概念级误用检测系统的认知能力研究

邹 涛[1,2],张 翠[1],田新广[1,2],张尔扬[1]
(1. 国防科学技术大学,湖南长沙 410073;2. 北京首信股份有限公司 IP 所,北京 100016)

摘 要: 无法检测到未知攻击以及不能自动更新知识库是现有误用检测系统的两大缺点. 概念级误用检测系统(CLMDS)中利用 SRRW 特征选取算法、CHCL 技术和独立双模型互训练结构极大地提升了系统的认知能力,有效地解决了上述问题. 文章从静态和动态两个层面对系统的认知能力进行了分析;实验结果表明:CLMDS 具有很强的认知能力,不但能检测到未知的攻击样式,而且还能实现知识库的自动更新.

关键词: 入侵检测系统; 特征选取; 机器学习; 误用检测
中图分类号: TP393.08 **文献标识码**: A **文章编号**: 0372-2112 (2004) 10-1694-04

A Study on Apperception Ability of Concept Level Misuse Detection System

ZOU Tao[1,2], ZHANG Cui[1], TIAN Xin-guang[1,2], ZHANG Er-yang[1]
(1. *National Univ. of Defense Technology, Changsha, Hunan* 410073, *China*;
2. *IP Network Technology Institute of Beijing Capitel Co., Ltd., Beijing* 100016, *China*)

Abstract: Current misuse detection systems are of little use for new attacks and they cannot automatically update their rule databases. SRRW, CHCL and the technology of co-training for independent dual-model greatly improve the apperception ability of CLMDS, and provide a good solution for the limitation of misuse detection systems. Apperception ability is analyzed from static aspect as well as dynamic aspect. Results of the experiments show that with powerful apperception ability CLMDS can not only detect new attacks but also update its own rule base automatically.

Key words: IDS; feature selection; machine learning; misuse detection

1 引言

无法检测未知的攻击以及无法实现知识库的自动更新是目前误用检测系统的两大缺陷. 传统的误用检测系统由安全专家对已知的攻击进行分析,选取攻击的特征并通过编码形成检测规则. 这样做存在的问题是:(1)检测性能取决于专家对于攻击的认知程度及编码方式,缺乏可靠性和通用性;(2)面对大量的攻击实例,人工方式归纳能力弱,使得误用检测系统缺乏对新的攻击实例和未知攻击样式的检测能力;(3)知识库更新慢,而且需要大量人工的烦琐劳动.

针对问题(1),已有研究者将智能算法,如:神经网络、数据挖掘、人工免疫学等技术应用于 IDS[1~3];同时,智能算法中的归纳学习技术也部分地解决了问题(2),使 IDS 能够在一定程度上识别已知攻击样式中的新实例. 为了实现知识库的自动更新,重训练及 IP 陷阱[4]等方法也被引入到 IDS 中来. 但问题并未因此而得以彻底解决:新的攻击方式依然无法被检测;依靠重训练来实现知识库更新的方法往往会被攻击者所利用,逐渐训练学习器,使之出现漏报. 而基于 IP 陷阱的方法,其关于新攻击的信息来源于针对 IP 陷阱的有限攻击. 一旦高明的攻击者避开了 IP 陷阱的诱惑,直接对目标机进行攻击,系统将因为没有新的训练实例而无法实现知识库更新.

上述问题的解决直接取决于 IDS 对于攻击的认知能力. 本文以国防科学技术大学与首信联合研制的概念级误用检测系统(CLMDS)原型为背景,通过介绍其所采用的 SRRW 算法、CHCL 技术和独立双模型互训练结构,从静态认知和动态认知两个层面分析了其所具有的优良认知性能. CLMDS 所具有的认知能力使其能够有效解决前文所述误用检测系统的主要缺点.

