the help of it, we can measure the damage which was caused by e-mail virus epidemics as well as we can learn how much would it cost to renovate appliances, and at last, we can set parameters which most effectively suppress epidemics prevalence.

3. Created e-mail virus epidemic model can offer the most effective way to identify virus epidemics. Working algorithm of system was created; artificial neural networks data preparation training peculiarities were analysed and parameters which can evaluate the quality of artificial neural network training were offered in order to use artificial neural networks in e-mail flow analysis.

4. The method of software security system effectiveness evaluation was analysed, which offers the optimal choice of software security system means. The evaluation of software security system and its demand of adequacy rate can be evaluated with the help of this method.

5. The experimental research was conducted to identify epidemics according the offered method. Optimal parameters of artificial neural networks were set, which offer the highest quality of classification using the lowest degree of input. The experiment proved that when classifying e-mail flow artificial neural networks, the correlation ratio of artificial neural networks is equal to 0, 99.

6. The achieved results in the course of the experimental research allow us to evaluate the effectiveness of the suggested method, and to diagnose, that classification errors would not have negative influence on quality of supplied QoS services.

REZIUMĖ

Temos aktualumas

Viena iš labiausiai galutinius telekomunikacijų paslaugų vartotojus kamuojančių problemų yra elektroninio pašto virusai (toliau bus vartojama tik „virusai“), kurių kiekis kiekvienais metais vis didėja. Šie virusai– žmonių sukurtos kenkėjiškos programos siuntinėjamos el.pašto protokolu. Paskleistos programos interneto tinkle vykdo joms užprogramuotas funkcijas, geba save kopijuoti be vartotojų įsikišimo. Kompiuterių virusai yra pripažįstami pirmaisiais dirbtinio intelekto atstovais nes gali plisti ir daugintis kaip ir biologiniai virusai.

Problemas sukelia ne tik naujų virusų programų skaičiaus didėjimas, bet ir virusų sklidimo tinklais spartėjimas. Tobulėjant technologijoms virusai tinklais skinda vis greičiau ir reikalauja vis spartesnės gynybinės reakcijos. Prognozuojama, jog ateityje daugiausia problemų sukels ne lėtai plintantys virusai, o labai greitai plintantys „nulinės dienos“ (zero day), tinklus perkraunantys virusai, sukeliantys epidemijas ir mažinantys paslaugų QoS (Quality of Service).

Darbo tikslai

1. Pasiūlyti būdą, leisiantį efektyviai aptikti žinomų ir nežinomų virusų epidemijas telekomunikacijų tinkluose.
2. Patikrinti sukurto būdo funkcionavimo efektyvumą.

Darbo uždaviniai

1. Atlikti tarptautinių saugumo organizacijų siūlomų virusų aptikimo sistemų analizę. Nustatyti, kokios rūšies virusų epidemijos kelia didžiausią žalą.
2. Susipažinti su egzistuojančiais biologines epidemijas aprašančiais modeliais ir jų parametrais. Pritaikyti pasirinktą virusų epidemijų modelį pavojingiausių telekomunikacijų tinklų virusų epidemijų modeliavimui.
3. Atlikti eksperimentinį modeliavimą, įvertinti, kokie epidemijos parametrai turi didžiausią įtaką mažinant epidemijų sukeliamą žalą ir pasiūlyti efektyviausią virusų epidemijų stabdymo sistemos veikimo būdą.
4. Pasiūlyti būdą, kuris leistų efektyviai valdyti epidemijos parametrus, labiausiai mažinančius virusų epidemijų mastus ir žalą.
5. Eksperimentiškai patikrinti sukurto būdo efektyvumą.

Mokslinis naujumas ir praktinė vertė

Disertacijoje pateikta el.pašto tinklų virusų epidemijų modeliavimo ir žalos nustatymo metodika. Pasiūlytas būdas, kuris leidžia aptikti virusų epidemijas stebint el.pašto vartotojų elgseną. Šis būdas, panaudojus dirbtinius neuroninius tinklus, geba atpažinti naujų iki šiol nežinomų virusų epidemijas.

