生物启发的网络安全理论与方法

郑瑞娟　张明川　著

科学出版社

生物启发的网络安全理论与方法

郑瑞娟　张明川　著

科学出版社

北京

内 容 简 介

本书在归纳分析国内外相关研究的基础上，从体系模型、态势感知、风险评估、自律调节和动态优化等全新角度研究了生物启发的网络安全理论与方法，包括①综合子网类型、群体规模和时序阶段，研究了多网安全体系模型；②采用改进的遗传神经网络研究了网络安全态势感知模型；③采用正态云模型研究了物联网安全风险自律评估方法；④研究了基于信赖域梯度的系统可信性自调节策略；⑤研究了鞅差逼近的系统服务动态优化与预测方法。

本书可以作为计算机科学与技术、信息安全等专业硕士研究生、博士研究生的专业课教材，也可作为从事信息安全、网络安全理论与方法、自律计算等研究领域的科技人员的参考书。

图书在版编目(CIP)数据

生物启发的网络安全理论与方法/郑瑞娟，张明川著. —北京：科学出版社，2013

ISBN 978-7-03-036333-6

Ⅰ. ①生… Ⅱ. ①郑…②张… Ⅲ. ①计算机网络-安全管理-研究 Ⅳ. ①TP393.08

中国版本图书馆 CIP 数据核字(2013)第 314378 号

责任编辑：刘宝莉 孙伯元 / 责任校对：邹慧卿

责任印制：张 倩 / 封面设计：蓝正设计

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

新科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2013 年 5 月第 一 版 开本：B5(720×1000)

2013 年 5 月第一次印刷 印张：15 1/4

字数：290 000

定价：60.00 元

(如有印装质量问题，我社负责调换)

序

随着人们安全意识的日益提高，虽然入侵检测系统、防病毒软件与防火墙等大量部署在网络中，但是网络安全问题依然没有改观。网络安全研究还存在很多亟待解决的问题。生物启发网络安全理论的提出，为网络安全的理论突破和方法改进提供了一种全新的思路。

完备的生物启发网络安全体系的研究与完善在保障网络和计算机系统安全上具有重要的理论及实用价值。生物启发网络安全模型和安全技术在保证网络信息的可靠性、提升网络系统的生存性、减少入侵事件和病毒带来的损失等方面具有重要意义。研究生物启发的网络安全理论与方法，是解决对国民经济发展和国防建设具有重大意义的网络与信息安全领域所涉及的重要科学问题和关键基础技术，其中所涉及的预防、检测与应答、容忍和恢复技术具备广阔的应用前景，可以为信息基础建设和高效安全运行提供科学基础。

本书内容将作者的研究成果进行梳理和扩展，完成对当前生物启发网络安全领域的最新进展进行有机整合，从内容和体系上突破相关领域知识点分散的局限，为读者提供一本具有自己鲜明特点的领域专著。

本书作者立足于网络安全研究的需求，清晰地陈述了生物启发网络安全理论的研究现状与存在不足，发表了多篇直接相关的学术论文、取得了多项科研成果，并在此基础上安排了合理的篇章内容，发掘了本书的主要创新点，具有较高的学术水平和内容组织能力。

王慧强
哈尔滨工程大学教授

前　言

目前，网络安全问题已经引起社会各方面的高度重视，各国政府投入了大量的人力、物力和财力进行网络安全相关理论和技术的研究，取得的研究成果广泛应用于商业和军事领域。在商业领域，网络安全产品已经成为 IT 行业中发展最快的部分；在军事领域，作为网络战、信息战的重要组成部分，网络攻击与防御技术受到重点关注。我国也将信息安全技术列为 21 世纪重点发展领域，作为国家 863 计划和国家自然科学基金的重点支持课题。

随着政府对网络安全研究扶持力度的逐渐加大和人们安全意识的日益提高，网络安全理论与技术不断取得进展，入侵检测系统、防病毒软件与防火墙等开始大量部署于网络系统中，但网络的安全状况依然没有改观，网络安全问题的根本解决更是无从谈起。

