一种基于 BP 神经网络的智能检测病毒方法

朱俚治
(南京航空航天大学 信息中心,江苏 南京 210016)

摘 要:病毒技术历经半个世纪的发展,如今已不是单纯的病毒技术,而是融合了其他黑客等技术,因此当今病毒的危害性和传播速度远远超过了病毒原始形态。病毒的传染性以及病毒的变种技术在病毒上的应用使得病毒呈现一定的智能性。因此,为了应对病毒的变种技术以及病毒其他方面所呈现出的智能性,文中参考已有的检测病毒方法之后,提出了一种具有智能性的检测病毒的方法。文中使用沙箱作为检测病毒的运行环境,并使用 BP 神经网络作为检测病毒的工具,然后给出了一种具有一定智能性的病毒检测方法。该方法能够对被怀疑的非法变化的程序中是否存在病毒做出判断。

关键词:病毒;BP 神经网络;沙箱
中图分类号:TP301 **文献标识码**:A **文章编号**:1673-629X(2014)10-0163-04
doi:10.3969/j.issn.1673-629X.2014.10.039

An Intelligent Virus Detection Method Based on BP Neural Network

ZHU Li-zhi
(Information Center, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: After several decades of development, the virus technology today is not merely the simple virus technology, but having combined with other technologies such as hacking and so on. So the hazard and transmission speed of virus today is over the original virus. The contagiousness of the virus and virus variant technology applied in virus has made them show certain intelligence. In order to deal with the intelligence viruses present, put forward an intelligent virus detection method after referring to existing methods. While using the sandbox as the operating environment of virus detection, also use the BP neural network as the tools. Afterwards, a new method of virus detection is given, which will be able to help to detect whether the suspected program exists viruses.

Key words: virus; BP neural network; sandbox

0 引 言

病毒的出现已有四十年左右时间,在这些年中因计算机病毒的破坏而造成的巨大损失给人们留下了深刻印象。病毒的出现是计算机技术开发人员和广大网络用户不愿见到的负面现象。因此为了减少病毒给人们带来的损失,于是反病毒人员开发出了各种各样的病毒检测方法。但目前,大部分的病毒检测方法都是非智能性的,在检测病毒的过程中需要人工干涉,而且目前使用单一的传统检测方法对病毒的查杀效果已十分不理想,因此如今在查杀病毒时,不得不将传统的杀毒方法综合使用才能达到较为理想的效果。但由于病毒技术的更新总比反病毒技术更新来得快,再由于目前部分病毒采用了病毒的变种技术,这些因素使得检测出未知种类的病毒变得更加困难。

事实上只有当病毒寄生在某些程序中后,才能破坏系统中其他的程序,因此,当病毒进入用户的系统之后,就能破坏系统中的程序并不断消耗系统资源等等。计算机用户为了避免染上病毒,首先必须检测出病毒[1]。但病毒技术发展迅速,并且采用了智能性技术。病毒的变种技术就是一种病毒的智能技术。因此,为了有效应对目前病毒技术的快速发展,反病毒技术必须融入人工智能技术,智能技术能够使得反病毒技术提高一个水平,并且智能技术在反病毒上的应用是反病毒技术将来的发展趋势。智能技术在查杀病毒上的

收稿日期:2013-11-08 **修回日期**:2014-02-13 **网络出版时间**:2014-07-17
基金项目:国家"863"高技术发展计划项目(2009AA043303);北京航空航天大学软件开发环境国家重点实验室开放基金资助项目(SKLSDE-2013KF)
作者简介:朱俚治(1980-),男,工程师,研究方向为计算机网络和信息安全。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20140717.1229.024.html

应用,能够使得杀毒软件具有智能性,使之能够更好地应对病毒的变种技术。杀毒软件的智能化能够使得网络用户的系统变得更加安全,减少病毒对用户系统的威胁。因此,查杀病毒技术的智能化,对遏制目前病毒技术的发展是十分有利的。基于上述原因文中使用 BP 神经网络作为检测病毒的工具,提出了一种具有一定智能性的病毒检测方法[2]。

