ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.11, November 2006, pp.2384−2394
DOI: 10.1360/jos172384

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

智能网格入侵检测系统*

魏宇欣+, 武穆清

(北京邮电大学 通信网络综合技术研究所,北京 100876)

An Intelligent Grid Intrusion Detection System

WEI Yu-Xin+, WU Mu-Qing

(Institute of Communication Networks Integrated Technique, Beijing University of Posts and Telecommunications, Beijing 100876, China)

+ Corresponding author: Phn: +86-10-62283128, Fax: +86-10-62282285, E-mail: weiyuxin@gmail.com, http://www.bupt.edu.cn

Wei YX, Wu MQ. An intelligent grid intrusion detection system. *Journal of Software*, 2006,17(11):2384−2394.
http://www.jos.org.cn/1000-9825/17/2384.htm

Abstract: Adopt the advanced distributed system combined with intelligent intrusion detection technology is one of the key technologies to be applied in the intrusion detection system. Through the study on grid and intelligent intrusion detection technology, an intelligent Grid Intrusion Detection System (GIDS) is proposed which deploys in the grid environment and uses neural network detection technology. In order to realize the load balance between the detection engines, a scheduling strategy is used based on the resource performance value. In addition, Multiplicative Increase Linear Decrease (MILD) is applied to alerts aggregation. The GIDS, which fully exploits resources in the grid and realizes load balance, has high effectiveness in detecting the malicious attacks under heavy network traffic environment. Experimental results show that the performance of GIDS is good.
Key words: intrusion detection system; grid; neural network; task scheduling; alerts aggregation

摘 要: 结合智能检测技术,并采用先进的分布式体系结构是当前入侵检测研究的一个主要方向.通过对网格与智能检测技术的深入研究,提出了一种智能网格入侵检测系统(intelligent grid intrusion detection system,简称 GIDS).该系统部署于网格环境并采用基于神经网络的检测技术;为了实现各数据分析引擎的负载平衡,采用基于资源可用度的调度算法决定任务的分配;为了减少告警数量,采用基于乘性递增线性递减(multiplicative increase linear decrease,简称 MILD)的动态窗口调整算法进行警报合成.该入侵检测系统不仅能够充分利用网格上的资源进行入侵行为的发现,而且实现了资源使用的负载均衡,在网络流量大的情况下能够获得较高的检测效率.最后介绍了相应的实验结果分析,表明了该系统的优越性.
关键词: 入侵检测;网格;神经网络;任务调度;告警合成
中图法分类号: TP393 文献标识码: A

计算机和网络技术的普及在给人们的生活带来极大便利的同时,也将安全隐患传播到整个网络中.作为保护网络安全重要手段之一的入侵检测系统(intrusion detection system,简称 IDS),一直被广大国内外学者所关注.

* Received 2006-05-14; Accepted 2006-08-07

随着网络规模的日益扩大,原来的单机入侵检测系统已经不能满足实际应用的需要,于是提出了分布式入侵检测系统的概念,相关的分布式入侵检测系统的原型也被开发出来.但是,这些分布式系统都存在处理数据量集中,速度较慢,不能满足实时处理要求的问题.

网格计算[1]是伴随着互联网技术快速发展起来的,目的是为了解决大规模的复杂计算问题.这种计算模式利用互联网把分散在不同地理位置的计算机组织成一个虚拟的超级计算机,其最大的特点就是数据处理能力强,并能充分利用网上的闲置处理能力.这种处理能力是传统的分布式系统所不能比及的.

将入侵检测系统部署到网格环境中,使得该入侵检测系统中各个检测节点的资源能够被充分利用,在保证较好的正确检测率的同时降低处理时延,达到整个系统的负载平衡.由此,我们提出了一种网格入侵检测系统.

本文首先介绍了入侵检测的研究现状和网格开发工具 Globus;然后设计了网格入侵检测系统的体系结构和相关算法,给出了相应的实验结果;文章最后对全文总结并讨论了未来的工作方向.

1 研究现状

对于入侵检测的研究可以追溯到 20 世纪 80 年代[2].入侵检测是指通过从计算机系统或网络中的若干关键点收集并分析信息,从中发现系统或网络中是否有遭到攻击的迹象并做出响应.根据采用检测技术的不同可以分为基于误用的入侵检测和基于异常的入侵检测.早期的开发主要集中在对单机入侵检测系统检测技术的研究方面.网络技术的进步为黑客技术的发展提供了条件,出现了分布式攻击,典型的拒绝服务攻击(denial-of-service,简称 DoS)和在其上演变而成的分布式拒绝服务攻击(distributed denial-of-service,简称 DDoS)都是通过向网络发送海量数据包,消耗系统资源,导致停止对合法用户提供正常的服务,进而使整个网络瘫痪.单机入侵检测系统已经不能有效防范这种攻击.于是,研究人员开始对分布式入侵检测系统进行研究和开发工作.分布式入侵检测系统通过分布采集、协同工作的方式,彼此交互网络信息进行关联分析,从而达到检测分布式攻击的目的.目前已经开发出来的典型分布式入侵检测系统有:(1) 美国加州大学 Davis 分校于 20 世纪 90 年代提出来的 DIDS(distributed intrusion detection system)[3],该系统采用分布采集、集中处理的方式,所有采集到的网络或主机数据将被传送到中心节点集中处理,判断是否存在攻击行为,这样,中心处理节点可能会成为系统瓶颈,在出现大量攻击时存在失效的威胁.(2) 为了克服 DIDS 集中分析的缺点,美国 Texas A&M 大学于 1996 年提出了 CSM(cooperating security managers)[4]系统.该系统采用对等体来组织系统,每个 CSM 就是一个入侵检测系统,各 CSM 之间通过交换信息来合作检测分布式入侵.但是,该系统在 CSM 数量大的情况下存在交互信息量大和综合判断能力不强的问题.(3) 由日本 IPA(information technology promotion agency)开发的 IDA(intrusion detection agent system)[5]是一个多主机检测系统,该系统采用两层的系统框架,其最大特点就是采用移动代理技术自动收集信息.但是,该系统只定义了某类特定事件,因此只适用于检测某一类分布式入侵,扩展性不强.(4) 在国内,对于分布式入侵检测系统的研究也有了一定的成果,如大规模分布式入侵检测系统(large-scale distributed intrusion detection system,简称 LDIDS)[6]采用分布采集、动态协调、集中管理的思想,采用树型的分层体系结构设计了一种大规模的分布式入侵检测系统.该系统具有很大的灵活性和可扩展性.但是,这些传统的分布式入侵检测系统由于处理数据量大且集中,负载不均衡,从而导致系统整体处理速率较慢,不能满足实时处理的要求.