究其根本原因，主要是传统网络安全理论和技术本身存在的两个致命缺陷，使其无法从根本上解决网络安全问题。一方面来源于网络安全防护措施本身。传统防护措施无法从根本上解决网络安全问题，如防火墙属于静态、被动的防御措施，均存在人为参与、版本更新、单层保护等缺陷，随着入侵网络和系统的技术和手段的复杂化，直接面向入侵者的攻防对战不可避免。另一方面来源于网络和计算机本身。计算机网络中所有计算机的软、硬件配置及互联协议几乎完全统一，这使得新病毒和未知入侵极有可能在安全措施失效的情况下迅速扩散，甚至造成整个网络系统的瘫痪。可见，在开放的分布式网络环境下，现有的传统网络安全理论与方法已无法适应日益恶化的网络生态环境，必须构建一种新的网络安全模型，建立强有力的防御体系，从根本上改变网络安全问题频繁发生的现状。因此，寻求新的网络安全理论和方法势在必行。

生物系统在多变的生存环境中展示了良好的鲁棒性与稳定性，由于生物系统与计算机网络的运行机理和安全机制高度相似，人们从中受到启发，提出了将生物安全机制应用于网络安全理论与方法研究的思路。

本书针对当前网络安全研究存在的缺陷，受生物系统安全机制的启发，在生物启发的网络安全理论与方法研究中取得了以下成果。

（1）全面阐述了生物启发网络安全的起源，归纳了生物启发的安全计算和容错计算两方面的现状；在深入研究生物启发网络安全现状的基础上，针对当前生物启发网络安全的研究的局限性，对生物启发网络安全的研究前景进行了总结。

（2）在阴性选择算法和诱饵技术的基础上，根据生物体的多层保护机制提出

了一种基于主机的局域网免疫系统模型。

(3) 构建 B-MNS 概念模型,论证其完备性和系统性。通过分析生物个体、群体间的安全机理和网络安全问题的拟合点,将二者关联并形式化,重点突出生物系统的层次性和多维性,从群体规模、时序阶段和子网类型三个角度建立准确合理的 B-MNS 概念模型;在此基础上,利用线性空间理论验证 B-MNS 概念模型网络状态间的跳变空间具有的线性空间特性。

(4) 借助 Markov 跳变理论将(1)中建立的 B-MNS 概念模型定量化、精确化,设定初始数学模型,并采用 HMM 构建合理完整的 B-MNS 数学模型。该模型具有智能性、进化性与鲁棒性等特性。

(5) 基于数学模型的形式化结果,提出 B-MNS 分级实现模式。实现模式的选择直接决定着 B-MNS 性能的发挥。实现模式包括初级模式、进化模式和高级模式三类。初级模式即传统安全模型的单网模式;进化模式借鉴了生物及其群体的 MNP 原理;高级模式在借鉴相同生物原理的基础上,结合了计算机物理链路的功能特性,实现 MNP 的综合模式。

(6) B-MNS 模型性能分析与验证。根据性能在模型中的从属关系,将 B-MNS 模型分为固有属性和服务属性两类。固有属性将通过理论推导与定量比较进行分析评价,服务属性则在此基础上采用 Web 协议和 Ftp 协议进行实验测试。

(7) 依据性能指标在模型中的从属关系,将 B-MNS 模型的性能指标划分为固有属性和服务属性两类。通过初值设定和理论推导,对各实现模式的单项性能分析与综合性能进行了定量分析。

(8) 借鉴自律计算的思想和自动控制中的反馈机制,构建了一个基于自律计算的网络安全态势感知模型,该模型具备了较好的自适应性,能够有效地获取态势信息,准确了解网络的当前安全状况,快速预测未来网络安全态势,能够动态智能地适应复杂环境并有效地指导未来的自主决策。从而减轻了管理员的负担,降低了管理成本,进一步解决网络安全管理复杂性问题。

