doi:10.3772/j.issn.1002-0470.2014.01.003

免疫入侵检测多形态检测算法①

席 亮② 张凤斌 刘海龙

(哈尔滨理工大学计算机科学与技术学院 哈尔滨 150080)

摘 要 针对基于免疫机制的入侵检测系统的单一形态检测器检测性能低下的问题,分析了二进制形态空间和实值形态空间各自的不足,借鉴免疫独特型网络理论和免疫危险理论的信号机制,提出了多形态检测算法,该算法使用二进制形态和实值形态两种检测器协同检测,通过协同信号判定事件是否异常。经实验表明,该多形态检测算法较单一形态检测算法在检测性能上有了极大的提高。

关键词 免疫入侵检测,形态空间,二进制,实值,检测器,多形态,否定选择算法(NSA)

0 引 言

入侵检测系统(intrusion detection system,IDS)作为新一代网络安全技术,已在网络安全体系中发挥着日益重要的作用。目前,IDS 以异常检测方法为主,主要有基于聚类的、基于协议分析的、基于人工免疫理论的异常检测方法等。由于基于免疫机制的异常检测是以正常样本来训练检测器,不需要捕获大量的异常事件来获取检测规则,并且还可以有效预防未知攻击,所以一直以来都受到相关学者的推崇。IDS 的功能特性与生物免疫系统一样,都是将观察物通过一定策略区分为自体(正常)与非自体(异常),而且生物免疫中的自适应控制、鲁棒控制等特性可以很好地作用于 IDS。这为 IDS 的研究提供了一个新的思路[1,2]。目前的人工免疫算法主要有否定选择算法、免疫网络和克隆选择算法[3]。其中,否定选择算方法是一种自体/非自体识别技术[4],是对免疫细胞的成熟过程的模拟,在可供选择的算法中具有明显作用且效果显著。算法通常将问题定义在形态空间中,主要有二进制空间和实值空间。二进制空间将问题定义为一个长度为 n 的字符串,一个属性占一位或几位;实值空间将问题定义为一个 n 维空间的超球体:一个 n 维空间的点及半径[5]。而检测算法以否定选择算法(negative selection algorithm,NSA)为基础并分为二进制否定选择算法(binary NSA,BNS)和实值否定选择算法(real-valued NSA,RNS)。无论何种形态空间,系统的检测性能主要依赖检测器的性能,即检测器对非自体空间的覆盖效果。随着研究的深入,虽然针对某一形态空间检测器的生成与分布优化算法有很多[6-8],但检测效果都不甚理想。这主要是因为单一形态检测器已不能适应实际系统性能的需要[9,10]。本文分析了两种形态空间检测器存在的问题,借鉴免疫独特型网络理论和免疫危险理论的信号机制,提出了一种新的异常检测算法——多形态检测算法。该算法根据不同的非自体区域特征生成不同形状的检测器,将各种形状检测器的优点相结合,达到最大化非自体空间覆盖和最小化检测器重叠的目的。

1 相关工作

1.1 形态空间、自体与检测器

形态空间 U 可分为自体子空间 U_S 和非自体子空间 U_N:$U = U_S \cup U_N$。其中,自体集合 $S \subseteq U_S$,检测器集合 $D \subseteq U_N$。自体和检测器分别分布于 S 和 D:$S = \{s_1, s_2, \cdots, s_{N_s}\}$ 和 $D = \{d_1, d_2, \cdots, d_{N_d}\}$,其中 N_s 和 N_d 分别为自体和检测器个数。一般通过否定选择算法进行区分。自体与检测器在不同的形态空间下表示方法也不同。

在二进制空间下,每个自体样本可以表示为

$$s_i = (s_{i1}, s_{i2}, \cdots, s_{iL}, l, OtherAttributes) \quad (1)$$

① 国家自然科学基金(60671049;61172168)资助项目。

② 男,1983 年生,博士;研究方向:网络与信息安全;联系人,E-mail:xiliang@hrbust.edu.cn

(收稿日期:2012-08-07)