网格的出现为入侵检测系统提供了新的工作环境,从而能够实现分布式的、负载均衡的入侵检测系统.我们所提出的网格入侵检测系统是基于 Globus 平台进行开发的.Globus 是美国 Argonne 国家实验室的研发项目,能够帮助规划和组建大型的网格实验平台.Globus 定义了在网格环境下进行互操作的一套通用协议,用来描述消息的格式和消息交换的规则,同时,在协议的基础上开发了系列的服务以及与服务功能相对应的 API (application programming interface)[7].

2 网格入侵检测系统框架

网格入侵检测系统由数据采集引擎、数据分析引擎、关联分析引擎和调度引擎构成.网格为系统提供工作环境.在该系统中,数据分析引擎是一种资源,其所提供的服务,也就是检测算法,需要在网格中进行注册才可能

被使用.数据采集引擎相当于网格中发起资源请求的用户,他们采集网络数据流,进行预处理操作,将生成的待处理文件提交给数据分析引擎.调度引擎担当资源管理器的角色,负责进行用户请求和资源之间的匹配,并根据一定的策略对满足要求的资源进行选择,同时对资源进行管理,实现该入侵检测系统的可扩展性.当分散在不同子网内的数据采集引擎采集到网络数据流后,提交请求给调度引擎,调度引擎首先根据资源需求描述信息在资源信息数据库中查找到提供相应服务的资源,然后根据一定的资源分配算法确定为其提供服务的数据分析引擎,并将相关信息返回给数据采集引擎.数据采集引擎将待处理文件传输给对应的数据分析引擎进行处理,生成的告警信息传输给关联分析引擎进行告警融合.系统体系结构如图 1 所示.

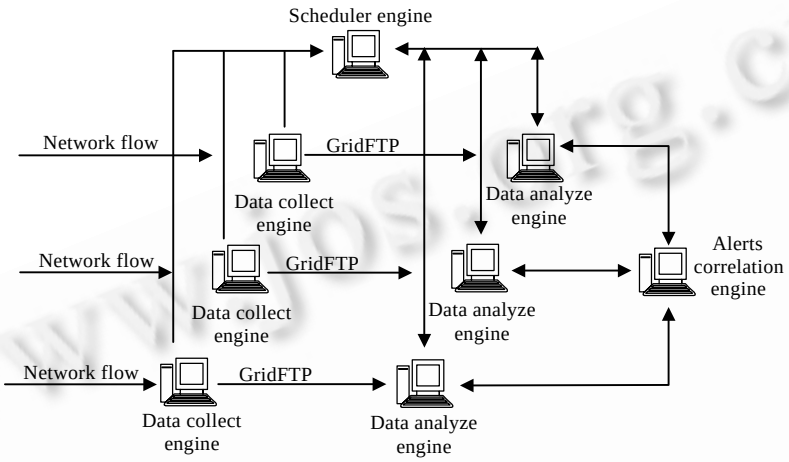

Fig.1 The architecture of grid IDS

图 1 网格入侵检测系统结构

2.1 数据采集引擎

数据采集引擎负责收集网络上的数据流,生成待处理文件,包括数据采集模块和数据预处理模块.其基本架构如图 2 所示.

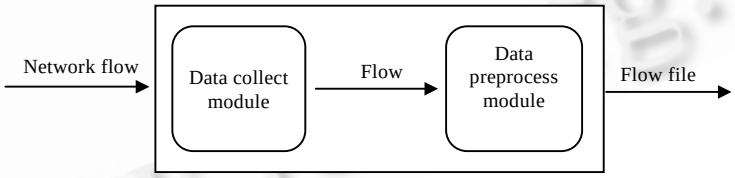

Fig.2 The architecture of data collect engine

图 2 数据采集引擎架构

数据采集引擎将采集到的网络数据包每 5 秒打包成一个文件.数据文件的大小与当时的网络流量相关.为了减少数据分析引擎的工作量,在数据采集引擎中加入了预处理模块.该模块主要进行网络数据的合并、去除冗余信息,记录每个数据包的源和目的 IP 地址及其在 5 秒内的通信量,同时,根据原始数据流提取下列信息交给数据分析引擎处理:协议类型(proto_type)、服务类型(service_type)、源到目的传送的字节数(bytes_src_to_dest)、目的到源传送的字节数(bytes_dest_to_src)、源与目的地址是否相同标识(land,相同为 1,否则为 0)、向相同主机发起的连接数(counts_to_dest)、向相同服务发起的连接数(counts_to_serv)、从相同的主机发起的连接数(counts_from_src)和从相同的服务发起的连接数(counts_from_serv).数据采集引擎将待处理文件的大小传送给调度引擎,请求分配数据分析引擎,得到调度引擎的处理结果后,使用 GridFTP 将待处理文件传送给数据分析引擎.GridFTP 是 Globus 借鉴 FTP 协议并在此基础上根据网格的特点进行扩展所形成的网格文件传输协

议.GridFTP 可以提供网格环境下安全传输、高效移动数据块的功能,满足网格计算环境不同的应用对广域范围分布的、大量的数据需求.同时,支持并行数据传输、缓冲区大小自动协商等,可以针对具体要传输的文件的大小,设置合适的缓冲区或窗口大小,同时,GridFTP 支持可靠的数据传输及数据重传[8].

