基于自由代理的入侵检测系统研究*)

张送保　张维明　肖卫东　蔡建国
(国防科技大学管理科学与工程系　长沙410073)

摘　要　本文针对多代理分布式入侵检测系统的单点失效和瓶颈等问题，给出了一种基于自由代理的入侵检测系统模型，该模型充分使用知识库中的知识对系统消息进行匹配，并遵循一定的移动策略，使得自由代理能够在系统网络中自由的移动，以便对某些繁忙代理或者失效代理及时提供功能上的替代和补充，从而提高整个入侵检测系统的性能。

关键词　自由代理，多代理，移动，知识库，入侵检测系统

A Research for Intrusion Detection System Based on Liberal Agent

ZHANG Song-Bao　ZHANG Wei-Ming　XIAO Wei-Dong　CAI Jian-Guo
(Dept. of Management Science and Engineering, National Univ. of Defence Technology, Changsha 410073)

Abstract　This paper proposes a type of system model about intrusion detection based on liberal agent, which aims at some problems of multi-agent distributed intrusion detection system such as single-point invalidation, bottlenecks, and so on. By richly using the knowledge in knowledge base to match the messages, and flexibly moving the liberal agents to the site where it is needed according to some measures, the system can replace the invalidated agents or support the busy agents to finish their tasks timely. Thus, the system will greatly improve its intrusion detection performance.

Keywords　Liberal agent, Multi-agent, Mobile, Knowledge base, Intrusion detection system

1　入侵检测系统介绍

网络安全在Internet中的重要性越来越明显，因此，作为网络安全措施一个环节的入侵检测系统(IDS)也越来越受到关注。入侵检测的含义为"检查并确认那些未经授权而使用计算机系统，以及可以合法访问系统但是滥用了权利的用户"。而入侵检测系统就是用来执行检测入侵任务的计算机程序。

原先，入侵检测系统(IDS)多针对单一对象，采用单一的检测技术，具有较大的局限性。而通过网络将主机的信息传送到中央分析单元的分布式IDS，其中央分析单元负载太重，难以避免单点故障，而且无法进行实时检测。

目前，智能化[1]、分布式和监视大规模网络是当前国际上许多入侵检测系统所强调的，例如Purdue大学所设计的自治代理入侵检测系统AAFID[5]和California大学所设计的基于图的入侵检测系统GrIDS，等等。

许多现行的基于多代理的分布式入侵检测系统(MADIDS)[6,8]模型，虽然解决了早期的检测系统将检测任务都集中在中心主机上所带来的系列问题(如中心主机的单点失效问题以及系统繁忙时中心主机的瓶颈问题等)，却也同时将这些问题从一定程度上转嫁到了检测系统的各个单机上。另外，采用新型的移动技术而设计的基于移动代理的入侵检测系统，由于系统中所包含的移动代理过多，因而无论是对系统的实时调控还是对网络带宽的要求都会大大增强，显然也不是非常理想的选择。

本文所提出的基于自由代理的入侵检测系统，充分考虑了以上的问题，在现有的MADIDS基础上，再为系统合适地设置一些自由代理(Liberal agent)[2,7]，从而在系统中某些代理繁忙或失效时，自由代理能够及时地给予补充和替代，以保障系统的正常运行。