(9) 针对物联网系统特征、安全信息及其变化的不确定性、不可预测性和模糊性,赋予物联网安全以自律特征,基于异构安全要素的动态融合结果,重点围绕系统安全风险的自主评估开展基于三维正态云的物联网安全风险自律评估算法研究,力图在物联网异构安全的自主保障机理研究中取得突破,为保障物联网在非确定环境下的应用服务安全提供新的解决方法和思路。

(10) 提出基于信赖域梯度计算的系统可信性自调节策略,通过可信度在线评估、可信度动态预测和自调节方案的自主寻求与调节实现系统可信性的持续增长,采用自调节策略的实时反馈完成先验知识的自主更新。

(11) 基于自律计算提出系统服务性能自优化机制,在内部环境变化时实施动态单次自优化;而在内部环境相对稳定时实施静态自优化预测。以系统内部参数

环境的变化为驱动因素，自主预测和调节系统服务性能的变化趋势，从而以服务性能最大化为目标，及时实施对系统参数的自主优化。

本书受到国家自然科学基金(No. 61142002，U1204614)的资助，由河南科技大学的郑瑞娟副教授和张明川老师撰写完成，其中郑瑞娟副教授完成了 21 万字，张明川老师完成了 8 万字。在本书的撰写过程中，得到了哈尔滨工程大学王慧强教授，河南科技大学的普杰信教授、吴庆涛副教授、杨春蕾老师和 507 研究室的张丹、朱丽娜、马政朝、李腾昊、张旭龙、刘娜娜和崔敏等研究生的支持与帮助，在这里一并表示感谢。

作　者

2013 年

目　录

第1章 绪 论

生物依赖环境的同时也在改变着环境，人们很早就意识到了这种相互依赖彼此依存的紧密关系。伴随着科学技术的发展，对于生物与环境之间复杂关系的研究逐步形成了生态学。关于什么是生态学，不同时期的学者给出的定义不尽相同，但已有的定义差别不大，通常可以概括如下：生态学是研究生物与环境及其相互作用的科学[1]。近几十年来，生态学的研究得到快速发展，与生态学相联系的一系列边缘学科相继产生，如生物化学、生物物理和生物经济学等。

基于生态学，文献[2]提出了关于单调控制系统这一新兴领域的机遇与挑战，主要包括以下几个方面：

(1) 生物测量与操纵控制和信号处理技术的核心内容主要集中在仪器设置方向。

(2) 生物研究的启发理论在控制工程领域的抽象与应用。近年来在基因研究方面所取得的成果，不断为系统提供更为详细的启发知识。传统的单调控制技术只能应用于自治动力系统，从某种意义上说，此种限制阻止了研究互联结构的可能性，特别是当研究的角度从组织严密的描述方法转换到相对分散的分析方法时，该局限性则表现得更加明显。但同时，需要分解的高复杂系统所拥有的细小分子是具有清晰特性的，这种转换也是必然的粗略地说，无论系统流的初始条件或进入该系统的外来信号具有单调特性，该系统都将是单调的。满足该要求的模型经常应用于多种不同领域，如分子生物学、生态学、经济学与化学等。

单调控制系统在生物中的应用尤为明显。近年来，生态学受到了越来越多的关注，而其中关注的一个重点是物种的共存即持续生存问题。即不同物种之间最基本的生存关系分为竞争关系、捕食-食饵关系、互惠关系等，但同时，许多物种之间的关系并不是单一不变的，而是要相对复杂和动态得多。

在研究多数复杂系统时，我们一般都会做出假定，即大量的简单元素通过大量汇聚与交互来表现系统的复杂行为。因为在多数情况下，系统的元素是均一而简单的。但生物系统是由大量的细胞、蛋白质以及基因组成，而这些元素或成分则无法用均一的或简单的概念予以描述和实现[3]。由此，生物系统的主要特征可以归纳为以下四点：

