

◎ 研发、设计、测试 ◎

采用行为分析的单机木马防护系统设计与实现

王泽东^{1,2}, 刘 宇^{1,2}, 朱随江^{1,2}, 刘宝旭², 潘 林^{1,2}WANG Zedong^{1,2}, LIU Yu^{1,2}, ZHU Suijiang^{1,2}, LIU Baoxu², PAN Lin^{1,2}

1. 中国科学院 研究生院, 北京 100049

2. 中国科学院 高能物理研究所 计算中心, 北京 100049

1. Graduate University of Chinese Academy of Sciences, Beijing 100049, China

2. Computing Center of Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

WANG Zedong, LIU Yu, ZHU Suijiang, et al. Design and implementation of Trojan horse protection system based on behaviors analysis. Computer Engineering and Applications, 2011, 47(11): 46-48.

Abstract: Based on behaviors analysis, this paper designs and realizes a Trojan horse protection system for PC. The system overcomes the shortage of Trojan horse detection techniques which based on static feature code and monitors suspicious behavior of the program, using Bayesian algorithm for analysis of the program behavior characteristics, Trojan horse program will be killed once found. The system repairs the software damaged and protects operating system better. Experiments show that the Trojan horse protection system significantly reduces false positive rate, with few impact on detection rate.

Key words: behaviors analysis; Bayesian algorithm; Trojan horse; protection

摘 要: 针对基于特征代码的静态木马检测技术的不足, 通过实时监控程序的可疑行为, 运用贝叶斯算法分析程序行为特征进而发现木马程序, 并对恶意木马程序的非授权操作进行修复, 设计并实现了一个基于行为分析的单机木马防护系统。实验表明: 该木马防护系统在对检测率影响较小的前提下, 显著降低了误报率。

关键词: 行为分析; 贝叶斯算法; 木马; 防护

DOI: 10.3778/j.issn.1002-8331.2011.11.014 文章编号: 1002-8331(2011)11-0046-03 文献标识码: A 中图分类号: TP393.08

1 引言

随着计算机网络技术的迅速发展, Internet 逐渐渗透到政府、工业、教育、国防领域, 网络在方便地带来大量信息的同时, 也带来了病毒、木马、蠕虫等诸多安全问题, 由此造成黑客攻击、信息失窃等安全事件的多发。特别是木马, 由于其攻击具有隐蔽性强、危害大的特点, 严重威胁着联网计算机的信息安全。设计并实现了一个针对单机的基于行为分析的木马防护系统, 动态实时监控计算机系统, 提高计算机系统的木马防护能力。

2 木马检测技术简介

木马是指附在应用程序中或独立存在的恶意程序, 主要采用客户端/服务器模式, 服务器端程序运行于被控制主机上, 远程客户端完成控制功能。客户端利用远程控制技术实现对感染木马的计算机进行控制, 进行恶意操作或窃取机密文件。近年来国内外在木马检测方面正做着积极的研究。Konrad Rieck^[1]等人提出了基于行为分析的恶意程序分类和检测方法, 较好地检测出未知恶意程序。YU-FENG LIU^[2]等人提

出了一种使用机器学习的基于行为的木马检测方法; 顾雨捷^[3]等人提出了一种新的基于多层模糊分类系统的反木马算法, 最终实现木马判别的局部高精度分类。

2.1 静态木马检测技术

静态木马检测技术通过检测木马程序在主机中驻留的静态特征进行判断, 包括木马程序在主机中放置的文件信息、注册表信息等。静态的木马检测需要用户主动运行程序实现检测。木马的静态特征分为: (1) 木马在目标系统中生成文件及木马原始文件的特征字符串; (2) 木马在目标系统中运行时进程的名称; (3) 木马运行时打开的固定 TCP/IP 端口号; (4) 木马在目标系统中具体的启动加载方式; (5) 木马植入后在目标系统中的生成文件名、文件大小及所在目录等。

