一种启发式木马查杀模型的设计与分析

单长虹　张焕国　孟庆树　彭国军
（武汉大学计算机学院，武汉 430072）
E-mail：ritatv@tom.com

摘　要　该文从特洛伊木马查杀模型的设计入手，引入了人工智能中的启发式分析技术，不但提供了对已知木马的查杀设计，还可以对未知木马进行启发式分析，最终实现对未知木马的判定和查杀。

关键词　特洛伊木马　文件特征码　行为特征　动态监控　启发式分析

文章编号 1002-8331-(2004)20-0130-03　**文献标识码** A　**中图分类号** TP393.08

Design and Analysis of a Kind of Model for Searching and Killing Trojan Horse Based on Heuristic Analysis

Shan Changhong　Zhang Huanguo　Meng Qingshu　Peng Guojun
(College of Computer Science, Wuhan University, Wuhan 430072)

Abstract: Heuristic analysis idea in artificial intelligence is introduced in designing the model that aims to search and kill Trojan horse.The new model not only can search and kill the known Trojan horse, but also can heuristically analyze unknown or suspected communication software and finally search and kill unknown Trojan horse.

Keywords: Trojan horse, file eigenvalue, behavior character, dynamical spy, heuristic analysis

1　引言

为了实现对未知病毒的查杀，计算机病毒防治领域出现了一种新的技术，这就是启发式分析技术。目前国内研究的还比较少，国外已经有人做过这方面的研究[1]。但是在特洛伊木马的防治领域，尚未见过这方面的文章，在这里，笔者结合自己的研究，提出一种启发式分析的木马实时查杀模型，并从模型的设计和技术实现的角度进行分析，介绍了它的工作原理，其中重点介绍了启发式分析模块。这种模型既可以实现对已知木马的查杀，又可以对未知木马进行启发式分析，并将分析数据提交专家系统，由专家系统对其判定，并实时对数据库进行更新，最终实现对未知木马的查杀。从总体来看，该文提出的模型是一个类似于智能病毒防御系统[2]的木马查杀模型。

2　特洛伊木马

特洛伊木马是一种基于 C/S 架构的网络通信软件，一般情况下可分为 server 端程序和 client 端程序两部分，其中 server 端在目标计算机上驻留，client 被控制者操纵。通常情况下，由 client 端程序通过某种方法，主动与 server 端程序连接并建立通信关系，对目标计算机进行操纵。现在还产生了一种反弹端口的木马，由 Server 端主动向 Client 发出建立连接的请求并建立通信关系。特洛伊木马可以通过多种方法与特定的程序进行关联，当目标计算机系统启动后，如果用户不幸打开了被木马用来关联的程序，则木马的 server 端程序将被激活，并开始打开端口（也有可能不使用新的端口，直接利用 ICMP 的 80 端口），建立与远程对应的 client 端程序的通信。通过打开的端口，client 端程序对目标计算机进行未授权的侦听。在侦听过程中，它将目标计算机的一些秘密信息，源源不断地送往 client 端，并可接收并执行来自 client 端的指令，执行一些诸如删除文件、重新启动计算机等非法操作。可见，特洛伊木马对网络上的计算机的安全性和信息的秘密性构成了极大的威胁。计算机一旦被植入木马，攻击者便可进行远程操纵，目标计算机将毫无安全和秘密可言。有的资料[3]对特洛伊木马进行了全面详细的介绍，有兴趣的读者可以查阅一下。

3　模型的基本模块设计分析

该模型主要有八大模块，分别为：静态检测、动态检测、动态监控、查杀引擎、隔离机、动态推理机、启发式分析机和专家系统，模型如图 1 所示。

3.1　模块工作原理

(1)静态扫描：用户通过用户界面接口对查杀引擎实施调用，再由查杀引擎调用静态检测模块。静态检测模块首先从静态特征码库中依次取出特征码，与被扫描文件的特征码进行比较，如果一致，则确定是木马文件，调用动态检测模块，查找对应文件正在运行的进程，杀死进程，将文件放入隔离库，交由用户处理。

