

基于自律计算的主动入侵检测模型

曹继邦, 吴庆涛, 郑瑞娟, 张聚伟

(河南科技大学 电子信息工程学院, 河南 洛阳 471003)

摘 要: 针对传统入侵检测系统的被动检测机制问题, 借鉴自律计算思想, 提出了一个具有自律特性的主动入侵检测模型。该模型以自主管理器为核心, 在 Agent 协同层引入拍卖机制感知系统环境变化和管理调配被管资源, 自主处理入侵信息, 实现对入侵行为的主动响应。实验结果表明, 通过合理的参数配置, 该模型能有效提高入侵检测系统的自适应性和检测准确率。

关键词: 入侵检测; 自律计算; 自主管理器; Agent 协同; 主动检测

中图分类号: TN47

文献标识码: A

文章编号: 1000-7180(2012)12-0031-05

Proactive Intrusion Detection Model Based on Autonomic Computing

CAO Ji-bang, WU Qing-tao, ZHENG Rui-juan, ZHANG Ju-wei

(Henan University of Science and Technology, Electronic and Information Engineering College, Luoyang 471003, China)

Abstract: Aiming at the defect of passive detection mechanism in traditional intrusion detection systems, a proactive intrusion detection model with autonomic characteristic is proposed in this paper. This model is based on autonomic manager and integrates multi-attribute Auction Mechanism into Agent coordination layer for perceiving the changing of the systems environment, adapting configuration management. Experimental results show that given an appropriate setting, the model can enhance the system's self-adaptive performance and obtain higher detection accuracy.

Key words: intrusion detection; autonomic computing; autonomic manager; Agent coordination; proactive detection

1 引言

入侵检测作为一种重要的安全检测技术, 已经被广泛的应用在网络安全领域。但传统入侵检测技术已经无法适应当前网络环境的需要。为了提高入侵检测的检测率、降低漏报率, 有效识别出攻击, 人工智能、移动代理、数据挖掘和信息关联等^[1-3]技术和方法均被应用到入侵检测中。

自律计算能够克服计算系统的异构性和复杂性, 被认为是实现系统自治、解决系统安全问题的新的有效途径。

本文借鉴自律计算的思想, 将自律计算与入侵检测相结合, 尝试解决传统入侵检测技术被动检测机制导致的自适应能力差的问题, 提出了一个具有

自律特性的主动入侵检测模型 (autonomic characteristic intrusion detection model, ACIDM)

2 自律入侵检测模型

主动入侵检测模型如图 1 所示。在该模型中, 被管资源 (Managed Resource, MR)^[4] 是各种物理资源和虚拟资源, 主要包括数据库、服务器、路由器、应用模块、Web 服务器和虚拟机器, 主机日志、网络数据包和防火墙报警信息等各种物理资源和虚拟资源。它们必须满足可管理性, 如状态可观测和状态可调整。资源的状态是指各种能够反映资源当前状态的数据 (事件), 包括各种日志事件和实时事件, 如资源的操作状态、性能状态 (吞吐量、资源利用率等) 和异常事件等。MR 由 Agent 协同层进行统一调度和

收稿日期: 2012-04-15; 修回日期: 2012-05-10

基金项目: 国家自然科学基金资助项目 (61003035, 61142002); 河南省科技攻关项目 (112102210187)

管理.

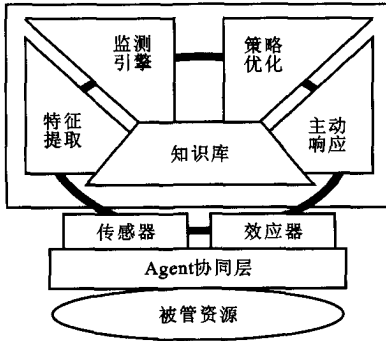


图1 自律入侵检测模型

在模型中,特征提取、检测引擎、策略优化、自主响应和知识库构成了自主管理器(Autonomic Manager,简称AM),四部分之间的连接线表明他们之间的消息和信息共享关系。即策略优化部分可能会要求检测引擎部分收集更多的信息,检测引擎部分也可能直接触发策略优化部分实施优化。四个部分之间通过异步通信技术进行协作。