其中 $i=1,2,\cdots,N_s$,s_{ij}非 0 即 1,$j=1,2,\cdots,L$,每个属性占若干位;l 为该自体样本判定阈值,*OtherAttributes* 为该自体样本其他属性,如年龄等。同理,每个检测器可以表示成

$$d_i=(d_{i1},d_{i2},\cdots,d_{iL},l,OtherAttributes) \quad (2)$$

其中 $i=1,2,\cdots,N_d$,d_{ij}非 0 即 1,$j=1,2,\cdots,L$,每个属性占若干位;l 为该检测器的判定阈值,*OtherAttributes* 为该检测器其他属性,如年龄等。

同样,每个自体样本在实值形态空间中可表示为

$$s_i=(s_{i1},s_{i2},\cdots,s_{iN},r,OtherAttributes) \quad (3)$$

其中 $i=1,2,\cdots,N_s$,s_{ij}为该自体样本第 j 维属性值,$j=1,2,\cdots,N$,r 为该自体样本训练半径,*OtherAttributes* 为该自体样本其他属性,如年龄等。同理,每个检测器可表示成

$$d_i=(d_{i1},d_{i2},\cdots,d_{iN},r,OtherAttributes) \quad (4)$$

其中 $i=1,2,\cdots,N_d$,d_{ij}为该检测器第 j 维属性值,$j=1,2,\cdots,N$,为该检测器的判定半径,*OtherAttributes* 为该检测器其他属性,如年龄等。

另外,为了便于计算,需要对自体与检测器进行数据的调整。对于二进制空间,由于样本每个属性是由二进制数组成的,但实际样本的属性值不一定都是二进制数,有可能是离散值或浮点数,所以,需要对其进行二进制数转换。对于实值空间,由于每维属性的值域范围不同,在进行后续检测处理时无法使用统一标准。所以,需要对其进行正规化处理,从而保证所有属性值在同一尺度下计算。本文的正规化方法采用如下策略:首先计算每一维属性值的方差 μ_j 和标准差 σ_j,再根据以下两式将样本进行正规化[11]:

$$s'_{ij}=(s_{ij}-\mu_j)/\sigma_j \quad (5)$$

$$s'_{ij}=(s_{ij}-\min_j)/(\max_j-\min_j) \quad (6)$$

其中,$\min_j$ 和 $\max_j$ 分别为该属性值域的上下界。

1.2　否定选择算法

否定选择算法是对免疫细胞的成熟过程的模拟。主要通过计算样本间的亲和力判断样本是否匹配。算法应用在两个阶段:训练检测器阶段和检测器检测阶段,如图 1 所示。训练阶段主要负责检测器的生成,过程为:先随机产生候选检测器并进行亲和力耐受训练,删除与自体亲和力较高(即匹配)的个体,并保留能检测非自体的个体作为成熟检测器,模拟成熟的免疫细胞;在检测阶段,事件依次与每个检测器进行亲和力计算,亲和力高的事件(即匹配检测器的事件)被认为是异常。

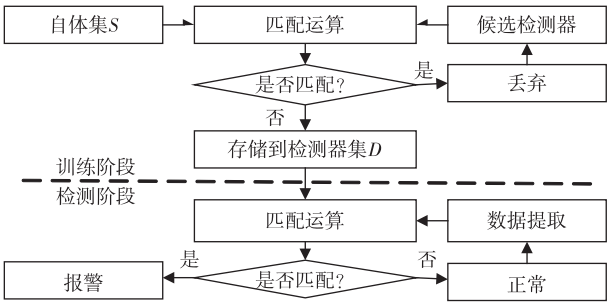

图 1　否定选择算法(训练阶段和检测阶段)

1.3　单一形态检测算法的弊端

实际应用中,不同的事件所包含的属性是不同的,不同的应用所需要的处理时间也是不同的,单一形态检测算法对这些属性一视同仁,无形中会造成许多有价值的信息的流失。

(1)二进制空间提供了正常和异常之间一个简单的“硬”的区别,而许多实际问题需要一个“软”的区别,即:检测的输出必须是一个异常度,而不是一个二进制的正常/异常、“非是即否”的输出。另外,许多问题是一个连续性的实值,用二进制表示则破坏了其本身的特性,而且二进制所采用的亲和力计算和匹配规则也不适用于这些问题。

