·计算机工程与应用·

一种基于多参数的IDS决策过程研究

杨 鹃，罗光春，卢显良

(电子科技大学信息中心 成都 610054)

【摘要】讨论了将数据融合技术运用到入侵检测系统中的方法，并提出了一个基于数据融合技术的入侵检测机制－DFIDM。在该机制中，有多个检测器搜集系统日志文件、网络流量信息、网络数据包等数据，这些数据在通过了本地决策、数据提取和对象提取阶段等预处理过程之后，传送到融合中心进行决策，重点研究了决策过程所涉及的多参数问题。为此，系统设计了检测器可靠性、时间因素、空间因素等五个主要因素参与融合与决策。最后通过实验证明，采用了该机制的入侵检测系统具有更好的准确性。

关 键 词 多参数; 入侵检测; 决策; 准确性
中图分类号 TP311.52 文献标识码 A

Research on Decision-Making of IDS with Multiple Parameters

YANG Kun，LUO Guang-chun，LU Xian-liang

(Information Centre, Univ. of Electron. Sci. & Tech.of China Chengdu 610054)

Abstract This paper introduces a method of intrude detection based on data fusion, and presents a new mechanism-DFIDM. In DFIDM, a few of sensors are configured to collect data, such as log file, information of network traffics and data package of network. After some pretreatments such as local decision-making, Data refinement, and object refinement, these data will be transferred to fusion center. In this paper, we mainly research the multiple parameters of final decision-making, such as reliability of sensors, the factor of time, the factor of space. As it showed by the research result this mechanism can improve the veracity of IDS.

Key words multiple parameters; intrusion detection; decision-making; veracity

1 数据融合技术简述

本文引入数据融合技术，提出了一种新型入侵检测机制DFIDM(Data Fusion Intrusion Detection Mechanism)，使用数据融合中心评估是否存在攻击。为提高检测性能，DFIDM设计了并行分布式检测决策融合系统模型，采用多检测器并行检测同一入侵行为的机制，并将各检测器的本地决策提交数据融合中心进行优化决策[1-2]。

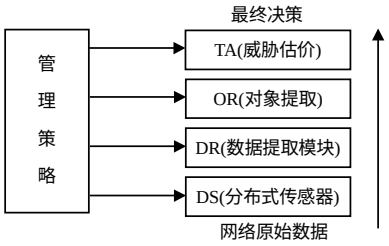

图1 DFIDM的体系结构

一个完整的基于多传感器数据融合的IDS应包括原始数据获取、数据提取、数据提取、对象提取、威胁估价和最终决策等部分，如图1所示。文献[3]详细介绍了该机制的体系结构和理论依据；文献[4]对数据提取和对象提取的工作机制，尤其是时空一致性问题进行了详细分析和设计。本文重点研究该机制最终决策过程中的各种参数和决策机制。

2 检测器的可靠性系数

DFIDM的一个主要设计目的是通过对各本地决策加以融合，降低单点检测误差，以及缺乏全网综合分析而引起的虚警和漏报，形成更为准确的决策。因此，有必要先讨论检测器的可靠性系数。

收稿日期：2005 - 01 - 04
基金项目：四川省科技厅应用基础研究资助项目(2006J13-070)
作者简介：杨 鹃(1981 -)，男，硕士生，主要从事网络安全方面的研究；罗光春(1974 -)，男，博士，副教授，主要从事计算机网络信息技术、信息系统安全方面的研究；卢显良(1943 -)，男，教授，博士生导师，主要从事计算机操作系统和计算机网络方面的研究.

(1) 可靠性系数的调整

由于需要检测多个入侵类型，DFIDM需要维护一个基于各检测器可靠性系数的二维矩阵 $\boldsymbol{\Omega}$。该二维矩阵(或数组)中，一维为本地检测器编号n=1,2,⋯,n；另一维为 m 种入侵行为编号m=1,2,⋯,m。一般地，Ω_{ij} 为第 i 个检测器对第 j 种入侵行为的可靠性系数，介于0到1之间，表示决策结果可靠性程度的数量变化。系统初始化时，所有 Ω_{ij} 统一赋值为1，随着系统的运行，将定期根据各检测器的虚警率和漏报率对可靠性系数进行重新评估和调整。系统考察每个检测器对每种入侵行为的检测结果，并预设一个可靠性系数 Ω_{ij} 的调整值 λ。初始化时的 $\boldsymbol{\Omega}$ 可表达为：

$$\boldsymbol{\Omega}=\begin{array}{c|cccc} & 1 & 2 & \cdots & n \\ 1 & 1 & 1 & \cdots & 1 \\ 2 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m & 1 & 1 & \cdots & 1 \end{array} \tag{1}$$

