文章编号:1001 - 9081(2005)08 - 1726 - 04

免疫学原理在入侵检测中的应用研究

赵林惠[1],戴亚平[1],徐立新[2]
(1. 北京理工大学 信息科学技术学院,北京 100081; 2. 北京理工大学 机电工程学院,北京 100081)
(dtree_mail@sohu.com)

摘 要:评述了近年来免疫原理在入侵检测中的应用,着重讨论抗体克隆选择学说、免疫网络学说和危险模式理论的应用研究现状,对所采用的免疫机制和算法进行了总结,并介绍了研究的热点问题。最后在分析入侵检测方法存在问题的基础上,探讨了今后研究的方向。

关键词:人工免疫系统;入侵检测;阴性选择;独特型网络;危险模式理论
中图分类号: TP393.08 **文献标识码**:A

Immunology principle and its application investigation in intrusion detection

ZHAO Lin-hui[1], DAI Ya-ping[1], XU Li-xin[2]
(1. *School of Information Science and Technology, Beijing Institute of Technology, Beijing* 100081, *China*;
2. *School of Mechatronic Engineering, Beijing Institute of Technology, Beijing* 100081, *China*)

Abstract: Immunology principle and its application in intrusion detection in recent years were reviewed. The research on antibody clonal selection theory, immune networks and danger theory were emphasized. The immune principle and algorithms used in intrusion detection were summarized and some of hotly debated issues were introduced. Based on the disadvantages of current methods, the development directions were discussed.

Key words: artificial immune system; intrusion detection; negative selection; idiotypic netwoks; danger theory

0 引言

近年来,随着计算机技术的不断发展,网络规模的不断扩大,系统遭受的入侵和攻击越来越多,网络与信息安全问题变得越来越突出。因此,检测与防范入侵攻击,保障计算机系统、网络系统及整个信息基础设施的安全已经成为刻不容缓的重要课题。对于入侵检测系统(Intrusion Detection System, IDS)来说,入侵的检测和分析模块是系统的核心。由于传统的入侵检测模型存在种种缺陷,而生物免疫系统(Human Immune System ,HIS)可以保护人类机体不受诸如病菌、病毒等各种病原体的侵害,且表现出了分布式保护、多样性、自组织、健壮性等良好特性,这种与入侵检测系统的功能和特性上的惊人相似,引起人们的高度重视,使得借鉴生物免疫系统的原理开发入侵检测技术成为研究热点。

本文针对近年来免疫原理在入侵检测中的应用以及新出现的免疫理论进行总结,指出了一些值得深入研究的问题。迄今为止,在免疫学中比较成熟并占主导地位的学说有:抗体克隆选择学说和免疫网络学说[1],基于这两个学说提出的人工免疫系统(Artificial Immune System, AIS)模型大致分为网络模型和阴性选择模型[2],其中后一模型广泛用于建立入侵检测系统[1],涉及到的免疫机制包括免疫应答、免疫系统的特异识别、模式识别、克隆选择和扩增、免疫记忆和自体与非自体区分等。最近,随着生物免疫学的丰富和完善,一个新的免疫理论——危险模式理论已经引起人们的兴趣,它在入侵检测中的应用研究已经展开。

1 抗体克隆选择学说

抗体克隆选择学说的要点是:外来抗原选择出原先处于静止状态的互补细胞克隆,被选择细胞克隆的激活、增殖和效应功能是免疫应答的细胞学过程,而针对自身抗原的细胞克隆则被抑制或消除,因而对外来抗原的识别是关键的因素[1]。基本原理是:免疫细胞即抗体的主要工作是区分本体细胞和异体细胞,其中本体细胞指的是正常的人体细胞;异体细胞是指有害的异质细胞,也称为抗原。每个淋巴细胞能够通过绑定机制来识别一定数量结构相似的抗原细胞。人体的淋巴细胞主要是在骨髓和胸腺部位产生的,在骨髓和胸腺部位存在着淋巴细胞的候选基因库,淋巴细胞就是通过从这些基因库中随机选择一些基因片断并进行组合来产生的。由于过程的随机性,产生的淋巴细胞就有可能将本体细胞误认为异体细胞,这是不允许的。为了减少这种可能性,随机产生的淋巴细胞在离开骨髓和胸腺之前必须经过一个阴性选择过程。在阴性选择过程中,淋巴细胞接触到大量的本体细胞,如果淋巴细胞能够绑定任一本体细胞,那么该淋巴细胞就被认为是无效而被删除。通过了否定选择过程的淋巴细胞,则认为它们是成熟的,会从骨髓和胸腺中释放出来,进入到血液中从事抗原检测任务。如果在有限的时间内能够绑定到数量超过某一阈值的抗原,那么淋巴细胞就会被启动以杀死抗原;反之,如果在一定的时间内没有被启动,那么该淋巴细胞就会死亡而代之以新的细胞。在淋巴细胞被启动之后,它进入克隆选择阶段。在这个阶段,被启动的淋巴细胞产生多个本身的

