一种新型入侵检测模型及其检测器生成算法

程永新，许家珀，陈 科

(电子科技大学应用数学学院 成都 610054)

【摘要】介绍了一种基于人工免疫原理的入侵检测模型，重点研究了否定选择算法模型中检测器集合的生成算法，提出了新的初始检测器的生成算法，并对算法性能进行了分析。结果表明：该算法模型可以对未知入侵行为和已知入侵行为的变异进行有效的识别。

关 键 词 人工免疫； 入侵检测； 检测器； 否定选择

中图分类号 TP393.08 **文献标识码** A

A Novel IDS Model and the Arithmetic to Get the Detection

CHENG Yong-xin，XU Jia-yi，CHEN Ke

(School of Applied Mathematics， UEST of China Chengdu 610054)

Abstract A new intrusion detection model that based on the theory of artificial immune system is introduced in this article. It researches emphases on the arithmetic of how to get the detector that included in the negative selection arithmetic. It also puts forward a new arithmetic of how to get the detector and then analyses the arithmetic. The result indicate that, the arithmetic model can do the effective identify variation between the unknown intrusion action and the known one.

Key words artificial immune; intrusion detection; detector; negative selection

入侵检测系统(Intrusion Detection System，IDS)作为当前安全技术的新热点，不仅越来越多的受到人们的关注，而且已经开始在各种不同的环境中发挥其关键作用。入侵检测一般分为误用检测(misuse detection)和异常检测(anomaly detection)两类。其中异常检测是通过建立用户正常行为模型，以是否显著偏离正常行为依据进行入侵检测，这种方法虽然误报率较高，但有可以发现新的攻击行为，是入侵检测研究的前沿和热点[1]。

免疫算法是一种新兴的智能计算技术，虽然起步晚，但已和神经网络、遗传算法并称为三大生物计算模型。它借鉴了生物免疫系统中抗体(Antibody)对抗原(Antigen)的识别原理以及抗体产生和进化的过程，将需要检测的异常状况设定为抗原，系统根据免疫算法原理产生抗体并以之对抗原进行识别[2]，这与入侵检测原理有着惊人的相似。本文将免疫学原理、分布式网络结构结合起来，设计了一种新的基于智能免疫系统的结构，为免疫学原理在IDS中的应用提供一种新的思路。

1 基于免疫原理的入侵检测模型

1.1 入侵检测网络模型

本文建立的入侵检测模型主要应用了免疫原理中的“自我/非我”理论来识别入侵行为。“自我”指正常的网络行为；“非我”指异常的网络行为即入侵行为。模型首先对网络信息进行信息解码，将网络信息转化成标准的格式。然后与所有合格检测器进行匹配。一旦与一个合格检测器匹配，则认为这个网络行为是入侵行为；否则认为是正常行为让其通过。如图1所示。

“入侵行为处理”模块对该入侵行为进行拦截，并将把识别出这个入侵信息的检测器的生命期延长，同时把这个检测器发布到“入侵信息黑板系统”。黑板系统则把这个检测器发送到其他主机IDS，使得局域

收稿日期：2005－10－21
基金项目：四川省科技厅基金资助项目(04JY029-017-1)
作者简介：程永新(1978－)，男，硕士生，主要从事人工免疫算法和入侵检测方面的研究.

网中所有主机都拥有这个合格检测器，从而识别出这种入侵行为。如图2所示。

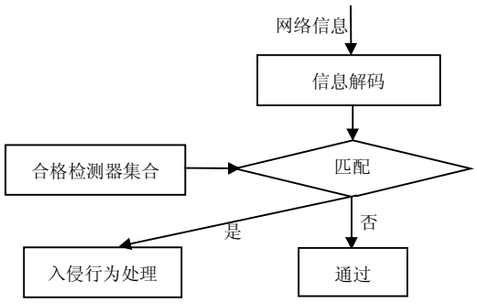

图1 主机IDS检测模型

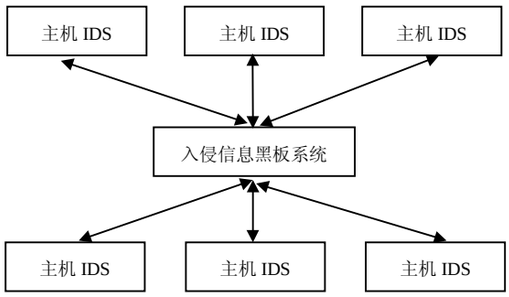

图2 入侵检测系统网络模型

1.2 检测器生成算法模型

合格检测器是IDS的核心部分，生成合格的检测器对IDS来说至关重要的。目前对合格检测器的生成算法已经有很多讨论。如：穷举检测器生成算法、线性检测器生成算法和贪心检测器生成算法等[3]，几种算法各有优缺点。但目前对如何生成初始检测器集合却未见报道。初始检测器集合的好坏直接影响生成合格检测器的效率和质量。本文提出高频变异器和低频变异器的概念来生成初始检测器集合。合格检测器生成算法模型，如图3所示。图中，各实体的功能如下：