2.2 数据分析引擎

数据分析引擎负责对待处理文件进行入侵检测,判断是否有入侵行为,包括数据分析模块、告警统计和管理模块以及监控和发现服务组件(monitoring and discovery service,简称 MDS),其框架结构如图 3 所示.

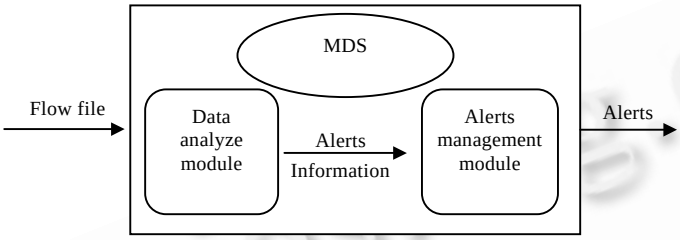

Fig.3 The architecture of data analyze engine

图 3 数据分析引擎架构

2.2.1 数据分析模块

数据分析引擎性能的优劣,取决于检测技术的选取.入侵检测技术的研究包括基于专家系统构造的技术、基于概率统计分析的技术和基于生物系统模拟的技术[9].专家的经验知识从逻辑上可以表示为产生式规则、层次树、状态转移图等形式.基于规则的入侵检测技术多应用于早期的入侵检测系统,如 P-BEST(production-based expert system toolset)[10].California 大学 Santa Barbara 分校的 STAT/NSTAT(state transition analysis tool/netstat)[11]是基于状态转移图的入侵检测系统.基于专家系统的检测技术具有误报少、准确性高的优点,但是它只能发现已知攻击,难以准确识别同一种攻击的变种,对未知的攻击不具备检测的能力.同时,规则库的建立与维护代价高,且容易出现冗余、矛盾、蕴含等问题.因此,目前基于专家系统构造技术的主要研究方向是用机器学习技术和数据挖掘技术使知识库的建造智能化.比如,基于数据挖掘的检测技术利用数据挖掘的关联分析、序列模式分析等算法提取入侵特征,从而达到识别入侵的目的.典型的系统是 Columbia 大学的 Lee 研究小组提出的 MADAM ID(mining audit data for automated models for intrusion detection)框架[12],但是,这种技术需要大量的审计数据,学习过程较慢,同时,对未知攻击的检测也不够理想.统计分析技术的关键是从描述系统或网络的行为和状态的属性中选择一组统计度量,并根据历史数据建立其正常的变化范围.这种方法的优点是不需要很多先验知识,有较为成熟的统计方法可以应用,但是,它往往对入侵发生的顺序不敏感,同时,对阈值难以确定.其中,支持向量机模型[13]应用于入侵检测取得了良好的效果.基于生物系统模拟技术的入侵检测技术主要研究将人工神经网络、进化计算和人工免疫系统应用于入侵检测中.比如,法国理工大学的 Ludovic 开发的 GASSATA(genetic algorithms for security audit trails analysis)是基于遗传算法的入侵检测系统[14],遗传算法的问题在于,交叉、变异和选择算子的设计目前主要依靠经验和实验的方法,若选择不合理,则会产生过早收敛的问题.人工免疫的检测技术是结合生物免疫系统的思想,将正常和异常的网络行为区分为自我和非我,从而达到判别入侵的目的,代表性的研究小组包括 New Mexico 大学的 Forrest 小组和 Memphis 大学的 Dasgupta 小组[15],这种检测技术具有极大的研究价值,但是仍然需要完善基于人工免疫的模型,同时,要找到合适的否定选择算法和克隆选择算法.

数据分析模块采用基于神经网络的检测技术.神经网络具有很好的学习能力和适应能力,能够处理噪声数据,而且实现简单、便于实用.我们采用的神经网络模型是 BP 网络.BP 网络是一种多层前馈神经网络,包括输入层、隐层和输出层.当学习样本提供给网络后,在输出层得到对输入的响应,按照减少目标输出与实际输出误差的方向,从输出层经过各隐层逐层修正各连接权值,以达到神经网络的实际输出与期望输出的最大拟和.对于入

侵的检测,实际上是一个二分类问题,通过对神经网络的训练,可以找到最佳分类面,达到识别攻击的目的.我们采用 3 层的网络设计,输入层由 9 个神经元组成,输出由 3 个神经元组成.对于隐层神经元数目的选择,采用经验公式与实验相结合的方式来确定.对于由 n 个输入神经元、m 个输出神经元的神经网络,其隐层神经元的个数可以根据如下公式进行选择:

$$h=\sqrt{n+m}+a,\ a\in[1,10] \tag{1}$$

每个数据分析引擎都安装了 MDS 组件,用于调度引擎查询当前资源的使用情况.调度引擎根据各分析引擎的处理器性能、内存大小以及当前资源使用情况和待处理的工作量大小进行统一的资源调度,实现各个数据分析引擎的负载平衡.

2.2.2 告警统计及管理模块

该模块负责对数据分析模块产生的告警进行汇总,方便关联分析引擎进一步的查询分析.告警统计及管理模块除了将告警流发送给关联分析引擎外,还根据源 IP 地址建立索引,记录连接的持续时间、协议、服务类型、传输的数据包数、传输的数据包字节数、连接数和攻击发现时间,提供给关联分析引擎确定是否发生攻击.

