ISSN 1392 - 1215 ELEKTRONIKA IR ELEKTROTECHNIKA. 2003. Nr. 4(46)

T 180 TELEKOMUNIKACIJŲ INŽINERIJA

Neuronų tinklų panaudojimo duomenų perdavimo apsaugai tyrimas

R. Laurutis
Telekomunikacijų katedra, Kauno technologijos universitetas
Studentų g. 50, LT - 3031 Kaunas, Lietuva, el.paštas remigijusl@aleja.lt

Įvadas

Kompiuterinėse tinklų sistemose dažnai plinta programos, kurios geba kurti savo kopijas. Dažnai jos pakenkia kompiuterinių sistemų veiklai. Tokios programos vadinamos kompiuterių virusais, arba piktadariškomis programomis. Dauguma (~80%) šių programų plinta internetu siunčiamais elektroniniais laiškais[1]. Piktadariškų programų padariniams pašalinti kasmet skiriama 10-15 milijardų dolerių.

Dažniausiai piktadariškoms programoms aptikti, komerciniai produktai naudoja šablonų atpažinimo metodus. Gaila, tačiau šiais metodais negalima sustabdyti ir padarytini nekenksmingomis naujai sukuriamų piktadariškų programų, kurių vis daugėja. 2002 metais buvo rasta 7000 naujų kompiuterinių virusų ir piktadariškų programų. Antivirusinių sistemų gamintojai pripažįsta, jog tradicinė kova su piktadariškomis programomis gali būti pralaimėta, jei jų skaičius ir stabdymo efektyvumas didės jiems būdingais greičiais. Asociacijos MessageLabs renkama statistika rodo, jog 2003 metų sausio mėnesį patikrinus 279,4 milijono el.laiškų, 1,3 milijono laiškų buvo rasta virusinių programų [2]. 85% kompiuteriu dirbančių žmonių, dalyvavusių apklausoje, buvo susidūrę su kompiuteriniais virusais [7].

Kovos su piktadariškomis programomis problemos

Dažnai manoma, jog aktualių sričių tiriant kenkėjiškas programas ir kovos būdus su jomis nėra, nes visos svarbiausios technologijos jau yra sukurtos ir šiuo metu lieka paprasta programavimo problema. Dar manoma, jog piktadariškų programų tyrinėjimas asocijuojasi su piktadariškų programų analizavimu. Tačiau taip nėra, nes literatūroje radome kelias kompiuterinių virusų tyrinėjimų ateityje gaires[3].

1. Daugėjant virusinių programų, reikia sukurti naujas, euristines, technologijas. Jų, kaip ir kitų prognozavimo sistemų trūkumas bus klaidingai atpažintų ir neatpažintų įvykių skaičius. Šį parametrą būtina maksimaliai sumažinti.

2. Reikia suprasti piktadariškų virusinių programų epidemiologiją. Apibūdinti ją sklidimo greičiu, gimstamumu, mirštamumu ir sklidimo terpe. Čia analogiją galima rasti su gerai ištirta biologine epidemiologija.

3. Reikia sukurti skaitmenines imunines sistemas, kurios aptinka, analizuoja ir skleidžia antivirusus automatiškai ir taip greitai, kaip greitai plinta virusinės programos. Dabartinė, centralizuota, sistema (aptinkami virusai siunčiami į laboratoriją, specialistai išanalizavę sukuria aptikimo šablonus) yra neperspektyvi.

4. Dabartinės virusines programas naikinančios technologijos yra reaktyvios (atsirado virusas sukuriamas antivirusas). Reikia kurti aktyviąsias ir intelektualiąsias sistemas kurios pačios neleistų plisti piktavališkoms programoms, o ne paprasčiausiai atsakytų į jų sukeltą reakciją [3].