(2)实值表示限制了检测亲和力函数的选择,目前较多使用的是 Minkowski 距离体系中的 Manhattan 距离和 Euclidean 距离,实值自体/非自体被定义为超球体。这样样本的每一维属性只是作为一个空间点的一维坐标,从而造成其属性特征被忽视,丧失了其独立性,从而对检测器的生成、更新策略带来许多问题。而且 Minkowski 距离体系的体积随空间维数增大而骤降的性质,使得实值表示无法应用于高维形态空间,限制了超球体检测器的应用范围[12]。文献[13]指出当维数超过 20 时,超球体的体积几乎降为零,如图 2 所示。这就限制了实值检测器属性域的选取。对于复杂的网络环境,要求检

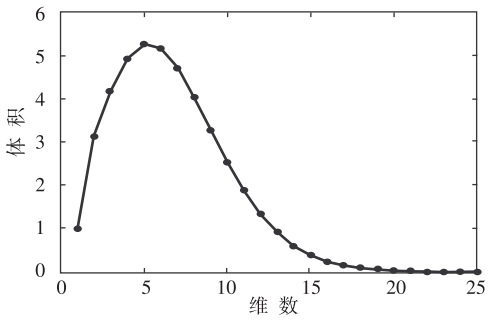

图 2　单位超立方体内切超球体的体积

测器以低维属性域来检测显然无法满足实际需要。另外,在较高维情况下,需要大量的检测器覆盖非自体空间从而增加算法的时空代价。

1.4 免疫独特型网络理论和危险理论信号机制

免疫独特型网络理论主要反映了免疫系统中多样的具有的抗原表位作用的独特型氨基酸促进 T 细胞、B 细胞等免疫细胞在一特定的免疫应答中相互作用这一特性,从而形成对抗原识别的相互连接的独特型网络[14]。

免疫危险理论(immune danger theory,IDT)认为引发机体免疫应答的关键是因有害物对机体产生损害而发出的危险信号,危险信号是由抗原提呈细胞(antigen presenting cells,APCs)识别的,并且在一定条件下 APCs 会发出一个协同刺激信号,该信号可以用来激活免疫细胞,为免疫应答提供有效的刺激[15]。危险理论模型中涉及到三种信号,包括危险信号(sig0)、抗原提呈信号(sig1)、协同刺激信号(sig2),通过三种信号的相互作用共同完成免疫应答,即如果同时接到 sig1 和 sig2 就激活免疫细胞,否则不激活(有 sig2 没有 sig1 忽略 sig2)[16]。图 3 显示的是三种信号的相互作用激发免疫应答。

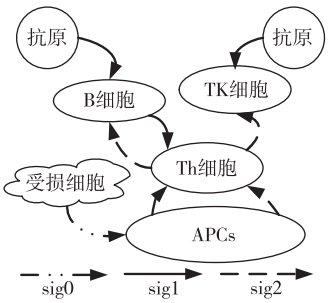

图 3 三种信号间的相互作用

2 多形态检测算法

以上分析表明,系统要考察的内容很多,单一形态检测器往往达不到要求,因此本文借鉴免疫特异性网络免疫元素多样性的特点提出一种多形态检测(multi-shape detection,MSD)算法来弥补单一形态检测算法的不足,辅以危险理论信号机制来判定当前环境是否异常。免疫独特型网络理论为 MSD 算法提供了理论支持,危险理论的信号机制为 MSD 算法的判定标准提供了方法支持。算法结合二进制和实值否定选择算法,在系统中同时加入 2 个检测模块:二进制检测模块(binary detection module,BDM)和实值检测模块(real-valued detection module,RDM)。在训练阶段时,两模块分别进行;在检测阶段,两个模块实时进行数据检测判定当前事件是否正常,并通过信号机制来实现两模块联合分析以判定当前是否有异常发生。算法模型见图 4。

