一种基于浓度调节的 RCB 算法

邵国金[1],景伟娜[1],吴春颖[2]
SHAO Guo-jin[1],JING Wei-na[1],WU Chun-ying[2]

1.河南城建学院 计算机科学与工程系,河南 平顶山 467001
2.中国人民武装警察部队学院,河北 廊坊 065000
1.Department of Computer Science,Henan Institute of Architecture and Technology,Pingdingshan,Henan 467001,China
2.The Chinese People's Armed Police Forces Academy,Langfang,Hebei 065000,China

SHAO Guo-jin,JING Wei-na,WU Chun-ying.Novel self-adapting *r* continuous bits matching algorithm based on hormone concentration.Computer Engineering and Applications,2009,45(23):104-106.

Abstract: The omission factor and fallout ratio of IPS are affected by matching algorithm,and dynamic matching algorithm of *r*-adjustable pay an import role in improving performance of IPS.The mechanism of endocrine system is analyzed,a dynamic balance model of hormone concentration is designed,and a dynamic matching algorithm based on hormone concentration is constructed.Simulation tests show that the matching algorithm given in this paper has advantages of good performance,and applying prospect.

Key words: hormone;concentration adjusting;Intrusion Prevention System(IPS);RCB matching algorithm

摘 要: 决定入侵防御系统漏检率和误检率的关键要素是模式匹配算法,改进 *r* 连续位匹配算法可以提高入侵检测系统的性能。受生物内分泌系统通过激素浓度调节适应内外环境机制的启发,设计了人工激素浓度的动态平衡模型,构造了一种基于浓度调节的 RCB 算法。实验数据表明,该算法能够根据网络状态变化,自适应地调整匹配参数,具有较好的应用价值。

关键词: 激素;浓度调节;入侵防御系统;*r* 连续位匹配算法
DOI: 10.3778/j.issn.1002-8331.2009.23.029 **文章编号:** 1002-8331(2009)23-0104-03 **文献标识码:** A **中图分类号:** TP18

匹配算法是影响入侵防御系统(IPS)性能的关键因素。字符串的模式匹配是一种重要的串运算,串匹配算法中最经典的算法是 KMP 算法和 BM 算法以及对 BM 的改进算法。这些算法在解决简单的模式匹配问题时有一定的优越性,但当应用于在线入侵检测的时候,特别是在处理入侵规模 N 值较大,编码串长 M 值较大时,算法会过于缓慢且需要较大的空间,甚至会造成计算瓶颈现象。

Forrest 等人在研究了 IPS 中检测器工作的内在机制后,提出了 *r* 连续位匹配算法,该算法的基本思想是[1]:对长度为 l 的串 T 和长度为 r 的模式串 P, 若串 T 中存在 r 连续的位构成的子串与模式串 P 相同,则称与模式串 *r* 连续位匹配。*r* 连续位匹配算法在 IPS 中得到了广泛应用。但随着网络带宽的加大和网络流量的剧增、网络攻击强度和频度的加大,基于 *r* 连续位匹配算法的 IPS 面临着两方面的挑战:一方面,若 *r* 过大,当网络流量突增时,由于来不及处理大量的待检测数据,造成 IPS 的漏检率较高,给企业的网络安全带来隐患;另一方面,若 *r* 过小,IPS 匹配了大量良性数据,造成 IPS 的误检率偏高。由此可知,*r* 值大小的设定成为决定匹配效果的一个关键参数。网络流量和网络遭受攻击情况都是动态的,固定的 *r* 连续位匹配算法很难适应。如何根据网络状态动态调整 *r* 的值,建立一个动态自适应的 *r* 连续位匹配算法, 已经引起国内外学者的关注,出现了一些动态匹配算法[2-4],但是这些动态匹配算法在 IPS 中应用效果不太理想。

生物体的内分泌是一个非常复杂的系统,对整个机体的生长、发育、代谢和生殖起着调节作用。内分泌系统含有成千上万的激素产生细胞(内分泌细胞),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, 使个体的运动向着有利于自己适应性的方向,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,AES),现已经发展成为一种解决复杂工程问题的有效方法,在工程领域得到了初步的应用[5-7]。基于此,借鉴内分泌系统中激素浓度自适应调控机制,提出了一种基于内分泌激素浓度调节的动态 *r* 连续位匹配算法。