在一次检测中，如果第i个检测器对第j种入侵行为的检测效果良好，对应的 Ω_{ij} 加 λ，但相加后的结果应小于1，如超过则取值为1；否则，每检查出一次检测失效应将可靠性系数 Ω_{ij} 减λ。随着系统的运行，各检测器的可靠性将动态地反映在二维矩阵 $\boldsymbol{\Omega}$ 中，为最终融合的可靠性提供依据。

(2) 可靠性系数的计算

由于不同检测器i对入侵类型j的检测可靠性存在差异，Ω_{ij} 是不同的。随着系统运行过程中对检测器自身性能进行检测和调优，特定检测器对特定入侵类型的可靠性系数也会发生变化，计算过程在每次融合时都要进行，计算公式为 $\Omega_j = g(\Omega_{1_j},\Omega_{2_j},\cdots,\Omega_{n_j})$，即系统将 n 个检测器在每次融合时，对入侵类型 j 的可靠性系数通过函数 g 运算来得出 Ω_j。从系统角度来看，考虑到各检测器的可靠性本身具有相同的可信度，为简化起见，FDIDM将函数 g 设定为算术平均和，即 $\Omega_j = \sum_{i=1}^{n}\Omega_{ij}/n$。

必要时，还可以根据实际需要对 g 进行调整。

3 参与决策的其他主要因素

除了检测器本身的可靠性，DFIDM在融合决策时还引入了空间因素[5]、时序因素、历史记录、人工加权等4个主要因素。限于篇幅，本文仅着重讨论空间因素。

空间因素是指目标系统各节点上，检测器对同一种入侵行为的联合预警，体现在当真实入侵发生后，由OR提交到TA的对象集合中的元素应该为多个而非单个。这些处于同一集合中的多个元素正是不同检测器同时检测到同种入侵，并提交给融合中心的。在DFIDM融合算法中，对于各检测器是否同时检测到同种入侵行为，定义了一个空间可靠性系数数组 Λ。该数组为一维，共有 m 个元素，分别对应于 m 种入侵类型。对于任一攻击类型 j，有 $0 \le \Lambda_j \le 1$，表达了入侵类型 j 被多个检测器共同检测到的量化程度，即 Λ_j 的值与检测到第 j 类攻击的检测器个数成正比。本文采用较为简单的方法计算 Λ_j，即选取h作为检测器个数的空间最小完全置信门限，当 h 个检测器共同检测到同一入侵类型 j 时，该类型入侵的可能性可以完全确认，即空间可靠性系数 $\Lambda_j = 1$；若检测到该类型入侵的检测器个数少于 h，将入侵类型 j 的存在表达为一个概率；如果检测到入侵类型的检测器个数多于 h，Λ_j 值大于1时则自动调整为1，其计算公式为：

$$\Lambda_j = k/h \quad 0 \le \Lambda_j \le 1 (k < h) \quad \text{or} \quad \Lambda_j = 1 \; k \ge h \tag{2}$$

4 融合与决策

以上介绍了最终融合所需要考虑的几个主要因素，即检测器可靠性系数矩阵 $\boldsymbol{\Omega}$、空间可靠性系数 Λ、时序可靠性系数 Ψ、历史记录影响系数 φ 和人为判断系数 Δ。TA在接收到来自OR的对象集合后，按照缓存的方式，结合以上各系数即可完成最终融合和决策。DFIDM的最终决策结果，表示为一个定性决策结果和起辅助作用的定量计算公式。