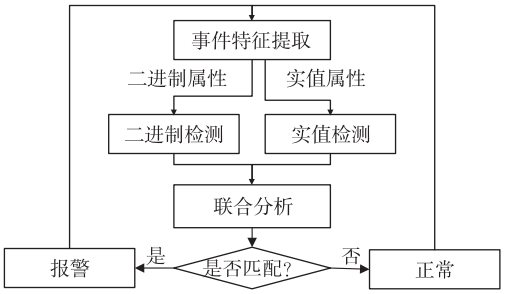

图 4 多形态检测模型

2.1 相关定义

定义 1:二进制空间自体/检测器长度为 n_{bi},实值空间维度为 n_{re}。

定义 2:多形态自体集定义为 $S=\{S_{bi},S_{re}\}$,其中 $S_{bi}=\{s_{bi}^1,s_{bi}^2,\cdots,s_{bi}^{N_s^{bi}}\}$表示二进制自体集,$N_s^{bi}$ 为二进制自体个数,每个自体为 $s_{bi}^i=[\delta_s^1,\delta_s^2,\cdots,\delta_s^{n_{bi}}]$,其中 δ_s^j $(j=1,2,\cdots,n_{bi})$ 非 0 即 1,设定二进制匹配阈值为 r_s^{bi};$S_{re}=\{s_{re}^1,s_{re}^2,\cdots,s_{re}^{N_s^{re}}\}$表示实值自体集,其中 N_s^{re} 为实值自体个数,每个自体为 $s_{re}^i=[s_{se}^{i1},s_{se}^{i2},\cdots,s_{se}^{in_{re}},r_s^i]$,其中 r_s^i 为其匹配半径。

定义 3:参照自体集的定义可以得到多形态检测器集的相关定义 $D=\{D_{bi},D_{re}\}$,其中 $D_{bi}=\{d_{bi}^1,d_{bi}^2,\cdots,d_{bi}^{N_d^{bi}}\}$表示二进制检测器集,$N_d^{bi}$ 为二进制检测器个数,每个检测器为 $d_{bi}^i=[\delta_s^1,\delta_s^2,\cdots,\delta_s^{n_{bi}}]$,其中 δ_s^j $(j=1,2,\cdots,n_{bi})$ 非 0 即 1,设定二进制检测阈值为 r_d^{bi};$D_{re}=\{d_{re}^1,d_{re}^2,\cdots,d_{re}^{N_d^{re}}\}$表示实值检测器集,其中 N_d^{re} 为实值检测器个数,每个检测器为 $d_{se}^i=[d_{se}^{i1},d_{se}^{i2},\cdots,d_{se}^{in_{re}},r_d^i]$,$r_d^i$ 为其检测半径。

2.2 亲和力计算

判断两个样本 s_i 和 s_j 是否相似主要是通过亲和力函数计算得到的。由于二进制空间和实值空间各自不同的特点,二进制检测算法和实值检测算法用到的亲和力函数也各不相同。

(1)二进制亲和力计算主要有简单匹配系数法、海明距离匹配法和 r 连续比特法(r-contiguous bytes,rcb)等。由于 rcb 更能反映样本间的相似程度,所以本文将采用它进行样本亲和力的衡量。rcb

通过计算两样本连续相同对应位的最大值是否超过匹配阈值来判定两样本是否足够相似:

$$d(s_i,s_j)=\begin{cases}1, & \exists l,m,m-l\geqslant r,s_{ik}=s_{jk},l\leqslant k\leqslant m\\ 0, & \text{其他}\end{cases} \tag{7}$$

为提高检测器生成效率,将二进制串按其各属性原有含义分片,使候选检测器的每个分片与每个自体样本对应分片进行匹配。阈值以 8 位片段为基准,根据分片大小乘以相应整数倍。例如,假定阈值为 3,则 32 位分片的匹配阈值为 12。另外,为提高匹配的效率,采用 KMP 算法进行样本间的匹配。