1 程序的运行环境

1.1 虚拟技术与虚拟机

系统虚拟机是通过虚拟技术来实现的,同样沙箱也是由虚拟技术实现,因此沙箱技术属于虚拟技术。系统虚拟机是由各种不同功能的程序相互融合成一个整体。系统虚拟机目的是采用虚拟技术模拟一个与真实系统一样的系统,系统虚拟机是一个对真实系统的仿真结果,虚拟技术可以模拟真实系统中各种硬件和软件。在系统虚拟机中的各种程序运行时,该系统与真实系统是相互隔离的,因此虚拟技术所仿真的系统对自身以外的系统是有隔离机制和安全机制的[3]。

1.2 虚拟技术与沙箱

由于沙箱技术来源于虚拟技术,在沙箱出现之前,系统虚拟技术在检测病毒方面就有了应用。程序在沙箱技术模拟的系统中运行不会对真实的系统产生影响,不会破坏真实系统中的软件和硬件资源。由于沙箱同样具有系统虚拟机的作用和功能,因此沙箱能够仿真一个系统虚拟机,采用沙箱技术仿真的系统虚拟机能够运行被怀疑的非法变化程序[4]。

文中提出的病毒检测方法在运行时,除了需要运行被怀疑的程序外,还需要运行若干原始程序。如果被怀疑的程序带有病毒,那么病毒也就处于运行中。病毒处于运行之中,就能够对用户的系统中程序进行传染和对程序造成破坏。因此,文中采用沙箱作为检测病毒时的运行环境。病毒在沙箱中运行时,带有病毒的程序对用户的真实系统没有破坏性,用户的真实系统是安全的。

2 人工神经网络的简介

2.1 人工神经网络技术

神经网络技术是一种智能技术,神经网络的工作方式由两个阶段组成,学习期和工作期[5]。

神经网络整体性能的三大要素:

(1)神经元(信息处理单元)的特性;

(2)神经元之间相互连接的形式--拓扑结构;

(3)为适应环境而改善性能的学习规则。

神经网络有以下基本特征[5]:

(1)知识与信息的存储;

(2)网络的学习和识别;

(3)具有联系记忆能力;

(4)网络的信息处理能力。

人工神经网络是开发人员通过对人类大脑的神经网络的模拟,利用模拟技术使得人工神经网络对外界具有智能性。因此人工神经网络是人工智能技术的一个分支,开发人工神经网络的目的是使得机器具有人类的智能。目前的神经网络对样本中的数据具有自我学习功能,网络中的神经元通过自我学习能够从这些数据中找出其中的规律,使得输出的数据具有预测性,同时也使得输出数据的特征具有样本数据的属性和特征。

2.2 BP 神经网络技术

神经网络发展至今已由单层的感知器,发展为多层网络。神经网络中的 BP 网络是前馈型网络[6-9],该网络由三层神经元组成(如图 1 所示)。样本由输入层传递给隐含层,再由隐含层传递给输出层。隐含层是 BP 网络的核心层,具有自我学习的功能。BP 神经网络通过不断调整神经网络中神经元的权值,使得样本的输出值逐渐逼近期望值[10]。BP 神经网络在病毒的检测这一方面已有应用,因此,文中再次将 BP 神经网络在检测病毒上进行应用,使得病毒检测方法具有一定的智能性。文中在此使用 BP 神经网络作为智能检测病毒的一种工具,再次说明了人工神经网络是一种智能技术。同时文中采用了人工神经网络这一工具,也再次说明了人工神经网络使得机器具有智能性这一方面的应用是广泛的。