从前面的讨论中可知，融合和决策过程涉及诸多因素，将这些因素完全量化后，通过计算得出以数值

形式表示的决策结果不一定是最准确的办法。其根本原因在于，如果将最终决策结果Z表达为各影响因素的函数，即 $Z=f(\boldsymbol{\Omega},\Lambda,\Psi,\varphi,\Delta)$。那么无论怎样对函数 f 进行设计， $Z=f(\boldsymbol{\Omega},\Lambda,\Psi,\varphi,\Delta)$ 这5个因素应该对决策结果施加的影响大小，实际上都难以被准确地反映为具体数值。也就是说，很可能由于其中一个因素的系数值选择失当，就会导致其他因素对决策的影响受到影响，甚至完全丧失。实际上，有很多种可能性存在，例如，各因素数值变化刻度选择不当导致一个因素的作用完全“掩盖”(Cover Up)其他因素；各因素在定量公式中的数学关系不能准确吻合它们在实际情况中的关系；真实入侵发生时，各因素针对不同情况起的作用也会发生一定变化，导致定量公式不一定能及时调整、吻合，等等。

正是基于以上考虑，DFIDM的最终决策结果以定性分析为主，定量计算为辅，其定量计算结果仅提供给系统管理者作为参考，或者说通过函数 $Z=f(\boldsymbol{\Omega},\Lambda,\Psi,\varphi,\Delta)$ 所计算出的Z值，只可看作对系统安全状况的宏观描述。

(1) 定性结果的描述。系统可规范地对融合过程中所涉及的因素进行描述，通过提交的报告，系统管理者能详细地了解目标系统中可能发生的各类入侵，其知识涉及融合的各个因素，并进行相应处理。

(2) 定量公式计算。设系统最终决策为对目标系统整体安全状况的描述(Target System Whole Security Situation Description)为Z，可用一维数组表达，其分量 Z_j 表示对第 j 种入侵行为的决策值。如果将所有因素等同考虑，则可得 $Z_j=(\Omega_j+\Lambda_j+\Psi_j+\varphi_j+\Delta_j)/q$，其中 q 为实际存在的融合因素的个数，例如当历史记录不存在时， $q=4$。各因素的值可根据前面的方法得到，其值介于0到1之间；而关于入侵类型 j 的最终决策为各因素的算术平均值。如果 $0\le Z_j\le 1$，最终决策Z是元素值介于0到1之间的一维数组，分别对应各类入侵在某个时刻 t_0 经过DFIDM融合后的发生概率。由于各个影响因素在入侵实际发生时，可能会对融合决策起到不同作用，故可以加入调整系数序列b_i，令i=1～5，对于不同因素进行调整，扩展的计算公式为：

$$Z_j=(b_1\Omega_j+b_2\Lambda_j+b_3\Psi_j+b_4\varphi_j+b_5\Delta_j)/q \tag{3}$$

式中 q介于0～5之间，为实际参与融合的因素个数。

实际应用中，对计算公式中系数序列b的各个分量的选取应根据实际情况进一步讨论得出，可考虑的思路有，根据经验选取系数序列，或根据影响估计大小选取系数序列，等等。

所以，可以通过DFIDM得到目标系统在任一时刻 t 对所存在的入侵行为的精确描述，进而通过连续的数据获取、本地决策、融合和最终决策，得出在任一时间段对入侵行为的精确描述；另外，还能得出一个表征各类入侵行为威胁程度高低的参考数组Z。凡此种种，系统就能准确地做出相关响应了。

5　实验结果与结论

为验证检测过程中的时空一致性，本文对DFIDM进行了一系列实验，实验环境为1个融合中心FC(Intel服务器1G)和5个节点(PC赛扬666)，全部置于同一局域网中，通过集线器连接，保证所有数据均同样流入5个节点；攻击数据由1台PC赛扬666提供；根据DFIDM的需要，设计并实现了DR、OR和TA所对应的数据库和功能模块；开发环境为Red Hat linux8.0，用ANSI C编写代码；数据库为My SQL 4.0；入侵类型选择了TCP Flood、UDP Flood、ICMP Flood、后门攻击、缓冲区溢出攻击等5种，通过下载相关攻击工具TFNZK，Netcat和WEBDAV实现。分别对同时攻击类型为1、3、5的情况进行3组实验，每组实验分别进行了1 000次。限于篇幅，这里仅给出同时攻击种类为5的漏报率和虚警率的比较结果，见表1。