(1) 自我库：即正常的网络数据特征。

(2) 高频变异器：从自我库中读取数据，通过高频变异算法产生新的数据，然后放入初始检测器集合。

(3) 合格检测器集合：用于存放可以识别入侵行为的检测器。

(4) 低频变异器：从合格检测器集合中读取数据，通过低频变异产生新的初始检测器集合。

(5) 入侵信息黑板系统：其他主机IDS把能识别入侵行为的合格检测器发送到黑板系统，黑板系统再将其送到各个主机的初始检测器集合当中。

(6) 初始检测器集合：把高频变异、低频变异和黑板系统的数据作为初始检测器集合

(7) 否定选择：用初始检测器集合的数据同自我库和合格检测器集合的数据进行匹配。匹配，说明能识别正常数据或者已经在合格检测器集合当中存在，删除；不匹配，说明能识别新的入侵行为，加注一个时间标记放入合格检测器集合。

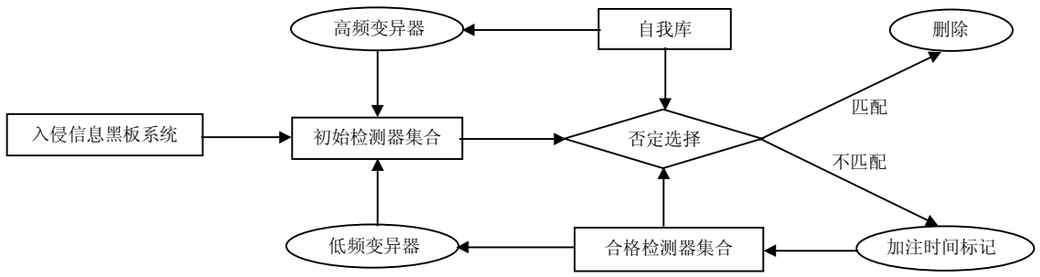

图3 合格检测器生成模型

2 变异算法原理和否定选择算法

2.1 变异算法原理

低频变异(Low Frequency Variation，LFV)原理：针对合格检测器集合(特别其中生命期较长的检测器)，采用位变异和交叉变异等方式。

位变异：在数据的关键位上进行变异。其原理如图4所示。

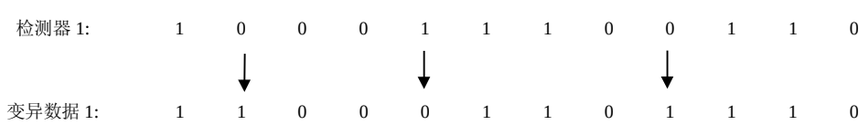

图4 位变异原理

交叉变异：两个(或多个)数据通过交叉互换自己的数据片段。其原理如图5所示：

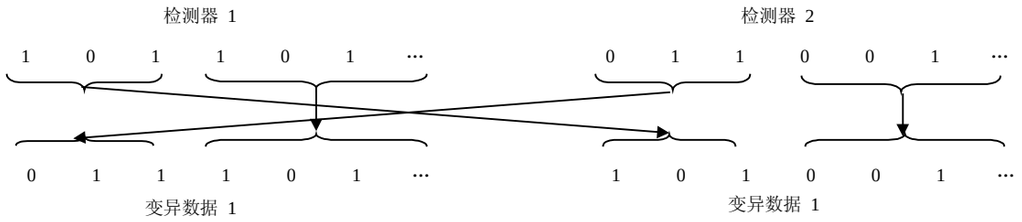

图5　交叉变异原理

高频变异(High Frequency Variation，HFV)原理：将选出的每个数据按照一定的规则打碎成多个数据片段，随机抽取一定的数据片段组成新的数据作为初始检测器。其原理如图6所示：

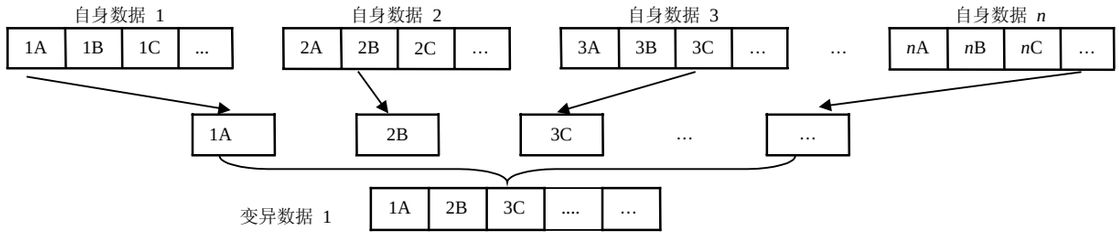

图6　高频变异原理

2.2　否定选择算法

否定选择算法(Negative Selection，NS)的目的是找出一个检测器集合R，它在不与S(自体集合)中元素匹配的前提下，能尽可能多的匹配M(非自体集合)中的元素。否定选择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集，使用这些检测元来进行入侵检测[4]。