2 认知能力分析

IDS 与攻击是一个不断斗争的过程. 面对已出现的攻击,IDS 需要一个好的检测模型,实现对已知及未知攻击的最大化检测;与此同时,一个训练好的 IDS 在面对新的攻击样式出现的情况下,还必须针对这些新攻击样式不断更新已有的知识库. 因此,根据 IDS 的应用背景,可以从两个方面定义其认知能力:

(1)在给定训练集(包含已知攻击和正常行为实例)的情况下,IDS 的静态认知能力;

收稿日期:2003-08-10;修回日期:2004-01-18
基金项目:普天首信重大科研项目(No. 021125)

(2)在 IDS 的工作阶段,系统对于新信息的动态认知能力;

静态认知能力决定了在相同已知条件下且不做知识更新时,IDS 系统对于攻击及正常行为的正确检测能力;而动态认知能力则反映了系统在给定静态基准点基础上,对工作环境的自适应能力.

首先以机器学习为背景分析静态认知能力.机器学习实际上可以看成一个搜索问题,即从特定的训练数据中对一个非常大的假设空间进行搜索,以确定一个能够将观察到的数据与学习器得到的知识拟合得最好的假设[5].

定义 1　训练样例集合

$$XC=\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1m} & c_1 \\ f_{21} & f_{22} & \cdots & f_{2m} & c_2 \\ \cdots & & & & \\ f_{n1} & f_{n2} & \cdots & f_{nm} & c_n \end{bmatrix}_{n\times(m+1)}=\begin{bmatrix} x_1 & c_1 \\ x_2 & c_2 \\ \cdots & \cdots \\ x_n & c_n \end{bmatrix},$$

其中 $x_i=[f_{i1},f_{i2},f_{i3},\cdots\cdots,f_{im}]$为第 i 个样例的特征矢量(为简化起见,也可不加区别地称为样例 x_i),f_{ij}为其第 j 个特征,c_i 为类别标识.记由所有样例构成的矩阵为 X,则训练样例集合可简记为 $C:X\to\{c_i\}$.

定义 2　一个假设 h 与样例集合 XC 一致,当且仅当:$h(x)=c(x)\ (\forall\langle x,c(x)\rangle\in XC)$.

静态认知能力直接影响假设 h 在未知测试集合 X 上的分类性能.静态认知能力越强,则假设 h 在未知测试集合 X' 上的分类性能越好.增强静态认知能力是解决误用检测系统无法检测未知攻击方式缺点的途径之一.

定义 3　给定在训练样例集合 XC 下得到的目标假设 h_0 和工作集合序列 X'_1、X'_2、……,学习器在时刻 t 的动态认知能力 $\mathrm{Prec}\,i_{h_t}(D)$ 定义为 t 时刻的假设 h_t 在样例分布 D 上的分类准确率:$\mathrm{Prec}\,i_{h_t}(D)=\mathrm{Func}(h_0,X'_1,X'_2,\cdots,X'_t)$.

目前的误用检测系统正是因为缺乏动态认知能力才无法实现知识库的自动更新.而 CLMDS 在具有较强的静态认知能力的基础之上通过独立双模型互训练结构实现了部分的动态认知.

3　静态认知能力

CLMDS 中通过两种手段来增强节点的静态认知能力:最优特征子集选取(Optimal Feature Subset Selection,OFSS)和攻击类别的概念层次化产生(Concept Hierarchy Generation for Labels,CHGL).

3.1　OFSS 分析

通过分析及实验可以发现,不同的攻击由于其所采用的手段不同,导致在检测阶段所需要的检测特征也不同.比如:land 攻击只与源地址和目的地址特征有关;teardrop 攻击与协议类型和错误分片数目特征有关;而扫描攻击则与一些基于时间的统计特征有关.因此,可以在建立误用检测模型之前先对每种攻击做 OFSS,以使学习更具有针对性.

CLMDS 中采用 SRRW 算法实现 OFSS[6].其基本思想是将遗传算法(GA)与学习算法打包在一起,共同完成 OFSS.其中,GA 的适应度函数直接用于评价特征子集的性能,是 SRRW 的关键.适应度函数的设计依据统计学习理论中的期望风险最小化原则.