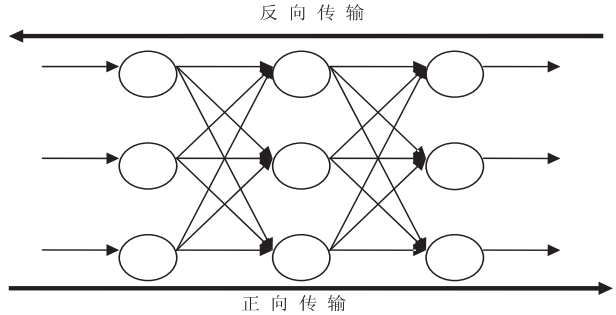

图 1 BP 神经网络示意图

3 病毒检测

3.1 病毒技术和特性简介

(1)病毒的复制和寄生。

病毒是一种具有某种智能性的恶意程序。病毒的恶意体现在传染性和破坏性,病毒的传染性是病毒消耗和破坏用户系统资源的前提条件,如果病毒不能寄生于某些程序,那么它就无法破坏被寄生的程序,因此病毒要破坏系统的程序必定要寄生于程序。当然除了病毒能够破坏系统的程序外,其他的恶意程序同样也

能消耗和破坏系统的程序和资源,但它们的破坏机理和破坏条件是有所差别的。

病毒在寄生某些程序时,由于被寄生的程序的不同,病毒采用不同的寄生方式。不同的寄生方式,对程序的破坏效果是有所区别的。病毒的寄生程序不同,病毒的检测方式也就不同。当今病毒为了躲避杀毒软件的查杀采用不同的寄生方法和寄生技术。如果按照病毒的寄生方式的不同,将病毒的寄生方式列举为以下几种:在程序的头部寄生;在程序的尾部寄生;在程序的中部寄生[11]。

当今病毒为了躲避杀毒软件的查杀,采用某些计算机技术使得病毒具有自我变种能力,具有变种能力病毒能够不断产生新的病毒,因此如果杀毒软件的病毒库没有及时更新,那么变种病毒就不能被杀毒软件查杀。不论是病毒的变种,还是病毒自身,亦是采用不同寄生方式的病毒,它们都是病毒,病毒就具有病毒所有的属性。既然所有的病毒都具有相同的属性,那它们就都具有传染性和破坏性,因此文中所采用的病毒的感染实验法能够检测出所有的病毒。

一种病毒进入用户系统之后,在它的周期中病毒具有以下的几种特性:传染性、潜伏性、发作期、破坏性[12]。病毒的特性决定着病毒的运行方式和工作机理,虽然有些恶意程序具有代码复制的功能,但不具备寄生能力。而病毒既能复制自身的代码又能将自身的代码和别的程序融合成一个完整的有机程序,因此病毒的传染性是病毒这一恶意程序所特有的。

(2)程序和病毒的运行。

程序要运行,就必须作为进程调入内存。因此,当携带有病毒的程序被执行时,病毒就会被加载到内存,作为进程进入内存的病毒,便开始寻找每一个符合传染条件的程序,只要条件满足,病毒就能进行传染。但由于病毒寄生于程序后,会引起原始程序的变化,因此,文中提出的病毒检测算法能根据原始程序发生的变化,并以此为根据来判断有变化的程序是否染上病毒。

3.2　BP 神经网络的输入

(1)提取被怀疑的程序;

(2)选取若干与被怀疑程序后缀名相同的原始测试程序,后缀相同的程序则为同一类程序;

(3)提取被怀疑程序中有变化的部分代码,作为样本输入 BP 神经网络;

(4)在原始测试程序中提取代表该类程序的特征代码,作为样本输入 BP 神经网络;

(5)将样本按固定长度 n 进行分割[13];

(6)将样本字符组成的片段输入堆栈,片段字符以 nbyte 的流量,做为样本输入 BP 神经网络[8]。

3.3　BP 神经网络的输出

检测的程序样本输入神经网络时,以二进制式的数据流输入 BP 神经网络,经过 BP 神经网络的学习和计算,最后以二进制的数据流输出。

文中将输出的数据流,按照固定长度 n 对数据流进行分割,将长度为 n 的数据流依次输入堆栈,最后将堆栈输出的数据流进行二进制与十进制的转换,这样就能计算出输出程序的大小。