(2)对于实值表示,两样本亲和力计算主要通过明考斯基距离体系求匹配率。其中用的较多的是 Manhattan 距离和 Euclidean 距离。为了后面的对比试验,本文采用常用的 Euclidean 距离:

$$d(s_i,s_j)=\sqrt{\sum_{k=1}^{n}|s_{ik}-s_{jk}|^2} \tag{8}$$

2.3　信号检测处理方式

对于两个模块的检测结果的处理需要设定一个时间间隔阈值 τ, 如果两个模块之间发出的报警信号时间差在 τ 内,则认为两模块同时报警;否则,认为两个模块为不同时间间隔内的独立报警。检测处理方式如下:当这 2 个检测模块都认为是正常的则系统就认为是正常的;当 2 个检测模块都报警则认为异常发生。根据模块重要性设定不同的报警级别,如表 1 所示。

表 1　报警级别

报警级别	BDM	RDM
0(安全状态)	否	否
1(潜在威胁)	是	否
2(出现危险)	否	是
3(发现攻击)	是	是

3　分析与实验

3.1　分析

多形态检测(MSD)算法的检测效果主要依据 BDM 和 RDM 的信号进行联合检测分析,由于机制简单易行,实现起来也较容易。相比单一形态检测机制的检测方法,MSD 的优势有两点:(1)二进制和实值检测器分别监视问题的不同特性的特征,更加适应实际环境;(2)多形态检测器使得实值检测器的维数在一定程度上得到降低,从而更加有效地发挥实值检测器的检测效能。

下面具体分析一下 MSD 的检测性能。首先定义 TP 为正确肯定次数、TN 为正确否定次数、FP 为错误肯定次数、FN 为错误否定次数,则

检测率 P:正确的检测概率,定义为

$$P=(TP/(TP+FN))\times 100\% \tag{9}$$

误报率 P_f:正确事件被误检测为异常的概率,定义为

$$P_f=FP/(TN+FP)\times 100\% \tag{10}$$

设 2 个检测模块分别为 A 和 B, 为简单起见,在此仅分析发现攻击状态的检测率和误报率,分别设为 P_A 、P_{fA} 和 P_B 、P_{fB}。系统总检测率和总误报率分别为 P 和 P_f, 则

$$\begin{aligned}(1-P)&=(1-P_A)+(1-P_B)(1-(1-P_A))\\&=(1-P_B)+(1-P_A)(1-(1-P_B))\end{aligned} \tag{11}$$

即 $P=P_A\cdot P_B$。因为 $0<P_A,P_B\leqslant 1$, 所以 $P\leqslant P_A$ 且 $P\leqslant P_B$。同理, $P_f=P_{fA}\cdot P_{fB}$, 则 $P_f\leqslant P_{fA}$ 且 $P_f\leqslant P_{fB}$。这就说明,2 个检测模块的协作可以有效地降低误报率,但同时也降低了检测率。但是由于 MSD 设置了不同的报警级别,所以实际检测率并不降低而是更加细化和准确。具体见下面的实验分析。

3.2　实验

本文选用 KDD CUP1999 数据集进行测试实验[17]。该数据集是本领域的权威数据集,每条记录有 41 个属性(32 个连续型属性和 9 个离散型属性)和一个标志位(1 个正常标志和 22 个攻击名称标志)。本实验选用其中一个 10% 的子集,包含 494021 条记录,其中 396473 条异常,97278 条正常。实验以正常记录为自体集训练检测器并使用这些检测器检测数据集中的异常。实验通过对比 BNS、RNS 和 MSD 的检测结果来验证的 MSD 的有效性。

(1)不同算法对比测试

实验前需要对数据集进行分析和预处理。由于第 7、8、9、11、15、18、20、21 维属性多数情况下为 0,最后 10 维属性涉及连接双方的属性,这些属性本实验不做考虑。而且经主成分分析(principal component analysis,PCA)后发现,数据集的前 14 维属性的累计贡献率已达到 87.151%,如图 5 所示。一般情况下,只需要选择累积贡献率超过 85% 的主成分,便可以保证在信息量足够的前提下,构造维数最低

