

深度防卫的自适应入侵检测系统

王 伟, 陈秀真, 管晓宏, 张响亮

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 为了全面检测黑客入侵和有效提高检测精度, 提出了一种深度防卫的自适应入侵检测系统模型. 该模型按照黑客入侵对系统影响的一般顺序, 使用不同方法对网络行为、用户行为和系统行为 3 个层次涉及到的网络数据包、键盘输入、命令序列、审计日志、文件系统和系统调用进行异常检测, 并利用信息融合技术来融合不同检测器的检测结果, 从而得到合理的入侵判定. 在此基础上, 提出了系统安全风险评估方法, 并由此制定了一种简单、高效的自适应入侵检测策略. 初步实验结果表明, 所提的深度防卫自适应入侵检测模型能够全面、有效地检测系统的异常行为, 可以自适应地动态调整系统安全与系统性能之间的平衡, 具有检测精度高、系统资源消耗小的优点.

关键词: 入侵检测; 深度防卫; 网络安全; 信息融合

中图分类号: TP393.08 **文献标识码:** A **文章编号:** 0253-987X(2005)04-0339-04

Defense-in-Depth Adaptive Intrusion Detection System

Wang Wei, Chen Xiuzhen, Guan Xiaohong, Zhang Xiangliang

(School of Electronics and Information Engineering, Xi an Jiaotong University, Xi an 710049, China)

Abstract: Aiming at detecting intrusions across-the-board and at improving detection accuracy, a novel model of defense-in-depth adaptive intrusion detection system (IDS) was presented. In this model, the behaviors in a computer system are monitored according to the general order of the impact of the attacks and divided into three layers including network behaviors, user behaviors and system behaviors. Various methods are then applied to process the data streams from network packages, keystrokes, audit trails, command sequences, file system and system calls obtained in the three layers for intrusion detection. The monitoring decision on intrusion is made by combining the six individual inferences based on information fusion technique. Based on the risk assessment method proposed in this paper, an efficient adaptive policy is drawn as well for IDS to reduce the expense of system resources. The model is tested and the results show that the model presented is effective to detect intrusions and to balance the system security and performance adaptively and dynamically. The model is promising as well in terms of detection accuracy, system resource requirement and implementation in practice.

Key words: intrusion detection; defense-in-depth; network security; information fusion

目前, 大多数入侵检测系统 (IDS) 只监控系统的某一个层面, 不能从整体上全面、深入地检测系统行为, 因而误报率高居不下, 给 IDS 的实际应用带来了较大的负面影响. 另一方面, 大部分 IDS 未能

兼顾系统安全与系统性能 2 个方面的要素, 过多地加强安全防卫将会造成系统资源浪费, 而盲目要求降低资源消耗则会减小入侵检测的力度, 背离 IDS 的初衷. 本文采取深度安全防护措施, 从系统的多个

收稿日期: 2004-06-21. 作者简介: 王 伟(1976~), 男, 博士生; 管晓宏(联系人), 男, 教授, 博士生导师. 基金项目: 国家杰出青年科学基金资助项目 (6970025); 国家自然科学基金资助项目 (60243001); 国家高技术研究发展计划资助项目 (2001AA140213, 2003AA142060).

层面检测入侵,并利用信息融合技术将多源的检测结果进行融合分析,最终得到一个总的入侵判定,以提高入侵检测的可靠性与准确性.同时,基于文中提出的评估方法制定了一套简单、易行的自适应安全策略,在系统性能与系统安全之间求得最优方案,来构建一种深度防卫的自适应 IDS.

1 深度防卫的自适应 IDS 模型

经验表明,一次完整的黑客攻击一般都要经历 3 个阶段,每个阶段都会产生相应的数据流.第 1 个阶段,在开始攻击主机时黑客首先要进行端口扫描以确定攻击入口,这个过程可以产生与主机相关的

网络行为数据流,其中包含网络连接数据包;第 2 个阶段,黑客将在主机上进行相关操作,此时产生用户行为数据流,其中包含键盘输入、命令序列以及系统审计日志数据;第 3 个阶段,黑客对主机进行一系列操作,这些操作会在主机的系统层面上反映出来,从而产生系统行为数据流,其中包含文件系统属性以及系统调用序列数据.