假设合格检测器中包含X个合格检测器，每个检测器包含P个基因片段，其中第j个基因片段共有Y种值，则第j个基因片段的信息熵为$S_j(X)$：

$$S_j(X)=\sum_{i=1}^{Y}-K_{i,j}\lg K_{i,j} \tag{1}$$

式中　$K_{i,j}$表示基因片段j为i的概率。所以合格检测器集合的平均信息熵为：

$$S\ (X)=\frac{1}{P}\sum_{j=1}^{P}S_j(X) \tag{2}$$

式中　每两个合格检测器之间的亲和力为：

$$\theta_{m,n}=\frac{1}{1+s(2)} \tag{3}$$

式中　当亲和力$\theta_{m,n}$大于一个域值δ的时候，则说明这个检测器识别自体集合，应当删除。

3　模型性能分析

模型通过高频变异，产生了大量的新型数据，这些数据有很大部分是自我库和合格检测器集合当中不包含的，所以对识别未知的入侵行为有很大帮助；而通过低频变异，保留原来合格检测器的特征，这样就可以很好的识别出已知入侵行为的变异体(比如病毒的变异等)。

由于网络特征是一直变化的，所以合格检测器集合当中的数据不可能一成不变，但也不可能无限扩充，因为庞大的合格检测器集合不仅对实时检测不利，而且那些从来没识别出入侵行为的数据就成了大量的冗余。因此，对每一个合格检测器数据加上时间标记。当一个合格的检测器在一定的时间内没有识别入侵行为，则自动死亡；如果识别了入侵行为，则自动延长生命期，并记录在案；同时系统管理员对记录在案的数据分析确认以后，可以标注为永久生命期。

模型运行一定的时间以后，将达到一个相对平衡的过程。这时候在一定时间内，肯定有一部分合格检测器自动死亡，从而使合格检测器集合总数量减小；由于合格检测器的数量减小，初始检测器数据在否定

选择过程中不匹配的数量将会相对增加，使得合格检测器数量增加。这样在一定时间内，一部分不能识别入侵行为的合格检测器死亡，同时又补充进来另一部分新的合格检测器，所以合格检测器的总数量将会在一个常量附近上下波动，从而达到一种动态平衡。

4 实 验

两个集合之间的识别本文采用r-连续位匹配规则。对于两个集合m，n，如果两个二进制串m，n相应位置上至少连续r位相同，那么两个二进制串是r-连续位匹配的。即：

$$f(m,n,r)=\begin{cases}1 & \exists i,j,j-i\geqslant r,m_k=n_k,i\leqslant k\leqslant j\\ 0 & \text{其他}\end{cases} \tag{4}$$

模拟模型运行的初期，经过多次实验，分别使用随机产生、高频变异、低频变异3种方法，多次产生100、200、400、1 000个初始检测器，取其平均值。在同等条件下3种方法能成为合格检测器的百分比如表1所示：

表1 随机产生和高/低频变异产生初始检测器的合格率比较

初始检测器个数/个	随机产生/(%)	高频变异/(%)	低频变异/(%)
100	17.5	19.7	20.1
200	17.2	19.4	19.9
400	16.4	18.8	19.4
1 000	15.1	18.2	18.7

实验结果表明，在系统运行的初期，随机产生的检测器成为合格检测器的成功率显然要低于其他两种方法。表明两种方法显然好于随机产生。而当系统运行到一定时间，这种模型的优势就会很快表现出来：及时淘汰合格检测器中不能识别入侵行为的检测器；通过低频变异实现对病毒变异体的识别；通过多台计算机的高频变异实现对未知行为的识别。从而实现对计算机网络的安全保护。

5 结 束 语

本文介绍了一种基于免疫原理的新型入侵检测模型，并提出了初始检测器的生成算法。模型具有以下优点：各个主机通过黑板系统交流合格检测器，相当于实现多台计算机共同检测入侵行为，具有了分布式的特征；就各个主机而言，通过高频变异和低频变异产生初始检测器，较好地解决了未知入侵行为和已知入侵行为的变异体的识别。

参 考 文 献

[1] Dasgupta D, Ji Z, Gonzalez F. Artificial Immune system (AIS) research in the last five years[J/OL]. http://lib.uestc.edu.cn IEEE 2003 O-7803-7804-0, 2005-06-25.

[2] Alessio G, Philippe C. Two models of immunization for time dependent optimization[J/OL]. http://lib.uestc.edu.cn IEEE 2000 O-7803-6583-6, 2005-07-28.

[3] 李 涛. 计算机免疫学[M]. 北京: 电子工业出版社, 2004.

[4] Forrest S, Perelson A S, Allen L, et al. Self-nonself discrimination in a computer[J/OL]. http://lib.uestc.edu.cn IEEE 1994 O-1063-7109, 2005-08-15.

编 辑 刘文珍