Elektroninio pašto virusų paplitimo priežastys

Išnagrinėjome terpes, kuriomis gali plisti kompiuterių virusai. Pirmieji virusai kaip plitimo terpe naudojo lanksčiuosius diskelius, todėl plisti galėjo tik ribotoje teritorijoje. Tobulėjant technologijoms, susikūrus tarptautiniam interneto tinklui, buvo sukurta galimybė perduoti duomenis labai įvairiais protokolais. Vienas iš jų- SMTP (angl. Simple mail transfer protocol) dažniausiai naudojamas interaktyviai keičiantis duomenimis. Apžvelgę literatūrą, radome šias el. pašto virusų paplitimo priežastis:

- beveik kiekvienas kompiuterio vartotojas turi el. pašto adresą;
- nemaža kompiuterių vartotojų nedaug tesupranta apie kompiuterinius el. pašto virusus;
- siųsti virusus el.pašto protokolu yra labai paprasta;
- el. paštas yra privatus, kaip ir paprastas paštas, todėl jo patikrinimo taisykles deklaruoja įstatymai [5].

Dėl šių priežasčių apie 80% šiuolaikinių kompiuterinių virusų plinta elektroniniais laiškais [1].

Kompiuterinių virusų plitimo modelis

Kompiuterinius virusus mokslininkai nagrinėja jau seniai. F. Cohen [4], šių sistemų tyrinėjimo pradininkas, suformavo teorinius kompiuterinių virusų pagrindus, dar tada, kai kompiuterinių virusų nebuvo. Modeliui sudaryti buvo naudojama analogija su biologine epidemiologija, kuri yra gana gerai ištyrinėta.

Virusų plitimo modeliai esti SIS ir SIR (angl. angl. susceptible - infectious - susceptible ir susceptible - infectious - resistant) tipų. SIS modelis (atsparus-infekuotas-atsparus) naudojamas tuo atveju, kai vienos rūšies virusu, vienas individas gali užsikrėsti kelis kartus, t.y. išgydytas neįgyja imuniteto. SIR (neatsparus-infekuotas-atsparus) modelis tinka tuo atveju, kai užkrėstas ir išgydytas individas tuo pačiu virusu jau nebeserga

(pavyzdžiui, kompiuteryje įdiegiama antivirusinė programa).

Atitinkamas n dydžio segmentas yra padalytas į X galimus infekuoti, Y infekuotus ir Z išgydytus kompiuterius. SIR modelį galime aprašyti taip:

$$\frac{dX}{dt} = -\lambda X \frac{Y}{X+Y}; \quad (1)$$

$$\frac{dY}{dt} = -\lambda X \frac{Y}{X+Y} - \delta Y; \quad (2)$$

$$\frac{dZ}{dt} = Z' = \delta Y; \quad (3)$$

čia $\lambda = c \cdot B$; c - vidutinis kontaktų tarp kompiuterių skaičius; B - tikimybė, jog kontakto metu bus perduotas virusas; λ - išgydymų tempas; t - laikas tarp matavimų keičiantis X, Y ir Z.

Matyti, jog santykis $1/\lambda$ reiškia vidutinį kompiuterio ligos laiką iki išgydymo. Fundamentali problema, apskaičiuojant ar epidemija prasidės, yra jautrumo slenksčio parametro reikšmė $R0$. $R0$ paprastai yra skaliarinė funkcija, aprašoma daugiadimensėje terpėje. Daugumoje formuluočių, jei $R0<1$, epidemijos nebus, tačiau jei $R0>1$, epidemija paplis visame segmente. SIR modelyje $R0$ galima aprašyti taip:

$$R0 = \frac{cB}{\delta} = \frac{\lambda}{\delta}; \quad (4)$$

Po tokio R0 aprašymo modelį galime perrašyti taip:

$$Y' = \delta Y\left(R0\frac{X}{X+Y} - 1\right); \quad (5)$$

$$R0 = \frac{\lambda}{\delta}\left(\frac{n}{n+a}\right); \quad (6)$$

čia n - segmento dydis; a- pradinis infekuotų kompiuterių skaičius.

Šios lygybės rodo jog dideliam neatspariam segmentui sąlyga, kai $R0=1$, yra riba tarp išgijimo ir epidemijos. Jei $R0<1$, tai $Y'<0$. Šiuo atveju $Y(t)$ mažėja, kai t didėja. Kai $R0>1$ ir pradinis infekuotų kompiuterių skaičius yra mažas, $Y'>0$. Tai rodo jog segmente plinta epidemija [6].