1) 元素的独特性

从遗传学的角度来看，生物系统由成千上万个基因组成，每一个基因具有不同的调控关系，而每一个基因的表达产物具有不同的功能。因此生物系统不能看作

是由多个均一成分组成的系统，同时，系统行为也无法简单地用均值行为来近似表达。

2）组分的复杂性

系统每一个元素均具有其本身的复杂性。每一个基因都包含着一个复杂的调控结构，而其转录产物主要包括自身的复杂结构和动力学行为。每一种蛋白质结构与功能各不相同，且蛋白质彼此之间存在着多种相互关联的关系。

3）组分之间相互作用的选择性

系统组分之间的相互作用具有高度的选择性。基因与基因之间的调控作用具有高度特异性，同时，蛋白质之间的相互作用也是高度特异的。

4）组分之间相互作用的网络化

系统组分之间的相互作用形成了高度复杂的生物化学网络。代谢作为最基本的生化反应网络，通过产生能量来驱动各种细胞的变化过程，从而降解并合成了许多不同的分子。代谢网络也被称为酶分子（蛋白质分子）网络。同时，系统内部还存在着蛋白质网络和基因调控网络。

通常情况下，生物系统的相互作用不仅仅表现在系统之间的相互作用上，其本身往往也要受到外界因素的影响，这种受到外界因素影响的表现称为随机力[4]。而随机力会破坏系统的有序性，因此系统的有序和无序的竞争则会表现出复杂的行为。值得注意的是，随机力并非只能起到消极作用。在生物非线性系统的分岔点附近，通过生物系统与随机力的作用，无序的能量可能会转化为有序的动力，从而促进系统新秩序的建立。近年来讨论的噪声诱导相变、随机共振、相干共振、分子马达运动、改变基因遗传选择平衡以及选择性控制细胞生长等一系列现象就属于典型的噪声对生物系统正面作用的例子。

生物系统本身就是一个自组织的进化系统，在生物与环境的交流和相互作用中，生物利用自身的冗余结构和冗余补充，使整个系统始终朝着有利于生物发展的方向进化。文献[5]采用 Type-2 模糊系统建立了生物群落冗余结构的数学模型，提出了生物群落稳定性的度量方法，研究了生物群落冗余结构与生物进化可靠性和稳定性的关系。

当前，网络应用日益呈现出高性能、大规模、多样性需求，用户对 Internet 网络提出了更高的分布式要求：需要以用户为中心的网络具有自扩充性、可移动性、可生存性、简单易操作性以及随着用户和网络环境的长期和短期变化而具备的自适应性等特点，因此对于网络应用的开发设计者和相关领域的研究人员来讲，有必要进一步优化 Internet 网络体系结构，并开发设计其应用。生物信息系统作为一个分布式自治系统，其运行机制能够为多个科学和工程领域提供富有成效的技术和方法，如人类社会、蜂群、生物免疫系统等大规模系统已形成许多重要的原理和机理，可用以启发人们采用科学方法和工程技术策略满足未来 Internet 的需求。

探寻系统生物学和生物信息学的运作机制和工作原理，则成为研究生物启发理论的基础与首要任务。

1.1 系统生物学

人们对生物体系统的研究多采用还原论法，完成对系统单个组成元素的独立分析与研究。目前为止，还原论的研究已经取得了大量成果，在细胞甚至分子层次都对生物体具有较深入的了解，但遗憾的是，相关成果对生物体的整体行为却难以给出系统的、全面的解释[6]。目前，生物科学还停留在科学实验的阶段，尚未形成一套完善的理论来描述生物体如何在整体上实现其功能行为。对子系统的研究成果互相割裂地罗列在一起，人们迫切需要将它们组合起来进行整体研究，从中发掘出更多信息，以期能够从宏观上深刻理解生物体系统的运行规律。近年来，生物技术的迅速发展，比如蛋白质芯片技术、DNA 微阵列技术，使得高通量地定量测定生物大分子成为可能，而且计算机网络技术日新月异的发展也从很大程度上促进了各类科学研究所产生的生物数据库之间的资源共享，这就为生物学理论的建立提供了大量的可用数据，使得理论的建立与检验成为可能。此外，人类基因组计划的完成在提供大量数据的同时，也宣告了全面挖掘人体系统机理的时代已经到来。正是在这种背景之下，系统生物学应运而生。

