一种基于生物免疫学原理的入侵检测方法

魏春英，刘培玉
（山东师范大学 信息管理学院，山东 济南 250014）

摘 要：论述了免疫原理在入侵检测技术中的应用，通过对目前负检测方法存在的局限性的分析，从实际应用的角度，对负检测方法进行了一定的改进，并对改进的可行性进行了理论上的分析与证明。该方法在一定条件下减少了待检模式与检测器的比较次数，提高了系统的检测效率。

关键词：免疫原理；入侵检测；负检测；检测元
中图分类号：TP393.08 **文献标识码**：A **文章编号**：1005－3751(2006)02－0226－03

An Approach of Immunity－Based Intrusion Detection

WEI Chun-ying, LIU Pei-yu
(Information and Management School, Shandong Normal University, Jinan 250014, China)

Abstract: Discusses about the application of immunity principle on intrusion detection. The localization of the negative detection approach is analyzed. From the point of practice, an improved approach of the negative detection has been given, then the feasibility of this improvement is testified theoretically. In a certain extent, the improvement approach reduces the times that the await detect patterns compare with the detect unit, and advances the detection efficiency.

Key words: immunity principle; intrusion detection; negative detection; detector

0 引 言

随着入侵检测技术研究的深入，人们逐渐认识到生物免疫系统在生物体中的角色类似于计算机安全研究领域中的入侵检测系统。入侵检测系统负责保护计算机系统不受来自内部和外部的入侵行为的侵害，而生物免疫系统能够保护人体不受细菌、病毒等外来病菌的侵害，即它们都要保护高度复杂的系统，以免受到恶意攻击。免疫学者传统上把免疫系统抵御病菌入侵描述为区分“自我”和“非我”。“自我”是人体内正常的细胞；而“非我”则是指那些外来的病菌。类似的，保护计算机系统不受恶意入侵也可以被视为区分“自我”和“非我”。但与人体免疫系统不同的是，在计算机系统中“自我”是指授权用户的合法行为、正常的通信模式等；“非我”则包括非授权用户的行为、病毒或蠕虫形式的外来代码、特洛伊木马和破坏性数据等。将生物免疫系统的原理、方法和体系结构用于计算机安全领域正成为当前的一个热点[1]。

1 生物免疫学原理概述

人类免疫系统的功能是通过大量不同类型细胞间的交互作用来实现的，在这些不同类型的细胞中，淋巴细胞（白细胞）处于主导地位。淋巴细胞也称为抗体（antibody），其主要工作是区分本体细胞（self）和异体细胞（nonself），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（antigen）。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负选择（negative selection）过程。在负选择过程中，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。

通过了负选择过程的淋巴细胞被认为是成熟的，它们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（clonal selection）阶段。在这个阶段，被启动的淋巴细胞产生多个

收稿日期：2005－05－23
作者简介：魏春英（1971—），女，山东成武人，讲师，硕士研究生，主要研究方向为网络信息安全；刘培玉，教授，硕士生导师，主要研究方向为网络信息安全、软件工程。

本身的复制，这些复制细胞可以称为内存细胞(memory cell)。内存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[2]。

2　免疫学原理在入侵检测技术中的应用

目前，利用免疫系统的免疫过程和原理进行入侵检测基本上围绕着两种方法来进行。第一种方法也是最直接和最早使用的方法，它是直接将正常的、合法的或可接受的操作模式定义为自我集，以此来构建正常行为模式库，再对操作模式进行监视，通过在正常模式库中搜索，若未找到相关记录就标记为不匹配，若找到相关记录就说明是合法操作或正常操作模式，不妨将这种方法称为正检测(positive detection)以与第二种方法相区别。另一种方法是所谓的负检测(negative detection)，由于该方法具有良好的分布性、健壮性、自适应性等优点，因此这也是目前在基于网络的入侵检测中使用最多的方法[3]。

这种方法最早由美国新墨西哥州立大学计算机系进化计算研究组 Forrest 于 1994 年提出，当时用于计算机病毒的检测，后来 Hofmeyr 等人将这个思想推广到广播型局域网的入侵检测中。负检测的核心思想是定义一个自我集作为训练集来产生不与自我集模式匹配的检测元(detector)集，使用这些检测元来进行入侵检测。显然，该方法的检测过程也分为训练阶段和测试阶段，在训练阶段，依据训练集模式来产生有效的检测元集，在测试阶段通过非我模式与检测元的匹配来检测异常入侵[4,5]。目前，对负检测的研究已成为计算机免疫学领域中的研究重点，研究内容主要包括自我集的构造、检测元的产生算法、检测系统的体系结构等方面[3,6]。