1 内分泌激素浓度的动态平衡模型

生物内分泌系统由内分泌细胞、内分泌细胞所释放的激素

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60575038,No.60502046)。
作者简介: 邵国金(1959-),男,副教授,主要研究方向为计算机应技术、网络安全、人工智能、计算智能等;景伟娜(1977-),女,讲师,主要研究方向为网络安全、自然计算、计算智能等;吴春颖(1977-),女,讲师,主要研究方向为计算机应用技术。
收稿日期: 2009-04-09 **修回日期:** 2009-06-12

(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[8]。

生物体中,某中激素浓度过高或过低,都会给生物体的健康带来伤害(例如,生长激素与骨的生长有关,幼年时期若如缺乏,则使长骨的生长中断,形成侏儒症;若过剩,则使全身长骨发育过盛,形成巨人症),值得庆幸的是,生物体内各种激素的浓度是动态变化的、自适应调整的。一种激素浓度过高时,机体会分泌另一种对该激素起抑制作用的激素,抑制该激素的分泌;另一方面,当一种激素的浓度偏低时,机体内对该激素起促进的作用的激素迅速产生,促进该激素的分泌。

设激素 i 在 t 时刻的浓度为 $f_i(t)$,则激素 i 浓度的动态平衡方程如下:

$$f_i(t+1)=w_i(t)\cdot f_i(t)+\delta\sum_{j=1}^{N-1}g_{ij}\cdot f_j(t)\cdot f_i(t) \tag{1}$$

其中,$w_i(t)$是激素 i 的分泌细胞在 t 时刻受到刺激强度,$w_i(t)\in(0,1)$;$f_j(t)$是激素 j 在 t 时刻的浓度;δ 是其他激素对激素 i 的制约系数,$\delta\geqslant 0$;g_{ij}是激素 j 对 i 的作用系数,定义如下:

$$g_{ij}=\begin{cases}1, j \text{ is stimulative for } i\\ 0, j \text{ is adiaphorous for } i\\ -1, j \text{ is inhibitory for } i\end{cases} \tag{2}$$

激素 j 对激素 i 起促进作用时,g_{ij}取 1;起抑制作用时,g_{ij}取-1;若无作用,则 g_{ij}取 0。

生物体内各种激素的浓度按照动态平衡方程动态调整,达到机体内外环境的平衡。这是一个自适应的动态平衡过程。这种自适应、自调节的机制,为后面设计的自适应 r 连续位匹配算法提供了很好的生物学支持。

2　基于激素浓度调节的动态 r 匹配算法

2.1　r 的动态平衡方程

网络系统和生物体系统都可看作是信息处理系统,IPS 和内分泌系统所扮演的角色具有惊人的相似性。生物体的内分泌子系统根据机体内环境情况动态调整各种激素的浓度,维持机体平衡。IPS 系统作为网络系统的一部分,也要根据网络的状况进行动态调整,为网络的安全提供保障。基于此,借鉴内分泌系统的激素浓度调节机制,提出动态 r 连续匹配算法,其思想为:r 值大小并不是一成不变的,而是根据网络自身的运行状态动态进行调整。

影响匹配性能的要素主要有:网络流量、网络攻击强度和安全需求强度。下面分别阐述各个要素对 r 值调整的作用。

(1)网络流量

网络流量,简言之就是网络上传输的数据量。网络流量具有突变性和非均匀分布性等特点。网络流量对 r 值调整影响很大,当网络流量增大时,若 r 值不变,则 IPS 因无能力处理数据不得不放弃对一些数据的检测,导致漏检率上升,对网络安全性带来危害。网络流量增大时,应把 r 值调小,使得 IPS 提高匹配速度,有能力处理较多的数据。当然,r 值不能调的过小,否则 IPS 区分正常数据和恶意数据的能力下降,导致误检率上升。

(2)网络攻击强度

网络攻击强度指网络正在遭受的来自外网的攻击数据数量和危害程度。当网络攻击强度较大时,应把 r 值调整较大些,这样就使得 IPS 有更好的能力区分正常数据和恶意数据,从而降低 IPS 误报率。同样的道理,r 值不能调的过大,否则 IPS 区分正常数据和恶意数据的能力下降,导致误检率上升。

(3)安全需求强度

一般来说,一个企业对网络的安全需求强度是不变的,但在特殊情况下(如单位发生紧急事件),安全需求也可能动态变化。这些变化也应该反映到 IPS 中。

基于上面对内分泌系统浓度自适应调节机制的分析,不妨把网络流量、网络攻击强度和安全需求强度均作为激素对待,把 r 也看作一种激素,r 值大小的调整看作激素浓度的调整。设在 t 时刻 r 的值为 $r(t)$,根据上面建立的激素浓度动态平衡模型,则在 $t+1$ 时刻 r 的动态方程为:

$$r(t+1)=w(t)\cdot r(t)+\delta\sum_{j=1}^{N-1}g_{ij}\cdot f_j(t)\cdot r(t) \tag{3}$$

