一种基于免疫的容侵触发器设计①

An Immunity-Based Trigger Model for Intrusion Tolerance System

王绍卜 (浙江万里学院 商学院 信息管理系 浙江 宁波 315100)

摘 要: 容侵技术是指系统在遭受攻击的情况下连续提供服务的能力。基于人工免疫思想，结合数据挖掘技术KNN，设计了一个基于免疫分类算法的容侵触发器模型，详细描述了其工作原理和模块结构。结果表明，该模型具有一定的价值。

关键词: 网络安全 生物免疫原理 容侵 数据挖掘 触发器

1 引言

随着计算机网络的快速发展和广泛应用，各行各业对网络信息系统的依赖程度也越来越高。与此同时，网络安全事件也逐年上升，网络安全问题正成为人们关注的焦点。传统的网络安全技术主要是通过保护和检测手段来保障网络的安全，随着网络信息系统规模日益庞大，并朝着高度的分布式方向发展(即无边界系统)，传统的安全技术日益显得力不从心。对于目前常用的检测技术如防火墙和入侵检测来说，要发现全部的攻击是不可能的，仍然有一些攻击会取得成功。因此，必须有新的安全理念来保证网络的安全性。

容侵技术就是认同安全问题的不可避免性.针对安全问题,不再将消除或者防堵作为第一重点,而把目光投射到如何在攻击之下系统仍能保证不间断地正常运行这一点.譬如一个金融系统在遭到网络攻击的情况下,仍旧保持正常的交易,不致因系统的崩溃造成巨大的损失.也就是说,在存在攻击,错误或突发事件的情况下,系统仍然可以及时地完成它的使命.自然界存在各种病毒和细菌，由于人体具有一定的免疫性，抵制和瓦解了大约 90％以上的进攻和入侵。换句话讲，由于人体的免疫性可以容忍大约 90%以上的入侵，人体就是一个很好的容侵系统。计算机网络的安全问题与生物免疫系统所遇到的问题具有惊人的相似性，两者都要在不断变化的环境中维持系统的稳定性，网络安全系统实际上具有免疫系统同样的目标和功能。因此，可以借鉴生物免疫系统的多样性、自适应、鲁棒性等特点来发展新的网络安全技术。

本文借鉴生物免疫原理，提出了一个基于免疫原理入侵容忍触发器模型，并对模型中的核心免疫聚类器进行了细致的描述，在一定程度上保障了网络的安全性。

2 生物免疫原理

生物免疫系统 BIS 是一个非常复杂的自然防御系统，具有分析、学习进入体内的外来物质 (抗原)，并通过产生抗体来消灭入侵抗原的特点，其基本功能就是区别自身和非自身，并将非自身分类清除。BIS 具有免疫识别，免疫记忆、免疫调节等功能特征，并保持机体本身的平衡，生存和发展。其中，免疫识别不仅能识别已知抗原，还能识别未知抗原。免疫记忆则能够对再次入侵的抗原发生快速反应 (即二次应答)。生物免疫系统(BIS)在不断变化的有机体环境中起到了保护自身、抵御外来入侵的作用。借鉴生物免疫原理，并根据已提出的克隆选择算法和免疫网络算法，利用记忆细胞对后续入侵的病毒菌产生特定、快速应答的特性，并结合数据挖掘中数据聚类及分类的概念，构造了一个对容侵系统中服务器运行状态进行实时监控的入侵容忍系统触发器[1]。

① 收稿时间:2009-01-14

3　基于免疫的容侵系统触发器

3.1 基本思想

在免疫触发器中，将衡量服务器所运行状态的数据集看作为抗原集 AG，将记忆细胞看作为抗体，于是该触发器的工作原理可看成为利用记忆细胞不断识别抗原 Ag，并最终形成用于分类的记忆细胞集 M。为此，将服务器运行可能出现的状态抽象表示为 $s=(s_1,s_2,...,s_k)$，其中 k 为正整数，表示状态类别数目。

3.2 免疫算法相关技术

3.2.1 抗原与抗体、亲和力

服务器的状态是由多个衡量系统的指标参数共同决定的，每个指标都代表了服务器某个部分的性能水平，所有指标之和就代表了整个服务器的状态，故将抗原集 AG 表示为：$AG=(Ag_1,Ag_2,Ag_3,...,Agn)^T$，n 表示抗原集的个数，即从服务器提取的数据集；其 l 表示抗原的维数，即构成服务器状态的指标个数。记忆细胞(抗体)的构造与抗原构造相同，它是由从抗原集 AG 中不分类别的随机抽取 m 个抗原组成，通过其与相同类别的抗原相互作用，不断识别抗原，最终形成用于分类的记忆细胞集 M 。

