第 34 卷 第 21 期 计 算 机 工 程 2008 年 11 月
Vol.34 No.21 Computer Engineering November 2008
·人工智能及识别技术· 文章编号：1000—3428(2008)21—0184—03 文献标识码：A 中图分类号：TP309.1

多 Agent 负载均衡在入侵检测系统中的应用

许素霞，傅秀芬，胡金霞，高保庆，苏 磊
(广东工业大学计算机学院，广州 510006)

摘 要：针对在高速网络环境中实现入侵检测系统的动态负载均衡问题，提出一种基于多 Agent 的负载识别和基于改进遗传算法的动态负载均衡策略，实现入侵检测系统中的智能负载均衡，给出相应的系统模型，并以实验结果证明其有效性。
关键词：负载均衡；遗传算法；耦合度；多 Agent 系统

Application of Load Balancing in Intrusion Detection System Based on Multi-Agent

XU Su-xia, FU Xiu-fen, HU Jin-xia, GAO Bao-qing, SU Lei
(Department of Computer, Guangdong University of Technology, Guangzhou 510006)

【Abstract】 Aiming at intrusion detection system dynamic load balancing in the high-speed network environment, this paper advances a strategy based on multi-agent load identification and Genetic Algorithms(GA) to improve the dynamic load balancing, achieves the intrusion detection system intelligent load balancing, presents the system corresponding model, and experiments show that the technology can be used to solve the problem of intrusion detection system load balancing in the high-speed network.
【Key words】 load balancing; Genetic Algorithms(GA); coupling degree; MAS

随着入侵手段的多样化和复杂化，仅有防御作用的静态安全技术已无法解决网络安全问题。入侵检测技术是一种能对网络安全实施实时监控，并对网络受到攻击时进行保护的动态策略。高速局域网和光纤通信等新技术的应用对网络性能和带宽流量也提出更高的要求。从最初 10M/100M 网络，发展到目前的高速网络，传统的入侵检测系统会出现严重的丢包/漏包现象。针对在高速网络环境中实现实时入侵检测的问题，本文从系统结构、负载调度策略、负载识别技术等各方面提出一种新的入侵检测方法，同时对入侵检测系统加以改进，使之符合高速环境的要求。

1 负载均衡技术

负载均衡指在集群环境中，保证各服务器(或处理器)的计算量与自身性能之比相等，从而在提高服务器利用率的基础上减少整体任务完成时间，可分为静态和动态 2 种[1]。由于静态负载均衡缺乏实时控制突发性事件的能力，本文基于动态负载平衡调度策略提出了一种集中式与分布式相结合的调度模型。

实现入侵检测中负载均衡需解决 3 个问题：服务器负载状况的定义，负载的识别和收集，负载的调度。负载识别是负载均衡的前提，只有识别出各种计算节点的负载情况，负载均衡才能实现。本文在研究分析现有均衡策略、负载识别技术的基础上，提出一种基于多 Agent 的负载识别和基于改进遗传算法的动态负载均衡策略。

1.1 基于多 Agent 的负载识别

Agent起源于人工智能领域，是在特定环境下能感知环境，并能自治运行代替其设计者或使用者实现一系列目标的计算实体或程序[2]。Agent具有主动性、智能性、反应性、通信性及移动性等特性，其基本结构如图 1 所示。

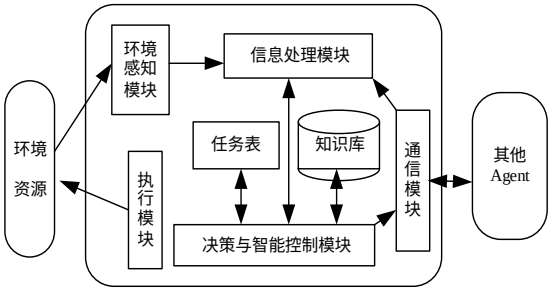

图 1 Agent 的基本结构

Agent 由环境感知模块、执行模块、通信模块、信息处理模块、决策与智能控制模块及知识库和任务表组成。其中，环境感知模块是 Agent 的核心，用于对服务器或网络设备处理能力的探测；信息处理模块依据知识库和决策与智能控制模块对感知和接收的信息进行处理；决策与智能控制模块是赋予 Agent 智能的关键，它运用知识库中的知识和其他 Agent 进行通讯或从任务表中选择适当的任务供执行模块执行；任务表为 Agent 所要完成的功能和任务。

当利用 Agent 识别和收集网络负载时，应将具有不同能力的 Agent 个体组成一个为完成某个任务而共同合作的智能整体，即多 Agent 系统(MAS)。本系统基于多 Agent 的负载