在综合考虑上述数据流涉及到的网络数据包、键盘输入、命令序列、审计日志、文件系统和系统调用等 6 种数据源的基础上,本文提出了一种深度防卫的自适应 IDS 模型,如图 1 所示.

图 1 深度防卫的自适应 IDS 模型

深度防卫的自适应 IDS 模型的功能模块主要包含数据收集、入侵检测分析和制定检测策略.其中:数据收集模块收集数据流中的 6 种数据;入侵检测分析模块针对这 6 种数据,使用相应的入侵检测方法进行实时的入侵检测分析,最后通过融合检测结果来提高入侵检测的可靠性与准确性,生成检测报告并通知管理员;制定检测策略模块通过网络行为分析和入侵检测报告,对系统进行安全评估,并根据评估结果生成自适应的检测策略.

2 深度防卫的 IDS 算法与实现

2.1 网络行为、用户行为的数据流入侵检测

2.1.1 网络连接数据包、命令序列和审计日志的入侵检测 在基于统计模型的 IDS 中,SRI 开发的新一代入侵检测专家系统(NIDES)取得了较好的检测效果^[1].为了构建一个统一的分析处理单元以降低系统资源的消耗,我们只考虑事件的频率特性,并

使用简化了的 NIDES 统计方法对网络连接数据包、命令序列和审计日志 3 种数据进行统计分析.以审计日志数据为例,在某个时间段 m 内,从系统监测开始计算,第 j 天审计日志中的某一事件(例如 CPU 使用时间)发生的相对概率记为 P_m ,而 B_m 表示 P_m 以及其他小于 P_m 的所有概率之和,那么 B_m 满足 $[0, 1]$ 的正态分布,由此可得某一事件的异常度 $s_m = \Phi^{-1}(1 - (B_m/2))$.考察审计日志信息中的 CPU 使用时间、用户登录信息等系统的 n 个行为特征 $S_i (1 \leq i \leq n)$,若不考虑事件的关联性,则时间段 m 内的审计日志统计值

$$T_m^2 = (S_{m1}^2 + S_{m2}^2 + \dots + S_{mn}^2) / n \quad (1)$$

若将统计值 T^2 作为审计日志事件中总的异常度,则该值越大,表明异常的可能性越大.

按照同样的方法,用网络连接数据包在某个时间段 m 内的频率相关特性来检测网络连接的异常状态,用命令序列在某个时间段 m 内的相关频率特

性来检测命令序列的异常状态.

2.1.2 键盘输入特性 在实际操作中,通过监测用户在键盘输入时键被按下与弹起之间的时间差(延迟时间)以及键弹起与下一次按下的时间差(间隔时间)来确定用户的输入模式.对每个用户而言,键盘输入的延迟时间与间隔时间均符合正态分布^[2].在某个时间段内,某个用户的键盘输入延迟时间和间隔时间的均值分别为 \bar{T}_1, \bar{T}_2 , 方差为 σ_1, σ_2 , 该用户在最近的某个时间段内的键盘输入延迟时间(或间隔时间)为 T_1, T_2, \dots, T_n , 那么延迟时间、间隔时间的异常度 H_1 和 H_2 的计算式为

$$H_j = C_j \prod_{i=1}^n \exp\left[-\frac{(T_i - \bar{T}_j)^2}{2\sigma_j^2}\right], \quad j = 1, 2 \quad (2)$$

式中: C_j 为加权系数,可通过实验确定.一旦发现用户的异常度 H_j 超过阈值,则认为发生了帐号冒用等内部攻击并上报结果,否则更新 \bar{T}_j 和 σ_j .

2.2 系统行为数据流分析

2.2.1 文件系统校验 在 Unix/Linux 环境下,文件系统的完整性校验可以免费使用开放源码的产品,其中功能较强的有 Tripwire^[3]和 Samhain.在实际环境中,可将 Tripwire 融合到系统中以检测文件的非法改动.为了降低系统的资源消耗,在配置 Tripwire 策略时只要求对文件安装、系统日志、系统文件属性等关键的文件进行完整性校验,并且采取定期检测的方法实现实时校验.