收稿日期:2005 - 01 - 18;**修订日期**:2005 - 04 - 29
基金项目:兵器预研支撑基金资助项目(YJ0467011);教育部留学回国人员科研基金资助项目(LXKYJJ200308)
作者简介:赵林惠(1973 -),女,河北沙河人,博士研究生,主要研究方向:网络入侵检测、神经网络; 戴亚平(1963 -),女,山东人,博士生导师,主要研究方向:智能系统、网络安全、信息融合; 徐立新(1969 -),男,黑龙江省人,副教授,博士,主要研究方向:目标探测与控制、模式识别与智能系统.

复制,这些复制细胞可以称为记忆细胞。记忆细胞与一般的淋巴细胞相比具有较小的阈值和较长的生命周期。这样,当人体中出现以前出现过的抗原时,这些复制细胞可以加速抗原的识别过程,从而保证了人体免疫系统的有效性。

基于此原理的入侵检测技术通常要建立一个检测子集合用以匹配抗原,检测子是随机生成的,然后要经过阴性选择阶段得到成熟的检测子。这些成熟的检测子监视网络中的数据,如果匹配到的异常超过预先设定的阈值,检测子则被激活,这时会向人工操作员报告并由其决定这是否是一次真正的入侵,如果是,检测子将提升为记忆检测子。一般将正常的用户或者网络行为视为自我,其他的都视为非我。构成自我和非我的基因用代表行为的属性来表达,例如 TCP SYN 请求包中的源 IP 地址、目的 IP 地址和服务端口等属性。检测子与抗原的特异性互补结合用匹配算法来实现[3]。

已提出的入侵检测系统主要有三个:1)美国 University of New Mexico 的 Forrest,Hofmeyr 小组提出的基于免疫耐受机制(阴性选择)的模型;2)美国 University of Memphis 的 Dasgupta 小组提出的基于多 Agent 的模型;3)英国 University College London 的 Kim ,Bentley 小组提出的主从 IDS 模型。这三个系统的详细描述和分析可以参考文献[4]。

这三个模型提出后,有关克隆选择和阴性选择机制在入侵检测中的应用研究多集中于对这三种模型的改进和完善,主要研究热点在以下四个方面:

阴性选择算法　文献[5]研究了阴性选择在用于网络入侵检测的 AIS 中的作用,实验结果证明在将该算法用于实际的网络流量数据时,存在严重的规模问题,即生成足够数量检测子所需的计算时间过长,根本无法用于实际的 IDS,从而得出结论:阴性选择最适合用作无效检测子的筛选器而不是成熟检测子的生成方法。实际上在这之前,他们已意识到由于随机生成检测子而带来的过量计算问题,因而提出了基于小生境的阴性选择算法[6]。另外,作为阴性选择方法的扩展,文献[7]提出了阴性表征(Negative Characterization,NC)方法,同时给出了阳性表征(Positive Characterization,PC)方法作为比较,实验结果表明尽管 PC 方法更准确,但需要过多的时间和空间资源,因此用 NC 方法产生检测子是可行的。

克隆选择算法　Kim 小组提出的主从 IDS 模型包含三个不同的阶段:阴性选择、克隆选择和基因库进化。为了克服单纯采用阴性选择算法存在的规模问题,文献[8]提出了一种带有阴性选择操作符的静态克隆选择算法,但这种方法却不适于不断变化的环境,因此又提出了动态克隆算法,不但可以学习变化的自我行为,而且可以预测非我的新模式[9,10]。

基因的表达方法　最初的阴性选择算法中基因的表达都采用二进制字符串,但逐渐发现这种方法不能准确表达亲和力间的关系[11],因此出现了采用实值表达基因的方法,如文献[12]提出的随机实值阴性选择算法,不但克服了上述问题而且能够很好地估计覆盖非我空间所需的检测子的最佳个数。将这种方法用于异常检测,实验结果证明优于二进制表达方法[13]。

