基于内分泌机制的防火墙自适应调控算法

朱思峰[1,2] 王华东[1] 魏荣华[3]
(周口师范学院数学与信息科学系 周口 466000)[1] (西安电子科技大学计算机学院 西安 710071)[2]
(河北工程技术高等专科学校计算机系 沧州 061001)[3]

摘 要 防火墙是常用的网络安全防御系统,防火墙参数自适应调控对提高网络性能有重要的作用,目前尚无自适应的调控算法。分析了内分泌系统对生物体内环境的调节机制,设计了工作在网络环境中的人工内分泌系统,构造了一种基于人工内分泌系统的防火墙自适应调控算法。仿真实验表明,该调控算法能够根据网络性能变化,自适应地调整防火墙参数,具有较好的应用价值。提出的算法经过适当修改,还可以应用于入侵检测系统的自适应调控。

关键词 内分泌系统,激素,防火墙,网络性能,自适应调控算法
中图法分类号 TP18,TP301 **文献标识码** A

Adaptive Adjusting Algorithm of Firewall Based on Endocrine System

ZHU Si-feng[1,2] WANG Hua-dong[1,2] WEI Rong-hua[3]
(Department of Mathematics & Information Science, Zhoukou Normal University, Zhoukou 466000, China)[1]
(School of Computer Science, Xidian University, Xi'an 710071, China)[2]
(Department of Computer Science, Hebei Engineering and Technical College, Cangzhou 061001, China)[3]

Abstract Firewall is a useful tool for protecting network, and adjusting its parameters adaptively is important to network performance. At present, there are no available adjusting algorithms for firewall. The mechanism of endocrine system was analyzed, a kind of artificial endocrine system (AES) was designed, and an adaptive adjusting algorithm based on AES was constructed. Simulation tests show that the adjusting algorithm given in this paper has advantages of good performance, and applying prospect. The algorithm with modification can be used in intrusion detection system.

Keywords Endocrine system, Hormone, Firewall, Network performance, Adaptive adjusting algorithm

防火墙作为网络安全防护的主要手段,在日益复杂的网络环境下,面临的主要问题是:一方面,日积月累的防火墙规则在提供网络安全性能的同时,由于大量过滤规则的模式匹配过程花费很多时间[1,2],使用户的上网速率下降,网络可用性下降,从而导致了网络用户的不满和投诉;另一方面,若对防火墙过滤规则宽松处理,可以提高网络的可用性,但又会发生漏检现象,给企业的网络安全带来隐患。如何平衡网络可用性和网络安全性并对防火墙参数进行动态调整,已经成为国内外学者的关注热点[3,4]。

生物体的内分泌是一个非常复杂的系统,它由内分泌腺体、内分泌细胞和相应的激素组成[5]。内分泌系统含有成千上万的激素产生细胞,每一类激素都影响着机体对内外环境的反应,同时,一定的环境刺激也能影响不同的腺体分泌一定量的激素,使其适应内外环境的变化,这种变化是一种机体的自我平衡,使个体的运动朝着有利于自己适应性的方向发展,从而增强生物体对环境的适应能力。

继人工神经网络、人工免疫系统之后,人工内分泌系统(Artificial Endocrine System,简称 AES),现已经发展成为一种解决复杂控制问题的有效方法,在工程领域得到了初步的应用[6-8]。基于此,借鉴内分泌系统对神经系统的高级调节作用,本文提出了一种基于内分泌调节的防火墙规则自适应优化机制。

1 生物内分泌系统

生物内分泌系统由内分泌细胞、内分泌细胞所释放的激素(荷尔蒙)和内分泌腺体组成。适宜的刺激可刺激生物体内分泌细胞产生适当种类和数量的激素,这些激素和神经系统的共同作用,维持着机体内环境的相对稳定,进而影响生物体的行为[9]。