在 IPS 系统中,在不同时刻 r 的变化只与网络流量、攻击强度和安全需求有关,根据这 3 种激素浓度对 r 的影响,可把式(3)简化为:

$$r(t+1)=[1-\alpha\cdot f_1(t)+\beta\cdot f_2(t)+\eta\cdot f_3(t)]\cdot r(t) \tag{4}$$

其中,$f_1(t)$表示在 t 时刻流量浓度,α 是浓度对 r 的制约权重;$f_2(t)$表示在 t 时刻攻击浓度,β 是攻击浓度对 r 的制约权重;$f_3(t)$表示在 t 时刻安全需求浓度,η 是安全需求浓度对 r 的制约权重。

需要说明的是:在 IP 网络中,以 32 位源、目的 IP 地址、16 位端口及 16 位协议标志构成的 l=96 的二进制串表示网络数据包的属性(正常、恶意),IPS 的作用就是区分正常数据和恶意数据。理论上,IPS 的匹配的连续位数 $r\in[1,96]$,但是若 r 的值过小导致 IPS 漏检率过大,r 的值过大导致 IPS 误漏检率过大,r 的值应在 15 与 45 之间[7]。因此,$[1-\alpha\cdot f_1(t)+\beta\cdot f_2(t)+\eta\cdot f_3(t)]\in[0,3]$。

2.2　r 的动态调控模型

内分泌系统中激素是动态调整的,激素之间的相互制约为生物体内外环境的平衡通过了保障。构建基于激素浓度调节的 r 值自适应调控环境如下:基于客户机/服务器工作模式,分布在网络各个节点中的多个担任不同功能的自主 agent 充当客户机进程,调控 agent 充当服务器进程,如图 1 所示。流量监控 agent 负责收集网络的流量状态信息;攻击监控 agent 负责收集攻击数据和危害程度;安全策略 agent 负责收集网络的安全需求变化和安全策略变化攻击数据和危害程度,这些 agent 定期把自己收集的信息传送给调控 agent;调控 agent 综合收集的信息动态调整 IPS 的 r 值。r 的动态平衡模型如图 1 所示。

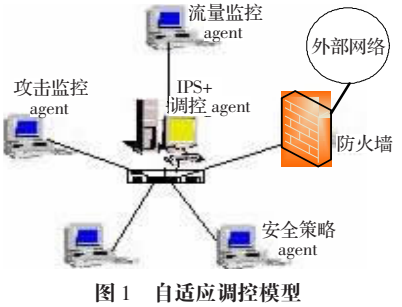

图 1　自适应调控模型

3　仿真实验

为了验证设计的调控机制的性能,组建的试验网络如图 1 所示。使用了开源社区下载的 Snort 开源代码,并在 linux 环境下进行了重新编译。在 C++环境下,自主开发了流量监控 a-

gent、安全策略 agent 和攻击监控 agent。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[9]。

为了验证设计的算法性能，依次进行了 2 组实验，每组实验分别做 5 次取平均结果：

(1)在时刻 t=0 至 t=60 这 60 秒内，每隔 5 秒利用测试数据集随机向测试网络注入测试数据，测试数据中正常数据和恶意数据的比为 100:1，依次测试固定 r 连续位匹配算法下的平均漏检率和误检率，取 $\alpha=\beta=\eta=1$ 和 $\alpha=1,\beta=\eta=0.5$ 动态 r 连续位匹配算法下的平均漏检率和误检率，实验结果如表 1 所示。

表 1 网络流量浓度动态变化下的实验结果

两种匹配算法的参数	平均漏检率/(%)	平均误检率/(%)
固定 r 连续位匹配算法(r 取固定值 25)	3.3	1.5
动态 r 连续位匹配算法(r 动态调整)($\alpha=\beta=\eta=1$)	1.6	0.9
动态 r 连续位匹配算法(r 动态调整)($\alpha=1,\beta=\eta=0.5$)	0.8	0.7

表 1 表明，在网络流量浓度动态变化的情况下，固定 r 连续位匹配算法下的平均漏检率和误检率分别为 3.3%和 1.5%，比动态 r 连续位匹配算法下的数据都要高很多。调大 α 在方程(4)中所占的权重，漏检率得到了明显的下降(从 1.6%下降为 0.8%)，这是因为此时网络流量激素浓度对 r 的影响较大，调大该激素权重，有利于 r 的动态适应。