亲和力是指记忆细胞和抗原之间的匹配程度，也就是免疫系统识别过程中利用抗原 Agj 寻找最大亲和力记忆细胞 mi 的过程。衡量亲和力大小有多种办法，考虑本文是基于实数编码，故采用欧氏距离来衡量亲和力[2]。

3.2.2 克隆选择和变异

记忆细胞的克隆选择和克隆变异是其与抗原的亲和力程度相关的，当亲和力大于某个事先定义的阈值时，记忆细胞便开始克隆阶段。在克隆阶段，记忆细胞的繁殖数目与其亲和力成正比，即当亲和力越大，该细胞繁殖的后代数目就越多；相反，亲和力越小，繁殖的后代数目就越少。克隆变异的规则和记忆细胞的亲和力大小相关的，即亲和力越大，变异率越小。克隆变异使得记忆细胞朝着更贴近抗原的方向变化，通过不断的克隆迭代，最终形成一个能够较好识别抗原的记忆细胞集，并用此进行数据分类。

变异记忆细胞成熟前还必须经过一个克隆抑制过程，该过程是通过亲和力和相似度作用于记忆细胞，以淘汰那些亲和力较低和相似度较高的记忆细胞，以产生亲和力较高且相似度较低的成熟记忆细胞，同时还可通过事先定义的阈值来控制整个记忆细胞集 M 的规模[3]。

3.3 免疫算法过程

在本文中，抗原集作为已知的样本，代表从服务器不同状态下获取的数据，而记忆细胞则作为该免疫算法所进化的候选聚类点，最终所获的记忆细胞即为样本组的聚类点。根据以上相关问题的定义，以及对免疫过程的描述，该算法的详细步骤如下：

(1) 输入样本数据集。即输入 n 个抗原 Agj，j=1,2,⋯,n；

(2) 产生初始记忆细胞群。确定克隆进化的代数 p，作为算法终止的条件。不分类别的随机产生初始记忆细胞群 M；

(3) 针对抗原集 AG 中所有抗原，对记忆细胞集 M 每次只提呈一个抗原 Agj，并进行下列运算：

Step1 亲和力计算。对于提呈的抗原 Agj,计算记忆细胞集中那些与其类别相同的记忆细胞亲和力；

Step2 克隆选择。根据记忆细胞的亲和力阈值，由大到小的选取一定的数目(t)作为克隆父代，计算克隆繁殖数目；

Step3 克隆突变。对于上步中的所有记忆细胞进行克隆突变，突变使其更加贴近抗原；

Step4 克隆成熟。对变异记忆细胞集再按亲和力大小，并选取一定的记忆细胞，进行相似度抑制.

Step5 内循环判断。若 j<n, 则置 j←j +1，返回 step3; 否则，进入下一步；

Step6 记忆细胞集 M 的构成及抑制。反复重复 Step3 之后，将得到的所有合并

Step7 外循环判断。若进化代数满足 p，则将形成的最后一代记忆细胞集 M 作为对于新抗原分类的记忆细胞。

该算法中，初始记忆细胞集根据提呈的抗原，自动完成对其自身的聚类，根据自身所带有的类别特征，确定不同的聚类类别，并且通过对各个调节参数的设定，最终达到以较为少量的记忆细胞来识别大量抗原的目的。其中，Step3 是算法的核心，通过其不断的重复，增强了每个聚类中的记忆细胞识别同类抗原的能力，保证了记忆细胞始终朝着更好识别抗原的方向变化。而通过 Step5 中的细胞抑制，则保证了各个聚类间的边界划分，使得聚类更为清晰[4]。