匹配规则　文献[14]指出大部分匹配函数可分成两类:距离和相似性,距离测量两个序列之间的区别有多大,而相似性测量它们之间的相似程度。文献[14]对距离或相似性的统计、物理和二进制度量共 12 种匹配规则进行了研究和比较。多数 AIS 采用海明距离或 r 连续位匹配规则[15,16],但是当字符串长度增加时,r 连续位匹配规则表现出匹配的效率低、给定覆盖率时所需的检测子的个数过多等缺点。基于此提出一种改进的 r 连续位匹配规则,也可以说是 r 连续位匹配规则的一种简化:r-contiguous templates,简称为 r-chunks。这种规则的优点在于不但能够解决上述问题,而且更易于进行数学分析[17]。

总的来说,基于阴性选择或自我—非我识别的 IDS 都存在误报较多以及难以用于实际的大型网络即规模问题等缺点,即使进行了一系列的改进和完善,要想根本解决问题必须求助于其他理论。

2　免疫独特型网络理论

免疫网络学说,主要是独特型网络假设,最早由文献[18]提出,其理论基础是抗体既能够识别抗原,也能够识别其他抗体。一般抗原具有能够被抗体识别并结合的抗原决定位,抗体除具有能够识别抗原的抗体决定位外,还具有能够被其他抗体识别并结合的抗原决定位,一般称为独特型抗原决定位,或称独特位。因此抗体由抗体决定位和独特位组成。这样,抗体可以识别另外的抗体,反过来又会被其他抗体识别,这种激励作用持续传播开来使得整个系统形成一种网络结构。抗体和抗原之间以及抗体和抗体之间的相互作用构成了独特型网络的调节机制。文献[19]对这种独特型作用进行了详细分析,认为这有助于解释对曾经发生过感染的记忆保持机制;此外,相似抗体间的抑制作用有利于保持系统中抗体的多样性。文献[20]提出了一个独特型作用公式:

$$\frac{dx_i}{dt} = c\left[\binom{antibodies}{recognised} - \binom{Iam}{recognised} + \binom{antigens}{recognised}\right] - \binom{death}{rate}$$

$$= c\left[\sum_{j=1}^{N} m_{ji}x_i x_j - k_1\sum_{j=1}^{N} m_{ij}x_i x_j + \sum_{j=1}^{n} m_{ji}x_i y_i\right] - k_2 x_i \qquad (1)$$

其中,N 是抗体的个数;n 是抗原的个数;x_i(或 x_j)是抗体 i(或 j)的浓度;y_j 是抗原 j 的浓度;c 是比例常数;k_1 是抑制作用,k_2 是死亡率;m_{ji} 是抗体 i 和抗体(或抗原)j 之间的匹配函数。公式中第一项是抗体识别其他抗体受到的激励作用;第二项是抗体被其他抗体识别受到的抑制作用;第三项是抗体识别抗原受到的激励作用。从中可看出独特型相互作用的性质可以是正的也可以是负的,即抗体的数量既可能增大也可能减少。实际应用中,可根据具体情况将公式做相应简化[19]。

最近,有关独特型网络理论的应用多见于优化问题如多模态函数优化、推荐系统以及建立免疫响应模型。

文献[21]的模拟抗体搜索机制,结合独特型网络调节理论,提出一种新的函数优化算法。该算法用抗体表示函数优化解的可能模式,用抗原表示待求解的多模态函数的各局部最优和全局最优模式,通过构造克隆选择算子(包括高变异克隆和优选两种操作)完成全局和局部最优解的搜索,利用 B 细胞网络保持多种抗体并存。

推荐系统是那些使用协同过滤技术来产生预测或者推荐结果的系统,目前多用于电子商务系统中,即模拟商店销售人员向用户提供商品推荐,帮助用户找到所需商品,从而顺利完成购买过程。Nottingham 大学的 Uwe Aickelin 小组提出了一种人工免疫系统推荐模型[22],将已知用户的偏好作为抗体而将待匹配的新偏好作为抗原。他们假设当能够提供良好匹配的抗体的浓度随时间不断增长时,应该将其作为具有良好匹配的抗体的子集而结束,也就是说,系统的目标不在于找到获得最佳匹配的抗体(即最优化问题),而在于找到一个抗体的集合,这些抗体具有相近的匹配而在同一时间彼此截然不同。