2.3 调度引擎

调度引擎是网格入侵检测系统的重要组成部分,整个系统的负载平衡能力通过调度引擎来调节.在网格入侵检测系统中,调度引擎完成的任务主要是收集各数据采集引擎经过预处理之后的文件大小,根据各数据分析引擎当前的负载情况和选择策略,将待处理文件分配给合适的数据分析引擎.

在网格环境下,解决任务调度问题也一直是研究人员关注的热点问题.任务调度包含调度模型、调度策略、目标函数和调度算法 4 个部分.就占用资源而言,调度策略包括:随机占用资源策略,这种策略具有最大的公平性,如用户直接指派 UDA(user directly assigning)算法;最早释放资源策略,这种策略要求尽早完成作业,释放资源,如最短完成时间 MCT(minimum completion time)算法;最大任务优先占用资源策略,这种策略要求资源优先分配给最大的任务,如 Max-Min 算法.此外,Beaumont 等[16]利用树型结构建模网格节点及其网络连接状况,提出网络带宽为中心的调度策略,并用来解决独立而等规模的任务调度.Buyya 等人提出的基于经济预算网格模型 Nimrod 提出了根据运行时限制和经济预算限制的调度算法[17].郑纬民教授等人针对 SMP 机群系统,从任务分配的角度进行了各个并行任务之间的通信优化研究[18].

我们结合网格入侵检测系统的实际情况,提出了结合资源可用度进行调度的策略.在网格入侵检测系统中,一个待处理文件是否分配给某个处理引擎,由该处理器上的性能综合指标决定,包括资源性能 P、资源忙碌程度 B 和资源执行作业的历史 H[19].一个大的待处理文件分配给综合指标高的处理器的概率将相对较高.处理器的性能综合指标在本文中被称为资源可用度,用 U 表示.

$$U_i=P_i\times B_i\times H_i\times U_{basei} \tag{2}$$

其中,U_i 是第 i 个分析引擎的资源可用度,P_i 是第 i 个分析引擎的资源性能,B_i 是第 i 个分析引擎的忙碌程度,H_i 是第 i 个引擎的作业执行历史,U_{basei} 是第 i 个分析引擎的资源可用度基准.

分析引擎上的资源性能由下面的公式来决定:

$$P_i=\sum_{j=1}^{n}Perf_{ij}\times W_j,\ \sum_{j=1}^{n}W_j=1 \tag{3}$$

其中,P_i 是第 i 个分析引擎的资源性能;$Perf_{ij}$ 是第 i 个分析引擎上第 j 个资源属性,如主频大小、内存大小、CPU 数目、CPU 型号等;W_j 是第 j 个属性所占的权重,代表设计者对各属性的重视程度.一旦分析引擎确定后,P_i 一般不需要改变.在网格入侵检测系统的设计中,只考虑主频大小和内存大小作为资源的性能指标.

资源忙碌程度由下面的公式决定:

$$B_i=\sum_{j=1}^{n}\left(1-\frac{U_{ij}}{T_{ij}}\right)\times W_j,\ \sum_{j=1}^{n}W_j=1 \tag{4}$$

其中,B_i 是第 i 个分析引擎的忙碌程度;U_{ij} 是第 i 个分析引擎上的第 j 类资源的使用量,比如 CPU 的使用、物理

内存的使用量等;T_{ij}是第 i 个分析引擎上的第 j 类资源的总量;W_j代表第 j 类资源使用量的权重.B_i是随时间不断变化的,因此,调度引擎需要定时地查询各分析引擎上各种资源的使用情况,这可以通过使用轻量级数据访问协议(lightweight directory access protocol,简称 LDAP)从 MDS 中获取.

资源的历史情况由下面的公式决定:

$$H_i = \frac{Size_i}{T} \tag{5}$$

其中:H_i是第 i 个分析引擎的历史情况;$Size_i$是在固定时间内处理完的文件大小(byte);T是一个固定时间;H_i记录了在一段时间内分析引擎的工作效率,该值的获得同样需要调度引擎与分析引擎的定时交互.

在调度引擎初始化时,会为每个分析引擎分配相同的资源可用度基准值.该值随着分析引擎不断地接收或完成一个文件任务而发生改变,其改变的过程如下:

当某个分析引擎接收一个任务时,

$$U_{basei} = U_{base}\left(1 - \frac{S_1}{S_{1i} + S_1}\right) \tag{6}$$

其中,U_{basei}是第 i 个分析引擎的资源可用度基准值,U_{base}是可用度初始值,S_1是分配给第 i 个分析引擎处理文件的大小,S_{1i}是第 i 个分析引擎上已有的待处理文件大小.

当某个分析引擎完成一个任务时,

$$U_{basei} = U_{base}\left(1 + \frac{S_2}{S_{2i} - S_2}\right) \tag{7}$$

其中,U_{basei}是第 i 个分析引擎的资源可用度基准值,U_{base}是可用度初始值,S_2是第 i 个分析引擎处理完成文件的大小,S_{2i}是第 i 个分析引擎上总共的待处理文件大小.若 $S_{2i}-S_2=0$,则 $U_{basei}=U_{base}$.

调度引擎按照上面的资源可用度计算方法为各分析引擎计算其资源可用度,决定待处理文件的分配方式.

2.4 关联分析引擎

数据分析引擎产生的告警数量很大,其中有相当一部分的告警都是针对同一攻击,这些相似的告警提供的信息量已经不大,还会导致一些有用的告警被淹没在告警洪流中,不易被管理员发现.为了提高网格入侵检测系统的总体性能,我们在设计时加入了关联分析引擎.关联分析引擎对告警进行聚集和关联分析,去除冗余告警,降低虚警率.它包括告警聚集模块和关联分析模块,其框架结构如图 4 所示.