算法在实例空间 D_n 上的经验风险泛函和期望风险泛函分别记为 $R_{\mathrm{emp}}(D_n)$ 和 $R(D_n)$.根据统计学习理论关于推广性的界的结论,经验风险和期望风险之间以至少 $1-\eta$ 的概率满足下式:

$$R(D_\eta)\le R_{\mathrm{emp}}(D_n)+\sqrt{\frac{VC(H)(\ln(2n/VC(H))+1)-\ln(\eta/4)}{n}} \tag{1}$$

公式(1)中的 $VC(H)$ 为假设空间的 VC 维、n 为训练集中样例的个数.定义在 D_n 上的假设空间 H 的 VC 维是可被 H 打散的 D_n 的最大有限子集的大小.如果 D_n 的任意有限大的子集可被 H 打散,那么 $VC(H)\equiv\infty$.根据 PAC 计算学习理论,学习样本复杂度满足式(2):

$$m\ge\frac{1}{\varepsilon}[4\log_2(2/\delta)+8VC(H)\log_2(13/\varepsilon)] \tag{2}$$

其中 m 为训练样例数目.式(2)表明训练样例个数 m 要大于等于公式右侧取值时,才能以 $1-\delta$ 的概率学习得到错误率不大于 ε 的目标概念.对于任意有限的 H,$VC(H)\le\log_2|H|$.而假设空间 H 的大小直接与特征的维数有关,特征维数越大,$|H|$ 一般也越大.因此,合理设计 SRRW 算法中的适应度函数不但可以提高机器学习的泛化能力,同时还能降低样本复杂度.根据上述分析,SRRW 算法中的适应度函数采用了 min_(features,$R_{\mathrm{emp}}(D_n)$)偏置.

3.2　CHGL 分析

攻击知识的学习可以在不同的概念层次进行.训练集中的每个样例都可以表示成$\langle x,c\rangle$的形式.其中 x 表示样例的特征矢量;c 表示样例的攻击标识.c 的层次决定了目标概念所处的层次.如果 c = 攻击 i,其正例为攻击 i 的不同取值实例,则目标概念为攻击 i.此时机器学习的泛化和推广能力体现于对属于攻击 i 的新实例的正确分类上.而如果 c = 攻击大类 i,其中攻击大类 i 是由多种相似的攻击共同组成,其正例为包含这些种攻击的不同取值实例,则目标概念为攻击大类 i.此时机器学习的泛化和推广能力体现于对属于攻击大类 i 的新实例的正确分类上.由于攻击大类 i 包含有多种相似的攻击,因此检测模型可以实现对未知但却与训练集中包含的攻击样式相似的新的攻击类型的检测.这种泛化能力正是现有误用检测系统所缺少的.将 c 的取值从具体的攻击类别上升为包含多种相似攻击类别的攻击大类,实际上是将攻击标识在目标概念的层面上加以提升.这个过程称为攻击类别的概念层次化产生[7](CHGL).

CHGL 增强了 IDS 的认知能力,使其具备了检测未知攻击样式的可能.实际工作中,对未知攻击的检测性能的好坏还取决于攻击大类中包含的具体攻击种类之间的相似性程度.虽然目前已经有不少针对攻击的分类学研究[8~10],但其出发点均不是以入侵检测的具体实施手段为基准.按其分类原则得到的攻击大类对于特定的误用检测系统来说因为不具备类内的攻击相似性而无助于提高系统的认知能力.CLMDS 中采用

相关特征聚类的方法有效地实现了 CHCL. 聚类算法中第 r 大类和第 s 大类之间的距离 $d(r,s)$ 按公式(3)计算:

$$d(r,s) = d(\bar{a}_r, \bar{a}_s) = \sum_{i=1}^{m} |co_{a_{ri}} - co_{a_{si}}| \quad (3)$$

其中, $co_{a_{ri}}$ 表示第 r 大类质心 $\bar{a}_r$ 的第 i 个特征的编码取值, $co_{a_{si}}$ 的含义与之类似, m 为特征个数. 若第 r 大类包含有 n_r 个点,则其质心 $\bar{a}_r$ 定义为该 n_r 个点的均值:

$$\bar{a}_r = \frac{1}{n_r} \sum_{i=1}^{n_r} a_{ri} \quad (4)$$

经过 CHCL 之后的误用检测方法称为概念级误用检测(CLMD)[11]. CLMD 的静态认知能力经由 CHCL 得以提升,具备了检测未知攻击样式的能力.