他们将此模型用于电影推荐系统中，抗原与抗体之间的相互作用用于匹配，而抗原与抗原之间的相互作用用于保持多样性。匹配算法采用的是最邻近算法（k-Nearest-Neighbour algorithm）。随后又将该模型进一步推广用于网站推荐系统[23]。除此之外，他们还对亲和力测量算法进行研究，比较了两种用于计算相关系数的方法：Kendall tau 和加权 Kappa 算法。比较结果说明加权 Kappa 算法更适合于电影推荐系统，同时证明该系统具有鲁棒性，只要选择合适的亲和力测量算法，结果都不错[24,25]。

文献[26]根据人类免疫系统的免疫响应过程和免疫学中独特型免疫网络的假设，从免疫响应中 B 细胞的结合、B 细胞的激励以及复制等方面入手，提出了建立人工免疫响应网络模型的新思想。其中抗体和抗原都是用一定长度的二进制字符串表示，抗体的性质由抗体的表位和对位共同决定，表位决定了其抗原性，对位决定了其结合抗原的性质。抗体可以表示为(p,e)，p 为其对位，e 为其表位。匹配算法采用的是测量两个字符串的 *Hamming* 距离，若该距离小于一个预先设定的阈值，则认为二者匹配。某时刻 B 细胞（抗体）所受到的总激励如下组公式所示：

$$\Delta S = (k_1 \sum_{i=1}^{m} S_{Ag_i} c_{Ag_i} + k_2 \sum_{i=1}^{n} S_{Ab_j} c_{Ab_j} - k_3 \sum_{k=1}^{n'} H_{Ab_k} c_{Ab_k}) c_{Ab_k} \tag{2}$$

$$S_{Ag} = \begin{cases} 0 & d(p,e_{Ag}) > D \\ 1 - d(p,e_{Ag}) & d(p,e_{Ag}) \leqslant D \end{cases} \tag{3}$$

$$\begin{cases} c = p \otimes e \\ d(p,e) = \sum_{i=1}^{N} c_i / N \end{cases} \tag{4}$$

$$S_{Ab} = \begin{cases} 0 & d(p,e_{Ab}) > D \\ 1 - d(p,e_{Ab}) & d(p,e_{Ab}) \leqslant D \end{cases} \tag{5}$$

$$H_{Ab} = \begin{cases} 0 & d(e,p_{Ab}) > D \\ 1 - d(p,e_{Ab}) & d(e,p_{Ab}) \leqslant D \end{cases} \tag{6}$$

其中，Ag_i、Ab_j 和 Ab_k 分别为 B 细胞周围的抗原、对 B 细胞起激励作用的抗体和起抑制作用的抗体；下角 c_{Ab} 为 B 细胞分泌抗体的浓度；c_{Ab_i}、c_{Ab_j} 和 c_{Ab_k} 分别为 Ag_i、Ab_j 和 Ab_k 的浓度；k_1、k_2 和 k_3 为加权系数；S_{Ag} 表示抗原 A_g 对 B 细胞的激励；e_{Ag} 表示抗原 Ag 的表位；$d(p,e_{Ag})$ 定义为在归一化的数据空间中的抗体的对位和抗原的表位的距离；D 为两者结合须达到的阈值；N 为 p 和 e 的位数；C_i 为 C 的第 i 位值；S_{Ab} 表示其他抗体对 B 细胞的激励；H_{Ab} 表示其他抗体对 B 细胞的抑制。即第一项为抗体与抗原结合受到的激励；第二项为抗体的对位与其他抗体的表位结合受到的激励；第三项为抗体的表位与其他抗体的对位结合受到的抑制。当周围抗原和抗体对 B 细胞的激励积累超过一定水平时，细胞会克隆扩增。克隆扩增特性是用 *Sigmoid* 函数来模拟的，激励水平比较低时复制水平也很低；当激励达到一定程度时，细胞发生克隆扩增，数量迅速增加；当抑制大于激励时细胞数量减少，达到一个新的稳定状态。

虽然目前还没有将独特型调节原理用于入侵检测的研究成果，但应该看到克隆选择算法也可以用来解决优化问题[27]，所以在 IDS 中用独特性调节原理代替克隆选择算法是可行的，因为它们所起的作用相似。

