毕业设计论文

题目：　 基于BP神经网络的网络安全态势预测技术研究

学员姓名

学　　号

所在单位

指导教师

完成日期

**摘要**

**内容摘要：**随着互联网的发展，随着互联网的发展，互联网已经对社会生产和生活的各个方面产生了巨大的影响，引发人们的普遍关注。这种情形下,网络安全态势感就诞生了,它通过针对互联网上的所有网络安全问题的研究,进行了有效的解决方案: 根据大量复杂的互联网信息和安全日志文件,综合剖析了各类威胁,从整个互联网的视角评估网络安全状况,并在此基础上预测网络安全状况的发展趋势。为可以比较准确的评估当前网络安全形势的发展变化,本章将时间序列模型结合BP神经网络,并利用陈秀真提出的层次化网络威胁量化评估方法[[1]](#endnote-1)[1]，对网络安全态势进行预估。

本文实验是在docker上搭建虚拟环境来采集所需分析的数据，进而对网络威胁进行赋值量化，通过Excel和Python工具进行计算得出网络安全态势评估值，最后利用BP神经网络以时间序列进行训练，当训练误差降低到符合要求即可进行网络安全态势预测。

**索引关键词** BP神经网络 网络安全 态势预测 层次化

目 录

[第一章 绪论 1](#_Toc102722722)

[1.1 研究背景及意义 1](#_Toc102722723)

[1.2 国内外研究现状 5](#_Toc102722724)

[1.2.1 态势预测研究现状 5](#_Toc102722725)

[1.2.2 预测方法的研究现状 7](#_Toc102722726)

[1.3本文的主要内容及结构安排 8](#_Toc102722727)

[第二章 网络安全态势概述 9](#_Toc102722728)

[2.1 网络安全概述 9](#_Toc102722729)

[2.2 国内外网络安全评价标准 11](#_Toc102722730)

[2.2.1 我国的信息安全评价标准（GB17859-1999） 11](#_Toc102722731)

[2.2.2 美国可信任计算机标准评价准则TCSEC 13](#_Toc102722732)

[2.2.3其他网络安全评价标准 16](#_Toc102722733)

[2.3 网络安全态势可预测性研究方法 17](#_Toc102722734)

[第三章 网络安全态势评估设计 18](#_Toc102722735)

[3.1 态势评估数据采集 18](#_Toc102722736)

[3.1.1 Docker介绍 19](#_Toc102722737)

[3.1.2 Snort介绍 21](#_Toc102722738)

[3.1.3 局域网搭建及数据采集 23](#_Toc102722739)

[3.2 网络安全态势评估模型 23](#_Toc102722740)

[3.2.1 服务级 24](#_Toc102722741)

[3.2.2 主机级 26](#_Toc102722742)

[3.2.3 系统级 27](#_Toc102722743)

[3.3 评估模型的实现 29](#_Toc102722744)

[3.3.1 各级态势值计算 29](#_Toc102722745)

[3.3.2 结果分析 33](#_Toc102722746)

[3.4 本章小结 33](#_Toc102722747)

[第四章 网络安全态势预测 34](#_Toc102722748)

[4.1 BP神经网络介绍 34](#_Toc102722749)

[4.1.1 基本原理 34](#_Toc102722750)

[4.1.2 BP神经网络的优缺点 35](#_Toc102722751)

[4.2 预测模型设计及训练 37](#_Toc102722752)

[4.3 预测实验 40](#_Toc102722753)

[第五章 总结与展望 41](#_Toc102722754)

[5.1 总结 41](#_Toc102722755)

[5.2 研究展望 43](#_Toc102722756)

[致谢 44](#_Toc102722757)

[参考文献 44](#_Toc102722758)

# 第一章 绪论

随着当前信息网络的高速经济发展,各种信息网络技术应用也更多地影响了人类的工作、学习和生活方式。根据中国互联网中心（CNNIC）于二零二一年八月二十七日发布的《第48次中国互联网发展统计报告》[[2]](#endnote-2)[2]，截至二零二一年六月，我国网络发展状况如下:我国网络用户数量已达十点一一亿,相较于上年末增长了二千一百八十五万,网络普及率已超过了百分之七十一点六。 比上年末增加0.7小时。