各种内分泌腺体形成一个闭合回路,它们之间通过复杂的相互作用维持着机体内环境的相对稳定。在闭合回路的基础上,中枢神经系统可接受外环境中的各种信息(声、光、温度、味等),通过下丘脑把内分泌系统与外环境联系起来,形成开口回路。下丘脑中的部分神经细胞可分泌神经激素,神经

到稿日期:2009-01-05 返修日期:2009-06-19 本文受国家自然科学基金项目(60575037,60502043),河南省自然科学研究项目(082400440260),河南省教育厅自然科学研究项目(2008A180041)资助。
朱思峰(1976-),男,博士研究生,副教授,CCF 会员,主要研究方向为人工智能、计算智能、智能信息处理等,E-mail:zhusifeng@163.com;王华东(1978-),博士研究生,讲师,主要研究方向为人工智能、计算智能、智能信息处理与进化算法等;魏荣华(1978-),女,助教,主要研究方向为计算机网络。

激素作用于垂体，影响垂体的分泌，进而影响靶腺（如甲状腺、肾上腺等）的分泌。靶腺激素反过来影响神经系统的发育及其活动，进而影响生物体的行为。正是因为内分泌系统与神经系统的密切配合，才使机体能更好地适应环境的变化。

由上面的简单论述，可知生物体的神经子系统和内分泌子系统是紧密联系和相互作用的，没有它们的密切配合，生物体的内环境就不能保持相对稳定，机体的功能就会受到影响甚至丧失。这为本文后面设计的调控系统提供了很好的生物学基础。内分泌激素对生物体神经系统的抑制和促进作用为防火墙参数调控策略的设计提供了思路。

2 设计思想

在生物体的内环境中，神经子系统和内分泌子系统的密切配合，使机体能根据环境的变化而自适应变化。本文把整个网络环境看作一个生物体，把防火墙系统看作生物体的神经子系统，把防火墙参数调控系统看作生物体的内分泌子系统。

1）内分泌细胞

为了简化，内分泌细胞和内分泌腺体不加区分，通称为内分泌细胞。这样，就可以把发布在网络中不同节点上的一个个自主 agent，看作一个个内分泌细胞。自主 agent 根据网络的状态（如网络的可用性、网络的安全性等）分泌激素。

2）核心内分泌细胞

不同于生物内分泌系统，在本文设计的人工内分泌系统中，把负责测算激素浓度和激素类型的功能交给一个核心内分泌细胞去完成（在实际的生物内分泌系统中是不存在的）。在网络环境中，自主 agent 分布在网络的多个节点上，调控 agent 只部署在与防火墙直接打交道的一个节点上，成为核心内分泌细胞，它负责收集网络的状态信息（激素浓度和激素类型），并根据这些信息直接调控防火墙参数。

3）激素

现代医学认为：在生物体内，激素是生物体受到适宜的刺激后，内分泌细胞分泌产生的一种类蛋白质液体，这种液体通过刺激神经系统的上导纤维管，来对神经系统起到促进和抑制作用。本文把自主 agent 根据网络性能变化而产生的对防火墙的反馈看作激素。当网络的可用性下降时（这说明防火墙控制过严），自主 agent 受到负面刺激分泌抑制激素（负反馈），这些激素对防火墙起到抑制作用（使得防火墙控制策略变为宽松）；当网络的安全性下降时（这说明防火墙控制过松），自主 agent 受到正面刺激，分泌促进激素（正反馈），这些激素对防火墙起到促进作用（使得防火墙控制策略变为严紧）。

网络可用性的定义如下：

$$f_{-}(usability)=\begin{cases}0, usability\leqslant\delta\\1, \text{otherwise}\end{cases} \tag{1}$$