3 危险模式理论

危险模式理论（Danger Theory，DT）最先由文献[28]提出，认为诱发机体免疫应答的关键因素是入侵者产生的危险信号的程度而不是入侵者的异己性，因而不论是自体因素发生改变，还是外界因素产生影响，只要出现供机体识别的危险信号就可以诱发效应细胞的活化。文献[29]中进一步指出，危险的外来入侵者会启动细胞应激或细胞死亡而导致危险信号的产生，危险信号由抗原提呈细胞（Antigen Presenting Cells，APCs）识别，共有两类：内部（如身体自身）生成的危险信号和来自入侵组织（如细菌）的外部信号。这两类信号都能够刺激 APC 并引发免疫应答。

Nottingham 大学的 Uwe Aickelin 小组近年一直致力于将危险模式理论应用于入侵检测中，目标是为危险模式理论建立一个计算模型，以定义、研究和发现危险信号，并根据该模型建立一些新的算法，用于构造具有较低误报率的入侵检测系统。文献[30]中指出现有的 AIS 都是基于自我—非我识别，即免疫细胞通过“阴性选择”的检查过程，使得免疫细胞只对“非我”成分的抗原作出免疫应答，对“自我”成分形成免疫耐受，不产生免疫应答，因此自我—非我识别是人工免疫系统的关键。但这种理论却不能解释一些现象，例如肠胃每天都接触大量的细菌，这些细菌都不会被归类为“自我”，但却不会引起免疫响应；自我—非我识别模型无法解释自我免疫疾病现象，如在多发性硬化例子中，免疫系统会对某些属于“自我”的细胞产生免疫应答。这些例子说明人类免疫系统不是仅仅依靠对自我—非我的识别结果来决定是否作出免疫响应的。现有的入侵检测模型都是基于自我—非我识别，因此存在误报、规模等问题，难以用于实际的大型网络。在详细分析了危险模式理论后，指出了该理论与 AIS 以及入侵检测之间的关联，并提出了一系列今后要研究的问题，例如怎样定义适当的危险信号。因为适当的危险信号可以克服自我—非我选择方法的局限性，可以将非我空间限制在一个可控的尺寸，还可以处理自我（非我）随时间改变的情况。

经过研究，文献[31,32]认为将危险模式理论引入入侵检测，不但可以检测已知和未知的攻击，而且可以降低误报率和解决阴性选择模型的规模问题。如前所述，细胞应激或细胞死亡会导致危险信号的产生，其中细胞死亡有两种方式：凋亡和坏死。凋亡是细胞死亡的正常方式，是有计划的细胞死亡，有定义好的细胞内通道和调节器；而坏死是细胞产生应激后的细胞异常死亡，将导致完全的细胞裂解，由于细胞的碎片引起发炎。从病理生理学角度看，凋亡被用于保持组织的动态平衡，在调节免疫响应中起重要作用。基于此，新的入侵检测系统将警报分成两类：凋亡警报和坏死警报。凋亡警报可定义为低级别的、噪声警报，其本身不会构成重大的不当行为，但经常作为攻击的先决条件，当引起系统重大损失的较严重攻击出现时，就会产生坏死警报。他们希望把警报同攻击行为联系起来，因此采用关联算法作为危险模式算法。凋亡有抑制作用，坏死有激励作用，尽管二者的区别不像现在认为的那样明显，但这两种作用之间的相关性是危险信号的基础。当引起最小损失的攻击出现后会产生危险信号，而且会被快速、自动地检测到，一旦危险信号被传送出去，那么免疫系统就会对发出危险信号附近的抗原作出响应。目前该小组正在进行的工作是研究如何在免疫系统中识别危险信号，该研究

结果可使我们清除了解在活的有机体内存在何种危险信号以及如何将他们转化为用于检测在计算机环境中的计算机系统中的危险[19]。

与基于 DT 的 IDS 相似的是文献[33]提出的双信号模型,即免疫系统的细胞需要两个信号产生激励:一个在检测到某些异常(即非我)时出现;另一个在检测到对身体产生损害时出现。因此免疫系统仅根据损害的百分比来设置响应,响应的力度与损害成比例。该损害—响应机制使免疫系统可以避免对误报作出响应,从而解决基于阴性选择的 IDS 误报率高的问题[20]。