静态木马检测技术的优点在于即使木马程序发生改变, 只要其特征不发生改变, 就能够很好地检测到木马程序, 误报率低。缺点是需要收集、提取大量的木马特征, 面对新的木马变种, 反应时间较长。

2.2 基于行为分析的木马检测技术

传统的基于静态特征的木马检测技术, 不仅面对已知木

基金项目: 国家科技支撑计划重点项目 (No. 2009BAH52B06); 中科院知识创新重点方向项目 (No. YYYJ-1013)。

作者简介: 王泽东 (1985—), 男, 硕士生, 主研方向: 网络信息安全; 刘宇, 博士生; 朱随江, 博士生; 刘宝旭, 博士, 副研究员; 潘林, 博士生。

E-mail: wangzd@ihep.ac.cn

收稿日期: 2010-09-08; 修回日期: 2011-01-17

马的各种隐蔽和变化检测能力不足, 对于未知的木马更是无能为力。行为分析的木马检测技术就是控制木马植入、隐蔽和恶意操作行为所需要的资源条件, 监控木马运行、启动的隐蔽行为和恶意操作^[9]。根据监控这些动态行为, 结合行为特征库对可疑程序是否是木马做出评估。主要监控的行为包括: (1) 对注册表操作监控; (2) 对系统端口和通信行为监控; (3) 对系统的进程进行监控; (4) 监视特定的 API 调用, 发现木马的隐蔽和恶意操作行为; (5) 对文件目录操作监控。

3 体系框架与基本原理

木马防护系统为系统的常驻进程, 实时保护系统, 对进程行为进行实时监控, 利用行为分析技术, 根据行为特征和基于贝叶斯的木马行为分析算法, 检测出可疑程序进程。系统结构示意图如图1所示。

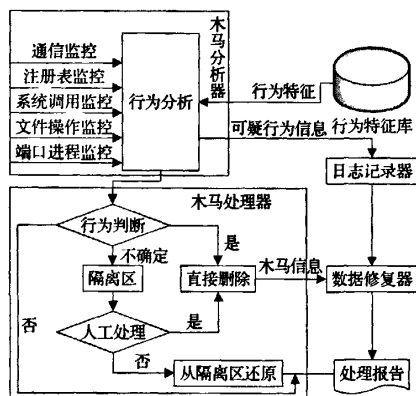


图1 系统结构示意图

系统结构主要包括木马分析器、日志记录器、木马处理器以及数据修复器、行为特征库等几个部分。

木马分析器首先对系统调用、文件目录操作、端口进程、注册表、网络通信等进行监控; 其次从行为特征库调用行为特征对监控到的行为进行贝叶斯分析。

木马处理器将对木马分析器的结果进行处理, 根据结果的不同, 有不同的处理方法: 确定是木马的程序系统采用直接删除的方式, 直接关闭木马进程、删除木马程序。对于可疑的程序, 系统无法判断, 可放置到隔离区, 供用户人工处理, 用户可以恢复或者删除之。

日志记录器将所有的可疑行为详细记录下来, 为后期数据修复提供日志文件。对于注册表的记录主要是被修改的项及其值。对于文件的记录, 除了文件名、大小、修改时间等基本属性外, 还需要记录文件的主文件表 (Master File Table, MFT)。

数据修复器根据日志记录对系统文件、注册表进行恢复。从日志数据库里面读出目标进程的日志文件, 根据日志记录的各种信息进行数据恢复。对于注册表的修改, 可以简单的写回操作就实现恢复。而对于文件、目录的修复需读取日志记录器记录的日志文件、目录的 MFT, 通过 MFT 里面的 FILE 项可以实现数据的修复, 将被删除、修改的文件或者目录恢复。数据修复完成后生成一份处理报告给用户。