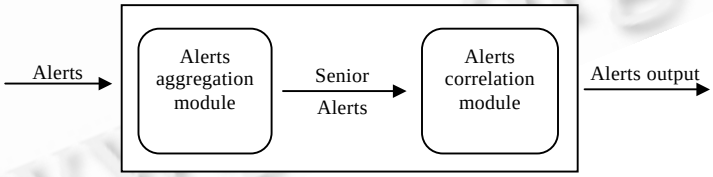

Fig.4 The architecture of alerts correlation engine

图 4 告警关联分析引擎架构

本文的告警合成算法借鉴了法国 MIRADOR 项目的关联模型 CRIM(cooperative module for intrusion detection systems)[20]和陈志文等人提出的警报合成算法[21].由告警聚集模块负责对告警进行汇聚,将多个相似的告警合并为一个高级告警集,合并规则如下:

规则 **1**. 只有相同攻击类型的告警才有可能合并为一个高级告警集.

规则 **2**. 对于相同攻击类型的告警,合并规则如下:每一个高级告警集都有其属性集合β记录其相关属性,包括该告警集的告警原型 p、该告警集中的告警集合 AS、当前告警集中的告警数量 n、该集合允许的告警数量 $N\in[N_{\min},N_{\max}]$、最近一次合并到该集合中告警的攻击发现时间戳 t、该告警集允许的等待时延 $T\in[T_{\min},T_{\max}]$、上一次相同高级告警集合并的告警数量 N_B和上一次相同高级告警集合并的等待时延 T_B.高级告警集组成的集

合称为告警基,还未被合并的警报称为孤立告警.

如果采用概率方法或专家系统的方法判定一个待处理的告警 p_{uk} 与某个高级告警集合的告警原型 p 相似,且 $n\neq0$,设该告警的攻击发现时间为 t_1.

1) 若 $n+1>N$,则将该告警集中的告警集合 AS 交给关联分析模块处理,删除该告警集,在新生成的告警集中保留删除的告警集中的告警原型 p,将 p_{uk} 插入到该高级告警集中,$N_B=N,T_B=T,n=1$.同时,将 N 和 T 按照如下公式进行更改:$N=\min[2N,N_{\max}]$和 $T=\min[2T,T_{\max}]$.

2) 若 $n+1\leq N$ 且 $t_1-t\leq T$,则将该告警合并到对应的告警集中,$n=n+1$.

3) 若 $t_1-t>T$,则将该告警集中的告警集合 AS 交给关联分析模块处理,在告警基中保留该高级告警的原型 $p,N_B=N,T_B=T,n=0$.

如果 p_{uk} 与某个孤立告警 p_s 相似,则生成一个新的高级告警集,$\beta=\{p_s,p_{uk},1,N,t_1,T,N,T\}$.

如果 p_{uk} 与某个告警集合的告警原型 p 相似,且 $n=0,\beta=\{p,p_{uk},1,\max[N_B-N_X,N_{\min}],t_1,\max[T_B-T_X,T_{\min}],N_B,T_B\}$,$N_X$ 和 T_X 可设定为小的常数.

如果 p_{uk} 不与任何一个告警相似,则将其加入孤立告警集中.

我们提出的这种告警合成算法,不仅能够有效地合成高级告警,减少告警量,而且通过当前的告警情况,按照乘性递增线性递减的方式动态调整 N 和 T 的设置,克服了 CRIM 告警模型中关于 N 和 T 难以选择的问题.在出现告警洪流的情况下,高级告警集的属性集合中保留了上次出现的告警洪流的相关信息,这样能够有效地避免文献[21]中同样的高级告警集再次产生告警洪流的情况.

关联分析模块主要对告警聚集模块产生的高级告警集进行关联分析,目前的关联方法主要有根据攻击过程进行关联[22]、根据攻击的因果关系进行关联[23]和基于入侵意图复合攻击检测[24]等.在本系统中,为了简化起见,对于 DoS 攻击的判定采用了设定阈值的方法,即当 DoS 攻击告警的数量超过一定的阈值时,即认为发生该种类型的攻击,产生告警输出.

3 实验及结果分析

在实验中,我们使用了 1 台数据采集引擎进行网络数据流的采集;使用 1 台调度引擎,4 台数据分析引擎进行攻击检测.这 4 台计算机的配置情况见表 1.

Table 1 Configuration of four data analyze engines

表 **1** 4 台数据分析引擎的配置情况

Data analyze engine	Processor	CPU clock speed (MHz)	Memory (MB)	Swap (MB)
Ⅰ	Pentium Ⅲ	800	1 024	2 048
Ⅱ	Pentium Ⅲ	800	256	512
Ⅲ	Pentium Ⅳ	2 400	512	1 024
Ⅳ	Pentium mobile	1 700	1 024	2 048

4 台计算机操作系统为 RedHat 9,均安装 Globus Toolkit 4.0.1,将这 4 台数据分析引擎组成一个集群.我们通过 WebMDS 统计各数据分析引擎的资源使用情况.图 5 是在调度引擎中注册的服务组信息.图 6 显示了数据分析引擎Ⅲ在运行检测任务时的资源使用情况.

数据分析引擎检测算法的网络结构是通过在实验中设置隐层神经元个数分别为 8,12,15 对网络进行训练而确定的.对于训练函数分别采用 LM(Levenberg-Marquardt)函数、一步正切函数(one step secant,简称 OSS)、动量及自适应 lrBP(learning rate BP)的梯度递减函数、Polak-Ribiere 连接梯度 BP(back propagation)函数、Fletcher-Powell 连接梯度 BP 函数和 Powell-Beale 连接梯度函数进行训练,根据实验结果确定.网络的隐层神经元传递函数为双曲正切 S 型函数,输出层神经元的传递函数为 S 型对数函数,当网络误差$\leq10^{-7}$时停止训练.对该神经网络的训练采用的数据集是 KDD CUP99,选择该数据集中与经过数据采集引擎预处理所提取信息相符合的 9 个字段,选取训练集中各 1 500 条正常连接和攻击连接进行训练.