2.2.2 系统调用分析 系统调用已经成为近年来的一种重要的入侵检测数据源.1996年,Forrest等人提出了一种简单、高效的短序列匹配算法,并取得了较好的检测效果^[4].在实际环境中,同样可使用这种短序列匹配方法为系统调用建模,检测系统中的异常行为.

2.3 检测结果融合

本文将每个数据源的入侵检测结果作为入侵事件的证据,使用 Dempster-Shafer 合成规则对 6 个检测器的入侵判定进行信息融合,从而得到最终的入侵判决,以提高入侵检测的准确性.

定义 设 Ω 为一有限集合,如果其中的各个元素都是相互排斥的,则称 Ω 为识别框架.如果集函数 $m: 2^\Omega \rightarrow [0, 1]$ 满足条件 $m(\emptyset) = 0, m(A) = 1$, 则称 m 为框架 Ω 上的概率分配函数^[5].

又设 m_1, m_2, \dots, m_n 为识别框架 Ω 上的 n 个基本概率分配函数,则 Dempster-Shafer 的合成概率分配函数^[5]

$$m_1 \oplus m_2 \oplus \dots \oplus m_n = \frac{m_1(A_1) m_2(A_2) \dots m_n(A_n)}{1 - \sum_{A_1, A_2, \dots, A_n = \emptyset} m_1(A_1) m_2(A_2) \dots m_n(A_n)} \quad (3)$$

式中: $(A_1, A_2, \dots, A_n, A) \subset \Omega$.

在深度防卫的入侵检测信息融合模型中,决策空间为{系统行为正常,系统行为异常}.通过系统数据流的 6 个数据源的入侵检测分析,可得到数据源对应的异常度,确定 $A \in [0, 1]$,由此得到数据源的概率分配函数 m (系统行为异常) = A .最后,使用 Dempster-Shafer 合成规则对这 6 个数据源的异常度进行融合,得到一个总的入侵判定并作为结果上报,以供管理员决策或作为自动实时响应的依据.

3 安全风险评估与自适应检测策略

3.1 安全风险评估

深度防卫的 IDS 需从多个层面对系统行为进行全面的监控,因此对系统资源的消耗较大.本文通过自适应的检测策略使得系统在性能与安全之间达成动态平衡,在保障安全的前提下发挥最大的潜力,而安全风险评估是制定与调整自适应策略的主要依据.安全风险是指资产的外部威胁因素和资产的固有漏洞引起资产损失的可能性,安全风险评估

$$R = AVT \quad (4)$$

式中: A 表示资产评估; V 表示漏洞评估; T 表示威胁评估.

3.1.1 资产评估 资产评估是从资产价值的角度来评估系统中的所有信息资产.本文采用半定量赋值的方法进行资产评估,即根据资产的实际情况赋以一个 0~4 之间的值,该值表示对资产的可用性、保密性、完整性的要求程度^[6]

$$A = f(U, C, I) = O_1 \{ \text{lb}[(2^U + 2^C + 2^I)/3] \} \quad (5)$$

式中: U, C 和 I 分别表示资产可用性、保密性和完整性的定量赋值;函数 O_1 表示是对计算结果进行四舍五入运算得到的整数值.

3.1.2 漏洞评估 根据计算机应急响应组织(CERT)统计得出的 446 个普通漏洞和风险(CVE),同时考虑 7 个因素相对于每个漏洞的赋值,赋值范围在数值 0~180 之间,分别表示不同严重程度等级的漏洞^[7].根据此值,再由

$$V = O_1 \left\{ \left[\sum_j (a_j/45) v_j \right] / 7 \right\} \quad (6)$$

来量化漏洞评估的结果.其中: a_j 是 CERT 对每个



漏洞赋予的分值,每个分值统一乘以系数 1/45,并将此分值的范围控制在 0~4 之间; v_j 是与 a_i 对应的漏洞数量.

