基于粗糙集理论的安全考试系统 *

张 萍, 王建忠, 吴 倩, 邹 帆, 高 轶
(四川师范大学 基础教学学院, 成都 610068)

摘 要: 为了进一步提高结合了传统静态安全技术和动态安全技术的计算机考试系统的安全性,提出了一种基于粗糙集理论的静态安全技术与动态安全技术相结合的高效低负荷的防御方法,用于监控进程的非正常行为。该方法从进程正常运行情况下产生的系统调用序列中提取出一个简单的预测规则模型,能有效地检测出进程的异常运行状态。同原有系统相比,用粗糙集理论建立正常模型要求的训练数据获取简单,而且得到的模型更适用于在线检测。实验结果表明,该系统的安全性优于原考试系统。

关键词: 粗糙集理论; 考试系统; 安全技术; 系统调用
中图分类号: TP3 **文献标志码**: A **文章编号**: 1001-3695(2010)03-1012-03
doi:10.3969/j.issn.1001-3695.2010.03.056

Secure examination system based on rough set theory

ZHANG Ping , WANG Jian-zhong, WU Qian, ZOU Fan, GAO Yi
(*College of Fundamental Education, Sichuan Normal University, Chengdu* 610068, *China*)

Abstract: Applied the traditional information secure technology and the rough set theory to the computer examination system in order to improve the system security. This paper proposed an effective method based on rough set theory which had low overhead and high efficiency, and used to monitor the abnormal behavior of processes. It extracted a set of detection rules with minimum size to form a normal behavior model from the record of system call sequences and could detect the abnormal operating status of a process. Compared with the old system, this method required a smaller size of training data set, less efforts to collect training data and more suitable for real-time detection. Experimental results show that the secure performance of this system is better than the old system's.

Key words: rough set theory; examination system; security technology; system call

随着网络的不断发展,威胁的不断增加,如何才能进一步提高考试系统的安全和检测效率以满足不断变化的考试系统的安全需求呢? 本文在采用传统信息安全技术的考试系统[1]基础上,提出了一种基于粗糙集理论的高效低负荷的检测防御方法,用于监控进程的非正常行为。该方法借助于粗糙集理论从进程正常运行情况下产生的系统调用序列中提取出一个简单的预测规则模型,能有效地检测出进程的异常运行状态。用粗糙集理论建立正常模型要求的训练数据获取简单,而且得到的模型更适用于在线检测。实验结果表明,该系统的安全性优于采用传统信息安全技术的考试系统。

1 相关的研究工作

现有的计算机考试系统一般都采用传统的信息安全技术,其基本安全主要由传统的静态安全技术和动态安全技术以及安全数据库技术构成。静态安全技术主要采用传统的信息安全技术如防火墙、加密技术、病毒防治和身份认证技术等,这些技术发展较为成熟,对于已知的攻击模式有很好的防御作用。但静态安全技术的防御能力是固定的,防御功能不能适应不断变化发展的攻击方式;各种静态安全技术之间相互独立,缺少相互之间的通信,对于复杂攻击的检测和防御也是有限的。其不足之处由动态安全技术弥补,如生物免疫技术、入侵检测、攻击陷阱等技术,它们改变消极被动的防御方式,积极主动地采取各种措施来保证系统的安全,有效弥补静态安全技术的不足,让整个系统能更好地防御外来攻击。但基于上述安全技术的考试系统也存在着很多缺陷,如具有仅可检测已知弱点的局限性,对于异常检测难以定量分析,对于未知的攻击存在错报和漏报的情况,而且不能采取及时的应对措施,致使系统损坏等。随着网络技术的发展,考试规模的发展壮大,只有提出新的安全技术才能满足考试系统的新的安全需求。

近几年来成为机器学习研究热点的粗糙集理论[2,3](rough set theory,RST)提供了一套比较完备的从小样本数据中寻找规律的系统方法。同其他学习方法比较,粗糙集方法有两个优点:a)粗糙集提供了一整套比较完整的小样本学习方法;b)用粗糙集能得到一组最小规则集。据此,用粗糙集理论从考试系统调用序列样本集得到的行为模型能更有效地逼近理想的简单正常行为模型。粗糙集理论中提供了一套完善和系统的方

收稿日期: 2009-07-24; **修回日期**: 2009-08-24 **基金项目**: 四川师范大学校级基金资助项目

作者简介:张萍(1971-),女,四川成都人,副教授,主要研究方向为计算机网络、信息安全等(zp81995251@163.com); 王建忠(1965-),男,四川岳池人,副教授,博士,主要研究方向为计算机网络等;吴倩(1978-),女,四川阆中人,讲师,硕士,主要研究方向为虚拟仿真;邹帆(1980-),女,四川成都人,讲师,硕士,主要研究方向为计算机科学技术、计算机技术在学校教育中的应用等;高轶(1973-),女,四川营山人,讲师,硕士,主要研究方向为计算机应用技术.