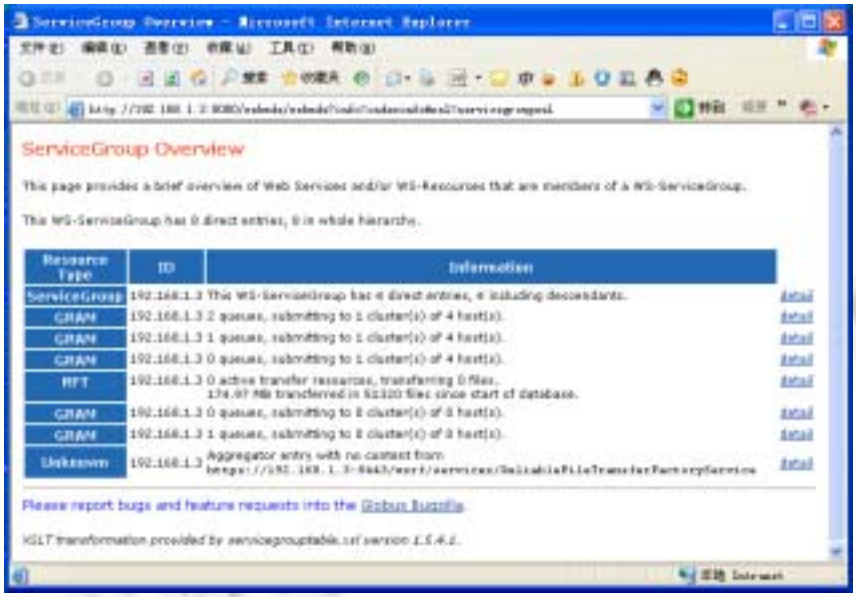

Fig.5 Service group overview

图 5 服务组信息

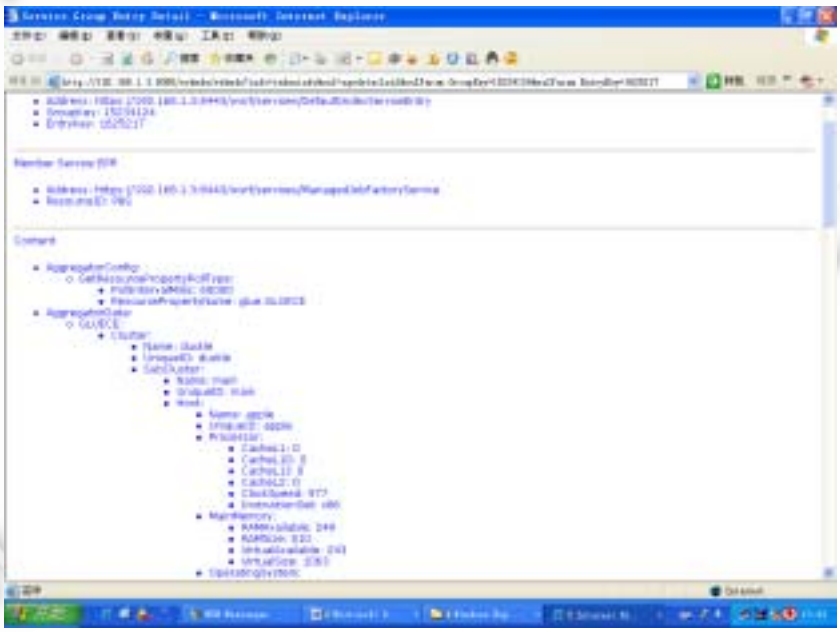

Fig.6 Detailed information of data analyze engine Ⅲ

图 6 数据分析引擎Ⅲ的详细信息

图 7 和图 8 分别是采用隐层神经元个数分别为 8,12,15 和 LM,OSS 训练函数的神经网络误差曲线.根据实验结果,采用 LM 算法与其他算法相比具有较快的收敛性,能够迅速收敛到目标误差,而当隐层神经元个数为 12 时,其误差较小,因此,决定该 BP 网络的隐层神经元个数为 12,训练函数采用 LM 函数.

为了验证数据分析引擎的检测性能,我们在实验中收集了 1 000 条正常连接和 1 000 条攻击连接,其中攻击

连接包括 DoS 攻击和普通攻击,攻击的速率在 200 包/秒~3 500 包/秒.实验结果的混淆矩阵见表 2.