4 动态认知能力

增强学习采用的"延时回报 -修正 -累计回报最大化"思想被广泛应用于机器人控制等领域. 但入侵检测领域的一大特点是缺乏延时回报. 系统不可能只有等待未知攻击对被保护对象造成了可觉察的危害之后才进行模型的修正. 比较理想的动态认知应该是:随着新实例的不断出现,入侵检测系统的知识和经验也在不断地增加,进而修正检测模型使检测性能不断提升. 图 1 所示的独立双模型互训练结构正是解决该问题的一种有效手段.

其中,SRRW 完成 OFSS,然后根据其结果将特征分裂为两个独立特征子集 OFS1 和 OFS2,分别训练两个学习器进而得到两个相互独立的 CLMD 检测器. 工作阶段,两个独立的 CLMD 检测器分别对工作序列集进行判决检测,选取判决结果中最可信的部分反馈给对方的学习器,完成在线学习. 具有较高准确率的检测器的检测结果作为输出,送给响应模块.

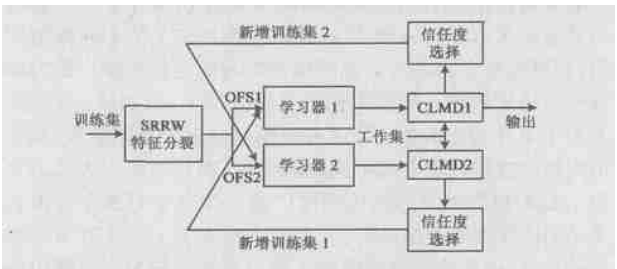

图 1 独立双模型互训练结构

与基于 ANN 和统计方法的重训练相比,独立双模型互训练结构由于在每个检测模型的输出部分增加了信任度选择,只选取检测结果最可靠的部分来为在线学习提供新的训练数据,使得其中出现错误标识的可能性大大减小,保证了在线学习的可靠性. 信任度选择同时也使攻击者逐渐训练 IDS 变得更加困难.

独立双模型互训练结构的动态认知能力取决于两个方面:(1)系统的静态认知能力;(2)双模型之间的互补性. 系统的静态认知能力决定了动态认知系统在每一轮新的训练中所能获得的知识量. CLMD 对于未知攻击的检测能力为系统的动态认知提供了基础. 双模型之间的互补性是决定动态认知系统是否能实现知识更新的另一个关键. 互补性保证了学习器在动态训练过程所需要的的训练数据中包含有"新"的成分. 而互补性则是由 SRRW 和特征分裂得以实现.

5 实验结果及分析

本文以 DARPA98 发布的入侵检测数据集为基础,分别对 CLMDS 中的 SRRW、CLMD 和独立双模型互训练结构进行实验,分析其对 CLMDS 静态及动态认知能力的贡献情况.

表 1 CLMD 对于静态认知能力的贡献

	攻击样例检测数	检测率	正常样例检测数	虚警率
误用	222,266	88.75 %	60,288	0.50 %
CLMD	228,233	91.13 %	60,270	0.53 %

图 2 显示了分别采用全部特征和 SRRW 算法输出的最优特征子集 F1 的情况下,IDS 在两组测试集 TS1 和 TS2 上的泛化能力. 可以看出,SRRW 增强了系统的认知能力,使其可以正确检测到更多的新实例. 表 1 为 CLMD 与常规误用检测系统的性能比较. 表中数据显示:与常规的误用检测系统相比,CLMD 的检测率提高 2.38 %,能够多检测到 5967 条攻击(包括 Apache2、processtable、mscan、saint 等 6 种未知攻击);但同时虚警率上升了 0.03 %,虚警数增加 18 条.

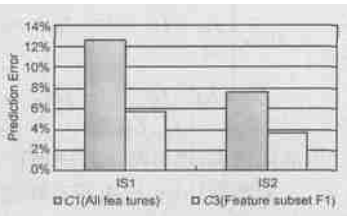

图 2 SRRW 对于静态认知能力的贡献

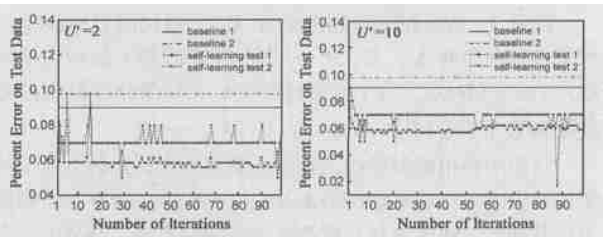

图 3 动态认知能力在两组测试集上采用不同步长的实验结果

图 3 显示了独立双模型互训练结构的动态学习效果. 其中,baseline 曲线是在没有做动态学习的情况下静态检测器在两个测试数据集上的检测性能比较基准. 而 self-learning 曲线则代表了更新步长分别为 2 和 10 的情况下,动态学习检测器的检测性能.