2.1 Agent 协同层

针对不同类型的MR,Agent协同层采用不同的智能Agent为AM提供数据支持。这些Agent均是能够独立运行的实体。Agent实体捕获MR信息并做预处理,去除冗余信息,最终把信息提交给AM。Agent协同层的另外一个主要功能是接收AM的信息反馈,自主调节系统环境,使其能够动态适应环境变化,以实现资源的动态配置、服务的动态合成、系统参数的动态校正。例如,当系统检测到入侵事件发生时,控制防火墙的Agent根据AM发来的警报与控制信息,动态更新防火墙的阻断策略,使防火墙在一定时间内根据IP、端口等信息阻断后续攻击事件的发生,从而形成一个完整动态自适应入侵检测系统。

Agent协同层中的各个Agent之间协商工作,构成多Agent系统(Multi-Agent System, MAS),MAS引入拍卖机制来解决任务分配、资源配置等问题,以优化系统性能^[5-6]。拍卖方法的种类很多,各种不同的拍卖方法服务于不同的拍卖环境。本文采用多属性拍卖方法。

定义1 多属性拍卖模型

$M = \langle A, B, S, V, C, Res \rangle$,其中 A 为所有物品的属性所组成的空间, $A = A_1 \times \dots \times A_m$ 。被拍卖的物品具有 m 个属性,用 $a_1, \dots, a_m$ 来表示,取值范围在 $A_1, \dots, A_m$ 之间, $a = (a_1, \dots, a_m)$ 称为由物品的属性多构成的向量,其中 $a \in A$ 。

在拍卖的过程中,只有一个买方为记为 B , B 要在拍卖中购买物品。

由拍卖的卖方集合组成的集合表示为 S ,包含 n 个卖方, $S = \{1, \dots, n\}$ 。

$V: A \rightarrow R$ 为 B 的属性权值函数(R 为实数集合),即 $V(a) \in R$ 表示买方 B 根据属性 a 对物品的评价。

$C = \{C_1, \dots, C_n\}$,其中 C_i 表示为物品成本函数,那么 $C_i(a) \in R$ 就是卖方根据属性 a 计算出的物品成本值。

$Result$ 为拍卖结果, $Result = (P, a)$,其中 $P \in R$ 表示为成交的价格, $a \in A$ 表示为由成交的属性所组成的向量。此时作为买方的 B 根据属性 a 对物品的收益定义为 $U = V(a) - P$ 。作为卖方的 S_i 根据属性 a 对物品的收益定义为 $U_i = P - C_i(a)$ 。

拍卖流程分为四个步骤:

①由买方公布估价函数 $V'(a)$ (V' 可与 V 有所区别)。

②每个卖方 i 进行暗标叫价 $B_i \geq 0$ 。

③确定成交卖方。首先买方确定备选成交卖方集合

$W = \{\omega | (\omega \in S) \wedge (B_\omega = \max_{i \in S} (B_i)) \wedge (B_\omega > 0)\}$
(B_ω 为 ω 的叫价),若 $W = \emptyset$,则没有成交卖方,拍卖结束。若 $W \neq \emptyset$,则随机产生 $\omega \in W$ 为成交卖方。并令

$$B^* = \max_{i \in S} (B_i)$$

其中 $\max_{i \in S} (B_i) \stackrel{\text{def}}{=} \min_{i \in S} (\max_{j \in S - \{i\}} (B_j))$,易知 $\max_{i \in S} (B_i) = \max_{i \in S - \{\omega\}} (B_i)$ 。 $\max_2$ 的直观含义为去除最大的一个元素后剩余元素中的最大值,例如 $\max_2(1, 2, 3) = 2$, $\max_2(1, 2, 3, 3) = 3$ 。 B^* 的直观含义为除成交卖方 ω 之外的其他卖方的最高叫价。

④由成交卖方提出成交方案 (P_i, a_i) ,合法提案需要满足 $V'(a_i) - P_i = B^*$,成交卖方与买方以此成交方案成交,拍卖结束。

2.2 传感器与效应器

系统的软件与硬件由不同的供应商提供,需要定义统一的标准接口来实现通信,屏蔽由于内部结构的不同所导致的异构性。为此,通过语义技术构建效应器与传感器来解决上述问题:

定义2 设 $T = \{t_1, \dots, t_n\}$ 是反映MR当前所处状态的特征向量; $V = \{v_1, \dots, v_m\}$ 是反映MR状态变化的集合; $O = (C, R)$ 是领域知识,其中, C 代表领域的概念所构成的集合, R 代表 C 上的一个关

系; $\xi=\{get, report\}$ 表示为一组对 MR 的操作,则传感器用一个四元组定义为: $Sensor=(T, V, O, \xi)$,其中,对于 $\forall t_i, v_j (1 \leq i \leq n, 1 \leq j \leq m)$,有 $t_i \in C, v_j \in C$.

从定义 2 中可以看出,MR 的特征向量与状态变化集合均遵从特定的领域知识,领域知识通过语义语言进行描述,以此支持机器推理、自主解释及状态的自我觉察. get 操作用于获取 MR 的特征向量, $get(t_i)$ 表示 AM 通过 Sensor 向 MR 取得特性 t_i ; $report$ 操作用于向 MR 报告自身状态的变化, $report(v_j)$ 表示 MR 向 AM 报告所发生的事件 v_j .

定义 3 设 $A=\{a_1, \dots, a_n\}$ 表示为对 MR 操作的动作序列; $Q=\{q_1, \dots, q_m\}$ 是由 AM 提供的、等待 MR 请求的动作序列集; $O=(C, R)$ 是领域知识,其中, C 代表领域的概念所构成的集合, R 代表 C 上的一个关系; $\phi=\{set, request\}$ 定义为一组操作的集合,则效应器用一个四元组定义为: $Effector=(A, Q, O, \phi)$,其中,对于 $\forall a_i, q_j (1 \leq i \leq n, 1 \leq j \leq m)$,有 $a_i \in C, q_j \in C$.

从定义 3 中可以看出, set 操作用来执行动作序列, $set(a_i)$ 表示 AM 通过效应器对 MR 执行动作 a_i ;操作 $request$ 用于向 AM 发送请求, $request(q_j)$ 表示 MR 向 AM 执行请求动作 q_j .

2.3 数据归一化

由于采集数据的多元异构性,要对其进行预处理.在预处理阶段,对异构数据源采用归一化^[7]进行处理,实现数据格式和类型的统一.在 IDS 中,特征向量之间的欧式距离计算结果会因为一个数值对另外一个数值的影响产生很大差异.数据经过归一化处理后,将有效降低这种差异性.归一化处理如下.首先,计算特征向量的各个特征属性的标准差和均值:

$$\text{mean}[j] = \frac{1}{n} \sum_{i=1}^n \text{instance}_i[j] \quad (1)$$

$$\text{standard}[j] = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\text{instance}_i[j] - \text{mean}[j])^2} \quad (2)$$

其中,公式(1)中的 $\text{instance}_i[j]$ 表示在训练样本 i 中,第 j 个属性的取值,样本的大小用 n 表示.然后,通过公式(3)将样本进行转换:

$$\text{newinstance}[j] = \frac{\text{instance}[j] - \text{mean}[j]}{\text{standard}[j]} \quad (3)$$

通过公式(3)将特征向量的每个属性值转换为这个属性的原值和均值的差与标准差的商,这样就实现了数据的归一化,把特征向量的属性值映射到

标准取值空间.

2.4 自主管理器(AM)

AM 根据 Agent 协同层提供的数据信息,由特征提取部件提取入侵特征,存入知识库.检测引擎根据知识库检测规则进行检测,若有入侵发生,则调用自主响应部件做出响应.自主响应部件会根据知识库中的策略知识,自主调节系统环境,使其能够动态适应环境变化,以实现资源的动态配置、服务的动态合成、系统参数的动态校正.