随着后基因组时代的来临，海量的生物数据不断产生。而生物芯片与质谱仪等高通量技术的日渐成熟，更加推进了系统生物理论的研究步伐，使其在信息收集、整合与数据挖掘的基础上，全方位地研究生命活动的规律。据此，以整体和相互关系为研究对象的系统生物学(systems biology)成为当今生物学研究领域中的新的研究热点[7]。系统生物学一词在美国 NIH 的 PubMed 文献库最早可检索到 Zieglgansberger 和 Tolle 于 1993 年发表的一篇神经系统疾病研究的论文摘要中，根据 1968 年国际系统理论与生物学(systems theory and biology) 会议(最早提出 systems biology 词汇)定义为采用系统论方法研究生物学，1989 年在美国召开的生物化学系统论与生物数学国际会议探讨了生物学的系统论与计算生物学模型研究。系统生物学的主要研究对象是生物有机体，生物有机体可以看成由基因、蛋白质和生化反应所组成的生物系统。不同于以往仅关心个别基因和蛋白质的分子生物学，系统生物学研究细胞信号传导和基因调控网路、生物系统组成之间相互关系的结构和系统功能的涌现。

按系统生物学的观点，可以将生物系统的特点概括为整体、动态、层次、整合[8]，其主旨在于从整体高度来理解生物系统，是生物学的一个新的研究领域，其出现是继人类基因组计划完成之后，生命科学转为整合性研究的一个新的高度。基于分子生物学对生命体研究的进展，系统生物学着重于以系统的联系性来阐释

生命现象[9]。

由此可见，系统生物学与传统生物学方法的区别在于，系统生物学从整合的角度来对生物学进行研究，而传统生物学更侧重于某一基因或其编码蛋白质的研究。系统生物学的观点认为，生物体内各个子系统间不是独立的，而是希望寻找新的方法来阐明子系统间交互作用的问题。即系统生物学和传统生物学的本质区别是前者强调用系统的观点研究生物，更注重各部分间的相互关系，即整体大于部分之和。

系统生物学的研究对象为有机体内无数个小网络通过层次与层次之间、网络与网络之间、系统与系统之间的联系来建立复杂的系统，而不是简单的系统叠加。这个复杂系统也会通过不同网络之间的信息传递和整合，使蛋白质最终具有生物功能，如此势必会出现一些单独系统所反映的新行为，这就是系统生物学的特征。因此也可以说，信息是生命赖以生存的关键因素。

但是，在生命发展的过程中，生命各个单元之间存在的相互影响和相互关系也是至关重要的，因此将生命体割裂成一个个彼此孤立的单元，仅从各个单元中找到它与整个生命体的关系的研究思路是片面的[10]。因此，系统生物学利用系统论的思想和方法，从整体的角度来分析研究生命的复杂特性，既要考虑多个层次、多种类型的生物信息，又同时要考虑时间因素。系统特性是由不同组成部分、不同层次间的相互作用而涌现出的新性质，因此，如果单纯研究某一组成部分或单一层次并不能真正准确地预测整体或高层次的行为，那么如何通过研究和整合去发现和理解涌现出的新的系统性质，则成为系统生物学面临的一个根本性挑战。为了应对这一挑战，系统生物学，特别是计算系统生物学必须建立有效的方法，通过整合系统各个层次的信息，建立能够反映该系统目前已知或可测量性质的物理、数学模型，并通过这样的模型来研究或预测目前未知的系统性状[11]。可以说，建模是系统生物学的最重要的研究手段之一，其通过建模与实验相结合的研究手段来定量阐述生命现象的本质规律。