(2)动态检测：负责对运行的进程进行检测。对正在运行的进程，可以通过使用内存扫描的方法实现，杀掉发现的木马进程，并在注册表中恢复被修改的键值，将磁盘上的对应文件进行隔离，并交由用户处理。

(3)动态监控：负责对文件、注册表和异常端口数据流量实施监控，这可以通过系统提供的 API 函数来实现。文件监控时，一旦发现磁盘上增加了新的可执行文件，则调用静态检测模

基金项目：国家自然科学基金（编号：90101005，66973034）；教育部博士点基金（编号：20020486046）
作者简介：单长虹，硕士研究生，研究方向计算机病毒与网络安全。张焕国，教授，博导，研究方向信息安全。孟庆树，博士，研究方向信息安全。彭国军，硕士研究生，研究方向计算机病毒与网络安全。

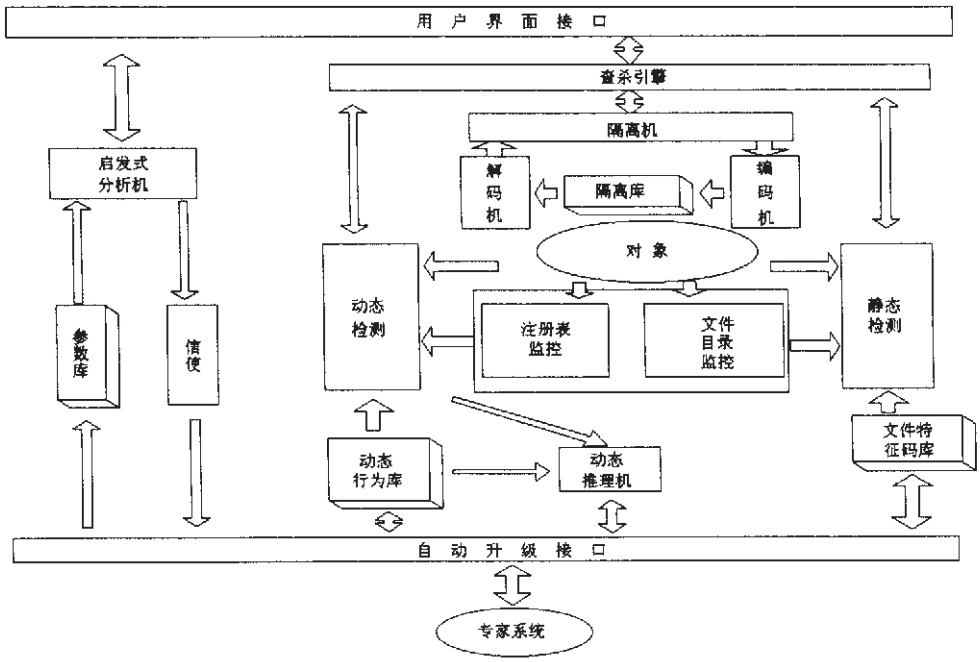

图1 基于启发式分析的木马实时查杀模型

块。注册表监控时,一旦发现注册表监控的键值发生了变化,则提取新键值,找到对应的进程,然后调用动态检测模块。如果发现异常端口数据流量偏大,也要调用动态检测模块。

(4)查杀引擎:负责处理用户界面接口和调度其它模块。

(5)隔离模块:对静态和动态监测模块提交的文件进行隔离。首先要对被隔离文件重新编码,并将其放入隔离库中,如果需要提取,则再对其解码。

(6)动态推理机:如果在已知木马特征库中找不到对应的可疑信息,则考虑与升级接口联系,更新特征码库,然后重新调用查杀引擎。

(7)启发式分析机:如果在使用了最新的特征码库之后,仍不能断定新的可疑进程是否为木马,则要进行启发式分析。并将分析数据提交给信使,由信使与自动升级接口进行联系,并将数据提交给专家系统处理,如果专家系统经过分析之后,认定其为木马,则更新各个数据库。