2　基于自由代理的入侵检测系统体系结构

基于自由代理的入侵检测系统体系结构如图1所示，它是在基于多代理的分布式入侵检测系统的基础上发展而来的。

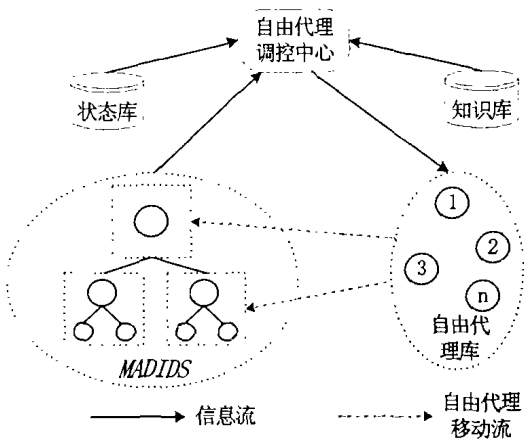

图1　基于自由代理的入侵检测系统结构

其中，MADIDS表示按层次组织的分布式多代理检测系统，主要由众多功能各不相同的代理按照一定的组织和层次组成一个有机的代理联合体，如：功能代理、主机检测代理、域检测代理和通信代理等，它主要负责系统正常运转时期主机和网络的实时分布式入侵检测。图中的自由代理库表示系统初始时期所设置的一系列自由代理，它主要在MADIDS系统部分代理繁忙或失效时及时的移动到该处，从而维持系统检测的持续稳定。而自由代理调控中心则相当于自由代理的守

*)基金项目：外协项目——入侵检测数据分析系统。**张送保**　博士研究生，主要研究方向是信息管理与智能决策；**张维明**　教授，博士生导师，主要研究方向是信息管理与智能决策；**肖卫东**　副教授；**蔡建国**　讲师。

护程序(daemon),它在 MADIDS 请求提供协助时,从其知识库[3]中搜索满足需求的系列代理,然后根据其状态库中所记录的自由代理实时状态选择最优的自由代理进行激活。

MADIDS 的具体结构及其工作机理见文[6],本文重点所论述的就是该体系结构中自由代理。

3 自由代理

3.1 自由代理调控中心子系统

自由代理调控中心在基于自由代理的入侵检测系统中处于核心的地位,其原理如图2所示。

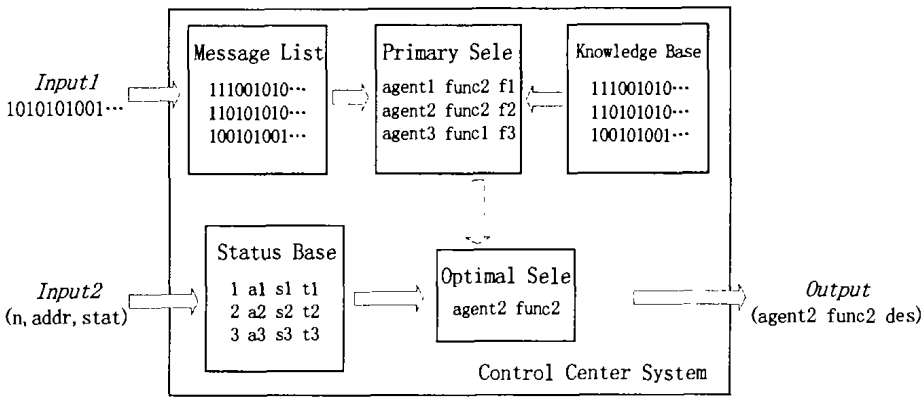

图2 自由代理调控中心子系统

调控中心从外部环境接收两种信息流:来自 MADIDS 的协助检测请求流和来自自由代理状态报告流。前者采用消息的形式,每条消息代表一个协助请求,其中包括所请求问题的简要描述以及所产生问题的位置,它在调控中心里和其它的请求消息一起以消息列表的形式组织;后者包括自由代理的编号、自由代理的位置及其当前状态,它进入系统之后存储在状态库中。状态库的结构如下:

Statebase:(*number*,*address*,*state*,*duringtime*)

其中 *state* 变量取四种值:1、2、3和0,分别表示补充、替代、移动和空闲;而 *duringtime* 则表示该自由代理(number)在该位置(address)所维持该状态(state)的时长,用它不仅可以表示该代理忙闲情况,而且也可表示系统某点的稳定状况和问题特征。