但是由于互联网的广泛应用,尤其是数字应用基础业务的扩大以及政务服务的网络化,信息安全正受到着巨严峻的考验,对个人进行信息安全和关键数据环境保护非常不足,个人信息泄漏的局面也十分严重。针对互联网和网络信息安全的威胁手段多种多样,如网络攻击、互联网金融欺诈、秘密窃取、病毒入侵等,都对网络和网络信息安全构成巨大威胁。而现有的如杀毒软件、防火墙等手段都只是从单个方面提供安全防护，不能很好的从宏观角度将网络上的各种信息、事件联系起来。

## 研究背景及意义

随着互联网的飞速发展壮大,要随时把握和了解互联网的安全状态,使管理人员可以适时制定措施,甚至制定必要的措施,关于互联网状态安全的科学技术研究显得日益关键。为了适应各种安全问题的发生,安全态势感知越来越成为预警、防止和反制安全问题的焦点,网络安全态势感知和预测技术被广泛研究和应用。态势感知一词，首先出现于1988年前美国空军首席科学家Mica R.Endsley提出的态势感知Endsley三级模型，如图1所示。并于1995年完善应用于美国空军。 [[3]](#endnote-3)[3]

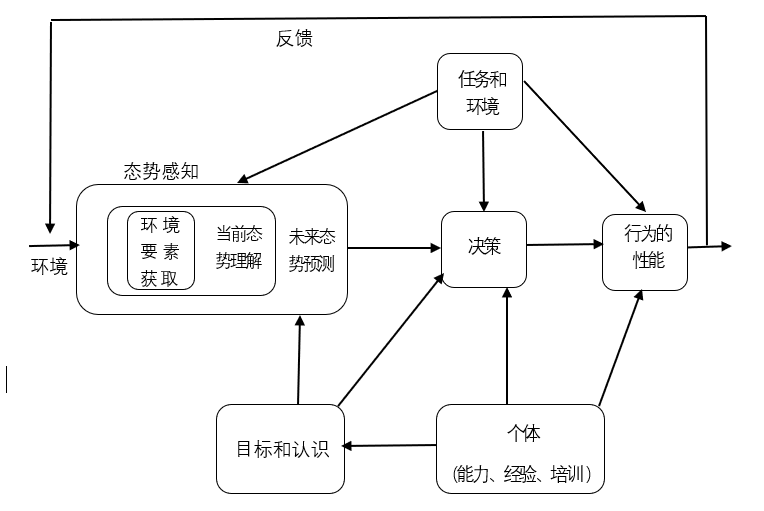


图1.1 Endsley态势感知三级模型

1999年，Tim Bass把态势感知带入了网络安全领域，提出了基于多传感器的态势感知框架为网络安全态势感知发展提供参考和借鉴，如图2所示。

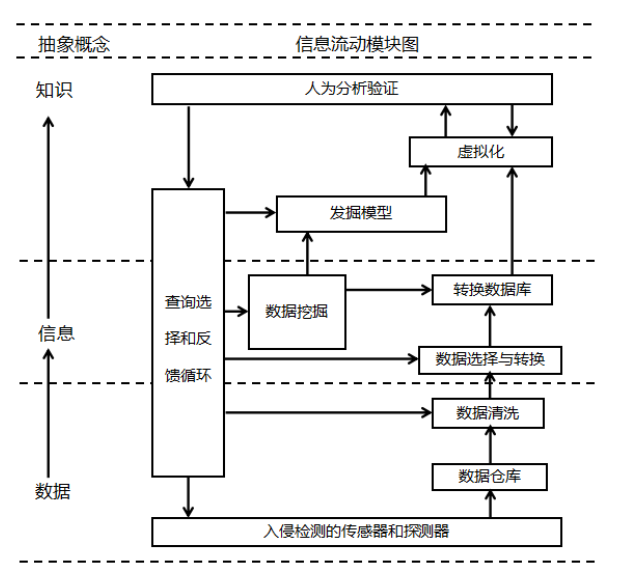


图1.2 Tim Bass态势感知框架

同样在1999年，Leopoldt 和Abasi通过将安全态势感知框架与空中交管态势感知框架类比，提出了态势感知的概念模型，此模型即为本文所用模型，如图3所示。

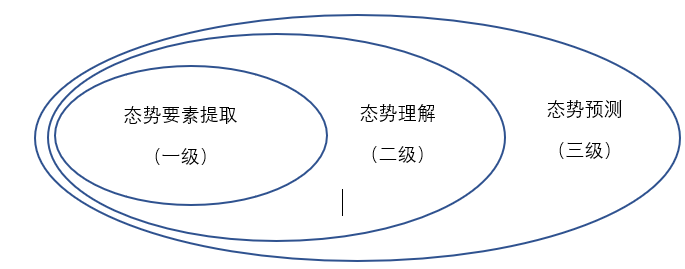


图1.3 态势感知概念模型

整个态势感知模型按照分析流程分为三个阶段，即态势要素提取、态势理解（或态势评估）和态势预测[[4]](#endnote-4)[5]，具体功能为：

（1）提取网络安全态势要素：指在海量的互联网数据和网络安全资讯中,全面、精准、有效的抽取出互联网信息安全趋势有关的因素，如状态、属性和动态信息等，并对其进行融合、整理等预处理，使其表现为需要的、方便理解的形式。要素提取是整个研究工作开展的前提和基础性工作。

（2）态势理解（或态势评估）：态势理解（或态势评估）是研究的核心。它建立在要素提取的基础上，通过构建相关指标和指标体系，对海量数据进行计算，将复杂多样的网络安全态势相关数据信息中的可帮助分析者和决策者理解网络状态的内质表现出来。