3.4　病毒检测函数

$y=f(x)=\frac{x_1}{x_2}$,x_1 = 原始输出程序的值-原始输入样本的值,x_2 = 被怀疑程序的值-未产生变化的原始程序的值。

病毒检测具体算法的过程如下:

(1)对被怀疑的程序做以下的运算:a = 被怀疑程序的值-未产生变化的原始程序的值;

(2)利用沙箱技术构造一个程序的运行系统;

(3)将被怀疑的程序和原始测试程序调入 BP 神经网络;

(4)计算输出原始测试程序值的大小 $b_1,c_1,e_1,\cdots$;

(5)将输出程序数据 $b_1,c_1,e_1,\cdots$与原始测试程序数据 $b,c,e,\cdots$进行相减,得 $b_2=b_1-b$,$c_2=c_1-c$,$e_2=e_1-e$,$\cdots$;

(6) $y_1=\frac{b_2}{a}$, $y_2=\frac{c_2}{a}$, $y_3=\frac{e_2}{a}$, $\cdots$, y_n;

(7)如果 y_1, y_2, $\cdots$, y_n 近似相等,并且都收敛于 1,则被怀疑的程序感染上病毒的几率增大;

(8)如果(7)不成立,则通过人工分析该被怀疑的程序是否未感染上病毒。

3.5　算法分析

要阻止病毒的传播,就要发现病毒。如何检测出病毒,是反病毒研究人员的重点研究方向。病毒寄生于程序后,就会引起原始程序的变化。由于病毒的寄生能够引起原始程序的变化,所以通过比较法能够找出变化的程序。但程序的变化有合法的,也有非法的,因此文中提出的新方法可以区分合法变化和非法变化的程序。并能检测变化程序中的病毒。

由于病毒检测技术的智能化,是今后病毒检测技术的重点方向之一。因此,文中提出一种智能性的病毒检测方法。该方法中使用了两种已有的技术:沙箱和 BP 神经网络。

沙箱能够在运行程序时提供一个安全的环境,使用 BP 神经网络能够使得检测方法具有智能性。但提出的智能性算法尚有局限性。因为检测病毒时,需要原始程序的存在。

在这里对 BP 神经网络输出的结果做以下的讨论：

(1)各自的输出程序与各自的输入程序做相减的运算,这时得到的差值就是各个程序变化的大小。

(2)被怀疑的程序与原始程序作相减运算,得到的差值就是被怀疑程序的变化大小。

如果由(1)得到程序变化的值大小与(2)得到程序变化的值大小相同,则被怀疑的程序染上了病毒。因此,文中使用 BP 神经网络对被怀疑程序进行检测,将得出输出程序值的大小与各自的原始程序进行比较,就可以判断被怀疑的程序是否染上病毒。

4 结束语

尽管病毒是一种有破坏性的恶意代码,但病毒的出现使得人们发现了程序存在的各种脆弱性,从而有了为程序打上补丁的概念,同时也有了查杀恶意程序的技术和经验,这样使各种程序变得更加安全和可信度更高。

如今由于病毒与黑客技术、蠕虫技术相互融合,使得进入计算机用户系统中的病毒更难被杀毒软件发现和清除,具有智能性和产生病毒的变种是当今病毒的一大特点,因此,为了提高检测病毒的效率,反病毒技术采用智能技术将使得杀毒软件更具有智能性和更高的效率。由于反病毒技术采用智能技术是反病毒技术发展的必然方向,智能技术在杀毒和查毒方面的逐步应用,笔者认为是一种创新上的应用,将是十分鼓舞人心的。

参考文献:

[1] 朱俚治. 病毒检测技术的研究与 0.5 级环[J]. 计算机技术与发展,2012,22(9):225-227.