4 系统的设计与实现

基于行为的木马检测方法主要是监控木马启动、运行、通

信等隐蔽行为和恶意操作, 再通过行为分析模型对这些行为操作进行检测判断, 数据流程如图2所示。

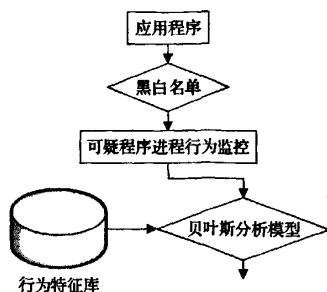


图2 数据流程示意图

本系统用到的两个关键技术, 分别是可疑程序行为监测和基于朴素贝叶斯的行为分析算法。

4.1 使用 API Hook 机制监测可疑程序行为

木马分析器首先对系统调用、文件目录操作、端口进程、注册表修改、网络通信等信息监控。对这些操作的监控采用的是用户级 API HOOK 技术^[9]。具体采用修改函数输入表 (Import Address Table, IAT) 方法, 该方法基于 Windows PE 文件, PE 文件的基本结构如图3所示。

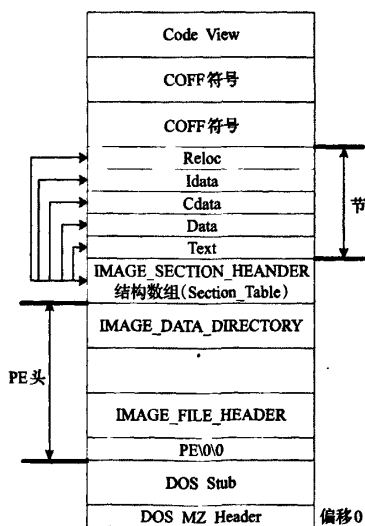


图3 Windows PE 文件格式示意图

当系统把一个进程模块加载到内存中时, PE 文件映射到进程的地址空间, 通过模块句柄 HModule 返回模块映像在内存中的地址。从模块句柄 HModule 返回开始, 使用 API 函数 Image Directory Entry To Data 定位 IAT 输入地址表 (Import Address Table, IAT)。找到 IAT 并遍历就可找到所需 API 地址, 然后用自定义函数地址覆盖表中函数调用地址, 所有关于这个 API 的调用将转到自定义的函数中, 从而成功实现 API 调用拦截。

4.2 木马行为分析模型

贝叶斯算法具有简单实用、计算高效, 概率表达良好的特点, 可以建立良好的分类模型, 被广泛应用各种检测系统和邮件分类中^[10]。为了提高系统性能, 降低误报率和漏报率, 本文将贝叶斯算法引入到对可疑程序行为检测结果的分析中。

朴素贝叶斯分类模型如图4所示, 设有变量集 $U=\{a_1, a_2,$

$a_3, \dots, a_n, C\}$, 其中 $a_1, a_2, a_3, \dots, a_n$ 是实例的属性变量, C 是取 m 个值的类变量。假设所有的属性都条件独立于类变量 C , 即每一个属性变量都以类变量作为唯一的父节点, 就得到朴素贝叶斯分类模型。朴素贝叶斯分类模型假定特征向量的各分量间相对于决策变量是相对独立的, 即各个变量独立地作用于决策变量。

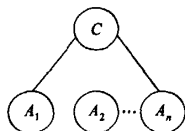


图4 朴素贝叶斯分类模型结构示意图

木马检测是一个分类的问题, 有木马程序、不确定程序、非木马程序三类。对于样本空间中的每一个样本程序 X , 将反映程序是否发生行为异常的特征用一组向量: $A=(a_1, a_2, a_3, \dots, a_n)$, a_i 为行为库中的行为特征。定义 C 为分类集, 即 {正常程序, 不确定程序, 木马程序}, 令 C_1 表示正常程序, C_2 表示不确定程序, C_3 表示木马程序, 它是一个随机变量。模型目标是: 在获得给定样本程序文件中的特征集 F 后, 判别出该样本是否木马程序的概率, 即计算 $P(C|A)$ 。根据朴素贝叶斯算法, 判断程序是否为木马的计算公式:

$$P(C|A) = \frac{P(A|C) \times P(C)}{P(A)}$$

因为特征向量 A 中包含分量 $a_1, a_2, a_3, \dots, a_n$, 所以上式又可以写成:

$$P(C|A) = \frac{\prod_{j=1}^n P(A_j|C) \times P(C)}{\prod_{j=1}^n P(A_j)}$$

最后根据 $P(C|A)$ 的最大值来确定 A 最大可能的类。即待测样本 X 的类别 C_x 为:

$$C_x = \arg \max_C (P(C) \prod_{j=1}^n P(A_j|C))$$

在木马查杀器中根据 C_x 的值来判断程序的类别。

4.3 木马处理器

木马处理器将对木马分析器的结果进行处理, 根据 C_x 值不同, 有不同的处理方法:

(1) 确定是木马的程序系统采用直接删除的方式, 直接关闭木马进程、删除木马程序。木马的删除必须按步骤进行删除, 首先终止其运行, 如果木马是以一个单独的进程运行, 则杀死此进程, 如果木马以一个进程的线程运行, 则取消线程占用的内核资源, 再停止线程的运行、释放其占用的进程资源, 否则木马文件在删除的时候会受到系统保护, 有些木马甚至会重新生成副本, 重新加载启动方式。其次, 删除宿主中的木马代码及其副本。最后解除木马的启动方式。

(2) 隔离可疑木马程序。对于可疑的程序, 系统无法判断, 可放置到隔离区, 供用户人工处理, 用户可以恢复或者删除之。首先需要建立一个隔离区, 将可疑进程关闭, 再将其程序转移到隔离区中, 并对其文件 MFT 进行修改, 使其无法运动。如果用户选择信任该程序, 再将其恢复。

(3) 对于确定不是木马程序的, 停止监控分析, 让程序正常运行。

5 实验及分析

本系统只对特定的程序进行监控, 对系统效率不会造成明显影响。为了验证系统的有效性, 从国家计算机病毒应急处理中心选择了 300 个程序 (200 个木马程序, 100 个正常程序) 作为测试样本, 木马行为特征库收集了 300 条木马行为特征, 主要是注册表的修改, 文件的修改删除, 通信操作。测试环境为 Windows XP SP3。为了便于对实验结果进行分析, 本文做以下定义:

木马程序被判定为木马程序的情况, 称为 TP (True Positive); 正常程序被判定为正常程序的情况, 称为 TN (True Negative); 正常程序被判定为木马程序的情况, 称为 FP (False Positive); 木马程序被判定为正常程序的情况, 称为 FN (False Negative); 测试指标包括检测率、误报率以及综合检测精度, 计算方式分别为:

$$\text{漏报率} = \frac{FN}{TP + FN} \quad (\text{木马程序被漏报为正常程序的比例})$$

$$\text{误报率} = \frac{FP}{TN + FP} \quad (\text{正常程序被误报为木马程序的比例})$$

$$\text{综合检测精度} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{正常程序被判定为木马程序以及木马程序被判定为木马程序的整体情况})$$

经过同国内几款木马查杀软件横向测试比较, 得出以下结果, 如表 1 所示。

表1 系统性能测试

软件名称	TP	TN	FP	FN	漏报率/(%)	误报率/(%)	综合检测精度/(%)
A	180	95	5	20	10.0	5	91.6
R	165	94	6	35	17.5	6	86.3
K	178	95	5	22	11.0	5	91.0
本系统	180	97	3	20	10.0	3	92.3