（3）态势预测：使用科学的方法和手段,在前一阶段工作(即态势理解)基础上,依据网络安全的历史状况信息与变化趋势,同时根据网络安全的当前状况信息,合理可靠地预测未来某一阶段的网络安全状态的变化,以便及时采取应对措施。

## 1.2 国内外研究现状

### 1.2.1 态势预测研究现状

网络安全态势预测是网络安全态势感知研究的重要环节，并且位于态势感知的顶层和整个研究的最高级阶段，是有针对性地采取网络安全应对措施的前提，有较高研究意义和价值。

飞速发展更新的攻击技术导致网络安全技术和概念也在不断发展完善甚至是更新换代。根据网络和网络信息的“五性”，即保密性、完整性、可控性、可用性和不可否认性，从计算机的出现至今，将网络信息安全理论发展分为三个阶段：

**计算机安全阶段（20世纪70年代至80年代）：**即将一个独立的微机控制系统作为主要防御目标，以确保微机控制系统中的软、硬件部分以及数据和信息在管理、存储、传递流程中的保密性、完整性和可用性。

**信息技术安全阶段（20世纪80年代至90年代中期）**：把直接连接电脑的网络系统和网络中的计算机系统作为主要研究对象,特别重视网络系统和计算机中互联网信息系统的保密性、完整性和可用性，并逐步扩展到信息的“五性”。

**信息保障阶段（20世纪90年代后期至今）：**研究整个网络系统中的所有网络基础设施、网络信息和各类数据等，确保整个网络系统和所有网络信息的安全。信息安全扩大到网络整体角度的信息保障安全。

在国内,尽管对这一领域的研究起步较晚,但是对互联网安全状态研究却受到了国内政府部门的大力支持,并且随着《网络安全法》和《国家网络安全战略》的颁布，对信息安全状态感知的研究也开始上升到了战略高度。

目前相关成果已有：上海交大的李建华研发综合处理系统，并提一种基于知识基的分析方法；西安交通大学的陈秀真等人，实现了基于IDS和防火墙的层次化网络安全威胁量化方法；中国科技大学韦勇提出基于信息融合及基于日志审计与性能修正算法的评估模型[[5]](#endnote-5)[7]等。

### 1.2.2 预测方法的研究现状

以网络攻击为主的网络安全事件的发生具有很强的随机性、突出性和不确定性，即安全态势变化趋势是一个非常复杂的非线性变化过程。正是如此，利用传统的数据工作进行预测将会产生较大偏差，人们开始研究寻求新的预测方法。现阶段常用方法有基于时间序列、支持向量机和神经网络的预测方法。

其中，研究以按时间序列的历史态势的变化，从而对未来一定时间内的态势做出延展预测，即为基于时间序列的预测方法。国内在这方面的研究，有王雪[[6]](#endnote-6)[8]提出的利用残差检验的基于时间序列的预测方法；文志诚、陈志刚[[7]](#endnote-7)[9]等在时间序列的基础上利用隐马尔可夫模型进行预测的方法。