(8)专家系统:负责对前台提交的数据进行分析,并适时更新各个数据库。

3.2 模型的工作过程与模块之间的调用

这里主要从对已知木马和未知木马的查杀两个方面来研究这个模型的工作过程。

(1)已知木马查杀模型工作流程

用户启动程序,调用查杀引擎,查杀引擎首先调用动态检测模块,如果发现木马,先记住木马文件存放位置,然后将木马进程杀掉,并清除注册表中的相关键值,然后调用静态检测模块,根据前面得到的木马存放位置,找到木马文件,并通过与查杀引擎交互,看是否需要隔离机的合作从而放入隔离库中。

(2)未知木马检测模型工作流程

用户启动程序,调用查杀引擎,查杀引擎首先调用动态检测模块,如果未发现与已知木马相符的信息,则调用动态推理机,通过一定的推理规则进行分析,并与专家系统联系,自动更新动态行为库和文件特征码库。有了新的库之后,系统将再次调用查杀引擎重复上面的步骤,如果发现木马,则进行4.1中的操作。如果这个时候,还不能够确定它是木马,则询问用户是否启用启发式分析机,如果用户同意则调用启发式分析机,对某些原始数据进行分析,通过一定的过程处理之后,如果发现确实可疑,则通过信使向专家系统发送文件信息、原始数据等信息。专家系统在对信使发送的数据实验分析之后,就可以做出一定的判断,如果确定为新木马,则需要对各个库里的数据进行更新。这样之后,用户就可以重复(1)中的步骤,对木马实施相应的处理。

4 启发式分析机

4.1 启发式分析基础

定义:启发式指的是具有自我发现的能力、运用某种方式和方法来判定事物的知识和技能。

定义:启发式分析就是利用特洛伊木马与正常程序和计算机病毒的显著不同,结合以往的知识和经验,对未知的可疑木马进行分析与识别。

启发式分析流程如图2所示:

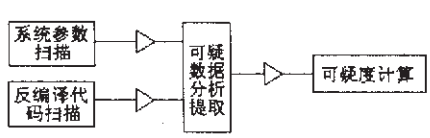

图2 启发式分析流程

启发式分析模块可以通过对某些系统参数进行分析,或者对代码进行反编译并对指令序列进行分析,从而获得感兴趣的数据。启发式分析技术中使用了启发式系统参数扫描和启发式代码扫描两种技术。通常情况下,木马在植入宿主计算机以后,会修改一些系统数据,如修改系统文件 win.ini 和 system.ini,修改注册表的键值,木马在运行起来以后,会自动打开某些端口进行通信。对于这样一些系统参数的变化,就可以使用启发式

系统参数扫描技术进行检测。木马的代码往往也与正常程序有所不同。木马作为一种网络通信软件和远程控制工具，必然在代码中要出现如远程文件管理、打开硬盘数据共享、进行远程屏幕抓取、调用远程关机或重启动、控制键盘和鼠标、进程隐藏等函数调用。启发式分析软件在对代码反编译以后，使用启发式扫描技术对反编译代码进行扫描，一旦发现诸如这样的函数调用，就有理由怀疑该驻留软件为可疑木马。

在具体实现上，启发式扫描技术是相当复杂的，通常这类木马检测软件不仅要对系统进行扫描，还要能够识别并检测许多可疑的程序代码指令序列，同时还要根据各个可疑指标的可疑程度，制定权值，并使用特定的规则进行计算，将得到的值与事先确定的临界值比较，如果比临界值大，则确定为木马。