实验得出本系统漏报率为 10%, 误报率为 3%, 综合测试精度为 92.3%, 并对木马程序的操作进行了完整性修复。测试结果表明, 本系统在漏报率及误报率指标上均优于 R 软件; 此外相比较于 A 和 K, 在漏报率相当的情况下, 显著降低了误报率; 本系统采用的贝叶斯算法模型比基于多层模糊分类系统的算法^[1]具有更高的综合检测精度。实验表明, 提出的基于行为分析的木马防护系统有良好的防护效果, 并克服了主流防护软件误报率高的缺点, 能对系统进行有效的保护。

6 结束语

针对基于特征检测方法的不足, 根据木马运行时的行为特征, 设计并实现了基于行为分析的木马防护系统。介绍了系统的框架结构和关键技术, 对其中的两个关键技术: API HOOK 技术和基于朴素贝叶斯的行为分析算法进行了描述。通过对大量木马的测试表明, 系统能准确监控到可疑程序的各种操作, 基于贝叶斯的行为分析算法能较为准确地检测出木马, 同时系统较好地对木马的各种操作进行修复。由于木马的复杂性和多态性, 下一步的工作需要通过实验提取更多的新型木马行为特征, 进一步完善木马行为库内容, 降低漏报率和误报率。

参考文献:

- [1] Rieck K, Holz T. Learning and classification of malware behavior[C]// Proceedings of 5th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, DIMVA, 2008.

(下转 148 页)

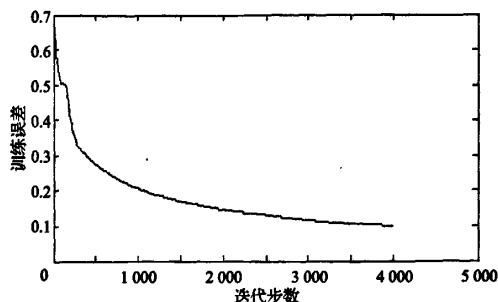


图5 训练误差函数曲线

和信息的利用率,采用相空间重构理论构造样本集。首先将实际385组时间序列采样数据按三次采油阶段分为三组,一次采油21组数据记为 $\{x_{1,k}\}_{k=1}^{21}$;二次采油327组数据记为 $\{x_{2,k}\}_{k=22}^{348}$;三次采油37组数据记为 $\{x_{3,k}\}_{k=349}^{385}$ 。按相空间重构理论分三个阶段进行时间延迟 τ 和最小嵌入维数 m 的计算。由式(17),计算结果为:一次采油阶段 $\tau_1=3$, $m_1=4$,共构造 $21-(m_1-1)\tau_1=12$ 组训练样本;二次采油阶段 $\tau_2=6$, $m_2=27$,共构造 $348-(m_2-1)\tau_2-21=171$ 组训练样本;三次采油阶段 $\tau_3=3$, $m_3=5$,共构造25组训练样本。

5.3 级联PNN的训练及预测结果

针对该区块油田实际开发情况,采用本文构建的输入/输出均为时变函数的3级级联过程神经网络来对三次采油过程中的采油速度进行拟合和预测。模型设计为:第1级子网的输入为地层压力随时间的变化函数,网络拓扑结构为1-5-1;第2级子网络的拓扑结构为3-12-1,输入为注入压力、注水量随时间变化的过程数据及前一级子网的输出;第3级子网络的拓扑结构为3-8-1,输入为注入压力、注聚量随时间变化的过程数据及前一级子网的输出。3个子网的输出均为区块采油速度函数。

将采用相空间重构理论构造出的训练样本集中的每个样本利用Walsh正交函数系^[11]拟合成一个分段函数。当Walsh基函数个数为32时,满足采样数据拟合精度要求。学习效率 $\alpha_1=0.60$, $\beta_1=0.50$, $\gamma_1=0.65$, $\alpha_2=0.55$, $\beta_2=0.50$, $\gamma_2=0.60$, $\alpha_3=0.50$, $\beta_3=0.70$, $\gamma_3=0.55$;学习精度 $\varepsilon=0.10$,网络学习3998次后收敛。训练过程和结果分别如图5、图6所示,取得了较好的实验结果。