Išanalizavę šiuo bei kitais modeliais [5,6] gautus modeliavimo rezultatus, darome tokias išvadas:

a) aukštas segmentacijos lygis mažina kompiuterinių virusų epidemijos galimybę ir mastus;

b) atskirų klasterio segmentų imunizavimas gerokai mažina epidemijos plitimo mastus;

c) hierarchinėje tinklo topologijoje (tokioje kaip interneto tinklas) imunizavus apie 5% mazgų, turinčių daugiausiai jungčių su kitais mazgais (domeninių serverių), epidemijos tikimybę galima sumažinti beveik dvigubai [5].

Intelektuali piktavališko elektroninio pašto stabdymo sistema

Įvertinus pateiktą informaciją, galima pasiūlyti eksperimentinį intelektualios antivirusinės sistemos modelį. Pagrindiniai jo veikimo principai turėtų būti šie:

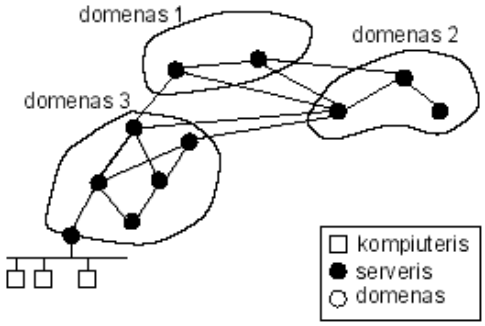
a) apsauga nuo piktavališkų programų plitimo efektyviausiai veikia ne galutinio vartotojo kompiuteryje, o mazge, turinčiame kuo daugiau ryšių su kitais. Kadangi dideli virusų kiekiai sklinda el. paštu, sisteamą tikslinga įdiegti el. paštą aptarnaujančiose stotyse;

b) apsaugos sistema turėtų būti ne reaktyvi, o aktyvi, t.y. stabdyti ne tik žinomas bet ir naujas piktavališkas programas. Čia naudotinos neuronų tinklų sistemos;

c) sistema turi veikti automatiškai, žmonėms ne įsikišant.

Praktikoje jau plinta panašiomis savybėmis pasižyminčios sistemos, tačiau dažniausiai jos veikia virusų šablonų atpažinimo principu. Žinomus kompiuterinius virusus jos stabdo, įtartinus siunčia į laboratorijas, kuriose specialistai juos ištiria ir priima sprendimą. Tai gana efektyvus, tačiau palyginti lėtas ir neperspektyvus metodas.

MET antivirusinę sistemą (angl. malicious email tracking) sudaro klientinė ir serverinė dalys [8]. Klientinė dalis integruojama į elektroninio pašto serverius domenuose.



1 pav. Domeninė interneto struktūra

MET serveryje įdiegiamas intelektualus mechanizmas, analizuojantis iš klientų gautą informaciją ir priimantis sprendimus.

Tokią sistemą pasiūlė Kolumbijos (JAV) universiteto mokslininkai. Tačiau jų pasiūlytoji eksperimentinė sistema veikia statistinių duomenų analizės principu.

Mūsų siūlomas principas - įdiegti į sistemą (MET klientą bei serverį) neuronų tinklą sprendimams priimti žmonėms neįsikišant.

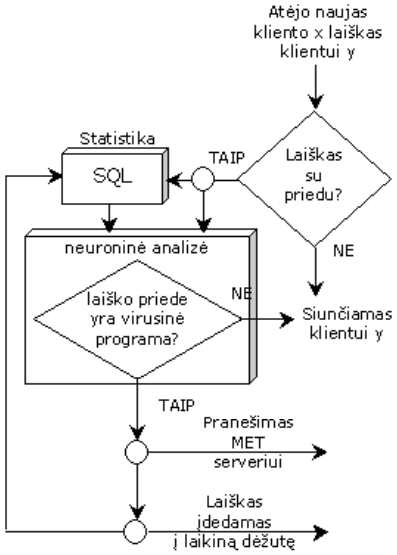
MET klientas

MET kliento užduotis - rinkti informaciją apie pašto serverio persiunčiamus laiškus ir jų priedus (angl. attachment), kuriuose gali būti kompiuterinių virusų. Surinkta informacija atiduodama analizuoti neuronų tinklui, kuris priima sprendimą, ar siunčiamas laiškas yra piktavališkas, ar ne.