法,可找到描述正常模型的最小预测规则集,有利于提高检测速度,适用于系统的实时检测。

2　粗糙集理论的引入

本文基于上述考试系统的安全研究,认为成熟的静态安全技术的不足之处可由动态安全技术弥补,动态安全技术的不足则由粗糙集理论来完善。因此提出了将粗糙集理论应用于考试系统中,以提高考试系统的安全。其基本架构如图 1 所示。

在基于粗糙集理论的安全考试系统架构中,本文着重阐述基于粗糙集理论的动态防御技术部分。在考试系统中,数据分为两种:一种为需要保护的考试数据,相对而言,需要保护的考试数据之外的其他任意的字符串称为非考试数据。一般情况下有基于主机、基于网络、基于序列[4~8]三种情况来定义考试数据与非考试数据集合。其中在基于主机方法中目前还没有一种拆分算法特别有效。基于网络方法对于考试系统内部的异常情况不易发现。基于序列方法是通过进程的系统调用序列或基于系统服务的访问序列进行考试数据集合的提取,用匹配方法进行识别。因为在进程正常执行时所产生的短序列局部连贯稳定, 而当非正常时会产生一些异常的系统调用序列,通过对序列的分析可以判断系统是否受到攻击。此种方法适宜于考试系统的情况。通过对三种情况的比较,本文采用了适用于考试系统的基于序列的方法。因此采用用户进程的系统调用来定义考试数据。在本文中,将粗糙集理论应用于考试系统正常模型的建模过程,从较小的正常系统调用序列样本中提取出预测规则集作为正常模型来进行系统的安全检测,建立进程的简单正常模型,进行高效的在线安全检测。

3　粗糙集理论的应用

目前, 基于粗糙集的规则获取主要有两种模式[9]。模式一[9]是由 Pawlak 教授提出,主要思想是通过寻找属性核及去掉多余的属性求出约简的决策表,并从最简决策表中获取相应的确定规则。模式二[9]是由 Wakulicz2Deja 等人提出,主要思想是直接从原始决策表中求取近似集, 并运用推理引擎,分别从下近似集中获取确定规则,从上近似集中获取可能规则。经过两种模式的比较,再结合考试系统,综合考虑后采用模式一,如图 2 所示。

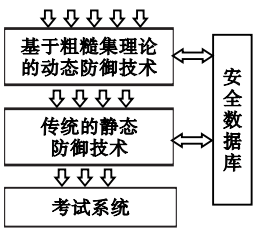

图 1　基于粗糙集理论的安全考试系统架构

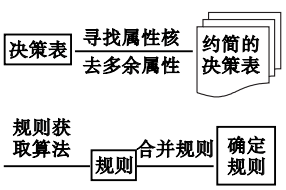

图2　基于粗糙集的规则推理模式

下面采用图2 中的模式方法,基于系统调用序列的进程正常行为模型,应用粗糙集理论建立进程正常模型 $M(L)$。设 X 是某一进程在正常运行状态下所能产生的所有系统调用序列的集合。用一个长度为 $L+1$ 的窗口沿着 X 中的每一条序列滑动,可以得到一系列长度为 $L+1$ 的序列段,所有的这些序列段组成了正常系统调用序列段集合 U。对于 U 中的每一个序列段,前 L 个系统调用被称为序列段的一般位置属性,它们组成了一般位置属性集 A;第 $L+1$ 个系统调用被称为序列段的末位置属性,末位置属性集表示为 D。A 和 D 的取值范围为所有系统调用组成的集合。建立正常行为模型的基本思想是根据 X 中的前 L 个系统调用预测第 $L+1$ 个系统调用,即找到 U 中的一般位置属性集 A 与末位置属性 D 之间的关系。基于系统调用行为的进程正常模型 $M(L)$ 是正常系统调用序列段(长度为 $L+1$)集合 U 中描述一般位置属性集 A 与末位置属性 D 之间关系的预测规则集。在实际情况中,通常只能收集到部分的正常系统调用序列作为样本,从中提取预测规则集。