4　结语

基于生物免疫的人工免疫系统在近几年得到迅速发展,由于它所具有的分布性、多样性、鲁棒性、适应性和特异识别等特性,正是入侵检测系统所希望具有的特性,因此一些免疫机制和免疫算法被用来实现入侵检测。目前对抗体克隆选择学说的研究已比较成熟,对其在入侵检测技术中的应用研究也最多,取得的成果也最丰富,但由于阴性选择算法只能处理简单问题,克隆选择算法处理动态问题的能力有限,因此基于自我—非我识别的入侵检测模型与实际入侵检测的要求还有一定差距,必须要借助其他的理论。免疫网络学说的发展也比较成熟,但对其的应用研究比较有限,目前集中于优化问题。应该看到,独特型调节理论与抗体克隆选择的区别仅在于抗体与抗体之间是否存在相互作用,而且克隆选择算法也可用于优化问题,因此将独特型调节理论和阴性选择算法相结合进行入侵检测值得一试。危险模式理论在生物免疫学界也属于较新的理论,发展还不完善,其中的一些原理还没有完全弄清,因此对其在入侵检测中的应用研究尚处于起步阶段,如何将其与入侵检测系统相对应,以及如何实现一个完整的入侵检测系统将是今后研究的方向。

参考文献:

[1] 焦李成,杜海峰. 人工免疫系统进展与展望[J]. 电子学报, 2003,31(10):1540 - 1548.

[2] AICKELIN U, GREENSMITH J, TWYCROSS J. Immune System Approaches to Intrusion Detection - A Review, Proceedings International Conference on Artificial Immune Systems[C]. Catania, Italy, 2004.316 - 329.

[3] HOFMEYR S, FORREST S. Architecture for an AIS[J]. Evolutionary Computation, 2000, 7(1): 1289 - 1296.

[4] 赵俊忠,黄厚宽,田盛丰. 免疫机制在计算机网络入侵检测中的应用研究[J]. 计算机研究与发展,2003,40(9):1293 - 1299.

[5] KIM J, BENTLEY P. Evaluation of negative selection in an artificial immune for network intrusion detection[A]. Proceedings of the GECCO[C]. 2001. 1330 - 1337.

[6] KIM J, BENTLEY P. Negative Selection and Niching by an Artificial Immune System for Network Intrusion Detection[A]. Proceedings of Genetic and Evolutionary Computation Conference (GECCO '99) [C]. Orlando, Florida, 1999. 149 - 158.

[7] DASGUPTA D, GONZALEZ F. An Immunity - Based Technique to Characterize Intrusions in Computer Networks[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(3): 1081 - 1088.

[8] KIM J, BENTLEY P. The Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with a Negative Selection Operator[A]. The Congress on Evolutionary Computation(CEC-2001)[C]. Seoul, Korea, 2001. 1244 - 1252.

[9] KIM J, BENTLEY P. Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Dynamic Clonal Selection[A]. The Congress on Evolutionary Computation[C]. 2002. 1015 - 1020.

[10] KIM J, BENTLEY P. Immune Memory in the Dynamic Clonal Selection Algorithm[A]. Proceedings of the First International Conference on Artificial Immune Systems (ICARIS) [C]. Canterbury, 2002. 57 - 65.

[11] GONZALEZ F, DASGUPTA D, GOMEZ J. The Effect of Binary matching Rules in Negative Selection[A]. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO) [C]. 2003. 196 - 206.

[12] GONZÁLEZ F, DASGUPTA D, NIÑO L. A Randomized Real - Valued Negative Selection Algorithm[A]. Proceedings of the 2nd International Conference on Artificial Immune Systems[C]. Edinburgh, UK, 2003. 261 - 272.

[13] GONZÁLEZ F, DASGUPTA D. Anomaly Detection Using Real-Valued Negative Selection[J]. Genetic Programming and Evolvable Machines, 2003, 4(4): 383 - 403.

[14] HARMER PK, WILLIAMS PD, GUNSCH GH, *et al*. An Artificial Immune System Architecture for Computer Security Applications [J]. IEEE Transaction on evolutionary computation, 2002, 6(3): 252 - 280.

[15] HARMER PK, WILLIAMS PD, GUNSCH GH, *et al*. An Artificial Immune System Architecture for Computer Security Applications [J]. IEEE Transaction on evolutionary computation, 2002, 6(3): 252 - 280.