MET serveris

MET serveris yra sudarytas iš komunikacinio mechanizmo su MET klientais duomenų bazės statistikai rinkti, bei valdymo modulio. Valdymo modulis, analizuodamas iš MET klientų gaunamą informaciją, priima sprendimus ir perduoda juos vykdyti MET klientams. Dėl to mažėja klaidų skaičius, nes, pavyzdžiui,

jei siuntėjas išsiuntė šimtui kolegų kvietimus, MET serveris matys, jog tai vienetinis atvejis internete, ir nestabdys laiškų. Šiuo atveju MET klientas, veikdamas autonomiškai ir neturėdamas duomenų apie kitų siuntėjų statistiką, gali priimti klaidingą sprendimą ir neišsiųsti tokių laiškų.



2 pav. MET kliento veikimo algoritmas

Eksperimentinis neuronų tinklo modelis

Vykdant piktavališkų programų epidemijos atpažinimo uždavinį, visų pirma reikia surinkti statistinę informaciją, kurią neuronų tinklas galės apdoroti. Siūlome rinkti 1 lentelėje pateiktus duomenis.

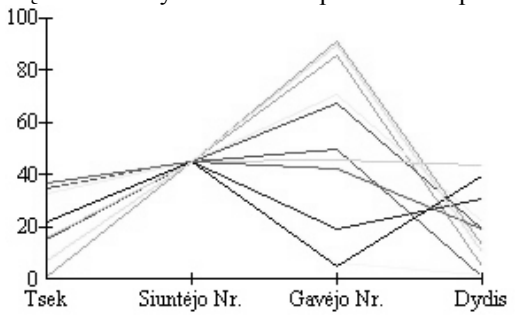
1 lentelė. Siūlomi rinkti statistiniai duomenys

Charakteristika	Paaiškinimas
Laiško identifikacinis numeris	Turi visi el. laiškai
Siuntėjo adresas	Statistiškai rinkti (Siuntėjo ID)
Gavėjo adresas	Statistiškai rinkti (Gavėjo ID)
Priedo dydis	Toks pats priedo dydis, pvz. šimte laiškų yra įtartinas
Laiško tema	Vieno viruso el. laiško temos dažnai būna vienodos
Laikas	Virusų dauginimosi greičiui skaičiuoti

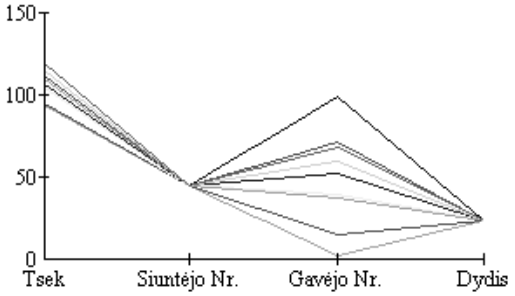
Eksperimentui sukūrėme 100 el.pašto siuntimo įvykių, iš jų 80 be virusų, 20 su jais. Kiekvienam įvykiui suformavome vektorių. Vektorių vaizdai pateikiami 3 ir 4 paveiksluose.

Iš 3 paveikslo matyti jog įvairiais laiko momentais Tsek siuntėjas Nr. 46 siuntė įvairiems gavėjams skirtingo dydžio laiškus. 4 paveiksle matyti, jog per trumpą laiko tarpą tas pats siuntėjas skirtingiems gavėjams išsiuntė daug vienodo dydžio laiškų. Neuronų tinklo užduotis šiuo atveju - atskirti tipinius ir netipinius vartotojų veiksmus.

Eksperimentui vykdyti su paketu NeuralSolutions sukūrėme neuronų tinklą, jį apmokėme atpažinti tipinius bei nenormalius siuntėjų veiksmus (kai virusas pradeda savaime daugintis). Eksperimentui buvo naudojamas vienas populiariausių praktikoje daugiasluoksnis perceptronas MLP, kurio įėjimo sluoksnyje yra 4, vidiniame 10 ir išėjimo sluoksnyje 1 neuronas [15]. Apmokymui buvo naudojamas Quickprop algoritmas. Neuronų tinklo mokymo rezultatai pateikiami 5 paveiksle.