调控中心的知识库所存储的是系统中所有自由代理能够执行功能的知识描述,其中每条知识分为两部分:〈知识〉:〈标识〉。知识部分由一个5元组(k_1,k_2,k_3,k_4,a)构成,其中,k_1~k_4 表示所要监控的 user、system、process 和 packet 四个层次,且 $k_i \in \{00,01,10,11\}$,其取值分别代表告警层次 normal、minimal、significant 和 dangerous,考虑到知识的通配特性,我们还在知识中采用了通配字符'#',表示此处或者匹配'0',或者匹配'1';另外,$a \in \{0,\cdots,7\}$,表示系统根据该条知识所要采取的不同行为。假设调控中心的输入模板消息队列中包含消息(m_1,m_2,m_3,m_4,*address*),那么该消息和知识(k_1,k_2,k_3,k_4,a)相匹配当且仅当:

$k_1 <= m_1$ 且 $k_2 <= m_2$ 且 $k_3 <= m_3$ 且 $k_4 <= m_4$,

也就是说,该知识的 k_1~k_4 部分代表不同层次的最小偏离值。

因此,如果我们有两条知识:$K=(k_1,k_2,k_3,k_4,a_k)$ 和 $L=(l_1,l_2,l_3,l_4,a_l)$,其中 $i=1,2,3,4, k_i <= l_i$,这就意味着 L 知识比 K 知识更为精确。所以此时任何和 L 匹配的消息肯定也将和 K 匹配。为了维持知识库的一致性,L 的行为(a_l)必须比 K 的行为(a_k)要更为强烈:$a_l >= a_k$。

而标识部分则表示包含该条知识的相应自由代理的对应功能编号。考虑到系统的健壮性,自由代理的设置采用了一定的冗余思想,因此,一条知识可能为多条自由代理所拥有,但它只属于该代理的某一个具体功能。

如此,通过实时消息在知识库中的匹配,就得到了满足条件的一系列初始选择(Primary Sele),并且系统在进行知识选取时,对所选择的知识还进行了对实时消息的适应度计算,Primary Sele 中的 $f1$、$f2$等就表示相应知识对实时消息的适应度(Fitness)。知识 K 对实时消息的适应度值简单计算如下:

$$Fintess(K)=1-\frac{1}{12}\sum_{i=1}^{4}|m_i-MinValue(k_i)| \quad (1)$$

其中,$MinValue(k_i)$按以下对应规则取值:

(00→0,01→1,10→2,11→3,0#→0,#0→0,1#→2,#1→1,##→0)

显然,$Fitness(K) \in [0,1]$,当消息和知识完全匹配时其值取1。如果只考虑知识匹配的情况,当然是适应度值 Fitness 越高,就越能更好地解决消息所表示的问题,但在实际应用中,由于此时所选取的自由代理也可能正处于检测即忙状态,或者说即使它处于空闲状态,但是离目的位置还有很远的一段距离,因此调控中心还需要进一步的选优处理,这时在 Primary Sele 表的基础上,结合自由代理状态表,得出最终的结果,显然,此过程可用以下函数表示:

$$StateWeigh=(3-state)/3 \quad (2)$$

$$Fitness'(K)=StateWeigh * Fitness(K)/(1+Distance) \quad (3)$$

$$OptiSele(m)=Max(Fitness'(K_i)) \quad (4)$$

其中,*StateWeigh* 的意义表明自由代理空闲时,其权取值为1,表示能够提供完善的服务;自由代理正在执行补充任务时,其权取值2/3,表示能够提供服务,但不是很完善;自由代理正在执行替代任务时,其权取值为1/3,表示该代理此时仍能提供很不完善的服务;而当自由代理处于移动状态时,其权取值0,表示该代理不能再提供额外的服务。而 *Distance* 值则为消息所反应的问题位置和自由代理当前位置的函数,它由自由代理的移动策略所决定。

经过以上的处理,调控中心选择出一个能够最好的解决消息所反应问题的自由代理,然后将其功能编号以及问题的目的位置以输出的形式发送给它,以便激活该代理,从而及时

地完善系统的检测功能。

3.2 自由代理的工作原理

自由代理设计的出发点就是解决系统的单点失效和瓶颈问题，其工作原理就如同排球场上的自由球员。自由球员既要具有场上其他各个角色的基本能力，也要能够在比赛的时候灵活机动地移动位置，担当起其它角色的任务。同样地，一个自由代理也被设计成具有系统其它几种关键代理的功能，当然，与其具体的功能代理相比，这儿的代理功能是粗糙的、非细腻的。其具体结构[4]如图3所示。