②在时刻 t=61 至 t=120 这 60 秒内，每隔 5 秒随机向测试网络注入测试数据，测试数据中正常数据和恶意数据的比为 100:50，依次测试固定 r 连续位匹配算法下的平均漏检率和误检率，取 $\alpha=\beta=\eta=1$ 和 $\alpha=0.5,\beta=\eta=0.5$ 动态 r 连续位匹配算法下的平均漏检率和误检率，实验结果如表 2 所示。

表 2 攻击强度浓度动态变化下的实验结果

两种匹配算法的参数	平均漏检率/(%)	平均误检率/(%)
固定 r 连续位匹配算法(r 取固定值 25)	2.3	2.7
动态 r 连续位匹配算法(r 动态调整)($\alpha=\beta=\eta=1$)	1.6	1.1
动态 r 连续位匹配算法(r 动态调整)($\alpha=0.5,\beta=1,\eta=0.5$)	1.2	0.5

表 2 表明，在攻击流量不断变化的情况下，固定 r 连续位匹配算法下的平均漏检率和误检率分别为 2.3%和 2.7%，比动态 r 连续位匹配算法下的数据都要高很多。调大 β 在方程(4)中所占的权重，误检率得到了明显的下降(从 1.1%下降为 0.5%)，这是因为此时攻击流量激素浓度对 r 的影响较大，调大该激素权重，有利于 r 的动态适应。

上述二组实验的结果表明，根据网络流量和遭受攻击强度的变化动态调整匹配连续位参数 r 的值，可以降低 IPS 的漏检率和误检率。这说明该文设计的基于激素浓度调节机制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。

4 结束语

动态 r 连续位匹配算法的性能决定 IPS 的漏检率和误检率，基于内分泌激素浓度构造的调控系统，可以根据网络情况动态自适应地调整 r 的数值，取得了较好的匹配效果。模拟内分泌激素浓度调节机制稍显简单，若能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，可以进一步提高匹配算法的性能，这是下一步研究的重点。另外，设计的动态匹配算法，在病毒检查、在线故障诊断等工程领域具有广泛的应用价值。

参考文献：

[1] Forrest S.Immunology as information processing[M].New York:Oxford University Press,2000.
[2] 李涛.一种基于免疫的动态入侵检测模型[J].科学通报,2005,50(17):1912-1919.
[3] 罗文坚,曹先彬,王煦法.检测器自适应生成算法研究[J].自动化学报,2005,31(6):907-916.
[4] 马莉,刘凤玉.一种改进的网络入侵检测器生成算法[J].计算机工程与应用,2007,43(21):150-152.
[5] 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.
[6] 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.
[7] 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 Flow shop 调度问题[J].系统仿真学报,2008,20(13):3425-3430.
[8] 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.
[9] KDDLib[EB/OL].[2009-03-02].http://kdd.ics.uci.edu/kddlib/kdd_up/.

(上接 69 页)

由式(23)得出，经过 24 周的测试之后，软件的剩余错误为：

$$N_c(t)=\approx 2\,247e^{-0.017\,9t^{1.635}}$$

由此可以预测，软件的错误总数为 2 247 个，在第 24 周测试结束后，软件的剩余错误数预计有 85 个。如果软件在第 24 周后交付使用，其失效率为 19.015 个/周，MTBF(软件平均寿命)为 0.052 6 周，如果要求软件交付使用后的失效率小于 0.6 个/周，则测试时间应大于 38.5 周，即要继续进行 14.5 周的测试才能满足要求。

参考文献：

[1] 黄锡滋.软件可靠性、安全性与质量保证[M].北京:电子工业出版社,2002.
[2] 刘志方,钟德明,曾福萍,等.软件可靠性测试的理论分析[J].测控技术,2008,27(10):62-64.
[3] 蔡开元,董昭,刘克.关于软件可靠性测试的若干问题[J].工程数学学报,2008,12(6):967-978.
[4] 于碧媛.软件可靠性模型及估值的介绍[J].导弹与航天运载技术,1994(1):55-60.
[5] 韦博成.近代非线性回归分析[M].南京:东南大学出版社,1989.
[6] Patton R.软件测试[M].北京:机械工业出版社,2003.
[7] 吴彩华,朱小冬,刘俊涛,等.基于可靠性增长模型的软件可靠性增长测试充分性准则[J].计算机科学,2008,35(11):281-283.