对两组不同步长的动态学习做 4 次实验并取平均,结果如表 2 所示:检测准确率均有较大提高. 由此不难看出:通过动态学习,可以增强系统的动态认知能力进而提高对未知实例的检测准确率.

表 2 动态认知能力 4 次实验的平均错误率结果比较

	$U'=2$		$U'=10$	
	Test 1	Test 2	Test 1	Test 2
Baseline	9.00 %	9.77 %	9.75 %	9.42 %
Self-learning	7.25 %	6.20 %	7.75 %	6.13 %

6 结束语

入侵检测系统对攻击行为的认知能力直接决定了其检测性能的好坏.本文针对现有误用检测系统的两大缺点,以CLMDS为背景,从静态和动态两个角度研究了系统的认知能力;分析了SRRW、CLMD和独立双模型互训练结构对系统认知能力的贡献;实验结果表明:CLMDS具有更强的认知能力,能够有效解决现有误用检测系统存在的主要问题.

参考文献:

[1] Cannady J. Applying neural networks to misuse detection[A]. Proceedings of the 21st National Information Systems Security Conference[C]. Crystal City:NISSC,1998. 368 - 381.

[2] R Lippmann, J Haines, et al. The 1999 DARPA off-line intrusion detection evaluation[J]. Computer Networks, 2000, 34(4): 579 - 595.

[3] Stephanie Forrest, S A Hofmeyr, et al. A sense of self for unix processes [A]. Proceedings of the 1996 IEEE Symposium on Security and Privacy [C]. Oakland: IEEE, 1996. 20 - 128.

[4] CHEN Shuo. A distributed intrusion detection system and its apperception ability[J]. Journal of Software, 2001, 12(2): 225 - 232.

[5] T Mitchell. Machine Learning[M]. New York: McGraw-Hill, 1997. 20 - 21.

[6] ZOU Tao. Data reduction in network based intrusion detection system [J]. Journal of National University of Defense Technology, 2003, 25(6): 16 - 20.

[7] ZOU Tao. Modeling a self-learning detection engine automatically for IDS[A]. Proceedings of the 2003 IEEE RISSP[C]. Changsha: IEEE RISSP, 2003. 10.

[8] S Axelsson. Intrusion detection systems: a survey and taxonomy[R]. Göteborg: Depart. of Computer Engineering, Chalmers University, 2000. 1 - 27.

[9] Peter G Neumann, Donn B Parker. A summary of computer misuse techniques[A]. Proceedings of the 12th National Computer Security Conference[C]. Baltimore, Maryland: NISSC, 1989. 396 - 407.

[10] Ulf Lindqvist, Erland Jonsson, How to systematically classify computer security intrusions[A]. Proceedings of the 1997 IEEE Symposium on Security and Privacy[C]. Oakland, California: IEEE, 1997. 154 - 163.

[11] 邹涛. 一种基于相关特征聚类的层次入侵检测系统[P]. 中国专利:03137094.2, 2003-06-28.

作者简介:

邹　涛　男,1974年11月出生于重庆,现为国防科技大学博士研究生,主要研究方向为网络安全、机器学习.

张　翠　女,1973年12月出生于山东日照,国防科技大学博士研究生,主要研究方向为模式识别与人工智能.

田新广　男,1976年1月出生于河北吴桥,国防科技大学博士研究生,主要研究方向为网络安全、数字信号处理.

张尔扬　男,1941年12月出生于浙江宁波,国防科技大学教授,博士生导师,中国通信学会会士,中国电子学会高级会员,中国密码学会理事,国家"863"航天领域专家委员会专家,享受政府特殊津贴,主要研究方向为信号处理、通信对抗.