一种基于多级否定选择的入侵检测器生成算法

冯艳华[1],钟 诚[1],李 智[1,2]
(1. 广西大学 计算机与电子信息学院,广西 南宁 530004;
2. 广西科技信息网络中心,广西 南宁 530012)

摘 要:文中给出一种改进的基于人工免疫入侵检测系统的否定选择算法。首先是用多级否定选择算法生成不同检测尺度的成熟检测器,然后为了模仿人体免疫系统中的第二次应答机制,引入了记忆检测器的概念及相应的算法,结合亲和力成熟与体细胞突变等方法,将成熟检测器提升为识别率极高的记忆检测器。

关键词:入侵检测;人工免疫系统;多级否定选择;克隆选择
中图分类号:TP301.6 **文献标识码**:A **文章编号**:1005-3751(2006)04-0234-03

An Algorithm of Generating Intrusion Detector Based on Multi-Level Negative Selection

FENG Yan-hua[1], ZHONG Cheng[1], LI Zhi[1,2]
(1. School of Computer and Electronics and Information, Guangxi University, Nanning 530004, China;
2. Guangxi Science and Technology Information Network Center, Nanning 530012, China)

Abstract: Presents an improved negative selection algorithm in intrusion detection system based on the artificial immune system. The different scales mature detectors are generated by the multi-level negative selection algorithm. In order to simulate the second respond mechanism of the human immune system, introduce the concept of memory detector and corresponding algorithm. The algorithm integrates the affinity maturation and somatic hypermutation to generate the memory detector which has the high recognition ability.

Key words: intrusion detection; artificial immune system; multi-level negative selection; clonal selection

0 引 言

随着 Internet 的飞速发展,计算机网络已经逐渐成为整个社会基础设施中最重要的一部分。为了确保计算机网络系统安全,入侵检测显得尤为迫切和重要。建立一个有效的以网络为基础的入侵检测系统 IDS,需要达到 3 个目标:分布、自组织和轻负担。而人体免疫系统的分布性、多样性、层次性、自适应性、自动应答和自我修复的特点正好满足了这些要求。

Forrest 在将人工免疫系统(AIS)应用到 IDS 的研究时,提出用否定选择算法(Negative Selection Algorithm)来产生成熟检测器[1]。但实验表明,此算法有几个缺点:

1)算法时间复杂度是指数级的,当问题空间太大时可行性不高[2,3];

2)它假定自我集被随机分布在整个空间,不太符合实际问题;

3)容易产生漏洞[4]。

针对上述这些问题,文中对基于人工免疫的入侵检测系统的否定选择算法进行改进。其主要思想是:首先采用多级否定选择算法来生成成熟检测器,然后再用相应的记忆检测器生成算法将成熟检测器提升为记忆检测器。

1 成熟检测器的生成

1.1 否定选择算法

Forrest 提出利用否定选择算法来处理各种异常检测问题。这种算法定义"Self"为一个网络监测系统中的正常行为模式。通过否定选择算法来生成用于检测异常的检测器,其基本思想可通过图 1、图 2 来描述。

1.2 漏洞的产生

根据已使用的匹配规则和串 S 的结构,用 rcb 匹配规则可能会产生一个检测器无法检测的非己串。现在,用一

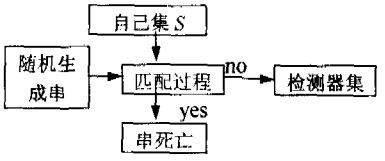

图 1 否定选择算法生成成熟检测器集过程

收稿日期:2005-07-29
基金项目:广西科学基金资助项目(桂科自 0339008);广西科技信息网络中心和广西大学博士科研基金资助课题(B0309031)
作者简介:冯艳华(1980-),女,湖北天门人,硕士研究生,研究方向为网络信息安全;钟 诚,博士,教授,研究方向为网络信息安全、并行计算。

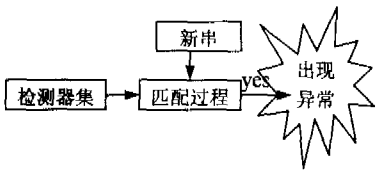

图 2　通过成熟检测器检测异常

个例子来说明漏洞的存在。假设有 3 个字符串 S1 = 001101,S2 = 111111,S3 = 001111,对于 r－连续位匹配(r = 3),匹配 S3 的子集有 $D3 = S_{001***} \cup S_{*011**} \cup S_{**111*} \cup S_{***111}$,这里 S_{001***} 是由模块 001＊＊＊产生的一个 6 位的字符串子集,其中＊表示 0 或 1 中的任意一个,D3 中所有的检测器都与 S1 或 S2 匹配。如果 S1,S2 是自己集的一部分,而 S3 是非己集,那么任何匹配 S3 的检测器都要匹配自己集,所以都不可能通过耐受训练。最终不可能有识别 S3 的检测器,S3 就成了一个漏洞。