[16] SINGH S. Anomaly detection using negative selection based on the r - contiguous matching rule[A]. Proceedings of the 1st International Conference on Artificial ImmuneSystems(ICARIS)[C]. University of Kent at Canterbury, 2002. 99 - 106.

[17] BALTHROP J, ESPONDA F, FORREST S, *et al*. Coverage and Generalization in an Artificial Immune System[A]. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002)[C]. Morgan Kaufmann, New York, 2002. 3 - 10.

[18] JERNE NK. Towards a network theory of the immune system[J]. Annals of Immunology, 1973, 125C: 373 - 389.

[19] CAYZER S, AICKELIN U. On the Effects of Idiotypic Interactions for Recommendation Communities in Artificial Immune Systems [A]. Proceedings of the 1st International Conference on Artificial Immune Systems(ICARIS-2002)[C]. Canterbury, UK, 2002. 154 - 160.

[20] FARMER JD, PACKARD NH, PERELSON AS. The immune system, adaptation, and machine learning[J]. Physica, 1986, 22D: 187 - 204.

[21] 徐雪松,褚静. 多模态函数优化的免疫算法[J]. 浙江大学学报(工学版),2004,38(5):530 - 533.

[22] CAYZER S, AICKELIN U. A Recommender System based on the Immune Network[A]. Proceedings CEC2002[C]. Honolulu, USA, 2002. 807 - 813.

[23] MORRISON T, AICKELIN U. An Artificial Immune System as a Recommender for Web Sites[A]. Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS-2002) [C]. Canterbury, UK, 2002. 161 - 169.

[24] AICKELIN U, CHEN Q. On Affinity Measures for Artificial Immune System Movie Recommenders[A]. Proceedings RASC-2004, The 5th International Conference on: Recent Advances in Soft Computing[C]. Nottingham, UK, 2004.

(下转第 1743 页)

图3所示为 *ref*max = 10 时,引入 RPSA 前、后的 P-Grid 系统的搜索成功率比较。此时的搜索成功率提高更加明显。

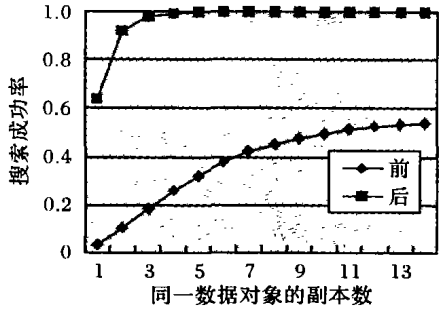

图3　ref*max* = 10 时搜索成功率比较

从图2和图3的比较还可以得出,在没有引入 RPSA 之前,系统搜索成功率随 *ref*max 变动很大;引入 RPSA 之后,系统搜索成功率基本不随 *ref*max 一起变动。而 *ref*max 的大小直接关系到路由表的大小及交换路由表信息时的网络流量,说明 RPSA 能在保证搜索成功率的前提下有效控制路由表的大小、减少交换路由表信息时的网络流量。

3.4　查询延迟分析

发起查询的节点必须等到查询成功结束、查询结果返回后才能下载所请求的数据对象,因此查询延迟对于搜索算法的性能来说也是一个重要指标,而且也反映了查询转发过程中的路由性能。

假设一条消息在网络上的每一跳需要花费一个 *timestep*,则悲观算法中查询成功时的最大延迟为 2 * *TTL* 个 *timestep*。当系统运行一段时间后,每个节点上都保持了已被修正的索引表,这时查询转发将根据索引表提供的信息,选择索引值较高的节点进行,即选择查询延迟较小的路径。例如 3.2 节中,若节点6再次转发查询"100",它会选择延迟为 4ms 的节点4,而不会选择延迟为 16ms 的节点5。可见 RPSA 确实能通过有效路由减少查询延迟,提高路由性能。

3.5　维护开销分析

本文所描述的 RPSA 能使节点加入和退出系统时的结构维护变得简单。当一个节点新加入系统时,只要在索引的各数据对象下都添加一个条目即可。新添条目的索引值可以取所有索引值的最大值,乐观地认为新加入的节点都是"最好"的节点,使其开始时具有很强的竞争力。当节点离开系统时就更简单了,甚至可以不做任何维护工作。因为如果节点不可达,查询就失败,则其索引值自然会降下来。同时,这也体现了改进后的算法在快速变化的网络环境下的健壮性。