支持向量机(SVM)是一种线性分类器,以监督学习,计算样本的最大边距超平面,对数据进行二元分类。南京邮电大学的王文思就基于支持向量机的态势预测提出了结合贝叶斯网络以提高预测预测精度的方法[[8]](#endnote-8)[10]。刘俊男等人将通过支持向量机提供一个融合了模拟退火和高斯干涉粒子群计算的预测模式[[9]](#endnote-9)[11]。

得益于计算机硬件的飞速发展，在网络安全态势感知领域基于神经网络的机器学习迅速发展，可针对复杂的网络安全环境进行灵活、准确的预测。常见的神经网络模型有BP神经网络和RBF神经网络，本文即采用的BP神经网络。

## 1.3本文的主要内容及结构安排

本文通过学习了安全状态感知方面的理论内容和神经网络概念,以及根据时间序列和层次化网络安全威胁量化方法,构建了合理的基于BP神经网络概念的网络安全状态预测模型。并在Docker上搭建虚拟环境，利用Snort入侵检测系统来采集所需数据经处理后进行训练和检验。本文结构大体如下：

第一章，绪论。本章首先介绍了选题背景和意义,随后简要介绍了国内安全态势预测与预测方法发展状况,最后主要叙述文章的主要内容和结构。

第二章，网络安全态势概述。本章介绍了国内外信息安全评估标准，为网络安全态势评估提供理论依据。浅述了时间序列的自相似性，提供了网络安全态势可预测性的基本理论。

第三章，网络安全态势评估设计。本章主要介绍了整个态势评估的过程包括三个部分。第一部分在Docker上搭建局域网采集数据，并利用Excel和Python工具进行预处理和特征提取；第二部分介绍态势评估模型，包括指标体系构建和计算方法。第三部分为态势评估值的计算，最终得出态势评估值的时间序列。

第四章，BP神经网络预测实验。本章首先从BP神经网路理论入手阐述,然后讲述了BP神经网络预测模型的建立过程及其针对该网络安全态势预测神经网络模型的训练方法并开展了相应实证检验,最后对实验结果进行了分析。

第五章，总结与展望。总结和讨论本文的总体内容，并对未来的研究工作进行展望。

# 第二章 网络安全态势概述

互联网的飞速普及，网络影响了社会生产各方面，但网络安全问题也令人担忧。随着网络安全事件逐渐增加，人们越来越注意到网络系统自身的安全性，网络安全问题已成近年来研究重点。

## 2.1 网络安全概述

在本文中,这里的"网络"特指现代网络。据维基百科,计算机和互联网系统是指使用通信装置或者电缆系统,把地理位置不同、但功能完全独立的或多台计算机技术连接起来,该系统的主要目的是通过使用完善的互联网软件,共用互联网的硬件、软件、资料和信息。具体来讲是指一种通过连接二个或者多台计算机技术,实现信息交流的网络系统。

一般而言, 网络安全[[10]](#endnote-10)[12]主要是指整个系统的安定,确保系统不被非法访问和使用。具体体现在数据与信息的保密性、完整性、真实性、可用性、可靠性、以及不可抵赖等方面。

（1）可用性, 可用性是指所有网络内容的授权使用者和服务实体都可以在权限范围内使用该信息的一种特性。即当必要时,或者说当网络系统的组成部分损坏而必须降级应用时,可以授予使用者个人或实体应用并继续向授权使用者提出更高效业务的功能。

（2）可靠性, 是指系统可以在一定环境条件和时限内,实现系统预期的目标的特性。可靠性是互联网信息安全最基础的要求之一,关系整个互联网信息系统的建立与运营安全。

（3）保密性,保密性是指未经授权的用户、实体或进程无法访问和利用信息的特性。

（4）完整性,完整性是指信息不被非法操作而发生变化的特性。确保资料不会遭到无意或故意的非法更改、删减、重放等手段。

（5）不可抵赖性, 在信息交互时,确保参与者与其所属信息的对应同一。也就是说,任何参与者都不可以否认或甚至违反他进行过的操作的承诺。

（6）可控性,可控性是指对网络消息的传递流程与内容有一定控制力的特征。

综上所述,信息安全的核心问题就是指利用电脑操作系统、计算机网络、密码技术和安全技术手段,保障人们在互联网信息系统中保存和传送的信息系统具体内容的“五性”。

## 2.2 国内外网络安全评价标准

### 2.2.1 我国的信息安全评价标准（GB17859-1999）

GB17859-1999作为在1999年10月由中国国家质量技术监督局颁布，并在国内强制实施的网络安全标准，是国内建立和评价网络安全比较通用的度量[[11]](#endnote-11)[13]。其中对计算机安全性保护分5个等级:

第1级 (GB1安全级别):用户的独立安全级别。该级别允许合法用户访问已被多种技术手段限制保护的数据，从而为用户提供了脱机保护的能力。这是最低级别的安全级别。

第2级(B2安全级别):用于系统监控的保护级别。除了第一级保护的所有功能外，还需要创建和维护所有用户操作的访问控制记录，以确保所有用户都对其行为的合法性负责。

第3级 (GB3安全级别):安全品牌的保护级别。除前一层安全保护的功能之外,还需要通过标记数据或访问调查对象的安全管理等级来控制访问者使用,以达到对使用对象的强制性防护。

第4级(GB4安全级别): 一个结构化的保护层,继承低级安全机制,并将系统分为关键与非关键部分,在访问对象与关键部分之间的提供直接监控,同时提高系统渗透性。

第5级(GB5安全级别):访问查看保护级别，它专门添加了一个访问查看功能，负责与访问访问访问对象相关的所有访问者操作。

中国也是国际标准化组织（ISO）的成员，因此中国的标准化工作也活跃在各个领域。自1980年代中期起，中国已陆续制定并颁布了若干信息安全规范。不过,也应当指出,构建国家安全技术标准体系需要丰富的使用经验和深度的技术调研。充分考虑到了这二个差距,目前我国在安全领域的标准化工作相对于发达国家在国际领域的现有工作水平而言,还不能受到充分的关注。

### 2.2.2 美国可信任计算机标准评价准则TCSEC

TCSEC是由美国国防部开发的计算机安全标准，又被称作网络安全橙皮书[[12]](#endnote-12)[14]。作为美国国防部网络安全评估标准，其被普遍地用于评估计算机的安全性。

自从1985年，美国国防部将安全橙皮书确立为其规范开始,近几年来与其他规范毫无改变,目前仍然是衡量美军以及其他国家和地方多客户端服务器和小型控制系统的方法。还有一些组成部分,如数据库管理系统和计算机,也总是通过安全橙皮书来说明并评价其可靠性。安全橙皮书把网络系统的等级由低至高共分为了D类、C类、B类和A类4个类型,而其中的每一类又分为了多个等级,如表2.1所示。

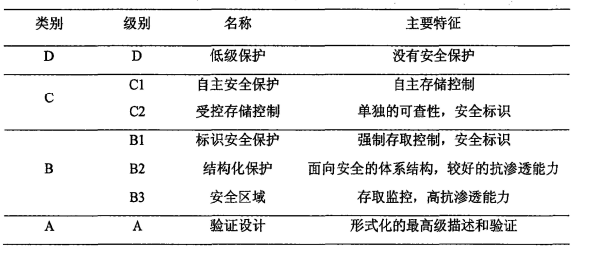


表2.1 网络橙皮书的计算机信息安全级别

D级是最低的系统安全分级,在这层级的控制系统中,是完全不信任的系统,谁都可能以最简单的方式进出，因此极其不安全。无法保护所有执行此级别操作系统的硬件;同时,也因为没有系统访问限制与数据存取限制,所以所有人都不需要帐户且不受限制地存取别人的系统,从而使操作系统变得脆弱。

C1是C类的一个安全子级。C1又称选择性安全防护系统。此类安全级别系统可为系统本身和用户的数据文件提供一定程度的保护。用户首先需要帐号和密码,由系统是否合法,并按帐号属性决定用户权限。

C2级在C1级的功能中又增加了访问控制环境的权限。在C2安全级别的访问访问控制条件中,增加了用户身份验证等级。此外，系统还会检查并记录系统日志包括启动时间、用户登录时间和位置等。因此，可以通过日志文件来检测入侵痕迹。除了审讯系统管理员执行的操作外，还为其他用户添加了一个身份验证级别，以通知用户操作。