1.3　MNS 算法

否定选择算法容易产生漏洞,并且假定自己集随机分布在整个空间,根据文献[5]可知,这不太符合现实。而 MNS(多级否定选择)算法则假定:自己集不是随意分布而是聚集在某个子空间且自己集仅仅占用了整个空间的一小部分。

多级否定选择算法的主要思想是综合克隆选择和否定选择生成不同尺度的检测器,克隆选择使用局部最佳技术产生更精确识别力的检测器。

文中采用多级否定选择算法来生成成熟检测器,具体算法描述如下:

```
Begin
{
  产生一个大尺度的检测器,使得识别器的识别空间总数覆盖整个空间;
  if (识别距离>阈值)
    {
      进行克隆选择,生成识别距离更小的检测器;
      经历否定选择;
    }
}
End
```

此算法中的阈值可保证所需的精度,所获得的检测器具有不同的尺度,即不同的检测能力。为了简化描述,以二元空间为例,假定检测器的坐标是 (x,y),识别距离是 dis,那么它的识别空间是一个正方形 $(x-\mathrm{dis}, y-\mathrm{dis})$,$(x-\mathrm{dis}, y+\mathrm{dis})$,$(x+\mathrm{dis}, y-\mathrm{dis})$,$(x+\mathrm{dis}, y+\mathrm{dis})$,若此检测器经否定选择不匹配自己模式,则克隆 4 种级别 $(x-\mathrm{dis}, y-\mathrm{dis})$,$(x-\mathrm{dis}, y+\mathrm{dis})$,$(x+\mathrm{dis}, y-\mathrm{dis})$,$(x+\mathrm{dis}, y+\mathrm{dis})$的检测器,其中 dis 变为 dis/2,即这 4 个检测器的识别距离变为 dis/2 且能覆盖父检测器的识别空间,然后 4 种检测器再经历否定选择,不匹配再分别对这 4 个检测器进行克隆,此过程不断继续下去,直到检测器识别距离达到设定的阈值为止。

2　记忆检测器的生成

2.1　生成记忆检测器的算法

由于人体免疫系统采用了“记忆免疫细胞”的方式快速识别再次出现的抗原,几乎在人体未能感知的情况下消除了有害入侵。在基于人工免疫的入侵检测系统中,也模仿了该机制来构造人工记忆免疫细胞即记忆检测器[6,7]。文中提出了记忆检测器的生成算法,先判断由多级否定选择产生的检测器是否被激活(即检测到入侵),然后对被激活的检测器进行一系列亲和力成熟与体细胞突变。

具体算法描述如下:

```
Begin
{
  while(还有训练数据未比较)
  {
  (1) 将一个训练数据与被激活的检测器进行比较,找到一个最好亲和力的检测器,将此检测器设为 MCmatch;
  (2) 根据亲和力以一定比率克隆和变异 MCmatch,并将克隆和交叉变异后的检测器放在 B 细胞池中;
  (3) 计算 B 细胞与此训练数据的亲和力;
  (4) If(B 细胞的平均亲和力>某一水平)
      {
      选择最好的 B 细胞作为候选记忆检测器,设为 MCcand;
      }
      else
        {
        克隆变异这些 B 细胞;
        goto (3);
        }
  (5) if (MCcand 亲和力 > MCmatch 亲和力) and (MCcand 与 MCmatch 之间的亲和力≤某一阈值)
      将 MCcand 提升为记忆检测器;
      }
      }
End
```

其中的克隆选择操作根据亲和力和浓度概率来决定,亲和力越大,则被选择的概率越大;个体浓度越大(即相似的个体越多),选择的概率就越小,变异操作时通过如下步骤进行:

1) 选择两个个体杂交产生后代 child;

2) if dist(child, parent1) ≤ dist(child, parent2) and fitness(child) > fitness(parent1) then

　　用 child 代替 parent1

　else if dist(child, parent1) ≥ dist(child, parent2) and fitness(child) > fitness(parent2) then

　　用 child 代替 parent2;