3　网络入侵检测系统的改进

从实际应用的角度，对基于免疫原理检测模型的网络入侵检测系统中的检测方法进行了改进，主要目标是为了提高入侵检测系统的检测效率和检测能力，克服其中的局限性。

3.1　网络入侵检测系统检测方法的局限性

从负检测方法的实质来看，系统中的检测元只是为了检测异常模式，而不是为了检测正常模式。另外，从负选择算法中可以看出，检测元在生成过程中要经历一个类似于免疫耐受的审查过程，只有与自我集的任何模式都不匹配的检测元才会成为有效检测元，由此可见，在检测过程中将代表正常连接的自我模式与检测元进行匹配试验是毫无意义的，因为这些自我模式根本就不可能与检测元集中的某个检测元匹配。因此，让大量的正常模式与检测元进行匹配试验就势必会导致系统在性能上存在一定的局限性，降低系统的检测效率。

其检测方法的流程如图 1 所示，其中待检模式就是测试集。

3.2　检测方法的改进

从上面的分析中可以看出，若能在待检模式进入检测器之前就已将其中的部分正常模式过滤掉，就可大大减少系统的整体匹配比较次数，因为要将所有的正常模式在检测器之前完全过滤掉是不可能的，也是不合实际的，因为正常模式间的匹配的判断一般只能使用完全匹配规则，因此，要过滤掉所有的正常模式就需要 $|S|$ 个模式。从实际情况来看，$|S|$ 远远大于 $|R|$，这也正是负选择算法的优势所在。

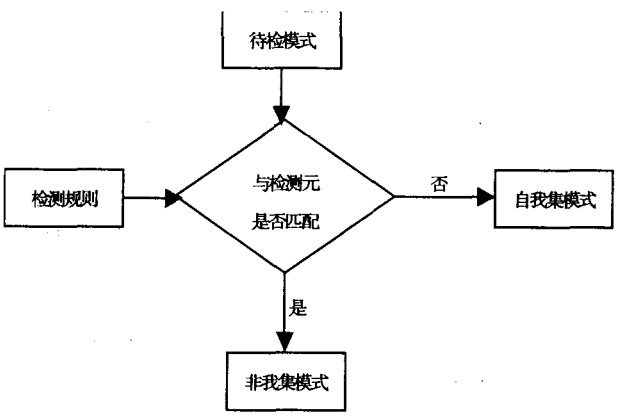

图 1　原检测方法

事实上，自我集 S 中的正常模式在出现频率上存在极大的差异，即正常模式的出现频率呈现出分散分布的特点，在这里，将出现频率高的正常连接模式称为高频连接模式。鉴于此，可使用训练阶段获得的自我集中的高频连接模式来构建过滤器 F，这样既可保证 $|F|$ 远小于 $|R|$，又可起到过滤掉在 S 中占较大比例的高频连接模式的作用。

检测方法改进后的流程如图 2 所示。

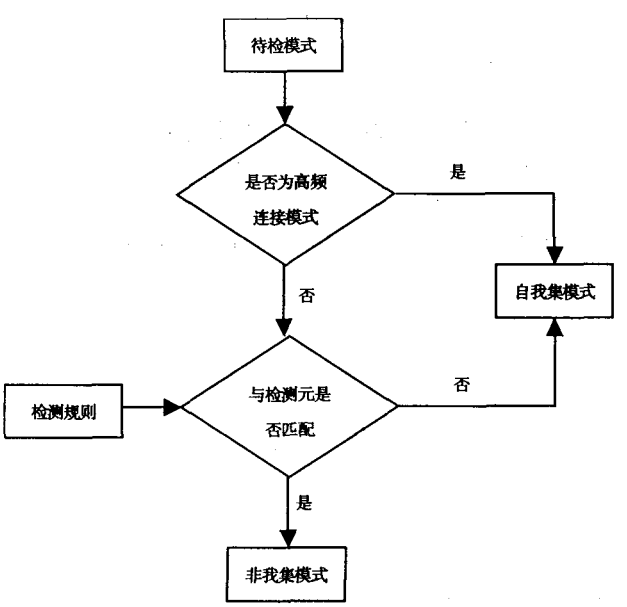

图 2　改进后的检测方法

4 两种检测方法检测效率的对比分析

4.1 几个设定和前提

为便于下面的分析,这里作如下定义:

(1) 设待检测模式只分为正常模式和异常模式。不能被检测元检测到的为正常模式;能被检测元检测到的为异常模式,且与检测元的平均比较次数为 P_m。

(2) 设检测器为 R。

(3) 设在时间[$£_1$,$£_2$]内所捕获的待检模式集为 P,其中正常模式的概率为 P_1。

(4) 由高频连接模式构成的过滤器为 F,且 $|F| \ll |R|$。

由于系统的性能主要体现在待检模式与检测元或过滤器的匹配比较过程的效率,因此,以下分析中,使用待检模式与检测元或过滤器的匹配比较次数来表示系统的检测效率或性能。

4.2 对比分析

由于异常模式远少于正常模式,因此,若能将这些正常模式在与检测元匹配前过滤,且过滤器远比检测元小,则从整体上来说,与检测元比较的模式就会大大减少。

下面对两种检测方法的检测效率进行对比分析。

(1)原检测方法。

a.对于正常模式:

从上面的分析中可以看出,正常模式与所有的检测元比较一次。待检模式中的正常模式数为 $|P|p_1$,因此检测正常模式所需的总比较次数为:

$$|P|p_1|R| \tag{1}$$

b.对于异常模式:

待检模式中的异常模式数为 $|P|(1-p_1)$,因此,检测异常模式所需的总比较次数为 $|P|(1-p_1)P_m$ (2)

综合上面两式,可得到检测待检模式总共所需的比较次数为:

$$|P|p_1|R|+|P|(1-p_1)P_m \tag{3}$$

(2) 改进后的检测方法。

a.对于正常模式:

由于设置的过滤器只是高频连接模式,而非所有的正常连接模式,因此将正常模式分为两类:

* 一类是可被过滤器识别的正常模式,此类模式仅需与过滤器比较,此时设正常模式属于该类模式的概率为 p_2,此类正常模式的数量为 $|P|p_1p_2$,因此检测这些模式所需的总比较次数为:

$$|P|p_1p_2|F| \tag{4}$$

由于 $|F| \ll |R|$,因此,此处未考虑不同正常模式与过滤器的比较次数差异,统一认为比较 $|F|$ 次即可。

* 另一类是不能被过滤器识别的正常模式,它经过过滤器之后还必须与检测器的每一个检测元进行比较。此类正常模式的数量为 $|P|p_1(1-p_2)$,因此,检测这些正常模式所需的总比较次数为:

$$|P|p_1(1-p_2)(|F|+|R|) \tag{5}$$

b.对于异常模式:

待检测模式中的异常模式数为 $|P|(1-p_1)$,因此,检测异常模式所需的总比较次数为:

$$|P|(1-p_1)(|F|+P_m) \tag{6}$$

综合上面 3 式,可得到检测待检模式总共所需的比较次数为:

$$|P|p_1p_2|F|+|P|p_1(1-p_2)(|F|+|R|)+|P|(1-p_1)(|F|+P_m) \tag{7}$$

将上式整理后得到:

$$|P|p_1|R|+|P|P_m-|P|p_1P_m+|P||F|-|P|p_1p_2|R| \tag{8}$$

将式(3) 整理后得到:

$$|P|p_1|R|+|P|P_m-|P|p_1P_m \tag{9}$$

可见,只要 $|P||F|-|P|p_1p_2|R|<0$,即不等式 $\frac{|F|}{|R|} < p_1p_2$ 成立时就可以使式(8) 小于式(9),即改进方法后的比较次数少于改进前的比较次数,达到提高系统性能的目的。而由于检测元的数量是可根据系统资源状况来进行调节,因此在 p_1 一定的情况下,一旦 p_2 确定,就可根据 p_1p_2 来调节 $|R|$。因此,要满足上式是完全可能的。

5 总 结

由于正常网络行为远多于异常连接,所以将大量的正常连接模式与检测器进行匹配是没有意义的。通过分析发现大量正常连接出现的频率远高于异常连接,本方法提出在待检模式与检测器匹配之前增加一个过滤器以过滤掉部分正常连接模式,从而在一定条件下减少了待检模式与检测器的比较次数,提高了系统的检测效率。

参考文献:

[1] 戚 勇,张 琨.基于生物免疫学的分布式入侵检测系统模型[J].计算机工程与设计,2004,25(4):481-482.

[2] 吴敏毓,刘恭植.医学免疫学(第 3 版)[M].合肥:中国科学技术大学出版社,1999.

[3] 鲁云平.基于免疫原理的网络入侵检测技术研究[D].重庆:重庆大学,2003.14-15.

[4] Hofmeyr S A, Forrest S. Immunity by design: An Artificial Immune System[A]. Proceedings of the Genetic and Evolutionary Computation Conference(GECCO)[C]. San Francisco, CA:[s.n.],1999.1289-1296.

[5] Hofmeyr S A. A Immunity-based Mode of Distributed Detection and its Application to Computer Security[D]. NM, US: University of New Mexico, 1999.

[6] 张彦超,阙喜戎,王文东.一种基于免疫原理的网络入侵检测模型[J].计算机工程与应用,2002,38(10):159-161.