其中，δ 为调控参数，$f_{-}(usability)$ 的值是 0，表示该内分泌细胞分泌的是抑制激素。

usability 的定义如下：

$$usability=1/\sum(packet\_losses/s+packet\_delays) \tag{2}$$

其中，*packet_losses/s* 表示平均每秒丢弃的数据包数；*packet_delays* 表示每个自主 agent 的数据包的平均延时。

网络的安全性的定义如下：

$$g_{+}(security)=\begin{cases}0, security\leqslant\zeta\\1, \text{otherwise}\end{cases} \tag{3}$$

其中，ζ 为调控参数，$g_{+}(security)$ 的值是 0，表示该内分泌细胞分泌的是促进激素。*security* 的定义如下：

$$security=1/\sum(leaks/s+abnormal\_packets) \tag{4}$$

其中，*leaks/s* 表示平均每秒漏检的异常数据包；*abnormal_packets* 表示每个 agent 发现的异常数据包。

4）内分泌系统

生物体的内分泌系统通过调整激素的分泌速率和激素种类对神经系统间接地起到促进和抑制作用。本文把防火墙的参数调控机制看作内分泌系统。下面阐述一下本文设计的调控机制。

普通的防火墙规则一般可以分解为 6 个字段：源地址、目的地址、源端口号、目的端口号、协议和动作，其中，源和目标地址分别代表发送者和接收者的 IP 地址；源端口和目的端口决定服务类型；协议主要为 TCP，UDP，ICMP 等；动作字段定义多为“允许”或“禁止”。如，一条典型的防火墙规则：

-s 193. 110. 96. 0-d 9. 181. 106. 126-p tcp—sport 1024：65535—dport 80-j DROP

传统的防火墙规则是以源主机 IP 地址为匹配的，若某个外部子网内的多个主机同时攻击某个单位内网，则在单位内网的防火墙中就会出现多条规则，造成防火墙规则表的快速膨胀，从而导致规则匹配总时间的增加。因此，若把对某个外部子网内的多个主机的拒绝规则改成对该外部子网的拒绝规则，则可给防火墙规则表瘦身，缩短规则匹配时间开销。例如：把拒绝来自 3 个主机（193. 110. 96. 1，193. 110. 96. 1，193. 110. 96. 1）的数据包的规则修改为拒绝来自外部网络 192. 110. 96. 0 数据包的规则。

-s 193. 110. 96. 1-d 9. 181. 106. 126-p tcp—sport 1024：65535—dport 80-j DROP

-s 193. 110. 96. 2-d 9. 181. 106. 126-p tcp—sport 1024：65535—dport 80-j DROP

-s 193. 110. 96. 3-d 9. 181. 106. 126-p tcp—sport 1024：65535—dport 80-j DROP

把 3 条规则调整为下面的 1 条规则：

-s 193. 110. 96. 0：255. 255. 255. 0-d 9. 181. 106. 126

-p tcp—sport 1024：65535—dport 80-j DROP

内分泌系统的作用就是根据分泌的激素种类和激素浓度来间接地作用于神经系统。在本文中，调控系统根据网络性能变化动态调整防火墙的规则。当促进型激素浓度超出某个阈值（反映的是网络安全性能下降）时，把防火墙规则调控为：由拒绝来自一个主机的数据包到拒绝来自一个外部子网内所有主机的数据包，实际上是把防火墙规则调控为严紧。反过来，当抑制型激素浓度超出某个阈值（反映的是网络可用性能下降）时，把防火墙规则调控为：若把允许来自某个主机的数据包改为允许来自一个外部子网所有机器的数据包，则是把防火墙规则调控为宽松。另外，一般防火墙使用线型列表来储存规则，并且从上至下逐条匹配。为了提高匹配效率，应当对规则排序，基于统计分析把常用的规则放在前面，把基于网络地址的规则放在基于主机地址的规则的前面，这样可以减少防火墙的平均匹配次数。