其中 dist(x,y) 表示 x 和 y 之间的距离,距离越小说明 x 和 y 越相似,fitness(x) 表示 x 的亲和力。

这样既可保留亲和力高的个体,又确保个体的多样性和分布均衡。在将候选检测器提升为记忆检测器时,对候选检测器设一个确认机制,若它的生命周期已到或生命周

期内确认机制确认其检测到一个自己模式时，则将该候选检测器删除；否则，执行步骤(5)。

2.2 记忆检测器的淘汰策略

当记忆检测器群已满时，需要一种淘汰策略。在人工免疫中使用得最多的就是随机淘汰策略[8]，它的最大缺陷是可能会删除最近添入的记忆检测器或记忆近期经常出现的入侵行为的记忆检测器，显然相应的入侵行为再次出现的概率是很高的，它的再出现就只能由成熟检测器去检测了，影响了入侵检测系统的反应速度。这里引入了亲和力技术来减少总的记忆细胞数。其主要思想是：在所有记忆检测器中计算它们之间的亲和力，如果两记忆检测器之间的亲和力小于亲和力阈值(满足公式(1))，那么只保留其中亲和力高的记忆检测器。

$$\text{affinity}(mc_i, mc_j,) < AT * ATS \tag{1}$$

$$AT = \frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}\text{affinity}(ag_i, ag_j)}{\frac{n(n-1)}{2}}$$

其中 mc_i 和 mc_j 代表两个不同的记忆检测器，AT 是亲和力阈值，ATS 是亲和力阈值标量(由用户设定)，ag_i 和 ag_j 代表两个不同的非己模式，n 为总的非己模式数。

3 算法分析

在多级否定选择算法中假定：自己集不是随意分布而是聚集在某个空间且自己集仅占用了整个空间的一小部分。下面，用典型的测试数据验证了这个假设。首先，设检测器 detector(x, y, z, distance)，x 为每秒钟的字节数，y 为每秒钟的 IP 包数，z 为每秒钟的 ICMP 包数，distance 为检测尺度，那么检测器的识别空间为一个立方体。开始时定义 detector(0.5,0.5,0.5,0.5)，此检测器经过否定选择，不匹配自己集时将生成 8 个新的检测器，它们的检测尺度为 0.25。这里定义门限值为 0.01，那么将克隆 5 种尺度的检测器，分别为 1/4，1/8，1/16，1/32，1/64。最后获得 120 个检测器，能识别的非己空间占 99.9573%，那么正常行为占整个空间的 0.0427%。这正好符合前面的假设。与传统的否定选择算法比较，多级否定选择算法能有效减少漏洞。漏洞的产生大部分是因为正常行为的边界不稳定，用单尺度检测器有些区域无法检测，而 MNS 算法能根据边界产生相应尺度的检测器，能识别更多的非己模式，而且其中小尺度检测器的产生能使检测精度提高。

关于记忆检测器生成算法，首先选出与训练数据亲和力最好的检测器进行克隆和变异，在其中提升一个成为候选记忆检测器，当候选记忆检测器满足一定条件时才提升为记忆检测器，因为记忆检测器与一般的成熟检测器相比具有较小的阈值和较长的生命周期，一旦同一入侵再次出现时，记忆检测器将迅速做出反应，这样既实现了记忆机制也提高了检测速度。

4 结束语

文中给出了一种否定选择算法的改进算法，以动态的方式生成有效地检测器，此方式生成的检测器的检测区域不相互重叠且有不同的识别尺度，有效地减少了漏洞，提高了识别率。为实现第二次应答机制还引入了记忆检测器，利用亲和力成熟算法使最终得到的记忆检测器具有最高的亲和力，提高了整个系统的检测率和速度。最后还给出一个记忆检测器的淘汰策略，使整个系统的性能进一步优化。

参考文献：

[1] Hofmeyr S, Forrest S. Architecture for an Artificial Immune System[J]. Evolutionary Computation, 2000,7(1):45-68.

[2] Harmer P K, Williams P O, Gunsch G H, et al. An artificial immune system architecture for computer security applications [J]. IEEE Transactions on Evolutionary Computation, 2002,6 (3):252-280.

[3] Kim J, Bentley P J. Evaluating Negative Selection in An Artificial Immune System for Network Intrusion Detection[A]. Genetic and Evolutionary Computation Conference 2001 (GECCO-2001)[C]. San Francisco:[s.n.],2001.1330-1337.

[4] Zhang jian, He hua-can, Zhao min. Hybrid Detector Set: Detectors with Different Affinity[A]. Conference '04[C]. Shanghai:[s.n.],2004.

[5] Heberlein L T, Dias G V, Levitt K N, et al. A Network Security Monitor[A]. Proc. Symposium on Research in Security and Privacy[C]. [s.l.]:IEEE Press,1990.

[6] Watkins A, Timmis J, Boggess L. Artificial immune recognition system (AIRS): An immune inspired supervised machine learning algorithm[J]. Genetic Programming and Evolvable Machines,2004,5:291-317.

[7] Watkins A, Timmis J. Exploiting Parallelism Inherent in AIRS, an Artificial Immune Classifier[EB/OL]. http://www.cs.kent.ac.uk/? abw5/,2004.

[8] Kim J, Bentley P. Immune Memory in the Dynamic Clonal Selection Algorithm[A]. Proceedings of the First International Conference on Artificial Immune[C]. Canterbury:[s.n.], 2002.57-65.