为了表述的方便，更为了用户和研究人员能直观地检测被测试程序中可疑功能调用的存在情况，按如下方法为不同的可疑功能调用设置标志。各个标志的含义如下：

(1)WinFile=对 win.ini 文件的特定项的修改。

(2)SysFile=对 system.ini 文件的特定项的修改。

(3)Reg=对注册表特定键值的修改，文件关联。

(4)Port=端口异常与在此端口上的数据流量。

(5)Disk=硬盘数据共享。

(6)Ftp=远程文件操作。

(7)Screen=远程抓屏操作。

(8)Power=计算机的关闭与启动。

(9)KeyMou=鼠标键盘的操纵。

(10)Hide=进程的隐藏操作。

(11)RemExe=远程执行可执行程序。

(12)Message=发送消息模块。

(13)Process=进程管理。

(14)Service=更改服务配置。

(15)FileShell=文件加壳。

(16)OperReg=注册表操作。

定义 1：$F=a_1$*WinFile+a_2*SysFile+a_3*Reg+……+a_{15}*OperReg，其中，a_i 表示相应参数的权值，满足 $a_1+a_2+a_3+\cdots\cdots+a_{15}=1$，$a_i$ 的取值体现了相应标志在体现木马特征中的重要程度。F 的值反映了程序为木马程序的可疑度，显然 F 的取值范围也必然是[0，1]。

另外，还必须指定一个临界值 ExcptLine，使得当 F 大于 ExcptLine 时，系统通过信使向后台的专家系统发信息，取得专家系统的帮助。当然，ExcptLine 的值的取法，不能够凭空臆断，也不可能通过严格的计算或证明得到，因为病毒和木马本身具有不可预见性。只有通过大量的实验，结合要进行分析的数据，综合考察之后，抽取各异常度，才能够获得这样一个至关重要的值。这个值的取值的好坏，会直接关系到这个启发式分析模块的工作效率。

4.2 启发式分析举例

下面这些木马将点亮如下一些标志：

冰河 2.2：Reg、Port、Disk、Ftp、Screen、Power、KeyMou、RemExe、Message、Process、Service、OperReg、FileShell

广外女生 1.0：Reg、Port、Disk、Ftp、Screen、RemExe、Message、Process

网络精灵：Reg、Port、Disk、Ftp、Screen、Power、Message

定义 2：公有特征是指所有木马都具有的特征。

定义 3：私有特征是指除了公有特征外木马还具有的其它特征。

在对权值赋值的时候，可以将在所有木马中出现的公有特征的权值取比较大的值，如上面列出的几个木马，一共有 6 个公有特征，这样就可以给其中的公有特征每个分配 0.1 的权值，其他 9 个特征共用 0.4 的权值，这样就可以将 ExcptLine 的值定为 0.6，当计算出来的 F>0.6 时，就可以认为该软件为木马。实验证明，类似于这种方法制订的启发式分析策略的误报率是很低的，几乎可以达到零误报率。

零误报率是人们追求的完美结局。有的人或许简单地认为，只要将上面所说的公有特征的权值赋得很高，就可以解决误报的问题。这恰恰忽略了一个事实，误报率的降低会导致漏报率的上升。如果木马的设计者获知了人们所确定的公有特征，他就完全可以在设计木马时故意去掉其中几个公有特征，从而使得启发式分析失效，导致漏报率上升。所以，公有特征的确定是启发式分析的重中之重，同样，公有特征的权值的测定也是启发式分析的关键所在。这两大问题的解决不能够通过单纯的理论分析来解决，而必须进行大量的实验，对现有的木马进行分析，才能找到最佳的解决办法，使得误报率和漏报率能同时达到极小值。

4.3 提高启发式分析的工作效率

为了使漏报率和误报率达到最小，必须努力做好以下几点：

(1)对木马行为的准确把握，精确定义可疑功能调用标志集合。结合大量的实验，准确定义公有特征，以及其对应的加权值。

(2)对常规程序代码和特定程序的识别能力。某些编译器提供运行时的实时解压或解码的功能，为了避免检测时的误报，应当在检测程序中加入识别这些情况的功能模块。

(3)对于特定的类似于木马的通信软件，只要用户不同意启发式分析，就不报警，并且将其特征写入数据库中，记忆下来。也就是说，启发式分析模块要设计出自学习的功能模块，记住那些非木马的文件，避免再次报警。