6 结论

针对非线性动态系统分阶段指标预测问题,提出了一种基于级联过程神经网络的动态预测模型和方法。模型能够

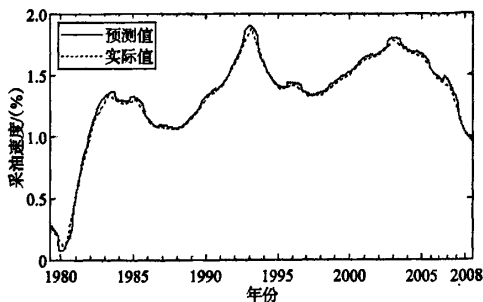


图6 实际输出与期望输出拟合曲线对比

反映系统在不同的阶段各项变量或指标可能具有不同的相互作用关系和信息变换机制,以及各阶段系统状态的连续性。采用相空间重构理论构造训练样本集,可弥补训练样本的不足和提高采样数据信息的利用率,增强网络的泛函逼近能力。基于过程神经网络的动态预测模型在机制上对于求解如非线性动态系统建模、系统辨识与过程控制等方面的实际问题也具有广泛的适用性。

参考文献:

- [1] 周新桂, 操成杰. 储层构造裂缝定量预测与油气渗流规律研究现状和进展[J]. 地球科学进展, 2003(3): 398-404.
- [2] He Xingui, Liang Jiuzhen. Procedure neural networks[C]//Proceedings of Conference on Intelligent Information Proceeding, 16th World Computer Congress. Beijing: Publishing House of Electronic Industry, 2000: 143-146.
- [3] Liang Jiuzhen, Zhou Jiaqing, He Xingui. Procedure neural networks with supervised learning[C]//9th International Conference on Neural Information Processing, Singapore, Nov, 2002: 523-527.
- [4] 何新贵, 梁久祺. 过程神经网络的若干理论问题[J]. 中国工程科学, 2000, 12(2): 40-44.
- [5] 许少华, 何新贵, 刘坤, 等. 关于连续过程神经网络的一些理论问题[J]. 电子学报, 2006, 34(10): 1838-1841.
- [6] 马千里, 郑启伦. 基于相空间重构理论与递归神经网络相结合的股票短期预测方法[J]. 计算机应用研究, 2007(4): 239-245.
- [7] 李玉震, 吴百海, 邢志鹏. 单变量时间序列相空间重构及应用研究[J]. 组合机床与自动化加工技术, 2004(2): 51-55.
- [8] 于宁莉, 易东云, 涂先勤, 等. 时间序列中自相关与偏相关函数分析[J]. 数学理论与应用, 2007(1): 54-57.
- [9] 党建武, 黄建国. 基于GP算法的关联维计算中参数取值的研究[J]. 计算机应用研究, 2004(1): 48-51.
- [10] 李梅霞. 国内外三次采油现状及发展趋势[J]. 当代石油石化, 2008(12): 19-25.
- [11] 关守平, 吕欣. 基于Walsh变换的过程神经网络建模及应用[J]. 控制工程, 2007(5): 473-478.

(上接48页)

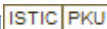
- [2] Liu Yufeng, Zhang Liwei. Detecting Trojan horse base on system behavior using machine learning method[C]//Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010.
- [3] 顾雨捷, 陈庆章. 用于行为分析反木马的模糊分类算法研究[D]. 杭

州: 浙江工业大学, 2008.

- [4] 朱明, 徐寿. 木马病毒分析及其检测方法研究[J]. 计算机工程与应用, 2003, 39(28): 176-179.
- [5] 朱建军, 熊兵. 网络安全防范手册[M]. 北京: 人民邮电出版社, 2007.
- [6] 陈福志, 史杏荣. 基于改进贝叶斯算法的信息安全模型[J]. 计算机工程, 2003, 29(20): 116-118.