知识库:此部件中的知识主要包含状态判定知识(KD)、策略知识(KP)、问题求解知识(KS)、检测规则知识(KR)四类.简单讲,知识库 $K=KD+KP+KS+KR$,其中, KD 主要用来监测被管资源和内外部环境的状态参数等; KP 主要包括 IT 管理者所定义的策略和通过机器学习所获得的策略,即从状态到动作的映射; KS 主要包括规划、配置、优化等知识,是在系统运行状态偏离预期时的问题求解; KR 主要包括基于误用入侵检测的特征库和基于异常入侵检测的行为模式库.其他所有部件都是在知识库的支持下运行的.

特征提取:传感器获取 Agent 协同层捕获的数据,通过对数据的分析、关联与融合,采用扩张矩阵理论来提取入侵特征^[8].该方法通过构造入侵子集和正常子集的扩张矩阵,建立其最优特征子集选择的整数规划模型,并利用简单遗传算法求解,最终生成可用于检测特定类型攻击的最优规则.

检测引擎:检测引擎是自主管理器的检测功能部件,可完成对入侵模式的判别.在该部件中,检测引擎采用混合检测技术.

策略优化:策略优化采用机器学习、智能规划等有关技术实现自主管理器的策略优化,以适应变化环境的需求.

自主响应:自主响应部件根据知识库的策略知识完成针对入侵行为的响应.

3 仿真实验及性能分析

3.1 实验数据集

为了验证本文提出模型的有效性,实验数据采用 KDD99 数据集.由于 KDD99 数据集庞大,采取对数据集进行抽样的方式,对训练数据集与测试数据集中的数据分别进行随机抽样,然后将随机抽样出的数据进行组合,生成实验用的训练数据子集和测试数据子集,其中,训练数据子集 21836 条,测试数据子集 36715 条.

3.2 实验方案

为更清晰、准确地验证模型的可行性与有效性,我们设计了以下两个实验.

实验 1:比较自主管理器检测引擎采用混合检测技术和滥用检测技术在检测准确率方面的性能.

实验 2:比较 ACIDM 与人工神经网络(Artificial Neural Network, ANN)方法和 SVM(Support Vector Machines, SVM)方法的入侵检测模型在检测准确率和检测时间方面的性能.

3.3 实验结果与分析

首先采用相同的数据集,AM 中的检测引擎采用混合检测技术和滥用检测技术检测准确率性能对比如表 1 所示.

表 1 采用滥用与混合检测技术性能对比

攻击方式	检测率(%)	
	滥用监测技术	混合检测技术
Normal	93.26	98.35
Dos	83.47	98.64
U2R	76.68	95.28
R2L	74.57	94.75
Probe	86.69	98.43

从表 1 可以看出,采用混合检测技术的检测性能明显高于采用滥用检测技术的性能.接下来,将 ACIDM 与应用 ANN 方法和 SVM 方法的入侵检测模型在检测准确率和检测时间方面的性能,试验结果如图 2、图 3 所示.

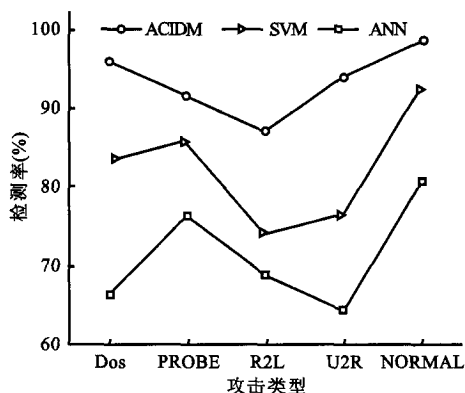


图 2 ACIDM、ANN 和 SVM 检测准确率

从图 2 中可以看出,ACIDM 在检测准确率方面的性能高于采用 ANN、SVM 的入侵检测系统模型,检测准确率有了很大的提高.检测时间性能对比如图 3 所示.