将粗糙集理论应用到预测规则集的提取中。在给定某进程的一个正常系统调用序列样本集中,用长度为 $L+1$ 的窗口在样本集中的序列上滑动,得到决策系统 D 和条件属性集 A。由正常系统调用序列样本集得到的进程的正常行为模型,就是决策系统 D 中由条件属性集 A 确定决策属性 D 的决策规则集。基于关于决策的约简可以得到最小决策规则集,对于由正常系统调用序列样本集转换得到的决策系统 D,根据关于决策的约简生成正常模型中规则的条件,可以得到最小决策规则集用做进程正常模型 $M(L)$。应用粗糙集理论可以从比较小的样本序列中提取预测规则集,得到的行为模型能有效逼近理想正常行为模型。

基于进程正常模型 $M(L)$ 的异常检测借助于训练得到的进程正常模型,可以根据该进程产生的系统调用序列判断出进程的运行状态是否为正常。一条规则同一个长度为 $L+1$ 的序列段匹配,应满足如下条件:a)规则适用的长度为 $L+1$;b)规则条件中描述的序列段各个位置上系统调用应与序列段中的情况相符。系统调用序列的异常度定义为序列中正常模型规则集预测失败的序列段个数与序列中序列段总数之比。

基于进程正常模型 $M(L)$ 的异常检测算法如下:

```
Y = {待分析的系统调用序列};
for (i = 1;i < = Y 的个数;i = i + L + 1)
{ W = 在 Y 中从 i 开始截取长度为 L + 1 的序列段;
根据 W 中前 L 个序列段在 M(L)形成的规则集 D 中查找规则;
if ( 没找到这样的规则)
{ W 中前 L 个系统调用的预测 = 训练集中最常出现的预测结果;
}
else
{ W 中前 L 个系统调用的预测 = 规则集 D 中出现次数最多的预测结果;}
if (W 中前 L 个系统调用的预测 = = W 中第 L + 1 个序列段)
{printf("%s","预测成功");}
else
{printf("%s","预测失败");}
}
```

4　实验测试与分析

基本粗糙集理论的方法需要有高质量的样本数据进行训练。在实际的考试系统应用中,能获得的训练样本数目不多,

而且绝大多数情况下只能得到正常的训练样本,有标定的异常数据很难获得。这样在实用时就会产生误报,在实用中也有一定的困难。本实验采用 UNM 提供的 Login、Ps、wu-ftpd 进程数据集[10]对考试系统中基于粗糙集理论的动态防御技术部分进行实验。实验时先对数据进行切割处理得到序列段,将其中 10% 的序列段作为正常训练集,剩余 90% 作为正常测试集,采用不同的窗口大小,不同步幅大小多次构建模型, 比较窗口大小与检测效率关系。实验采用的部分数据集如表 1 所示,实验中程序的部分异常度比较如表 2 所示。从表 2 中可以看出异常序列的异常度都明显高于正常序列的异常度。根据异常度的差异, 选择适当的阈值可以准确区分异常状态。实验结果表明,在窗口长度取 21 时预测相对最稳定,基于粗糙集理论的检测方法能更快速区分正常和异常行为,考试系统的安全相应地得到了改善。

表 1 实验用的数据集

	正常序列	异常序列
login	两条 login 正常序列	两条木马程序(recovered Trojan login data, homegrown Trojan login data)
ps	两条 ps 正常序列	两条木马程序(recovered Trojan login data,homegrown Trojan login data)
named	一条 named 正常序列	一条 buffer overflow 攻击序列
wu-ftpd	两条 ftpd2.4 正常进程	一条远程单字节溢出攻击序列

表 2 程序的异常度比较

	窗口大小	7	11	15	21	25
login	normal	0.31	1.18	0.31	1.57	0.83
	homegrown	1.39	4.82	3.75	7.50	3.54
	recovered	9.27	11.23	5.48	9.27	7.82
ps	normal	0.41	0.45	0.63	1.90	1.2
	homegrown	1.93	2.38	2.49	5.78	4.85
	recovered	3.66	6.10	4.56	6.50	5.29
named	normal	1.34	0.84	1.59	0.92	1.21
	abnormal	83.24	80.5	80.07	81.6	80.09
ftpd	normal	1.53	2.26	3.67	2.40	2.97
	abnormal	38.33	34.56	34.54	36.03	35.51