3.1.3 威胁评估 威胁评估是指对系统及资产构成潜在破坏力的可能性因素或事件.通过对 IDS 的上报事件进行取样、统计和分析,可以获得一段时间内的所有攻击的类型、强度和频度的概率分布.本文将上报的攻击事件分成 5 类:Discovery 类,包含 IP Sweep、DNS Zone Transfer 等;Scan 类,包含 Que-so-Scan、Port-Scan、Santan 等;Escalation 类,包含 Buffer Overflow、Dictionary 等;DoS 类,包含 SYN Flood、Smurf 等;Stealth 类,包含 IP Spoof 等.根据系统的安全目标及各类攻击可能造成的危害,为每类攻击分别定义了一个 0~16 之间的威胁因子 r_i ,并用

$$T = O_i \left\{ \left[\prod_j r_i k_j \right] / \prod_j k_j \right\}$$

$i \quad (1,2,3,4,5), \quad j \quad (t-t, t) \quad (7)$

计算时间 t 内系统所受威胁总和的平均值,从而对系统的威胁做出量化评估.其中: k_j 表示截止到当前时刻 t 之前的 t 时间段内的所有攻击.

3.2 自适应入侵检测策略

所谓自适应检测是指 IDS 根据环境变量自行确定入侵检测策略.本文将主机系统某个时间段内的安全风险评估结果作为条件,启停检测器作为自适应策略,先收集某个时间段内 IDS 的报警事件,再按照式(4)计算得到安全风险评估的结果.主机风险按 2 的倍数分为 4 个等级,则可制定出一种如表 1 所示的一种简单、实用的自适应入侵检测策略.

表 1 自适应入侵检测策略

主机 风险 等级	检测策略					
	网络连 接监控	键盘输 入监控	命令序 列监控	审计日 志监控	文件系 统校验	系统调 用监控
0~8	Y	N	N	Y	N	N
9~32	Y	N	N	Y	Y	N
33~128	Y	Y	Y	Y	Y	N
129~	Y	Y	Y	Y	Y	Y

注:Y 表示启动监测器,N 表示关闭监测器.

4 应用

目前,我们已经开发并实现了 IDS 的击键分析、命令序列分析、审计日志分析、文件系统校验以及系统调用序列分析 5 大功能模块.采用本文方法进行入侵检测,在所有检测器保持开启的状态下,每个检测模块将会实时上报系统行为的异常度,利用

信息融合技术最后可得到一个总的入侵判定.

系统在实际运行中的某个时段内,每个模块上报的异常度如表 2 所示,其中的入侵概率为 0.86,它可通过式(3)计算得到,这个判定结果表明入侵的可能性很大,如果该判定值大于报警阈值时,则系统进行报警处理.

采用深度防卫技术中的多个模块进行入侵检测,利用信息融合技术对检测结果进行融合,可以使入侵判定的推理更理性化,在一定程度上可减少误报.测试结果表明,该系统能够成功地检测出缓冲区溢出攻击和计算机系统的滥用或误用,但系统的总体性能会受到一定的影响.同时还发现,在安全环境比较好的情况下,仅使用一个入侵检测器就能满足入侵检测的要求,可以说在总体上提高了系统的性能.通过实验,我们确定了一种如表 1 所示的简单、实用的自适应入侵检测策略,该策略的可操作性很强,能使机器在性能与安全之间达成动态平衡,实用价值较高.

表 2 系统在某个时间段内的 5 个检测模块上报的异常度

入侵判定	异常度				
	键盘 输入	命令 序列	审计 日志	文件 系统	系统 调用
0.86	0.7	0.4	0.5	0.8	0.5

5 结论

深度防卫的自适应 IDS 按照黑客攻击的一般顺序,从数据流入手,对 6 种数据源进行入侵分析,最后利用信息融合方法得到一个总的入侵判定,以提高入侵检测的准确率.基于文中提出的安全风险评估计算方法制定的自适应入侵检测策略,能够动态调整系统性能与安全之间的平衡,用它可构造出一种实用、高效的深度防卫的自适应 IDS.今后,我们将开展如下 3 个方面的工作:考虑各个数据源之间的相关性与权重关系,利用相应的信息融合技术得到更合理的入侵判定;深入研究主机风险评估方法,提出更有效的自适应入侵检测策略;针对不同的数据源,重点研究更有效的入侵检测方法.