3.6　存储开销分析

每个节点都在某一层为每个被请求的数据对象保持索引。所以,如果有 R 个数据对象,则需要 $O(R * ref\max)$ 个条目,对于一个典型的节点来说,这并不是负担。而对于存储容量有限的节点,可以定期删除那些长时间不用的索引。比如可以给条目添加一属性 lastusetime(上次使用时间),这样如果超过一段时间没有使用,即可将其删除。另外,考虑到对同一文件的查询数满足幂规律,即多数查询都是针对少数文件的,所以节点保持的索引的使用率很高。

4　结语

论文提出了一种对现有 DHT 算法进行改进的算法——RPSA,并以 P-Grid 算法为例进行验证。分析表明 RPSA 能在低存储开销的情况下精确定位搜索方向,选择"最短"(时间上最短)路径进行查询转发,有效提高并保证P-Grid算法的搜索成功率,很好地解决了 P-Grid 算法的路由性能问题以及节点加入和退出时的结构维护问题。当然,RPSA 还需进一步完善,进一步研究并验证其在更广泛的 DHT 算法中的应用等问题。

参考文献:

[1] MILOJICICDS, KALOGERAKIV, LUKOSER, *et al*. Peer-to-Peer Computing[R]. Technical Report, HPL-2000-57. HP Laboratories Palo Alto, 2002.

[2] Napster. [EB/OL] http://www.napster.com, 2002-02.

[3] Gnutella[EB/OL]. http://gnutella.org, 2002-04.

[4] BALAKRISHNAN H, KAASHOEK MF, KARGER D, *et al*. Looking up Data in P2P System[J]. Communications of the ACM, 2003, 46(2):43-48.

[5] P-Grid in a nutshell[EB/OL]. http://www.p-grid.org, 2004.

[6] Barkai D. An Introduction to Peer-to-Peer Computing[R]. Intel Developer Update Magazine, 2002.

[7] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology[A]. Processing of the ACM SIGCOMM'99[C]. New York: ACM Press, 1999.251-262.

[8] BARBOSA MW, COSTA MM, ALMEIDA JM, *et al*. Using Locality of Reference to Improve Performance of Peer-to-Peer Applications[J]. ACM SIGSOFT Software Engineering Notes archive, 2004, 29(1).

[9] ABERER K. P-Grid: A Self-organizing Access Structure for P2P Information Systems[A]. Sixth International Conference on Cooperative Information Systems(CoopIS 2001)[C]. 2001.

[10] 夏素贞,杨德仁,曹静霞. 基于 P-Grid 的 P2P 信息共享系统的设计和实现[J]. 计算机应用,2004,24(增刊).

(上接第 1729 页)

[25] AICKELIN U, CHEN Q. Movie Recommendation Systems Using An Artificial Immune System[A]. 6th International Conference in Adaptive Computing in Design and Manufacture[C]. Bristol, UK, 2004.

[26] 孙勇智,戴晓晖,韦巍. 人工免疫响应的模型研究[J]. 浙江大学学报(工学版),2004,38(6):682-686.

[27] NUNES L, FERNANDO J. The Clonal Selection Algorithm with Engineering Applications [A]. In Workshop Proceedings of GECCO'00, Workshop on Artificial Immune Systems and Their Applications[C]. Las Vegas, USA, 2000.36-37.

[28] MATZINGER P. Tolerance Danger and the Extended Family[J]. Annual reviews of Immunology, 1994, 12:991-1045.

[29] MATZINGER P. The Danger Model: A Renewed Sense of Self[J]. Science, 2002, 296:301-305.

[30] AICKELIN U, CAYZER S. The Danger Theory and Its Application to Artificial Immune Systems[A]. Proceedings of the 1st International Conference on Artificial Immune Systems(ICARIS-2002)[C]. Canterbury, UK, 2002:141-148.

[31] AICKELIN U, BENTLEY P, CAYZER S, *et al*. Danger Theory: The Link between AIS and IDS?[A]. Proceedings ICARIS-2003, 2nd International Conference on Artificial Immune Systems[C]. 2003.147-155.

[32] GREENSMITH J, AICKELIN U, TWYCROSS J. Detecting Danger: Applying a Novel Immunological Concept to Intrusion Detection Systems[A]. 6th International Conference in Adaptive Computing in Design and Manufacture[C]. Bristol, UK, 2004.1135-1142.

[33] HOFMEYR S. The implications of immunology for secure systems design[J]. Computers & Security, 2004, 23:453-455.