当然，不管采用什么样的措施，虚警谎报现象总是要存在的。用户的参与对查杀软件来说也是必不可少的，他们要在某些报警信息出现时做出自己的判断。也许会有人说“我怎么知道被报警的程序到底是木马还是属于误报？”大多数人在问及这个问题的第一反应是“谁也无法证明和判断。”事实上是有办法做出最终判决的，但是这还要取决于应用启发式代码分析检测技术查木马程序的具体解释。

假如检测软件仅仅给出“发现可疑木马功能调用”这样简单的警告信息，而没有更多的辅助信息，对于用户来说几乎没有什么帮助。如果告诉用户“可能是木马”，似乎永远没错，不必担负任何责任，而用户不希望得到这样模棱两可的解释。相反地，如果检测软件把更为具体和实际的信息报告给用户，比如“警告，当前被检测程序含有屏幕抓取和文件操作的功能”，类似的方法更能帮助用户弄清该采取怎样应对措施。这样一来，报警的可疑木马常用功能调用都能得到合理的解释，也能得到用户更好的支持，取得更满意的结果。让任何用户能从同样的报警信息中推理出“木马”还是“非木马”，并不是件简单的事情，除非用户很有经验。因此，如果把这类软件设计成有某种学习记忆的能力，在第一次扫描时由有经验的用户逐一对有疑问的报警信息作好“是”与“非”的判断，而在以后的各次扫描检测时，由于软件学习并记忆了第一次检测时处理结果，将不再出现同样的烦人的提示警报。因为不论在什么情况下，偶尔请教一下某个有经验的“高手”并不难，难的是每次就同样的问题去麻烦别人。

(下转 139 页)

在现有网络体系中进一步的数据控制中还可以加入路由器协同控制，目的主要是防止欺骗攻击(Spoofing)、拒绝服务攻击(DoS)和基于 ICMP 的攻击。通过配置使路由器只允许以 Honeynet 的 IP 地址作为源 IP 地址的数据包通过，这样可以防止欺骗攻击，如 SYN 泛洪或 SMURF 攻击，同时还阻塞了向外的 ICMP 连接，弥补有些防火墙不能对 ICMP 数据包进行跟踪的缺陷。

前面提到，文中实现的 Honeynet 环境允许任何从 Internet 到 Honeynet 的连接。这与一些没有防火墙或过滤机制的组织所处的环境相同。如果要模拟其它一些有防火墙或过滤机制的生产环境，可以根据需要过滤任何对内连接，这取决于人们想要了解什么。这只要复制具体环境中的防火墙规则到陷阱网络的防火墙。由此可以识别可能存在于具体产品网络的风险，从而反馈到网络安全策略，使其重新考虑它愿意承担的风险，并更好的实现防火墙。

进一步的工作包括实现防火墙与路由器的协同控制，新一代陷阱网络在数据链路层的数据控制等。(收稿日期：2004 年 1 月)

参考文献

1.Chunming Rong，Geng Yang.Honeypots in Blackhat Mode and its Implications[C].In：Proceedings of 4th Int Conf on Parallel and Distributed Computing(PDCAT'03)，Chengdu，China，2003：185~188

2.Lance Spitzner.The Honeynet Project.Know Your Enemy[M].Addison Wesley，2001

3.Lance Spitzner.Honeypots，Tracking Hackers[M].Addison Wesley，2002

4.http：//www.honeynet.org/tools/rc.firewall

5.北京启明星辰信息技术有限公司.防火墙原理与实用技术[M].北京：电子工业出版社，2002

6.http：//www-900.ibm.com/developerWorks/cn/linux/network/s-netip/index.shtml

(上接 132 页)