从图 3 中可以看出,虽然 ACIDM 在检测准确率方面的性能高于采用 ANN、SVM 的入侵检测系

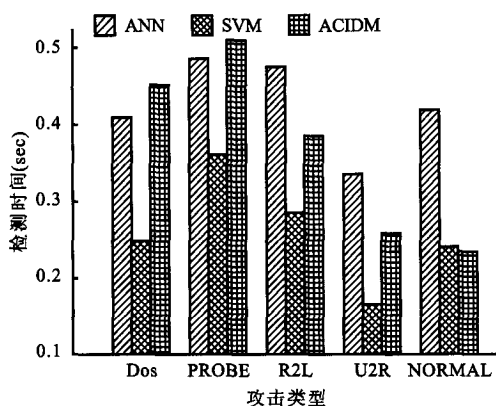


图 3 ACIDM、ANN 和 SVM 检测时间

统模型,但是在检测时间性能方面却不理想,原因在于 ACIDM 在检测过程中,为了提高系统的自适应性,不仅采用混合入侵检测技术,而且添加了自主响应和响应策略优化,导致检测时间较长,采用 SVM 的入侵检测系统模型检测时间最短.

根据实验结果可以得出如下结论:

(1)检测引擎采用混合检测技术比滥用检测技术在检测准确率方面的性能有较大提高.

(2)ACIDM 比 ANN、SVM 的检测准确率高,但是检测时间较长.

4 结束语

本文提出的 ACIDM 是一个基于自律计算技术的入侵检测模型,主要为解决入侵检测系统自适应性的问题而提出的,它结合了各种主要的入侵检测技术的优点,混合使用了误入侵检测和异常入侵检测技术,在检测准确率和自适应性方面有较大提高. ACIDM 表现出很高的检测性能,但检测时间较长,这也是下一步研究和改进的重点.

参考文献:

- [1] 李涵. 一种改进的聚类方法在异常检测中的应用[J]. 微电子学与计算机, 2010, 27(8): 66-69.
- [2] 张家超,孔媛媛. 结合 SVM 与免疫遗传算法设计 IDS 的检测算法[J]. 微电子学与计算机, 2008, 25(10): 207-209.
- [3] 严巧. 基于免疫机理的入侵检测系统的数学描述[J]. 计算机科学, 2009, 36(6): 79-84.
- [4] 廖备水,李石坚等. 自主计算概念模型及实现方法[J]. 软件学报, 2008, 19(4): 779-802.
- [5] 张海俊. 基于主体的自主计算研究[D]. 北京:中国科学院计算技术研究所, 2005.
- [6] 石纯一,张伟. 基于 Agent 的计算[M]. 北京:清华大学出版社, 2007.

- [7] 庞清乐,孙同景,杨福刚,等. 基于粗集理论的归一化方法[J]. 计算机工程,2007,35(8):36-38.

- [8] 陈志贤,黄皓. 应用扩张矩阵理论的攻击特征提取[J]. 计算机科学,2010,37(4):49-51.

作者简介:



曹继邦 男,(1987—),硕士研究生. 研究方向为网络安全.



郑瑞娟 女,(1980—),博士,副教授. 研究方向为网络信息安全、入侵容忍.



吴庆涛 男,(1975—),博士,副教授,硕士生导师. 研究方向为网络与信息安全、智能计算.



张聚伟 男,(1978—),博士,副教授. 研究方向为无线传感器网络.

(上接第 30 页)

- [3] Oliveira R, Jagirdar A, Chakraborty T J. A TMR scheme for SEU mitigation in scan flip-flops[C]// Proceedings of the 8th International Symposium on Quality Electronic Design (ISQED'07). San Jose: 2007:905-910.
- [4] Calin T, Nicolaidis M, Velazco R. Upset hardened memory design for submicron CMOS technology[J]. Nuclear Science, 1996,43(6):2876-2878.
- [5] Mongkolkachit P, Bhuva B. Design technique for mitigation of alpha-particle-induced single-event transients in combinational logic[J]. IEEE Transactions on

Device and Materials Reliability, 2003,3(3):89-92.

作者简介:

宋丽丽 女,(1986—),硕士研究生. 研究方向为超大规模集成电路设计.

来逢昌 男,(1961—),副教授. 研究方向为超大规模集成电路设计、深亚微米芯片物理实现技术.

肖立伊 女,(1961—),教授,博士生导师. 研究方向为VHDL-AMS混合信号模拟器设计、SoC/IP核设计与标准化.