[2] 朱俚治. 一种防病毒智能网络接口的方案设计[J]. 中国科技信息,2011(1):73-74.

[3] Smith J E, Nair R. 虚拟机[M]. 安 虹,张 昱,吴俊敏,译. 北京:机械工业出版社,2011.

[4] 彭 晖,常 乐,沈亚军. Internet 环境下沙箱问题的一种解决方法[J]. 电脑开发和应用,2002,15(8):5-6.

[5] 徐丽娜. 神经网络控制[M]. 北京:电子工业出版社,2009.

[6] Sadeghi B H M. A BP-neural network predictor model for plastic injection molding process[J]. Journal of Materials Processing Technology,2000,103(3):411-416.

[7] Jin Wen, Zhao Jiali, Luo Siwei, et al. The improvements of BP neural network learning algorithm[C]//Proc of 5th international conference on signal processing. Beijing: IEEE, 2000: 1647-1649.

[8] 吴宏伟. 基于改进 BP 神经网络的分布式入侵检测模型研究[D]. 哈尔滨:哈尔滨理工大学,2005.

[9] Bennett K P, Mangasarian O L. Neural network training via linear programming[M]//Advances in optimization and parallel computing. North Holland: Amsterdam, 1992.

[10] 韩 丽. 神经网络结构优化方法及应用[M]. 北京:机械工业出版社,2012.

[11] 刘功申. 计算机病毒及其防范技术[M]. 第 2 版. 北京:清华大学出版社,2011.

[12] 李雪萍,汪晓兰. 基于 BP 神经网络的图书馆网络病毒检测方法[J]. 农业图书情报学刊,2005,17(11):10-12.

[13] 郭 晨,梁家荣,梁美莲. 基于 BP 神经网络的病毒检测方法[J]. 计算机工程,2005,31(2):152-153.

(上接第 162 页)

rence in dynamic scenes[C]//Proc of 21st international conference on pattern recognition. [s. l.]:[s. n.],2012:3172-3175.

[5] Lee D S. Effective Gaussian mixture learning for video background subtraction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2005,27(5):827-832.

[6] Zhang Z, Lipton A J, Venetianer P L, et al. Background modeling with feature blocks[P]. USA: US8150103,2012-04-03.

[7] Elgammal A, Duraiswami R, Harwood D, et al. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance[J]. Proceedings of IEEE, 2002,90(7):1151-1163.

[8] Guan Yepeng, Du Jinhui, Zhang Changqi. Improved HSV-based Gaussian mixture modeling for moving foreground segmentation[M]//Advances on digital television and wireless multimedia communications. Berlin: Springer, 2012: 52-58.

[9] Zhang Ruolin, Ding Jian. Object tracking and detecting based on adaptive background subtraction[J]. Procedia Engineering,2012,29:1351-1355.

[10] Wang Yang, Loe K F, Wu Jiankang. A dynamic conditional random field model for foreground and shadow segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2006,28(2):279-289.

[11] Farcas D, Marghes C, Bouwmans T. Background subtraction via incremental maximum margin criterion: a discriminative subspace approach[J]. Machine Vision and Applications, 2012,23(6):1083-1101.

[12] System and method for human detection and counting using background modeling, hog and haar features[P]. European: EP2518661,2012-10-31.

[13] Wu Mingjun, Peng Xianrong. Spatiotemporal context for codebook-based dynamic background subtraction[J]. International Journal of Electronics and Communications,2010,64(8): 739-747.

[14] Tu Q, Xu Y, Zhou M. Box-based codebook model for real-time objects detection[C]//Proceedings of the 7th world congress on intelligent control and automation. Chongqing, China: IEEE,2008.

一种基于BP神经网络的智能检测病毒方法

作者： 朱俚治， ZHU Li-zhi
作者单位： 南京航空航天大学 信息中心,江苏 南京,210016
刊名： 计算机技术与发展 ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2014(10)