不管启发式分析有怎样的缺点和不足，和其它的扫描分析技术相比，启发式代码分析扫描技术几乎总能提供足够的辅助判断信息，让人们最终判定被检测的目标对象是否为木马。启发式分析技术仍然是一种正在发展和不断完善中的新技术，但已经在大量优秀的反病毒软件中得到迅速的推广和应用。按照最保守的估计，一个精心设计的算法支持的启发式扫描软件，在不依赖任何对木马预先的学习和了解的辅助信息，如特征代码、校验和与行为特征等等的支持下，可以毫不费力地检查出 95%以上的对它来说是完全未知的新木马。可能会出现误报、谎报的情况，适当加以控制，这种误报的概率可以很容易地被降低在 0.1%以下。

启发式扫描技术，难免存在对木马的误报和漏报。所以在实际的应用中，最好将其与传统的扫描技术相结合，这样才能够更有效地对木马实施检测，从而提高检出率。

4.4 启发式分析技术与传统扫描技术结合

前面论述了许多启发式分析技术的优点和长处，会不会引起某些人的误解，以为传统的检测扫描技术就可以丢弃了呢？情况当然不是这样。从实际应用的效果看来，传统的手法由于基于对已知病毒的分析和研究，在检测时能够更准确，减少误报；但如果是对此前根本没有见过的新木马，由于传统手段的知识库并不存在该种木马的特征数据，则有可能产生漏报的严重后果。而这时基于规则和定义的启发式代码分析技术则正好可以大显身手，使这类新木马不至成为漏网之鱼。传统与启发式技术的结合使用，可以使病毒检测软件的检出率提高到前所未有的水平，而另一方面，又大大降低了总的误报率。详见以下测试实验结果对比数据：

启发式判定结果	传统式判定结果	真正结果
干净	干净	极有可能是干净的
干净	木马	很可能误报
木马	干净	很可能是木马
木马	木马	极有可能是木马

两种技术结合使用时，得到的测试数据：

	启发式判定结果	传统式判定结果	可能的真正结果
误报率	5%	0.1%	0.1%
漏报率	1%	0.001%	0.0001%

可见，某种木马能够同时逃脱传统和启发式扫描分析的几率是非常小的，如果两种分析的结论相一致，那么真实的结果往往就与其判断结论一致。两种不同技术对同一检测分析所得结论不一致的情况几率很小，这种情形下需借助专家系统去得出最后结论。

随着人们研究的逐步深入，技术也在不断进步。一方面绝大多数反病毒厂家的产品中还未能引入一个较为成功和可靠的启发式检测技术的内核，另一方面，即使是在少数知名反病毒产品中，这项技术也需要不断地完善和发展。任何的改良都会使效率有不同程度的提高，但是绝对不能企望在没有误报为代价的前提下使检出率达到 100%。恐怕在今后相当长的时间里，虚报和漏报的概率不可能达到 0%。100%正确的检测结果之所以不存在，是因为有相当一部分程序(或代码)介乎于木马与非木马之间，即便对于人脑来说，合乎逻辑又合乎木马定义的结论往往会截然相反。举个例子来说，Windows2000 的远程终端服务，可以实现的功能之强大，已经远远超出了现有的任何木马，而且其对系统主要性能如 CPU、内存的占有率相当低，它一旦运行起来，几乎很难察觉。对于这样的软件，它的功能与木马非常相似，这里的查杀软件是否也要像对待木马一样将其控制住呢，显然行不通！

5 结论

笔者在对木马的研究过程中，总结了这样一个模型，模型中的启发式分析模块不仅适用于木马查杀软件的设计，还可以应用于防治病毒软件的设计，具体的参数，还需要您结合自己的系统的具体情况具体分析。抛开启发式代码分析技术实现的具体细节和不同手法不谈，这种代表着未来反病毒技术发展的必然趋势、并且具备某种人工智能特点的反木马技术，有诸多传统技术无法企及的强大优势。在新病毒新木马不断涌现的今天，这种新技术的产生和应用将具有重要的意义。