Sukurtas virusų epidemijų aptikimo telekomunikacijų tinkluose būdas leidžia sumažinti populiariausių virusų epidemijų žalą. Pasiūlyta apsaugos sistema padeda išvengti pagrindinio dabartinių apsaugos sistemų trūkumo– nesugebėjimo atpažinti naujų virusų. Įdiegus apsaugos sistemas, naudojančias šį būdą, sumažėtų tinklų apkrovimas virusų epidemijų metu, būtų apsaugoti vartotojai, padidėtų teikiamų paslaugų patikimumas bei pagerinta vartotojams teikiama tinklų paslaugų kokybė QoS. Svarbiausia– būtų sumažinta virusų epidemijų sukeliama žala. Pasiūlytam epidemijų aptikimo būdui pasitvirtinus, galėtų būti inicijuota naujo tipo antivirusinės įrangos kūrimo pradžia.

Darbo aprobavimas

Disertacijos tyrimų tema paskelbti 6 moksliniai straipsniai, 4 iš jų recenzuojamuose Lietuvos moksliniuose leidiniuose, įtrauktuose į Mokslo ir studijų departamento patvirtintą sąrašą.

Pagrindiniai mokslinių ir eksperimentinių tyrimų rezultatai aprobuoti tarptautinėse „Elektronika" konferencijose Kaune (2000m – 2004m.), konferencijoje „Kompiuterininkų dienos 2001" KTU, konferencijoje „Informacinės technologijos 2002" KTU, Lietuvos katalikų mokslo akademijos XIX konferencijoje (2003) ŠU.

Išvados

1. Virusų aptikimo metodų analizė parodo, kad atpažįstant naujus virusus efektyvesnės yra vartotojų elgsenos pasikeitimus analizuojančios sistemos. Jos leidžia uždaryti apsaugos sistemų valdymo kontūrą ir taip išvengti santykinai lėto žmonių įsikišimo į jų valdymą.

2. Pasiūlyta telekomunikacijų virusų epidemijų modeliavimui panaudoti biologijoje naudojamus epidemijų modelius. Šių modelių pagrindu sudarytas el.pašto virusų epidemijų modelis. Jo pagalba galima įvertinti el.pašto virusų epidemijų žalą bei įtaisų atstatymo kaštus, nustatyti parametrus, kurie efektyviausiai įtakoja epidemijų plitimą.

3. Sudarytas el.pašto virusų epidemijų modelis leido pasiūlyti efektyviausią būdą virusų epidemijų aptikimui. Norint dirbtinius neuroninius tinklus panaudoti el.pašto srauto analizei, sukurtas apsaugos būdo algoritmas, išnagrinėti duomenų paruošimo dirbtinių neuroninių tinklų apmokymui ypatumai, pasiūlyti parametrai leidžiantys įvertinti dirbtinio neuroninio tinklo apmokymo kokybę.

4. Parinkta ir išnagrinėta apsaugos sistemų efektyvumo nustatymo metodika, leidžianti optimizuoti apsaugos sistemos priemonių parinkimą. Šios metodikos pagalba galima įvertinti pasiūlytos apsaugos sistemos naudą ir jos poreikių atitikties laipsnį.

5. Atliktas pasiūlyto būdo epidemijų aptikimo eksperimentinis tyrimas. Nustatyti optimalūs dirbtinių neuroninių tinklų parametrai, leidžiantys užtikrinti geriausią klasifikavimo kokybę mažiausiomis sąnaudomis. Eksperimentu įrodyta, kad klasifikuojant el.pašto srautą dirbtiniais neuroniniais tinklais, koreliacijos koeficiento reikšmė siekia 0,99.

6. Eksperimentinio tyrimo metu gauti rezultatai rodo, kad naudojant pasiūlytą būdą, virusų epidemiją galima aptikti išanalizavus 3–10-ties el.laiškų parametrus. Eksperimentiškai nustatyta, kad pasiūlyto būdo epidemijų aptikimo klaidos, neturės neigiamos įtakos vartotojams teikiamų paslaugų kokybei QoS.