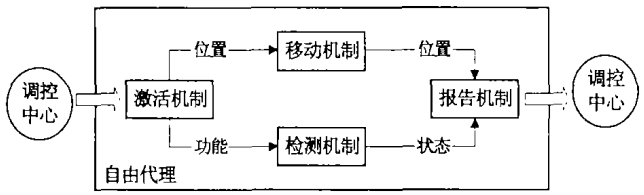

图3 自由代理工作原理

其中，激活机制主要从调控中心接收目的位置参数和功能选项参数，然后将位置参数传递给移动机制单元，将功能选项参数传递给检测机制单元，从而进行该代理的激活处理。一个自由代理也只有同时接收到这两个参数，才可能会被激活，也就是说此两参数为自由代理的点火条件。

自由代理的移动机制为自由代理在系统中的自由移动提供策略支持，而此移动策略也就是计算自由代理当前位置和目的位置距离的主要依据。考虑到 MADIDS 系统层次组织结构的就近移动迅速和功能相似特性，我们采用树结构（如图4所示）来实现移动策略，从而对系统中自由代理的移动自由度进行有效的控制与管理。按照移动策略树思想，自由代理的调用移动原则如下：

（1）各单机优先调用位于本机中的自由代理，此时其距离值 *Distance* 为0；如果本地调用失效或本地调用仍然解决不了问题时，才进行异地调用。

（2）异地调用时，优先调用兄弟节点间的自由代理，此时其距离值 *Distance* 为0.5；其次再考虑父子节点间的自由代理调用，此时其距离值 *Distance* 为1；再其次考虑祖孙节点间或者是叔侄节点间的自由代理调用，此时其距离值 *Distance* 为1.5；除此以外的自由代理调用（不包括原则3）时，其距离值 *Distance* 为3。

（3）从自由代理库调用未用的自由代理时，其距离值 *Distance* 为2。

（4）自由代理移动到系统的某点完成检测任务后，就在该处进行休眠。

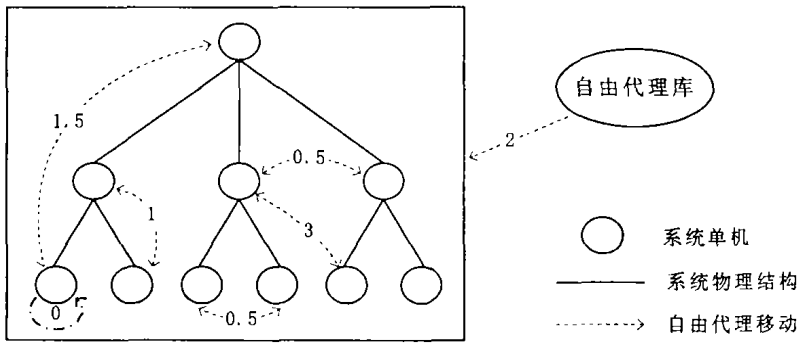

图4 自由代理称动策略树

考虑到系统失效位置一定程度的随机性，因此自由代理在完成相应的检测任务后仍然在该位置休眠，从而从整个系统来看，这些自由代理便随机性地遍布在系统的各处。另外，自由代理的就地休眠也符合“发生问题的位置在下一个时间段仍然具有最大可能的问题发生”特性，从而使系统更为有效地维持其正常运转。

自由代理的检测机制封装了该代理的一组功能函数，每个功能函数对应一个具体的检测功能，其所依据的检测原理（包括 user、system、process 和 packet 四个层次）由知识库中相对应的知识所描述。接收到激活机制所传来的功能选项参数并移动到目的位置之后，该检测机制就会依据功能选项参数将相应的功能函数设置为活态，并进行相应的检测分析。