基金项目：广东省自然科学基金资助项目(06021484, 500184)
作者简介：许素霞(1982－)，女，硕士研究生，主研方向：网络安全，计算机网络，人工智能；傅秀芬，教授；胡金霞、高保庆、苏 磊，硕士研究生
收稿日期：2007-12-26 **E-mail：**xusuxia7@163.com

识别和收集技术，可实时动态探测识别出各节点的负载能力和负载状况，还可及时侦测出节点故障并进行负载的再分配。

1.2 基于改进遗传算法的负载均衡调度算法

负载调度算法是负载均衡技术中的核心，直接影响系统的性能。本文利用遗传算法堆网络请求调度进行全局搜索，设计一个基于改进遗传算法的负载均衡系统[3]。

1.2.1 种群的确定

将问题空间中的参数编码，转换成遗传空间根据基因结构组成的染色体或个体。设有$m(30\leq m\leq 160)$个节点参与负载均衡计算，其中有j个节点负载失衡，则有$m-j$个节点作为迁移接收方。n个Agent分别为$\{a_1, a_2,\cdots, a_n\}$，有k个Agent选定参与负载均衡计算。$\{l_1, l_2,\cdots, l_k\}$分别为是节点$\{m_1, m_2,\cdots, m_j\}$上Agent$\{a_1, a_2,\cdots, a_k\}$的负载，$\{l_{k+1}, l_{k+2},\cdots, l_{k+(m-j)}\}$分别是节点$\{m_{j+1}, m_{j+2},\cdots, m_m\}$上其余Agent的负载。节点$i$的负载是该节点上所有Agent负载之和，用$L_i$表示。规定：需特殊的处理器、输入输出设备、或有大量数据放置在本机上的Agent以及作为接收方的机器不参与迁移，但后者参与负载均衡的计算。因此初始种群为$\{l_1, l_2,\cdots, l_j,\cdots, l_m\}$。

1.2.2 适应度函数

在设计适应度函数时，需要做一个等效转换将节点的总体负载和服务器的负载联系起来。个体适应度函数单调递减函数，即服务器负载值越小，个体适应度函数值越大；服务器负载值越大，个体适应度函数值越小。

定义 1 Agent的信誉值[4]C_i, C_i为

$$C_i=-x_1w_i-x_2u_i \tag{1}$$

其中，x_1, x_2为非负实数，由实际应用决定；w_i为a_i运算负载；u_i为 a_i 通信负载；w_i与u_i均由时钟计算所得。每个Agent的信誉与Agent系统和Agent之间作用的行为保持一致，反映一个Agent和节点之间的密切关系，而节点之间通信对信誉值起反面作用，故在式中用负号。Agent的信誉值受以下行为影响：该Agent的工作量、与同节点其他Agent的通信、与不同节点上其他Agent的通信及该Agent的移动性能。

定义 2 正在执行的 Agent 的通信负载为

$$u_i=\sum_{M(a_i)\neq M(a_j)} c(a_i,\ a_j) \tag{2}$$

其中，$c(a_i,\ a_j)$表示a_i和a_j之间通信的时钟数，由每个Agent使用的信息大小所得。

定义 3 节点m_k的负载L_k定义为该节点上所有Agent的负载之和，表示为

$$L_k=\sum_{M(a_i)=k} c(w_i+u_i) \tag{3}$$

定义 4 Agent a_i 的耦合度 A_i 分别由通信负载和通信负载与计算负载比值决定，表示为

$$A_i=\alpha(\beta_1\times(u_i/w_i)/\max_{j\in[1,n]}(u_i/w_i)+\beta_2\times u_i/\max_{j\in[1,n]}u_j) \tag{4}$$

其中，u_i为通信负载；w_i为运算负载；$\max\limits_{j\in[1,n]}u_j$ 为所有通信负载中的最大值；$\max\limits_{j\in[1,n]}(u_i/w_i)$ 为所有通信负载与运算负载比值的最大值；β_1, β_2, α,均为参数，α决定通信负载和运算负载在整体运算中所占的比重，称为放大因子，在同一系统中取固定值。α值越大，则整个系统中每个Agent的通信负载所占的比重越大。

A_i有 2 个极端情况：该值足够大时，运算负载可忽略不计，此时只考虑通信负载的影响；该值足够小时，通信负载可忽略不计，此时只考虑运算负载的影响。因此，A_i值是动态更新的。