5 结束语

本文在将传统静态安全防御技术与动态防御技术相结合的基础上,引入粗糙集理论来更快更有效地识别考试系统的正常和异常行为,保护考试系统的安全,使系统能更好地防御外来的异常情况。该系统在实时检测中对系统性能的影响很小,是一种高效低负荷的实时检测方法。不过怎样提高检测的准确性和及时性,出现了报警怎样对其进行分析, 用户如何处理报警,如何更好地实现安全系统的动态平衡等都是以后需要解决的问题。

参考文献:

[1] 张萍,王建忠,余 ,等.免疫网络安全考试系统[J].计算机应用研究,2007,24(8):162-164.

[2] PAWLAK Z. Rough sets-theoretical aspect of reasoning about data [M]. Dordrecht: Kluwer Academic Publishers,1991.

[3] 张文修,吴伟志,等.粗糙集理论与方法[M].北京:科学出版社,2001.

[4] FORREST S, HOFMEYR S A, SOMAYAJI A. Computer immunology[J]. Communications of the ACM, 1997,40(10):88-96.

[5] HOFMEYR S A, FORREST S. Immunity by design:an artificial immune system[C]//Proc of GECCO'99.1999.

[6] HOFMEYR S A, FORREST S. Architecture for an artificial immune system[J]. Evolutionary Computation, 2000, 8(4):443-473.

[7] FORREST S,HOFMEYR S A. Immunlogy as processing[C]//SEGAL L A,COHEN I R. Proc of Design Principles for Immune Systems & Other Distributed Autonomous Systems. Oxford: Oxford University Press, 2000.

[8] HOFMEYR. An immunological model of distributed detection and its application to computer security [D]. Mexico: University of New Mexico,1999.

[9] LEE W, STOLFO S, CHAN P. Learning patterns from unix process executiontraces for intrusion detection[C]//Proc of AAAI'97 Workshop on AI Methods in Fraud and Risk Management. 1997: 50-56.

[10] [EB/OL]. http: //www. cs. num. edu/ ~immsec/.

(上接第 1011 页)

[9] 王伟. 广义预测控制理论及其应用[M]. 北京:科学出版社,1998: 46-64.

[10] 胡耀华,贾欣乐. 广义预测控制的直接算法[J]. 控制与决策,2000,15(2):221-223.

[11] RAMOND G, DUMUR D, BOUCHER P. Direct adaptive constrained receding horizon predictive control with conditional to motor drives with variable inertia[C]//Proc of the 38th IEEE Conference on Decision and Control. 1999:755-760.

[12] RAMOND G, DUMUR D,BOUCHER P. Application of direct adaptive generalized predictive control to an automatic gear box with a continuous variable transmission[C]//Proc of IEEE International Conference on Control Applications. 2000:303-308.

[13] PIMENTA K B, ROSARIO J M,DUMUR D. Application of direct adaptive generalized predictive control (GPCAD) to a robotic joint [J]. IEEE International Symposium on Industrial Electronics, 2003,2:1011-1016.

[14] 师五喜,霍伟,吴宏鑫.一类未知非线性离散系统的直接自适应模糊预测控制[J].自动化学报,2004,30(5):664-670.

[15] 师五喜. 未知参数多变量线性系统自适应模糊广义预测控制[J].控制与决策,2009,34(2):23-26.

[16] 师五喜. 自适应模糊广义预测控制研究[D].北京:北京航空航天大学,2003:53.

[17] 陈志旺, 王洪瑞. 非线性模糊自适应直接广义预测控制[J].电机与控制学报,2007,11(1): 55-59.

[18] 师五喜. 基于跟踪误差调节的非线性自适应模糊预测控制[J].控制与决策,2006,21(3):297-304.

[19] 王洪瑞,陈志旺,李建雄.非线性系统参数自适应直接广义预测控制[J].自动化学报,2007,33(10):1110-1114.

[20] GOODWIN G C, SIN K S. Adaptive filtering, predictive and control [M]. Englewood Cliffs, New Jersey: Prentice-Hall,1984:91-94.