4　容侵触发器模型与分析

4.1 容侵触发器模型

该触发器结构如下图所示：

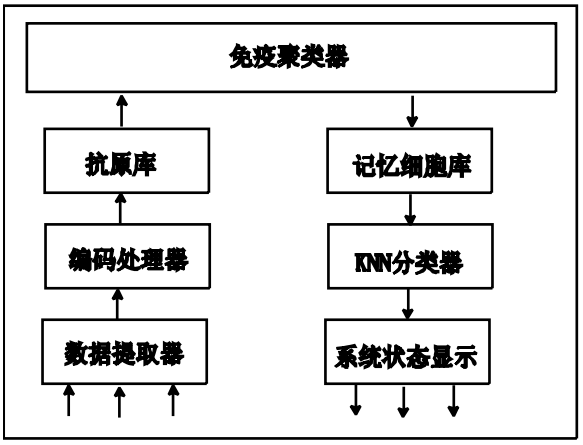

图 1 容侵触发器模型

该容侵触发器是一个实时的模块，对其所在的服务器状态进行实时地、动态地监控，容侵系统就可以根据该触发器所反应出的当前服务器状态，提供不同级别的服务和执行不同策略的安全保护。该触发器主要由以下几个子模块组成：

(1) 数据提取模块，主要负责收集能衡量服务器系统实时运行状态的数据。数据提取的范围包括 CPU 消耗、内存占用率、网络宽带资源消耗、系统核心文件长度等；提取的方式按照时间序列离散取值，将某一时刻所提取的数据构成一个原始抗原；

(2) 数据编码模块，对从数据提取器中递交来的原始抗原，按照免疫算法的要求进行标准化处理；

(3) 抗原库，将标准化处理后的抗原统一集中在抗原库中，为免疫聚类模块提供数据输入；

(4) 免疫聚类模块，对抗原库中的抗原进行克隆繁殖、变异，生成成熟记忆细胞；

(5) 记忆细胞库，储存聚类后的记忆细胞，为 KNN(K 最近邻法)分类模块提供对新抗原分类的标准；

(6) KNN 分类模块，给新提呈的抗原集做出类别标记；

(7) 系统状态显示模块，根据(6)中的标记，判断并显示系统当前所运行的状态[5]。

4.2 容侵触发器的工作流程

该容侵触发器的工作分为两个阶段。第一阶段为容侵触发器的测试学习阶段，称之为免疫聚类，即通过对服务器上提取的不同状态下样本数据的学习，经历一系列的生物免疫变化过程，产生对抗原具有记忆识别作用的记忆细胞集，并形成对具有不同类别的记忆细胞聚类集；第二阶段为实时运行阶段，称之为状态分类，即利用第一阶段中产生的记忆细胞聚类集，结合 KNN 分类特性，对系统运行中实时提取的抗原数据做出类别的鉴定，并由此反映出该抗原提取时间段中服务器的运行状态。

4.3 容侵触发器的特性分析

该触发器是在融合了常用的误用检测和异常检测的思想方法上，借鉴生物免疫系统的多种免疫机制，结合数据挖掘技术中相关算法来设计和实现的，其特性总结如下：

(1) 触发器监控的方式在于对服务器自身性能的监控，而与攻击方式无关；

(2) 触发器结构被设计成为一个集成于监控服务器之上的模块，并通过对免疫聚类中记忆细胞生成规模的控制，可以大大节省服务器系统资源和网络资源的消耗；

(3) 通过对生物免疫系统过程的模拟，使触发器具有自我学习聚类，自动调整相关参数的能力，有较强的实用性。

5 结语

本文借鉴生物免疫系统的多种有效机制，提出了一个基于免疫原理入侵容忍触发器模型，并对模型中的核心免疫聚类器进行了细致的描述。该模型采用模块化设计，嵌套于服务器之上，运行简洁，自适应性强。当然，系统还有很多问题需要解决，如怎样优化提取的状态数据，改进亲和力阈值的取值，以提高触发器的准确率等。下一步我们将进行进一步研究。

参考文献

1 莫宏伟.人工免疫系统原理与应用.哈尔滨:哈尔滨工业大学出版社, 2002.
2 李涛.计算机免疫学.北京:电子工业出版社, 2004.
3 位耀光,郑德玲,付冬梅,周颖.基于生物免疫系统克隆选择机理和免疫网络理论的免疫算法.北京科技大学学报, 2008,27(2):245－249.
4 张著洪.人工免疫系统中智能优化及免疫网络算法理论与应用研究.[博士学位论文].重庆:重庆大学, 2007.
5 傅涛,孙文静,孙亚民,等.基于免疫学原理的混合入侵检测系统的设计与实现.计算机科学, 2008,35(6):63－66.