的主成分空间[18]。所以,对于 RNS,选取 PCA 处理后按贡献度大小排序后的前 14 维属性构成样本并

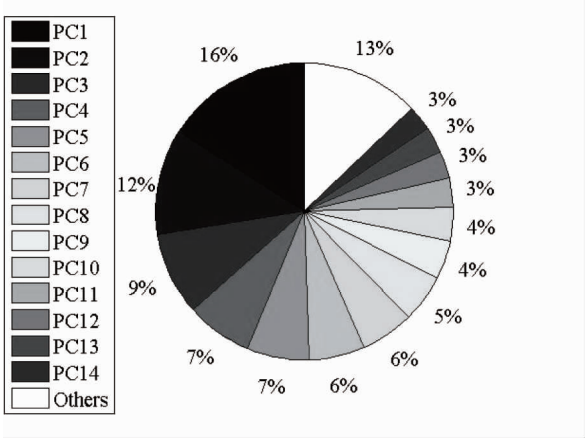

图 5 KDD CUP 1999 数据集的部分主成分的贡献率

正规化处理;对于 BNS,除了数据集中原有的 9 维二进制属性外,选取实值属性 PCA 后的前 5 维属性分别离散化后转换为二进制属性来构成 BNS 的属性域空间;对于 MSD 中的 BDM 选取样本的 9 维二进制属性组成,选取 PCA 后的前 5 维属性分别组成 RDM 的样本属性域并正规化处理。

实验对三种算法分别生成 5000 检测器。为保持一致,对于 BNS,设定 r_s 和 r_d 均为 $|0.6 \cdot n_s^j|$;对于 RNS,设定 r_s 和 r_d 均为 0.05;对于 MSD,设定 r_s^{bi} 和 r_d^{bi} 均为 $|0.6 \cdot n_s^j|$,r_s^i 和 r_d^i 均为 0.05。由于数据集为离线数据,所以在此设定 τ 为每个事件的检测时间记为一个周期。实验总的检测结果见表 2。

表 2 检测结果

算法	攻击数	误报数	误报率(%)	检测数	检测率(%)	威胁数	危险数	攻击数
BNS		95897	19.41	110876	27.97	—	—	—
RNS	396473	11381	2.30	285867	72.10	—	—	—
MSD		4326	0.88	371445	93.69	47344	43451	305678

从表中结果对比可以看出:① BNS 的检测率最低,而误报率却很高,从中可以推测造成检测效果如此不好的主要原因是 BNS 的二进制属性的检测性能本身的问题,而且丢弃过多的实值属性对算法的影响很大;② RNS 的检测结果较 BNS 好许多,这主要是因为由 PCA 得到的主成分构成检测器可以保留足够的信息量,从而在维度合适的情况保持较高的检测性能;③ MSD 的检测结果相比 BNS 和 RNS,无论是检测率还是误报率,效果都很好,这就看出,虽然 BDM 属性选取的是原始数据集的离散型属性,RDM 选取的主成分属性也不多,但二者相互合作,从而可以保证系统具有较高的检测性能,而且,依据不同情况设置报警级别也可以为系统或管理员提供更加细化的参考,从而及时做出正确的处理。

(2)算法对不同异常的检测实验

为了进一步考察 MSD 算法对于不同攻击类型的检测效果和处理级别,在此考察数据集中所占比例排序前 8 的攻击类型的检测效果,分别是:back、ipsweep、neptune、portsweep、satan、smurf、teardrop、warezclient。为了对比检测效果,同样采用 BNS 和 RNS 进行对比试验。属性域选取策略、各种参数设置、测试方案设计等同上。记录被检测为异常的事件,统计其中为这 8 类攻击类型的事件的数量,检测结果见表 3。