定义 5 个体适应度函数是节点负载值单调递减的函数，一个染色体 i 的适应度为

$$F(i)_{\text{fitness}}=\frac{\max\limits_{i=1}^{m}L_i-L_i}{\frac{1}{m}\sum\limits_{i=1}^{m}L_i}\times\frac{\text{Average}(|C_i|)}{\text{Average}(A_i)} \tag{4}$$

其中，$\frac{1}{m}\sum\limits_{i=1}^{m}L_i$ 为所有节点的平均负载；$\max\limits_{i=1}^{m}L_i$ 为节点最大负载；Average($|C_i|$) 为节点所有 Agent 平均信誉值；Average(A_i) 为节点所有 Agent 的平均耦合度。

1.2.3 改进 GA 算法自身参数的设定

(1)选择算子

用点搜索法(μ,λ)选择，在(μ,λ)即 ES 中，比值 μ/λ 用以控制群体多样性和选择力度，它决定算法的收敛速度，是一个重要参数。实验证明，当 μ/λ=1/5（μ>15）时，进化策略最优。

(2)交叉算子

随机设定两个交叉点，个体中两交叉点中间部分将被用于交叉。

(3)变异算子

采用交换变异，随机选取一对变异交换相应的值。

2 基于负载均衡的入侵检测系统设计

本文设计一种高速网络环境下基于动态流量负载均衡的分层式入侵检测系统模型，该模型主要由数据捕获模块、负载均衡模块、转发代理、探测器及控制中心等组成[5]，其结构如图 2 所示。

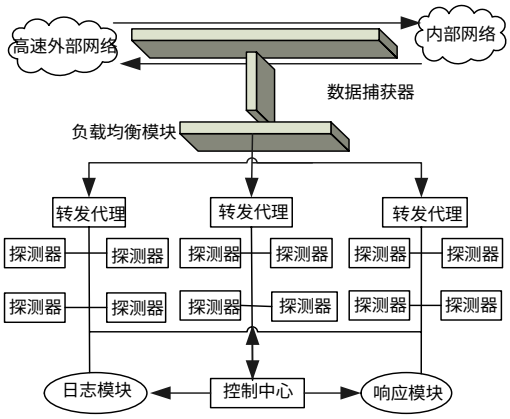

图 2 基于负载均衡的入侵检测系统结构

数据捕获模块由多个移动 Agent 组成，负责从监控网络收集流量，并转发给负载均衡模块。在具体的网络拓扑下数据捕获模块可接到交换机的监听端口及其镜像端口处。

负载均衡模块根据负载均衡算法将流量转发到与其相连的所有探测器，保证其下探测器只负责部分流量的检测，完成第一层分流。

转发代理接收均衡代理转发的流量，并将其转发给探测器。与同一个转发代理相连接的所有探测器均并行工作。

探测器接收到来自转发代理的数据流量，根据自身的检测规则及检测功能，利用过滤功能对流量进行第 2 次分流，进行流量检测，并将报警信息以二进制的形式通过控制中心进行相应的响应及存储。

控制中心模块由多个 Agent 组成，负责检测探测器的工作状态，收集探测器的负载信息。当有探测器发生故障或系统内探测器的负载变化超过设定阀值时，控制中心会通知负载均衡模块重新对流量分配比例进行调整。

本系统的负载均衡整体算法描述如下:

```
Initialization:初始化系统中所有参数;
While (true) do {
负载收集:通信 Agent 获得每个功能 Agent 的负载值;
 负载统计:通信 Agent 统计每个节点的负载;
 if(节点 S 负载>域值) {
        通知中央 Agent;
        中央 Agent 收集全局负载信息;
        确定初始种群;
        用改进遗传算法进行负载均衡计算;}
        Sleep (T);
              }
```

3 仿真实验

硬件环境：负载均衡器为一台 PC 机（Pentium Ⅳ 2.0 GHz, 512 MB, 80 GB 硬盘）；探测器为若干台 PC 机（Celeron(R)CPU2.40 GHz 或 Pentium Ⅲ1.7 GHz, 256 MB 内存，40 GB 硬盘）；转发代理为 100BASE-T 8 口 HUB。

系统中每个工作Agent给定一个负载范围，用于创建工作负载值包括通信负载值和计算负载值，在实验中须调节参数 β_1, β_2 和 α 的值。测试数据集使用MIT林肯实验室的DARPA1999 IDS测试数据集中的第5周中周4和周5的数据，使用TCPREPLAY按不同速度回放。为证明本文设计的负载均衡调度算法的优越性，进行测试，结果如表 1 所示。