3 pav. Dešimties el. laiškų be virusų siuntimo įvykių vektoriai



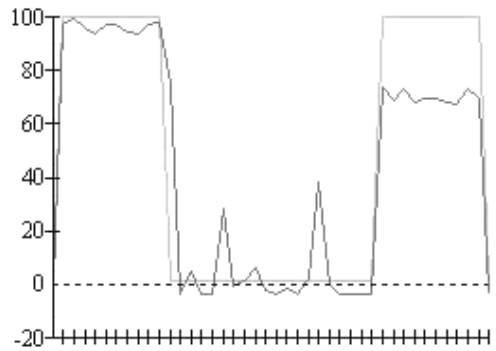
4 pav. Dešimties el. laiškų su virusais siuntimo įvykių vektoriai

5 pav. Neuronų tinklo apmokymo klaidų grafikas

Matome, jog tinklas buvo sėkmingai apmokomas, klaidų mokymo procese vis mažėjo. Norėdami nepermokyti tinklo, mokymui skyrėme 1000 epochų (ciklų).

Su apmokytu tinklu atlikome testavimą. Sukūrėme dar vieną duomenų aibę, kurią turėjo apdoroti apmokytas neuronų tinklas. Sugeneruotus įvykių (laiškų siuntimo) duomenis pateikdavome apdoroti neuronų tinklui. Neuronų tinklas išanalizavęs ar duomenys turi infekcinį pobūdį gražindavo tikimybę ar laiškas užkrėstas virusu ar ne. Pateikiame lauktų ir gautų rezultatų palyginimo grafiką. Ašyje y pavaizduota procentinis idealių ir gautų rezultatų palyginimas, jog laiškas užkrėstas virusu, ašyje x - laiškų seka laiko eigoje. Dvi smailės rodo du gana klaidingus neuronų tinklo atsakymus.

Turėdamas tokius rezultatus, MET klientas gali nuspręsti, ar laiškas yra užkrėstas, ar ne.



6 pav. Lauktų ir gautų eksperimento rezultatų palyginimas

Klaidas (dvi smailės grafike) turėtų minimizuoti MET serveris, kuris lygindamas daugelio MET klientų rezultatus nuspręstų ar įtartinas laiškas iš tikro yra kenkėjiškas ir jo dauginimosi parametrų įverčiai siekia epidemijos lygį.

Išvados

1. Atlikta statistinė analizė parodė, jog dauguma šiuolaikinių virusų plinta el.pašto protokolu, todėl tikslinga sukurti el. pašto apsaugos sistemas.
2. Analizė ir matematinio modeliavimo išvados leido suformuoti pagrindinius intelektualios antivirusinės sistemos kūrimo principus.
3. Pasiūlyta MET kliento - serverio architektūra naudojanti neuronų tinklą, leido sistemą padaryti intelektualią, suteikti jai nuolatinio mokymosi funkciją.
4. Eksperimento metu gauti rezultatai parodė, jog neuronų tinklas gali klasifikuoti saugius ir nesaugius laiškus analizuodamas statistinius duomenis.
5. Saugių ir virusu užkrėstų laiškų atpažinimo klaidas tikimasi kompensuoti MET serverio renkama statistine informacija.