3 基于内分泌的自适应调控算法

3.1 人工内分泌系统的工作原理

内分泌系统的作用就是根据分泌的激素种类和激素浓度来间接地作用于神经系统。本文设计的人工内分泌系统是基于客户机/服务器模式工作的，分布在网络各个节点中的自主 agent 充当客户机进程，调控 agent 充当服务器进程，如图 1 所示。自主 agent 负责收集网络的状态信息，并定期把自己收集的信息传送给调控 agent；调控 agent 对收到的信息进行综合处理后，根据网络性能变化动态调整防火墙的规则。

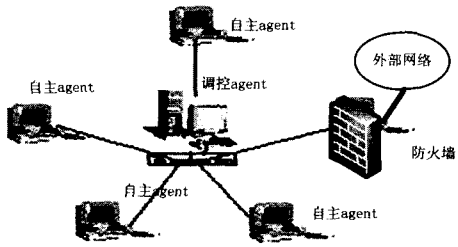

图 1 人工内分泌系统的工作原理

3.2 自适应调控算法

本文设计的基于内分泌的防火墙参数自适应调控算法如图 2 所示。

```
Procedure control_algorithm
Begin
  While (1)
    Begin
      自主 agent_i 计算网络的可用性；
      If (! f_(usability) Then
        分泌 A_i 个单位的抑制激素；
      自主 agent_i 计算网络的安全性；
      If (! g+(sec urity) Then
        分泌 B_i 个单位的促进激素；
      调控 agent 收集激素；
      If (∑_{i=1}^{N} A_i > ∑_{i=1}^{N} B_i Then
          对防火墙参数进行调松操作；
        Else
          对防火墙参数进行调紧操作；
    End
End
```

图 2 算法流程

4 实验

为了验证本文设计的调控机制的性能，组建了试验网络，如图 1 所示。本文使用了费尔实验室提供的防火墙系统开源代码，并在 windows 环境下用 C++进行了重新编译实现。在本文设计的调控系统中，自主 agent 具有网络可用性和网络安全性评估功能，其中可用性评估是用计算测试数据包的平均延迟来衡量的，这个功能使用作者自己开发的一个程序来实现，网络安全性评估功能使用从开源社区下载的 Snort 开源代码加以改进实现。为了模拟外网访问内网的情形，在外网的某个主机上部署了美国加利福尼亚大学 UCI KDD 实验室的 KDD CUP 测试数据集。

实验中参数的设定：选取抑制激素的分泌阈值 $\delta=0.25$，促进激素的分泌阈值 $\xi=0.25$。分 3 种情况进行实验，以测试调控系统中促进激素和抑制激素的浓度变换情况和防火墙参数的调整情况。

1）当网络运行一段时间后，外网以 Nbps 的恒定速率产生定长的正常数据包并发送给内网，此时，内网中促进激素和抑制激素的浓度变换情况如图 3 所示。

图 3 表明，随着时间的推移，内网中抑制激素的浓度逐步增大，而促进激素的浓度逐步减小。事实上，由于大量正常数据包在内网上流动，使内网的负载加重，内网的延迟增大，可用性下降，需要把防火墙调松。这说明实验数据与实际情况是一致的。再查看防火墙的规则表，发现规则表中出现了多个允许来自外部子网数据包的规则，这说明防火墙确实进行了调松操作。

2）当网络运行一段时间后，外网以 Mbps 的恒定速率产生定长的恶意数据包并发送给内网，此时，内网中促进激素和抑制激素的浓度变换情况如图 4 所示。

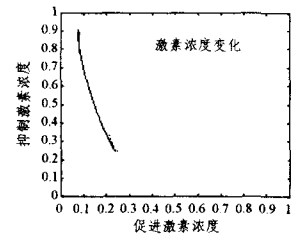

图 3 抑制激素浓度增大趋势

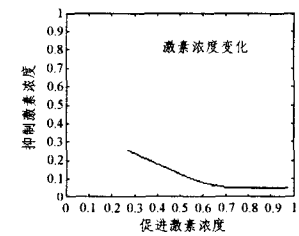

图 4 促进激素浓度增大趋势

图 4 表明，随着时间的推移，内网中促进激素的浓度逐步增大，而抑制激素的浓度逐步减小。事实上，由于大量恶意数据包在内网上流动，使内网的安全性下降，需要把防火墙调紧。这说明实验数据与实际情况是一致的。再查看防火墙的规则表，发现规则表中出现了多个拒绝来自外部子网数据包的规则，这说明防火墙确实进行了调紧操作。