参考文献:

[1] Anderson D, Frivold T, Valdes S. Next-generation intrusion detection expert system: a summary [R]. Technical Report, SRI-CSL-95-07. Menlo Park, USA: Computer Science Laboratory, SRI International, 1995. 1-52.

(下转第 346 页)



图5 取值对聚类正确率的影响

较小时,聚类的正确率较低;随着 α 的增加,聚类的正确率会有所提高,但同时计算的复杂性也会增加.综合评价结果是,当 α 在 0.80 左右时,EIGRSCA 在保证聚类结果具有较高的正确率的同时,计算的复杂度也不是很高,因此 α 可以作为 EIGRSCA 聚类正确率折衷的控制参数.

另外,通过实验可以对聚类结果进行粒度分析,以考察聚类效果.从中我们发现:信息粒度在某种程度上取决于聚类参数 α ;颗粒基数越小,粒度值越大,则聚类数目越少,体积越大,聚类结果越粗糙;颗粒基数越大,粒度值越小,则聚类数目越多,体积越小,聚类结果越细致.

3 结束语

本文主要讨论了具有混合属性特征的数据聚类问题.在分析聚类和信息粒度原理的基础上,将熵、信息粒度与粗糙理论结合起来进行聚类,实验结果表明,该方法能够有效地发现数据中的聚类结构,并

在处理混合属性上有较好的聚类性能.

参考文献:

- [1] 史忠植. 知识发现 [M]. 北京:清华大学出版社, 2002.
- [2] Han Jianwei, Kamber M. Data mining: concepts and techniques [M]. San Francisco: Morgan Kaufmann Publisher, 2000.
- [3] Grabmeier J, Rudolph A. Techniques of cluster algorithm in data mining [J]. Data Mining and Knowledge Discovery, 2002, 6(4): 303-360.
- [4] 卜东波,白 硕,李国杰. 聚类/分类中的粒度原理 [J]. 计算机学报, 2002, 25(8): 810-850.
- [5] Pawalk Z. Rough sets [J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356.
- [6] Pawlak Z. Rough set: theoretical aspects of reasoning about data [M]. Norwell, Netherlands: Kluwer Academic Publisher, 1991.
- [7] Skowron A, Peters J. Rough sets: trends and challenges [A]. Proceedings of the 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing [C]. Berlin: Springer-Verlag, 2003. 25-34.
- [8] Hirano S, Tsumoto S, Okuzaki T, et al. A clustering method for nominal and numerical data based on rough set theory [J]. Bulletin of International Rough Set Society, 2001, 5(1-2): 211-216.

(编辑 苗 凌)

(上接第 342 页)

- [2] Song D, Venable P, Perrig A. User recognition by keystroke latency pattern analysis [EB/OL]. <http://citeseer.nj.nec.com/song97user.html>, 2003-10-05.
- [3] Tripwire Inc. Tripwire [EB/OL]. <http://www.tripwire.org>, 2003-08-11.
- [4] Forrest S, Hofmeyr S A, Somayaji A, et al. A sense of self for Unix processes [A]. Proceedings of the 1996 IEEE Symposium on Research in Security and Privacy [C]. Los Alamos, USA: IEEE Computer Society Press, 1996. 120-128.
- [5] 张文修,梁 怡. 不确定性推理原理 [M]. 西安:西安交通大学出版社, 1996.
- [6] 陈秀真,郑庆华,管晓宏,等. 网络化系统安全态势评估的研究 [J]. 西安交通大学学报, 2004, 38(4): 404-408.
- [7] Carnegie Mellon University. CERT/CC vulnerability note field descriptions [EB/OL]. <http://www.kb.cert.org/vuls/html/fieldhelp#metric>, 2004-02-04.

(编辑 苗 凌)