Literatūra

1. **Best Don.** Virus Protection Via Postini's Email Pre-processing Infrastructure, October 2000. Prieiga per internetą <http://www.postini.com/company /pr/pr100200.html>.
2. MessageLabs VirusEye Volume 2 Issue 1.- January 2003. Prieiga per internetą <www.MessageLabs.com/ VirusEye>.
3. **White Steve R.** Open Problems in Computer Virus Research // IBM Thomas J. Watson Research Center.- Presented at Virus Bulletin Conference. - 1998.
4. **Cohen F.** Computer Viruses:Theory and Experiments.- 1984. Prieiga per internetą <http://vx.netlux.org/ lib/afc01.html>.
5. **Zou Cliff Changchun, Towsley Don, Gong Weibo.** Email Virus Propagation Modeling and Analysis // University of Massachusetts, Amherst.- 2002.
6. **Gallop Robert J.** Modeling General Epidemics: SIR MODEL // University of Pennsylvania, Philadelphia, PA. - 1996.
7. **Rabinovitch Eddie.** Securing your Internet connection // IEEE Communication magasine. - June 2002.
8. **Bhattacharyya Manasi, Schultz Matthew G., Eskin Eleazar, Hershkop Shlomo, Stolfo Salvatore J.**. MET: An Experimental System for Malicious Email Tracking// Department of Computer Science, Columbia University. - 2002.
9. **Bishop Cristopher M.** Neural nework for Pattern Recognition. Oxford University Press Inc. - 1998. – P. 16-34.

Pateikta spaudai 2003 02 02

R. Laurutis. Neuronų tinklų panaudojimas duomenų apsaugai tyrimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2003. – Nr. 4(46). – P. 61-64.
Apie 80 % kompiuterinių virusų plinta elektroniniu paštu. Jų sukeltiems padariniams likviduoti kasmet išleidžiama 10-15 milijardų dolerių. Kasmet kompiuterinių virusų kiekis vis didėja, o šiuolaikinės sistemos nemoka efektyviai sustabdyti naujai atsirandančių kenkėjiškų programų. Šiame darbe bandoma suprojektuoti biologinės epidemiologijos stabdymo principais veikiančią sistemą, valdomą neuronų tinklo. Darbo metu buvo padarytos išvados jog tikslinga antivirusines sistemas diegti ne galinių klientų kompiuteriuose, o kolektyvinėse el.pašto serveriuose, taip suskaidant tinklą į kuo mažesnius segmentus. Atliktas sistemos projektavimas, atliktas užkrėstų ir neužkrėstų el.laiškų klasifikavimo panaudojant neuronų tinklus eksperimentas. Eksperimentas parodė šios technologijos lankstumą ir perspektyvumą. Il. 6, bibl. 9 (lietuvių kalba; santraukos lietuvių, anglų, rusų k.).

R. Laurutis. Application of Neural Networks for Data Protection Research // Electronics and Electrical Engineering. – Kaunas: Technologija, 2003. – No. 4(46). – P. 61-64.
About 80% of computer viruses are distributed by e-mail. Approximately 10-15 billion US dollars are spent annually for the liquidation of their outcomes. With every year, the quantity of computer viruses is increasing, and the contemporary systems are not able to stop this newly appearing damaging software. In this work, the attempts are taken to design a system that operates basing on the principles of biological epidemiology arrest and is controlled by the neural network. Conclusions are made in this work that it is advisable to install antiviral systems in common e-mail servers instead of the computers of final users, and also to group the network into the smallest possible segments. System design has been carried out and experiment of classification of infected and not infected messages using neural networks has been done. The experiment has shown the flexibility and prospects of this technology. Ill. 6, bibl. 9 (in English, summaries in Lithuanian, English, Russian).

Р. Лаурутис. Исследование использования нейроновых сетей для защиты данных // Электроника и электротехника. – Каунас: Технология 2003. – № 4(46). - С. 61-64.
Около 80% компьютерных вирусов распространяются по электронной почте. Для ликвидации последствий, вызванных такими вирусами, ежегодно тратится 10-15 миллиардов долларов. Число компьютерных вирусов растет с каждым годом, а современные системы не умеют эффективно остановить сново и сново появляющихся вредительских программ. Эта работа - попытка проектирования системы, действующей на принципах биологической остановки эпидемии, управляемой нейроновой сетью. В ходе работы сделаны выводы, что целесообразно внедрять антивирусные системы не в компьютерах у конечных пользователей, а на коллективных серверах электронной почты, путем расчленения сети на сегменты по возможности меньшего размера. Выполнено проектирование системы, проведен эксперимент классификации зараженных и незараженных электронных писем с применением нейроновых сетей. Эксперимент показал гибкость и перспективу такой технологии. Ил. 6, библ. 9 (на английском языке; рефераты на литовском, английском и русском яз.).