作者: [王泽东](#), [刘宇](#), [朱随江](#), [刘宝旭](#), [潘林](#), [WANG Zedong](#), [LIU Yu](#), [ZHU Suijiang](#),
[LIU Baoxu](#), [PAN Lin](#)

作者单位: [王泽东, 刘宇, 朱随江, 潘林, WANG Zedong, LIU Yu, ZHU Suijiang, PAN Lin\(中国科学院研究生院, 北京, 100049; 中国科学院, 高能物理研究所, 计算中心, 北京, 100049\)](#), [刘宝旭, LIU Baoxu\(中国科学院, 高能物理研究所, 计算中心, 北京, 100049\)](#)

刊名: [计算机工程与应用](#) 

英文刊名: [COMPUTER ENGINEERING AND APPLICATIONS](#)

年, 卷(期): 2011, 47(11)

参考文献(6条)

1. [Rieck K; Holz T](#) [Learning and classification of malware behavior](#) 2008
2. [Liu Yufeng; Zhang Liwei](#) [Detecting Trojan horse base on system behavior using machine learning method](#) 2010
3. [顾雨捷; 陈庆章](#) [用于行为分析反木马的模糊分类算法研究](#)[学位论文] 2008
4. [朱明; 徐骞](#) [木马病毒分析及其检测方法研究](#)[期刊论文]-[计算机工程与应用](#) 2003(28)
5. [朱建军; 熊兵](#) [网络安全防范手册](#) 2007
6. [陈福志; 史杏荣](#) [基于改进贝叶斯算法的信息安全模型](#)[期刊论文]-[计算机工程](#) 2003(20)

本文读者也读过(10条)

1. [颜会娟, 秦杰, YAN Hui-juan, QIN Jie](#) [基于非线性SVM模型的木马检测方法](#)[期刊论文]-[计算机工程](#)2011, 37(8)
2. [贺小伟, 余景景, 王淼, HE Xiao-wei, YU Jing-jing, WANG Miao](#) [基于SPI技术漏洞的新型木马的防范方法](#)[期刊论文]-[西北大学学报\(自然科学版\)](#) 2006, 36(6)
3. [胡光俊, 宋伟航, 徐国爱, HU Guang-jun, SONG Wei-hang, XU Guo-ai](#) [基于行为序列灰色模糊判定的计算机木马检测方法](#)[期刊论文]-[北京理工大学学报](#)2011, 31(5)
4. [X卧底手机监听软件: 可窃听通话窃取短信](#)[期刊论文]-[IT时代周刊](#)2011(14)
5. [缪海英](#) [浅谈程序正当性](#)[期刊论文]-[法制与经济](#)2011(11)
6. [王海宾, 杨引明, 杨雅薇, 朱雪松, WANG Hai-bin, YANG Yin-ming, YANG Ya-wei, ZHU Xue-song](#) [上海世博会全球对地观测三维可视化系统设计与实现](#)[期刊论文]-[计算机应用研究](#)2011, 28(5)
7. [时慧娟, 王国俊, SHI Huixian, WANG Guojun](#) [逻辑系统MTL \$\nabla\$ 及其完备性](#)[期刊论文]-[计算机工程与应用](#) 2011, 47(6)
8. [刘志都, 程新党, 廖湖声, LIU Zhi-du, CHENG Xin-dang, LIAO Hu-sheng](#) [多功能组合木马架构的研究](#)[期刊论文]-[海军工程大学学报](#)2008, 20(4)
9. [陈繁, 管群, CHEN Fan, GUAN Qun](#) [模糊决策在农村劳动力转移安置中的应用](#)[期刊论文]-[计算机工程与应用](#) 2011, 47(10)
10. [于西昌, 谭桂梅, YU Xichang, TAN Guimei](#) [几种逻辑系统中的概率真度](#)[期刊论文]-[计算机工程与应用](#) 2011, 47(12)

本文链接: http://d.wanfangdata.com.cn/Periodical_jsjgcyyy201111014.aspx