(收稿日期：2004 年 1 月)

参考文献

1.张小磊编著.计算机病毒诊断与防治[M].中国环境科学出版社

2.祁明等.SCUT 智能反病毒体系(IVDS)设计[J].计算机工程与应用，2002，38(17)：43~45

3.张友生等编著.计算机病毒与木马程序剖析[M].北京科海电子出版社

一种启发式木马查杀模型的设计与分析

作者： 单长虹， 张焕国， 孟庆树， 彭国军
作者单位： 武汉大学计算机学院,武汉,430072
刊名： 计算机工程与应用 ISTIC PKU
英文刊名： COMPUTER ENGINEERING AND APPLICATIONS
年，卷(期)： 2004,40(20)
被引用次数： 9次

参考文献(3条)

1. 张小磊 计算机病毒诊断与防治

2. 祁明 SCUT智能反病毒体系(IVDS)设计[期刊论文]-计算机工程与应用 2002(17)

3. 张友生 计算机病毒与木马程序剖析

本文读者也读过(10条)

1. 刘刘 可牛免费杀毒专杀工具:对症下药杀木马[期刊论文]-网络与信息2010(9)

2. 张净 木马病毒的辨别与查杀方法[期刊论文]-中国科技信息2009(19)

3. 杜瑞颖.张焕国.彭国军.黄传河.傅杰 武汉大学信息安全专业实践教学体系的探索与研究[期刊论文]-计算机教育2007(19)

4. 彭国军.张焕国.刘丹 实验教学与信息安全本科生实践创新能力培养[期刊论文]-计算机教育2007(22)

5. 洪宁.颜晖.陆慧娟.周苏.彭国军.朱江华 应用型院校的毕业设计到底应该如何加强[期刊论文]-计算机教育2009(3)

6. 彭国军.张焕国 关于计算机病毒教学的几点建议[期刊论文]-计算机教育2006(7)

7. 张焕国.王丽娜.黄传河.杜瑞颖.傅建明 武汉大学信息安全学科建设与人才培养的探索与实践[期刊论文]-计算机教育2007(23)

8. 傅昌建.FU Chang-Jian 负DG范畴的导出范畴上的t-结构[期刊论文]-四川大学学报（自然科学版）2009,46(1)

9. 陈丽.石勇国.CHEN Li.SHI Yong Guo 函数方程f[m]=1/f的实解[期刊论文]-数学研究与评论2008,28(2)

10. 王健 浅谈WINDOWS下如何查杀木马病毒[期刊论文]-跨世纪（学术版）2008,16(5)

引证文献(9条)

1. 谭云松 一种启发式反病毒技术的研究[期刊论文]-网络安全技术与应用 2006(11)

2. 彭迎春.谭汉松 基于DLL的特洛伊木马隐藏技术研究[期刊论文]-信息技术 2005(12)

3. 罗平.徐倩华 网络游戏外挂技术及检测[期刊论文]-计算机工程与设计 2007(6)

4. 杨玲.孟传良 基于启发式分析的木马检测技术研究[期刊论文]-现代机械 2006(4)

5. 梁意文.曹玲林.蔡瀛 危险感知的数字微分初步[期刊论文]-哈尔滨工程大学学报 2006(z1)

6. 杨建召.程彦.范伟琦.刘棣华 基于主机的入侵检测系统设计与实现[期刊论文]-长春工业大学学报（自然科学版）2005(4)

7. 陈俊峰 计算机操作行为分析与数据安全管理系统[学位论文]硕士 2005

8. 单长虹 计算机远程控制技术研究[学位论文]硕士 2004

9. 唐树刚 基于文件静态特征的木马检测研究[学位论文]硕士 2005