从表中结果对比可以看出:① BNS 检测效果较好的攻击类型在用 RNS 检测时效果很差(如 back 等),反之亦然(如 neptune 等),这就说明单一形态检测各有其擅长的领域,也各有其欠缺的地方;②某些攻击用单一形态检测方式效果均不理想(如 teardrop、smurf 等),这就说明单一形态检测的不完备性,也间接说明进行多形态检测的必要性,从而达到"1 +1 >2"的效果;③ MSD 对于某些危害性很大的攻击类型(如 neptune 等,属于 DoS 攻击,危险性极高)和一些危害性较小的攻击类型(如 satan,属于 Probing 攻击,危害性较小)判定精确,可以准确有效地判定当前环境的危险程度,这是 BNS 和 RNS 简单的判定机制所无法达到的;④ 从中也发现对于有些攻击类型三种方法都不理想(如 ipsweep,也属于 Probing 攻击),这就说明某些类型的攻击对于亲和力计算类型的检测规则而言,其攻击特征不明显,需要进一步优化算法。

综合以上实验可以看出,MSD 的检测效果不仅从整体上优于单一形态检测方法,而且对于各种攻击的检测效果也得到了大幅改善,并且不同的危险级别也可以更有利于系统或管理员的进一步处理。

表 3　检测结果

算法	攻击	攻击数	检测数	检测率(%)	威胁数	危险数	攻击数
BNS			1903	86.38	—	—	—
RNS	back	2203	132	5.99	—	—	—
MSD			2056	93.33	287	579	1190
BNS			593	47.55	—	—	—
RNS	ipsweep	1247	341	27.35	—	—	—
MSD			579	46.43	398	157	24
BNS			204	0.19	—	—	—
RNS	neptune	107201	107201	100	—	—	—
MSD			107201	100	6647	56822	43732
BNS			370	35.58	—	—	—
RNS	portsweep	1040	1037	99.71	—	—	—
MSD			1037	99.71	682	303	52
BNS			386	29.33	—	—	—
RNS	satan	1589	1475	92.83	—	—	—
MSD			1475	92.83	1043	353	79
BNS			9	0.003	—	—	—
RNS	smurf	280790	54	0.02	—	—	—
MSD			280410	9986	12324	17939	250147
BNS			246	25.13	—	—	—
RNS	teardrop	979	199	20.33	—	—	—
MSD			479	48.93	39	98	342
BNS			491	48.14	—	—	—
RNS	warezclient	1020	204	20.00	—	—	—
MSD			683	66.96	471	106	6

4　结　论

本文分析了基于免疫机制的入侵检测中单一形态检测器存在的问题,借鉴免疫特异性网络理论和免疫危险理论中的信号机制,提出了多形态检测(MSD)算法,设计其中的不同检测模块:BDM 和 RDM,并给出检测方法。MSD 算法的检测效果主要依据 BDM 和 RDM 的信号联合检测分析,相比单一形态检测机制,MSD 的主要优势可以体现在以下几点:(1)二进制和实值检测器分别监视问题的不同特性的特征,弥补对方在检测方面的不足,从而使系统更加地适应实际环境;(2)BDM 可以使 RDM 的属性域处于一个更低的水平,使得实值检测器的维数在一定程度上进一步降低,从而更加有效地发挥实值检测器的检测效能;(3)通过信号机制实现检测模块的联合检测,操作简单易行,实现起来也较容易。实验也表明,模型的两个检测模块可以很好地进行协作,从而更好地服务系统,提高系统的检测性能。在过去,国内外学者把主要精力集中在单一形态检测器的研究上,很少涉足多形态检测器,本文提出的算法弥补了这一空白,但也还有很多不足,比如更全面的实验分析,特别是对某些攻击检测效果不理想的分析等,这也是我们下一步要研究的重点,拟通过提取更合适的属性、亲和力计算方法的设计或者增加起补丁作用的算子来解决。

参考文献

[1] Dal D, Abraham S, Abraham A, et al. Evolutionary induced secondary immunity: an artificial immune systems based intrusion detection systems. In: Proceedings of the 7th Computer Information Systems and Industrial Management Applications, Ostrava, Czech Republic, 2008. 65-70