系统生物学的主要研究目标是了解一个生物复杂系统中所有组成成分的构成及在特定条件下这些组分间的相互关系，分析该系统在一定时间内的动力学过程，即从大量的生物学数据中得到一个尽可能接近真正生物复杂系统的理论模型，根据模型的预测或假设，设定和实施新的改变系统状态的实验，不断地通过实验数据对模型进行修订和精练，使其理论预测能够反映出生物系统的真实性。

现有生物学的研究依据所用方法可以分为两大类：一种是依据所观察到的生命活动现象提出假设，建立相关的模型，然后进行模拟，并设计实验来验证这个模型是否正确，这种途径可称之为 top-down；另一种方式是经过大量的实验，积累相关的原始数据，再对数据进行分析、比较，此基础上，建立生物学模型，并对生物学模型进行模拟，再以实验来验证模型，这种途径称之为 bottom-up[12]。目前，系统

生物学主要采用的是后一种研究方式，其研究对象是生物体内的网络结构，而不是简单的组成这个网络的单个分子组成成分。

通过对系统生物学研究机理的分析，目前对系统生物学的研究主要集中在以下四个方面：

(1) 系统结构的识别和研究，主要包括基因的相互作用以及这些相互作用以何种机制调节生物系统的研究。

(2) 系统的动态特征，主要研究生物系统在不同状态下的行为。

(3)系统的控制方法，主要研究系统如何控制细胞的状态，能够使细胞功能受损最小，从而为疾病治疗提供潜在的药物靶标。

(4)系统的总体设计方法。

目前，对系统生物学的研究虽然取得了一定的进展，但仍面临着一些问题：

(1) 系统生物学中的数据挖掘问题。系统生物学对生物系统中不同性质的数据进行整合，从基因到细胞、到组织、到个体的各个层次[13]。大量组分数据的收集来自于实验室(湿数据)和公共数据资源(干数据)。但这些数据存在着许多不利于处理分析的因素，如数据的类型差异，数据库中存在大量数据冗余以及数据错误；存储信息的数据结构也存在很大的差异，包括文本文件、关系数据库以及面向对象数据库等；缺乏统一的数据描述标准，信息查询方面大相径庭；许多数据信息是描述性的信息，而不是结构化的信息表示。如何快速地从包含了错误数据在内的海量数据中获取系统生物学建模所需的正确数据模式和关系是数据挖掘的主要任务。

(2) 系统生物学中数据的有效整合。文献[14]提出利用高通量技术系统研究各种分子网络，强调以网络、相互作用、动态行为等整体论观点，并结合数据与方法论的整合性观点，探索如何对复杂生命现象进行不同层面的理解和诠释。例如基因芯片、酵母双杂交系统、免疫共沉淀和染色体免疫沉淀芯片方法。通过实验手段得到的数据，提供了细胞或复杂生命系统的快照式轮廓，从而可以用于推断或重建蛋白质相互作用的网络和基因调控网络。此外，来源于不同角度的各种大通量数据中的信息，使获得更为完整的分子网络成为可能。通过对网络整体性和动态性的深入研究，不仅可以帮助生物学家理解复杂的生化现象，也可以从系统角度对细胞系统的根本机制或者本质规律进行诠释。因此，这种系统生物学研究方法在短短几年中得到迅速发展，成为生物学研究中的热点。

细胞中的蛋白质并不是孤立存在的，或者说很少单独地表达特定的功能，而是通过彼此之间复杂的相互作用来完成特定的生命活动[15]。蛋白质相互作用在生命活动中起核心作用，它不仅是正常生理过程如 DNA 复制、转录、翻译、物质代谢、信号传导以及细胞周期控制的基础，也在病理过程中起着重要的作用。可以说，几乎所有的生物过程都是通过蛋白质间的相互作用来精确地执行的。一个生