如果C2系统上的用户不是系统超级用户，则只需使用额外的身份验证即可执行管理任务。许可证分组允许系统管理员将用户分组，并允许他们访问特定程序或目录。。

B级包括了三种安全等级:B1、B2和B3,其中的B1等级规定了当对象处在强制访问或控制状态时不允许改变文件权限。B2级规定必须标记计算机系统中的所有对象，并根据设备在系统中的重要性为每个对象指定一个或多个安全级别。B3级是用硬件提高域的安全。因此,存储管理硬件可以用来保护域免遭非法活动的攻击。同时，此层还要求用户通过受信任的工具连接到系统。

A级被称为设计验证级别，是最高安全级别，包括低级别的所有功能,同时其安全控制被严格设计。首先系统要在数学角度上进行理论验证,进而分析其内部秘密通道和可靠分布,确保安全。

橙皮书虽然被用作美国国家标准，但也有不少缺陷。首先,TCSEC仅评估隔离的计算机,尤其是中小型电脑和主机系统。其次，假定对计算机系统有必要的物理保护,那么标准仅适合于政府部门和军方,不宜于一般公司。最后，它仅仅只是一个静态模型。

### 2.2.3其他网络安全评价标准

其他团家或者组织也先后根据国际标准化组织CSO) 的相关准则,和于一千九百八十五在美国国伤郡委员会提出的国际可信计算机标准与评价原则,进一步建立了一些准则,如：

（1）信息技术安全性评估通用准则简称CC,是以美国为首包括加拿大、法、英等国家和部分欧洲组织联合协商起草的。它参考了国际标准化组织标准，逐渐形成了国际安全评估事实上的标准规范体系，代替了许多发达国家的己有评价准则于1999年被正式颁布发行。

（2）信息安全管理指南[[13]](#endnote-13)[15]是由SO/正CJTC1制定的技术报告，是一个信息安全管理方面的指导性标准，为更有效地进行IT的安全管理提出建议。

## 2.3 网络安全态势可预测性研究方法

数据具有可预测性才有预测研究的意义和价值。数据可预测指的是相邻数据点之间有联系。建立在时间序列上的安全态势是具备可预测性的，其重点在于安全态势必须是用科学的评估手段得到的，并按照它们的时间顺序排列。

时间序列模型就是通过研究以时间为序的对象，从而研究对象的发展趋势，并类推或延伸到未来，从而未来时间的状态做出可信的预测。本文以BP神经网络结合时间序列模型来对安全态势值进行预测。

自相似性是常用于判断一个时间序列可不可以进行预测。自相似性是指,在整体中抽取出局部仍可表现整体的特征。

在之前已经有相关研究：Leland等人用收集和统计分析以太网流量数据的方法，证明了以太网流量数据中普遍存在的自相似性[[14]](#endnote-14)[16]。Popescu根据大量研究实验指出，分形自相似模型与网络业务流模型的数据流本质特征十分贴近。分形自相似性在网络连接数据的预测研究中地位非凡。

# 第三章 网络安全态势评估设计

态势感知(或态势评估)是网络安全态势感知模型中的第二级也是预测的前提。网络信息和安全日志的海量数据难以直观地反映网络安全态势，态势评估是了解信息的相关性，从宏观的角度是将其转化为基于网络信息和日志处理的网络安全态势的量化且直观数值。

关于数据整合算法，目前在网络安全评估中广泛采用的方法大致分为基于数学模型、基于逻辑关系、基于规则归纳和基于概率统计的方法。本文应用陈秀真提出的分层加权网络安全威胁量化评估方法，这是基于数学模型的最具代表性的方法。

## 3.1 态势评估数据采集

有关态势数据采集，本文设计为在Docker上搭建局域网，模拟正常用户访问和网络攻击行为，利用Snort入侵检测系统对相关数据进行采集。

### 3.1.1 Docker介绍

Docker作为一种开源的容器引擎，属于容器的一种封闭方式，。具体功能是提供了一个容器接口。特点是应用程序和依赖项包可以打包到可在底层Linux或Windows操作系统上发布或虚拟化的可移植映像中。由于容器充分利用沙箱机制，因此不需要接口。

（1）Docker的组成

完整的Docker通常由三部分组成:仓库、镜像和容器。在Docker中，镜像是容器运行的先决条件，仓库是存储镜像的位置。

镜像（Image）, 一种比较特别的文档管理系统,给出了执行容器时所需要的文档管理系统,里面包括了一种配置参数。镜像中不包括动态数据,因此建立某个镜像后,其内容就无法修改。