实验结果表明，DFIDM能有效降低检测的虚警率和漏报率，较大地提高了检测的准确性，见表2。

表1　攻击种类为5时的漏报率结果比较

攻击类型	攻击实现工具	漏报率/(%)		
		Snort	Bro	DFIDM
TCP Flood	TFN2K、Netcat WEBDAV	5.1	4.9	2.1
UDP Flood		4.8	4.1	1.6
ICMP Flood		4.7	4.7	1.9
后门攻击		3.9	4.5	1.8
缓冲区溢出攻击		4.3	5.0	2.0

表2　虚警率的比较/(%)

Snort	Bro	DFIDM
2.2	3.1	0.6

(下转第810页)

；BEGIN：
WDTCR:(插入实现定时功能的代码构件)
；END;
；BEGIN：
LTEST：(插入实现缩时功能的代码构件)
；END;
；BEGIN：
ERROR: (插入实现出错处理的代码构件)
；END;
…
；BEGIN：
；END;

(2) 框架构件描述块信息填写为：

框架构件描述块 = {(XXXXXX、东芝__空调__框架构件、1)，
(空调、东芝TMP87C846、东芝汇编语言、东芝编译器)，
(该构件版本号为X，构件作者为XXXX，入库时间为XXXX，修改情况)
}

通过以上步骤，就可在框架规范指导下完成框架构件的生成。

4 结 束 语

与传统的构件规范相比，本文提出的框架构件规范具有以下特点：(1) 该规范基于家电控制器常用MCU体系结构、程序设计语言、常用家电功能及外设驱动源码研究，设计面向智能家电嵌入式软件构件化，针对性强；(2) 该规范已经应用于某研究开发课题之中，具有可实现性；(3) 该规范抽象于具体应用领域和程序设计语言，适用于智能家电嵌入式软件构件开发，具有很强的适用性和可扩展性。

参 考 文 献

[1] 费玉奎, 王志坚. 构件技术发展综述[J] . 河海大学学报(自然科学版), 2004, 32(6): 696-699.
[2] 江 峰, 陈文智, 吴朝晖. Liquid-构件化嵌入式操作系统[J]. 计算机工程, 2005, 31(4): 77-87.
[3] 古幼鹏, 熊光泽, 桑 楠. 基于构件的嵌入式软件仿真开发环境模型研究[J]. 系统工程与电子技术, 2004, 26(10): 1495-1499.
[4] 胡文蕙, 赵 文, 张世琨, 等. 基于构件技术的应用框架元模型的研究[J]. 软件学报, 2004, 15(1):1-8.
[5] 马 亮, 孙艳春. 软件构件概念的变迁[J]. 计算机科学, 2002, 29(4): 28-30.
[6] 肖 忠. 构件软件工程研究[D]. 成都: 四川大学, 2005.
[7] 徐拥军. 基于构件的软件开发方法及其支撑平台[J]. 软件工程与标准化, 2005, 3: 37-42.
[8] 谷今杰, 莫继红. 基于构建的软件复用技术研究[J]. 科学技术与工程, 2005, 5(12): 824-827.

编 辑 熊思亮

(上接第803页)

参 考 文 献

[1] Kam M, Zhu Q, Gray W W. Optimal data fusion of correlated local decisions in multiple sensor detection systems[J]. IEEE Trans. AES., 1988, 18(5): 916-920.
[2] Nilsson N J. Artificial intelligence: A new synthesis[M]. 北京: 机械工业出版社, 1999.
[3] 罗光春, 卢显良, 张 骏, 等. 一种基于多传感器数据融合的入侵检测机制[J]. 电子科技大学学报(自然版), 2004, 33(4): 71-74.
[4] 罗光春, 卢显良. IDS决策过程中的时空一致性研究[J]. 计算机科学, 2005, 32(6): 121-123,139.
[5] Bass T. Intrusion detection systems and multisensor data fusion: creating cyberspace situational awareness[J]. Communications of the ACM, 2000, 43(4): 99-105.
[6] Baek W, Bommareddy S. Optimal m-ary data fusion with distributed sensors[J]. IEEE Trans. AES. 1995, 31(3): 1150-1152.
[7] 郑 辉. Internet蠕虫研究[D]. 天津: 南开大学信息技术科学学院, 2003.

编 辑 熊思亮