自由代理的状态报告机制采用系统同步时钟进行，每隔一个预设的时间段，它便会自动地向自由代理调控中心报告自身的状态信息，然后调控中心便会动态刷新自由代理的状态库。

3.3 自由代理系统的安全及其有效性

系统引入自由代理，目的就是为了解决入侵检测系统中繁忙或失效问题，尽管自由代理系统重新引入了一个新的中心主机（自由代理调控中心），但是它位于入侵检测系统 MADIDS 结构之外，因而遭受外部攻击（单点失效）的可能性非常小，而它所调控的自由代理数量也是相当有限的，因而也不会增添新的瓶颈问题；单个的自由代理由于是随机性的分布在检测系统的各处（移动机制中已论述），从而遭受攻击的可能性就更小了。总之，自由代理系统的引入对原入侵检测系统不会带来新的安全性问题。

自由代理主要是为了短期内补充和替代某些关键代理的功能，其功能只能是粗糙的，也只能包括几种关键代理的功能，否则，如果其功能过于庞大全面，那么它就会因为本身的数据量太多而不适于在系统中灵活的移动。

另外，系统初始阶段所设置的自由代理数目也必须适中。自由代理只是在非常时刻作为其它代理的辅助而工作的，其它代理仍然是检测系统的主体，如果在系统中设计的自由代理数目过多，那么大量的自由代理在系统网络中频繁地来回移动，无疑又另外大大增加了系统的数据流，从而引起新的瓶颈。

4 系统实现与分析

本系统的实现是在多代理分布式入侵检测系统基础上再集成自由代理系统来完成的。目前，代理的移动技术大都采用与平台无关的语言，如 Java、Tcl/Tk 语言等，由于 Java 语言

（下转第60页）

确定，也可以设计算法，从大量训练数据中得到。

7）根据关联结果，管理员将可以容易地看出，来自192.168.1.240的黑客，首先登录到218的主机，采用NMAP工具对网络进行扫描，当他发现239主机存在Wu-ftp漏洞后，使用权限提升工具Exploit_Wuftpd对239主机发起攻击，获得root权限后黑客从239主机下载Passwd文件，破译分析正常用户的口令。最后该黑客telnet到239主机，使用SU命令，根据破译后的root口令将自己变为超级用户身份。

结束语 本文提出了一种面向分布式多源传感器实时事件数据采集及关联分析的方法，运用Bayesian算法网络进行事件分类处理，并对事件进行过滤、归约、格式转换，消除冗余事件，形成基于IDMEF的统一格式，为上层的入侵判定提供证据。本文所提的模型已经在我们承担的教育部重点科技课题课题中得到实际应用，这种方法在复杂网络环境下能有效地分析攻击报警事件，帮助管理员从大量数据中找到有效的信息。当然，本文所提的方法也存在不足和改进之处，表现为：1）这种方法需要预先提取黑客的攻击特征模式，具有足够的知识库和特征相似度矩阵；2）具有自适应的相似度计算能力，以适应复杂的分布式协同攻击判定；3）采用朴素贝叶斯算法可以定量地进行入侵事件的关联，初步实现了入侵事件的多角度分析，但其相对来说比较简单，对于复杂条件下的事件关联准确度不够，同时缺乏自学习的功能。因此需要找到更好的关联算法，同时要收集足够的训练数据，提高关联的准确度。

参考文献

1 McHugh J, Christie A, Allen J. Defending Yourself: The Role of Intrusion Detection Systems. IEEE SOFTWARE Sep./Oct. 2000

2 http://www.sdl.sri.com/projects/emerald/project.html

3 http://www.ietf.org/internet-drafts/draft-ietf-idwg-idmef-xml-06.txt

4 http://www.ietf.org/html.charters/idwg-charter.html

5 Han Jiawei, Kambr M. Data Mining: Concepts and Techniques. Copyright 2001by Morgan Kaufmann Publishers, Inc.

6 Valdes A. Skinner K. Probabilistic Alert Correlation. RAID 2001, LNCS 2212, 2001. 54～68

7 Lindqvist U, Porras P A. Detecting Computer and Network Misuse Through the Production-based Expert System Toolset (P-BEST). In: Proc. of the 1999 IEEE Symposium on Security and Privacy, Oakland, California, 1999. 9～12