3）当网络运行一段时间后，外网以 NMbps 的恒定速率，以 t 秒间隔依次产生正常包和恶意包，此时，内网中促进激素和抑制激素的浓度变换情况如图 5 所示。

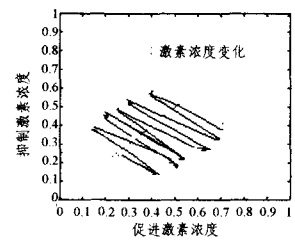

图 5 两种激素浓度增大趋势

图 5 表明，内网中促进激素浓度和抑制激素浓度呈螺旋状变化，一种激素浓度的增长伴随着另一种激素浓度的下降。事实上，由于大量正常数据包和恶意数据包的间隔出现，内网可用性和安全性也出现间隔性变化，防火墙也需要不断地调整。以 t 秒间隔抽查防火墙规则表，发现规则表中的规则按照相同的时间间隔不断地进行调紧调松操作。若时间间隔 t 设置过小，防火墙参数调整过于频繁，会导致防火墙系统的性能下降。

上述 3 组实验的结果表明，防火墙能够根据网络可用性和安全性的变化自适应地调整参数。这说明本文基于内分泌机制设计的调控系统是有效的。实验是在模拟超恶劣环境下

进行的，如模拟网络瞬间遭到大量的恶意数据包攻击，若本文设计的调控系统工作在真实的网络环境中，性能表现会比实验时要好一些。

结束语 本文基于内分泌系统构造的调控系统，仅用两种激素模拟网络的可用性和安全性，事实上，正如内分泌系统本身的复杂性一样，网络的性能评价也是很复杂的。这就需要：一方面对内分泌系统进行更深入的研究，从而提取更优的调节机制用于设计基于内分泌系统的调控算法；另一方面对网络性能评估指标进行深入的研究，设计出更科学合理的评价方案。这些都是本文下一步研究的重点。

参 考 文 献

[1] 刘克龙，蒙杨，卿斯汉. 一种新型的防火墙系统[J]. 计算机学报，2000，23(3)：231-236

[2] 张磊，卿斯汉. 一个基于 Agent 的防火墙系统的设计与实现[J]. 软件学报，2000，11(5)：642-645

[3] Fulp E W. Optimization of network firewall policies using ordered sets and directed a cyclical graphs，1265[R]. Winston-Salem，USA：Wake Forest University，2004

[4] Chen Wenhui，Wang Weiping，Li Zhepeng. Dynamic update of firewall policy based on MFDT，2375 [R]. Winston-Salem，USA：Wake Forest University，2006

[5] 张万会，王复周. 神经、免疫及内分泌系统间的关系[J]. 生理科学进展，1993(03)：261-268

[6] 陈得宝，赵春霞. 基于内分泌调节机制的粒子群算法[J]. 控制理论与应用，2007，24(6)：1005-1010

[7] 刘宝，丁永生，王君红. 一种基于内分泌超短反馈机制的智能控制器[J]. 计算机仿真，2008，25(1)：188-191

[8] 王祎，陈为栋，顾幸生. 基于内分泌激素调节机制的免疫算法的 Flowshop 调度问题[J]. 系统仿真学报，2008，20(13)：3425-3430

[9] 逄曙光. 内分泌与代谢病的免疫学发病机制研究[D]. 济南：山东大学，2008

(上接第 203 页)