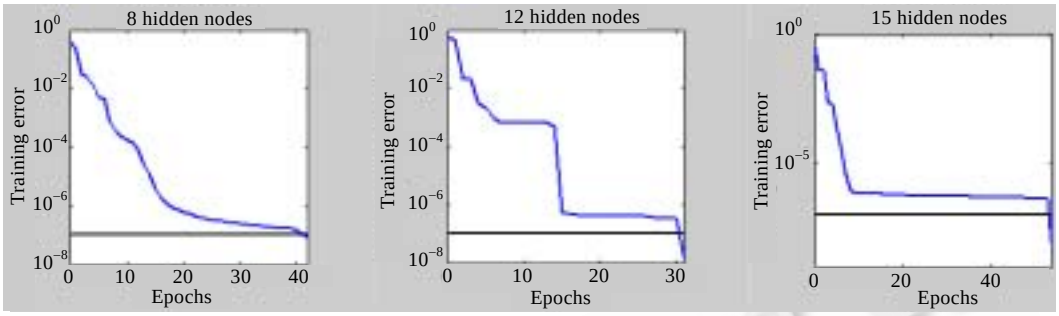

Fig.7 Error curve of LM with 8, 12, 15 neurons in hidden layer

图 7 隐层神经元个数分别为 8,12,15 的 LM 算法的误差曲线

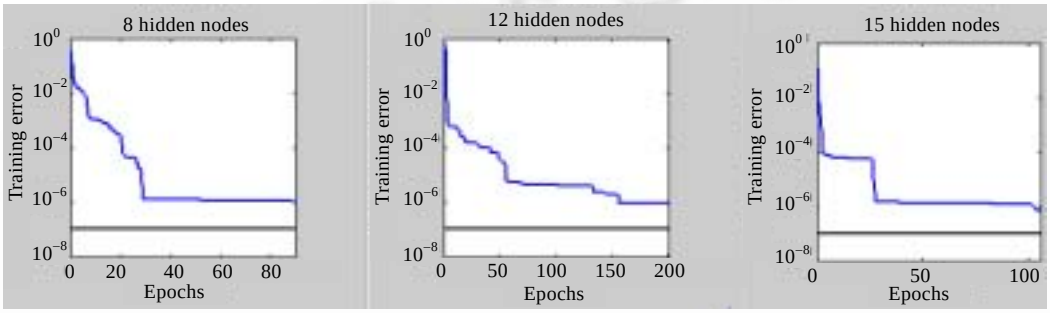

Fig.8 Error curve of OSS with 8, 12, 15 neurons in hidden layer

图 8 隐层神经元个数分别为 8,12,15 的 OSS 算法的误差曲线

Table 2 Confusion matrix

表 **2** 混淆矩阵

	Normal	Attack	Total
Normal	928	72	1 000
Attack	87	913	1 000

由实验结果可知:数据分析引擎的检测正确率为 92.05%,漏警率为 8.7%,虚警率为 7.2%.检测性能基本满足要求,但是仍然需要考虑优化算法进一步提高正确检测率和降低漏警率.

我们分别测试了采用单机入侵检测系统(single intrusion detection system,简称 SIDS)、采用 Round-Robin 方法分配分析引擎的网格入侵检测系统(round robin GIDS,简称 RR-GIDS)和基于资源可用度调度算法的网格入侵检测系统(Perf-GIDS)在相同攻击条件下的性能.攻击包括 DoS 攻击和普通攻击,攻击的平均速率为 3 200 包/秒,所产生的待处理文件的大小平均为 8MB.其结果见表 3,其中各统计项的说明情况如下:各分析引擎 CUP 负载的最大值和最小值(max/min PL)、各分析引擎的 CPU 负载平均值(average processor load,简称 APL)、各分析引擎内存使用的最大值和最小值(max/min AL)、各分析引擎内存使用的平均值(average memory load,简称 AML)、系统平均响应时间(average response time,简称 ART)和系统平均检测时间(average detection time,简称 ADT).

Table 3 Results of the experiment

表 **3** 实验结果

	Max/Min PL (%)	APL (%)	Max/Min AL (%)	AML (%)	ART (s)	ADT (s)
SIDS	78.2/69.5	74.3	28.1/10.8	19.3	48.9	13.7
RR-GIDS	89.7/54.2	71.6	35.2/7.6	18.5	25.2	18.9
Perf-GIDS	72.1/20.4	43.7	26.0/3.9	6.8	2.2	8.6

从实验结果可知,Perf-GIDS 在整体的 CPU 负载和内存负载方面的性能明显优于 RR-GIDS 和 SIDS.在 CPU 的负载方面,Perf-GIDS 只有 SIDS 和 RR-GIDS 负载的 58.%和 61.0%;而在内存的使用方面,Perf-GIDS 只有 SIDS 和 RR-GIDS 内存负载的 35.2%和 36.8%;在响应延迟方面,Perf-GIDS 与 SIDS 和 RR-GIDS 相比具有较低的延迟,平均的响应延迟仅为 2.2 秒;而检测时间的性能也优于 SIDS 和 RR-GIDS,仅需 8.6 秒就可以检测出攻击.该系统能够解决传统的分布式入侵检测系统由于处理的数据量大、负载不均衡而产生的检测效率下降的问题,为实现实时入侵检测提供了良好的解决方案.

4 结 论

本文的研究目的在于,将智能检测技术与先进的分布式体系结构相结合,建立新的入侵检测系统架构.为了实现此目的,本文提出了基于神经网络检测技术的智能网格入侵检测系统.基于神经网络的检测技术具有学习能力强的优点,具有较好的正确检测率;文中提出的报警合成算法基于乘性递增线性递减的方法实现动态窗口的调整,能够防止高级告警集再次产生告警洪流.实验表明,该系统采用基于资源可用度的调度算法在进行资源调度时,将待处理的大文件分配给具有较高资源可用度的分析引擎,这不仅缩短了攻击检测时间,而且实现了负载均衡,在检测性能和检测能力上明显优于 RR-GIDS 和 SIDS.

本课题下一步的研究方向包括:1) 将 P2P 与网格结合,利用优势互补建立基于 P2P-Grid 的入侵检测系统;2) 研究基于免疫的和基于支持向量机的入侵检测技术,实现采用不同检测算法的数据分析引擎,并根据条件进行合理选择;3) 研究根据当前的网络流量预测待处理文件大小的算法,进一步缩短处理的响应时延;4) 研究报警关联算法,主要针对复合攻击进行关联和入侵意图的预测,进一步提高系统性能.

References:

[1] Plaszczak P, Wellner R. Grid Computing. San Francisco: Morgan Kaufmann Publishers, 2005.

[2] Anderson JP. Computer security threat monitoring and surveillance. Technical Report, 79F296400, Fort Washington: James P. Anderson Company, 1980.