8 Debar H, Wespi A. Aggregation and Correlation of Intrusion-Detection Alerts. RAID 2001, LNCS 2212, 2001. 85～103

（上接第56页）

能提供其程序运行的平台无关性和良好的安全机制，因此本系统采用Java语言来实现自由代理的移动技术。图5所示为该系统实现时某个时刻的总控界面，它向用户实时的显现出了系统总的入侵检测情况统计以及各个入侵事件的细节。

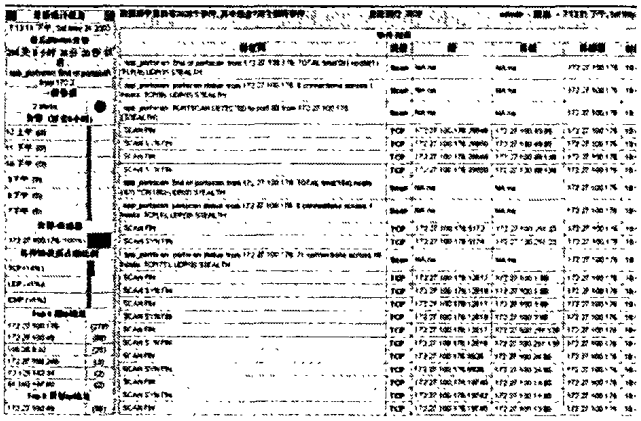

图5 基于自由代理的入侵检测系统实现

对系统结构及其构建分析，我们将自由代理应用到网络入侵检测系统中，可具有这样一些优点：能实现全局范围内的入侵检测功能，具有清晰的系统结构和良好的可扩展性；具有优良的可移植性能；对网络系统和主机资源的占用较低，能够优化网络通信和计算资源，实现系统的负载平衡，从而减少了出现瓶颈的可能。

对系统的运行性能及其结果分析，通过对检测系统模拟攻击的实验，我们发现单纯的MADIDS系统通常会出现因批量或是大规模攻击而使得系统告警发生明显的时延甚至是漏报或错报，其中也出现过由于某点失效而导致整个系统紊乱甚至崩溃的情况，但是在相同的实验环境下，本系统的检测性能则得到了大大的增强，在等同的规模攻击面前，其告警时延已不再那么明显，而且至今还未出现过由于等同攻击而发生崩溃的情形。

结束语 本系统在充分利用多代理分布式入侵检测系统的优点的基础上，利用自由代理的技术，解决了目前多代理分布式以及其它入侵检测系统中单点失效和处理瓶颈等方面的问题，使二者很好地结合在一起，从而提高入侵检测系统的整体性能。当然，如果能将智能性集成到自由代理中，使其具有一定的自学习能力，那么对于其知识库的动态更新以及系统检测性能的进一步提高都将是一个很大的改进。

参考文献

1 Pagurek B, White T, Bieszezad A. A Network modeling for management applications using intelligent mobile agents. Journal of Network and Systems Management, Sep. 1999

2 Selliah S. Mobile Agent Based Attack Resistant Architecture for Distributed Intrusion Detection System. Master Thesis to College of Engineering and Mineral Resources at West Virginia University, 2001

3 Dasgupta D. Gonzalez F A. An Intelligent Decision Support System for Intrusion Detection and Response. In: the Proc. of Intl. Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS), May 2001

4 Padgham L, Winikoff M. Prometheus: A Methodology for Developing Intelligent Agents. RMIT University, GPO Box 2476V, Melbourne, AUSTRALIA, 2001

5 韩东海，王超，李群. 入侵检测系统实例剖析. 清华大学出版社，2002

6 王锋波，曾昭苏. 一种基于多代理技术的分布式入侵检测系统. 计算机工程与科学，2000

7 黄雷，王闽. 一种基于移动代理的入侵检测系统. 福州大学学报，2001(8)

8 段海新，吴建平. 一种分布式协同入侵检测系统的设计与实现. 软件学报，2001(12)