息。文献[7]把过去共现过的命名实体都扩充进模型，门槛过低，导致引入大量噪音，淹没了真正的话题相关信息，很难再对扩充后的报道话题关联性进行判断。

2)扩充信息种类：通过实验[11]发现，表示模型转换过程中若有信息丢失，系统性能会受损。例如："印度古吉拉特邦发生地震"，其中"地震"不是实体词却与该话题关系密切。文献[7]的方法只使用命名实体表示报道，是方法性能较差的原因之一。文献[5]及本文所提方法都使用整个相关报道进行扩充，后者还对扩充信息做了进一步分析。

3)扩充信息量：扩充信息并不是越多越好，相反扩充信息越多，噪音就越多，尤其是相关信息较少的情况下，加大扩充信息量会带来系统性能的急剧下降[7]。本文方法使用最新话题相关报道扩充的性能就优于使用所有话题相关报道进行扩充[8]，但对信息累积精化却带来了性能损失。在实现文献[5]中方法时也对扩充报道个数进行了实验，实验表明随着扩充报道个数的增多，系统性能逐渐下降。

结束语 针对报道表示中存在的稀疏问题和话题动态演化问题，本文提出一种信息动态扩充方法，并对模型中的核心特征、名实体特征、依存名词特征 3 类信息进行精化，以进一步改进话题表示模型。该方法用于改进话题关联识别研究中的报道表示。实验表明，无论是动态扩充方法还是 3 种特征精化都能够较好地改进系统性能，尤其是扩充技术和核心特征精化对降低误判率和丢失率都有较大的影响，是改进识别效果的两个有效途径。同时发现，对重要信息的定位及组合精化策略都还需要进一步研究。在知识表示方面也要认识到：虽然使用了比词法更进一步的句法知识，但仍然停留在较浅的知识层次，应该进一步挖掘。

参 考 文 献

[1] James A，et al. Introduction to Topic Detection and Tracking in Topic Detection and Tracking：Event-based Information Organization[M]. Kluwer Academic Publishers，2002：1-16

[2] Wayne，Charles L. Topic Detection and Tracking (TDT)：Overview & Perspective[C]// Proceedings of the Broadcast News Transcription and Understanding Workshop. Lansdowne，Virginia，1998

[3] Margaret C，Ao F，Giridhar K，et al. UMass at TDT 2004[C]// Proceedings of the 7th Topic Detection and Tracking (TDT2004). Gaithersbury，2004

[4] Francine C，Ayman F，Thorsten F. Multiple Similarity Measures and Source-Pair Information in Story Link Detection[C]// HLT-NAACL 2004. Boston，2004：313-320

[5] Victor L，James A，Edward D，et al. Relevance models for topic detection and tracking[C]// Proceedings of Human Language Technology Conference (HLT). California，2002

[6] Ramesh N. Semantic language models for topic detection and tracking[C]// Proceedings of the HLT-NAACL 2003 student research workshop. Edmonton，2003

[7] Chirag S，Bruce C W，David J. Representing documents with named entities for story link detection (SLD)[C]// CIKM 2006. Virginia，2006

[8] Zhang Xiaoyan，Wang Ting，Chen Huowang. Stroy Link Detection based on Dynamic Information Extending[C]// Proceedings of the International Conference on The Third International Joint Conference on Natural Language Processing (IJCNLP2008). Hyderabad，2008

[9] Thorsten B，Francine C，Ioannis T. Topic-Based Document Segmentation with Probabilistic Latent Semantic analysis[C]// Proceedings of the International conference on Information and Knowledge Management (CIKM). McLean，2002

[10] The 2003 Topic Detection and Tracking (TDT2003) Task Definition and Evaluation Plan [OL]. http：// www. nist. gov/speech/tests/tdt/tdt2003/evalplan. htm

[11] 张晓艳，王挺，陈火旺. 基于 SVM 的多向量文本表示模型话题关联识别研究[C]// 第七届中文信息处理国际会议. 武汉，2007