[2] Dasgupta D, Yu S, Nino F. Recent advances in artificial immune systems: models and applications. *Applied Soft Computing*, 2011, 11(2): 1574-1587

[3] Gonzalez F, Dasgupta D, Gomez J. The effect of binary matching rules in negative selection. In: Proceedings of 2003 the Genetic and Evolutionary Computation Confer-

ence, Heidelberg: Springer Berlin, 2003. 198-209

[4] Chmielewski A, Wierzchon S T. Simple method of increasing the coverage of nonself region for negative selection algorithms. In: Proceedings of the 6th International Conference on Computer Information Systems and Industrial Management Applications, Elk, Poland, 2007. 155-160

[5] Gonzalez F, Dasgupta D. Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines*, 2003, 4(4): 383-403

[6] Elberfeld M, Textor J. Negative selection algorithms on strings with efficient training and linear-time classification. *Theoretical Computer Science*, 2011, 412(6): 534-542

[7] 王辉, 毕晓君, 于立君等. 基于疫苗理论的变阈值免疫阴性选择算法. 哈尔滨工程大学学报, 2011, 32(1): 69-72.

[8] 张凤斌, 席亮, 王胜文等. V-detector 优化算法. 高技术通讯, 2012, 22(5): 449-454

[9] Shapiro J M, Lamont G B, Gilbert L. P. An evolutionary algorithm to generate Hyper-ellipse detectors for negative selection. In: Proceedings of 2005 Genetic and Evolutionary Computation Conference, Washington, D. C., USA, 2005. 145-156

[10] Balachandran S, Dasgupta D, Nino F, et al. A framework for evolving multi-shaped detectors in negative selection. In: Proceedings of 2007 IEEE Symposium on Foundations of Computational Intelligence, Honolulu, USA, 2007. 401-408

[11] Zhou J, Dasgupta D. Revisiting negative selection algorithms. *Evolutionary Computation*, 2007, 15(2): 223-251

[12] Stibor T, Timmis J. Commnets on real-valed negative selection vs. real-valued positive selection and one-class SVM. In: Proceedings of 2007 IEEE Congress on Evolutionary Computation, Singapore, 2007. 3727-3734

[13] Stibor T, Timmis J, Eckert C. On the use of hyperspheres in artificial immune systems as antibody recognition regions. In: Preceeding of the 5th International Conference on Artificial Immune Systems, Oeiras, Portugal, 2006. 215-228

[14] Jerne N K. Towards a network theory of the immune system. *Annual Immunology*, 1974, 125C: 373-489

[15] Greensmith J, Aickelin U, Twycross J. Detecting danger: applying a novel immunological concept to intrusion detection systems. In Proceedings of the 6th International Conference in Adaptive Computing in Design and Manufacture, Bristol, UK, 2004. 1-3

[16] Matzinger P. The danger model: a renewed sense of self. *Science*, 2002, 296: 301-305

[17] KDD Cup 1999 Data. http://kdd. ics. uci. edu/databases/ kddcup99/ kddcup99. html, 2011

[18] 王大伟. 基于免疫的入侵检测系统中检测器性能研究[D]. 哈尔滨理工大学博士学位论文, 2010: 57-58

A multi-shape detection algorithm for immunity-based intrusion detection systems

Xi Liang, Zhang Fengbin, Liu Hailong

(College of Computer Science and Technology, Harbin University of Science and Technology, 150080 Harbin)

Abstract

To solve immunity-based intrusion detection systems' lower detection performence caused by their detectors' shape kinds of the binary and the real-valued, the defects of the two kinds of shape were analyzed, and a new multi-shape detection (MSD) algorithm was presented according to the idiotypic immune network theory and the signal mechanism in the immune danger theory. The MSD algorithm uses binary and real-valued detectors to monitor the events by collaborative signals. The experimental results show that the new detection algorithm can improve the detective performance of detectors.

Keywords: immunity-based intrusion detection, shape-space, binary, real-valued, detector, multi-shape, negative selection algorithm (NSA)