表 1　基于负载均衡调度策略的响应率

网络负载/(Mb·s^{-1})	遗传算法/(%)	改进遗传算法/(%)
50	97.2	99.2
80	96.5	97.5
100	94.3	96.1

可见，改进遗传算法的响应率优于遗传算法调度策略。由于调度遗传算法的时间不能太长，否则会影响到调度的效率，因此必须限制每次调度遗传算法要完成的任务量、每次调度遗传算法的换代次数以及 GA 算法自身参数的设定。

上述限制会影响调度遗传算法的效果，并且随网络负载的加重，降低基于遗传算法和基于改进遗传算法的负载均衡策略的响应率。通过上述实验，可取得调度遗传算法最佳参数，使改进遗传算法可在相同调度时间内得到最佳调度方案。

4 结束语

本文研究基于负载均衡的入侵检测系统，提出基于多 Agent 的动态智能负载均衡策略。在高速网络环境的入侵检测系统中，采用基于 MAS 的负载识别和收集及利用改进遗传算法进行负载调度。尽管 GA 是一种有效的最优解搜索方法，但其自身参数的确定很复杂，下一步研究工作是通过仿真实验对基于 GA 算法的负载均衡调度算法与其他负载均衡算法进行性能比较分析，以发现 GA 调度算法的不足，进一步优化基于遗传算法的负载均衡调度算法。

参考文献

[1] Hac A, Jin Xiaowei. Dynamic Load Balancing in a Distributed System Using a Sender-Initiated Algorithm[C]//Proc. of the 13th Conf. on Local Computer Networks. N Y, USA: [s. n.], 1988: 172-180.

[2] Wooldridge M J, Jennings N R. Agents Theories, Architectures and Languages: A Survey[C]//Proc. of the Electronic Cultural Atlas Initiative Workshop on Agent Theories, Architectures and Languages. Berlin, Germany: Springer Verlag, 1995: 1-32.

[3] Zomaya A Y, Yee-Hwei Teh. The Observations on Using Genetic Algorithms for Dynamic Load-balancing[J]. IEEE Trans. on Parallel and Distributed Systems, 2001, 12(9): 899-911.

[4] Chow K P, Kwok Y K. On Loading Balancing for Distributed Multi-agent Computing[J]. IEEE Trans. on Parallel and Distributed Systems, 2002, 13(8): 787-801.

[5] NSS Group. Intrusion Detection and Vulnerability Assessment[R]. NSS, Oakwood House, Wennington, Cambridge Shire, UK, Tech. Rep.: 2009-215, 2000.

（上接第 183 页）

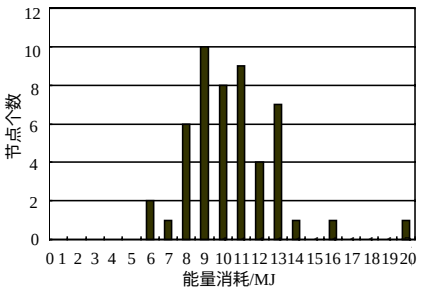

图 6　10 轮拓扑控制中的节点能量消耗分布

4 结束语

本文结合煤矿井下的实际情况，借鉴神经网络能从环境中学习，进而改善其行为和结构以适应环境要求的特点，提出一种基于竞争学习的聚类算法，阐述了采用该算法的传感器节点在 6 种状态间的转换和簇的形成过程。仿真试验表明，该算法具有网络能耗均衡、动态组网速度快等特点，满足井下灾害发生情况下传感器节点损坏和救援人员所携带传感器节点加入网络等动态组网的需求，在煤矿井下人员定位以及救援过程中具有重要的实际意义。

参考文献

[1] Heinzelman W B, Chandrakasan A P, Balakrishnan H. An Application Specific Protocol Architecture for Wireless Microsensor Networks[J]. IEEE Transaction on Wireless Communications, 2002, 1(4): 660-670.

[2] Chan Haowen, Perrig A. ACE: An Emergent Algorithm for Highly Uniform Cluster Formation[C]//Proc. of the 1st European Workshop on Wireless Sensor Networks. Berlin, Germany: Springer-Verlag, 2004.

[3] Estrin D, Govindan R, Heidemann J. Next Century Challenges: Scalable Coordination in Sensor Networks[C]//Proceedings of MobiCom'99. New York, USA: ACM Press, 1999.

[4] 朱大奇，史　慧. 人工神经网络原理与应用[M]. 北京：科学出版社, 2006.

[5] 姜国彬，张世永，钟亦平. 一种带有自维护功能的无线传感器网络聚类算法[J]. 计算机工程, 2006, 32(10): 99-101.