[3] Snapp SR, Brentano J, Dias GV, Goan TL, Heberlein LT, Ho CL, Levitt KN, Mukherjee B, Smaha SE, Grance T, Teal DM, Mansur D. DIDS (distributed intrusion detection system)—Motivation, architecture, and an early prototype. In: Proc. of the 14th National Computer Security Conf., Vol 10. Washington, 1991. 167−176.

[4] White GB, Fisch EA, Pooch UW. Cooperating security managers: A peer-based intrusion detection system. IEEE Network, 1996, 10(1):20−23.

[5] Asaka M, Taguchi A, Goto S. The implementation of IDA: An intrusion detection agent system. In: Proc. of the 11th FIRST Conf. 1999. Brisbane, 1999.

[6] Chu YG. Researches on large-scale distributed intrusion detection system [Ph.D. Thesis]. Beijing University of Posts and Telecommunications, 2005. 65−72 (in Chinese with English abstract).

[7] A globus primer. 2005. http://www.globus.org/toolkit/docs/

[8] Xu ZW, Feng BM, Li W. Grid Computing Technology. Beijing: Publishing House of Electronics Industry, 2005. 140−142 (in Chinese).

[9] Zhao JZ. Research on intrusion detection system model based on immunity system [Ph.D. Thesis]. Beijing Jiaotong University, 2003 (in Chinese with English abstract).

[10] Lindqvist U, Porras PA. Detecting computer and network misuse through the production-based expert system toolset (P-BEST). In: Proc. of the 1999 IEEE Symp. on Security and Privacy. Oakland, 1999. 146−161. http:// ieeexplore.ieee.org

[11] Ilgun K, Kemmerer RA, Porras PA. State transition analysis: A rule-based intrusion detection approach. IEEE Trans. on Software Engineering, 1995,21(3):181−189.

[12] Lee W, Stolfo SJ, Mok KW. A data mining framework for building intrusion detection models. In: Proc. of the 1999 IEEE Symp. on Security and Privacy. Oakland, 1999. 120−132. http://ieeexplore.ieee.org

[13] Batur C, Zhou L, Chan CC. Support vector machines for fault detection. In: Proc. of the 41st IEEE Conf. on Detection and Control. Las Vegas, 2002. 1355−1356. http:// ieeexplore.ieee.org

[14] Ludovic M. Genetic algorithm, a biologically inspired approach for security audit trails analysis. In: Proc. of the 12th Int'l Conf. On Computer Safety. 1993.

[15] Dasgupta D. Immunity-Based intrusion detection system: A general framework. In: Proc. of the 22nd NISSC. 1999.

[16] Beaumont O, Carter L, Ferrante J, Legrand A, Robert Y. Bandwidth-Centric allocation of independent tasks on heterogeneours platforms. In: Proc. of the Int'l Parallel and Distributed Processing Symp. 2002.

[17] Buyya R, Abramson D, Giddy J. Nimrod/G: An architecture for a resource management and scheduling system in a global computational grid. In: Proc. of the High Performance Computing in the Asia-Pacific Region. 2000.

[18] Zheng WM, Yang B, Lin WJ, Li ZG. An improved communication of parallel task scheduling on SMP system. Science in China (Series E), 2001,31(5):442−454 (in Chinese with English abstract).

[19] Zhang SD, Cao YD, Liao LJ. Job scheduling algorithm based on credit model in cluster environment. Mini-Micro System, 2005,26(12):2140−2143 (in Chinese with English abstract).

[20] Cuppens F. Managing alerts in a multi-intrusion environment. In: Proc. of the 17th Annual Computer Security Applications Conf. 2001.

[21] Chen ZW, Wang KY, Jiang JG. Design of alert merging algorithm of network-based intrusion detection system. Information and Electronic Engineering, 2005,3(3):182−185 (in Chinese with English abstract).

[22] Valeur F, Vigna G, Kruegel C, Kemmerer R. A comprehensive approach to intrusion detection alert correlation. IEEE Trans. on Dependable and Secure Computing, 2004,1(3):146−169.

[23] Ning P, Cui Y. An intrusion alert correlator based on prerequisites of intrusion. Technical Report, TR-2002-01, Department of Computer Science, North Carolina State University, 2002.

[24] Bao XH, Dai YX, Feng PH, Zhu PF, Wei J. A detection and forecast algorithm for multi-step attack based on intrusion intention. Journal of Software, 2005,16(12):2132−2138 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/16/2132.htm

附中文参考文献:

[6] 褚永刚.大规模分布式入侵检测系统关键技术研究[博士学位论文].北京:北京邮电大学,2005.65−72.

[8] 徐志伟,冯百明,李伟.网格计算技术.北京:电子工业出版社,2005.140−142.

[9] 赵俊忠.基于免疫机制的入侵检测系统模型研究[博士学位论文].北京:北京交通大学,2005.

[18] 郑纬民,杨博,林伟坚,李志光.SMP 机群系统上优化通信的并行任务调度.中国科学(E 辑),2001,31(5):442−454.

[19] 张树东,曹元大,廖乐健.资源调度中的资源信度模型和调度算法.小型微型计算机系统,2005,26(12):2140−2143.

[21] 陈志文,王开云,姜建国.网络入侵检测系统的警报合成算法设计.信息与电子工程,2005,3(3):182−185.

[24] 鲍旭华,戴英侠,冯萍慧,朱鹏飞,魏军.基于入侵意图的复合攻击检测和预测算法.软件学报,2005,16(12):2132−2138. http://www.jos.org.cn/1000-9825/16/2132.htm

魏宇欣(1982－),女,江西南昌人,博士生,主要研究领域为网络安全,人工智能,Ad Hoc 网络.

武穆清(1963－),男,博士,教授,博士生导师,主要研究领域为网络安全,宽带接入理论与新技术.