容器（Container），与镜像具有基本相同的内涵，最大的区别是容器可以修改，这意味着它可以写入到最顶层。等于是只要给镜子增加一个可以读写的顶层就行了。

仓库（Repository），由两部分组成:仓库注册服务器和仓库。前者可以为镜像提供集中存储和分发服务，后者可以为镜像文件提供集中存储。

（2）Docker的架构

Docker 是客户端/服务器结构，如图3.1所示。客户端直接申请服务器，服务器承担容器的建立、操作与分配。客户端既可能与服务器执行在同一Host上,也可使用socket或REST API远程通信。

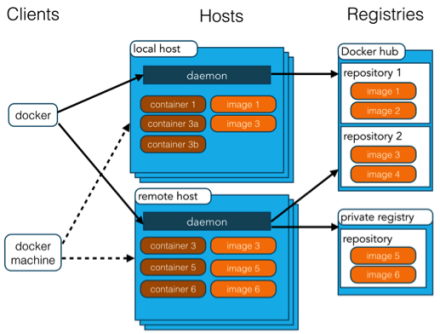


图3.1 Docker的C/S架构

Docker的架构由四个部分组成：客户端、守护进程、镜像和容器。

其中，镜子和容器已经在上面描述过了。客户端提供了命令行工具，这是用户与Docker交互的主要方式。其主要是创建和退出应用程序。守护进程是服务器的一个组件，是Docker的中央后台进程，其主要任务是执行系统调用和容器管理操作以响应客户端请求。

### 3.1.2 Snort介绍

Snort是一种功能强大的互联网攻击监测/防范技术,具备多服务、即时流量分析、IP数据包查询(Pocket)等功能。

Snort系统包括三个基本运行方式:Sniffer系统、数据信息包标记器、以及互联网入侵监测系统。Sniffer系统仅需从互联网读取数据包,并在终端中将数据表示成连续流。包将日志模式的包写入磁盘。网络入侵测试模型是最复杂、且可选择的模型。我们能够允许snort解析网络数据流以符合特定的自定义准则,并针对结果选择了某些方法。

1. 工作原理

Snort能够根据捕获分析网络中的包，但它与其他探测器不同，因为它可以根据特定规则进行响应和处理。通过分析收到的数据包，在分析规则后，根据规则链，可以使用激活机制(标记并启动另一个动态规则链)、Dynamic(从另一个规则包调用)、Alert(标记)、Pass(跳过)、Log(不标记而是记录网络通信)五个响应。

Snort具备了数据流探测器、解析、数据流监测、反馈管理等多项主要功能,通过各种模组完成了各种各样的功用,各个模组都以插件方法和Snort技术相结合,功能易于扩展。例如,预处理插件的主要功能包含:启动直到规则被滥用;重新组织键入的片段;实现如HTTP解码、Telnet编码等新功能;处理插件以执行验证日志字段;结束链接;响应攻击；以及其他正确记录或警告输出插件的功能。

1. 工作过程

Snort可以从5级TCP/IP网络结构的连接级别捕捉网络数据包,但数据包捕捉功能必须与无线网卡放在混合模式,也可以通过libpcap功能或winpcap功能直接在互联网中抓取数据包,但具体取决于操作系统。之后,在解码器中解码数据包,并且解码器解码格式是统一的、为Snort所认知的。然后发包预处理，包括重组分片和处理明显错误等，该过程由插件进行。

数据包经过编码、筛选,经过预处理后加入主要Snort环,建立规范并按照规定实施管理。规则控制也是Snort的主要组成部分,其功能在于判断外包是否存在入侵行为。该规则的语法涉及各种元素，例如类型、内容、长度、标题等。如果您使用规则档案，请使用3D清单储存有关与后续套件协调规则的资讯，而3D清单在建立后，会以某种方式搜寻3D清单并执行对应和回应。规则控制的性能要求Snort计算机根据其性能受网络负载和其他因素影响的规则数工作。最后一步是输出模式，检测后，软件包需要将结果以各种形式输出到警告文件、另一个日志文件、UNIX数据库域或套接字等。

### 3.1.3 局域网搭建及数据采集

在Docker上以图3.2所示的拓扑图搭建局域网，并在每台主机和路由器上安装布置好Snort用来记录主机和网络的安全日志。具体步骤有拉取镜像、启动elastic、启动kibana、启动logstash、启动suricata等。

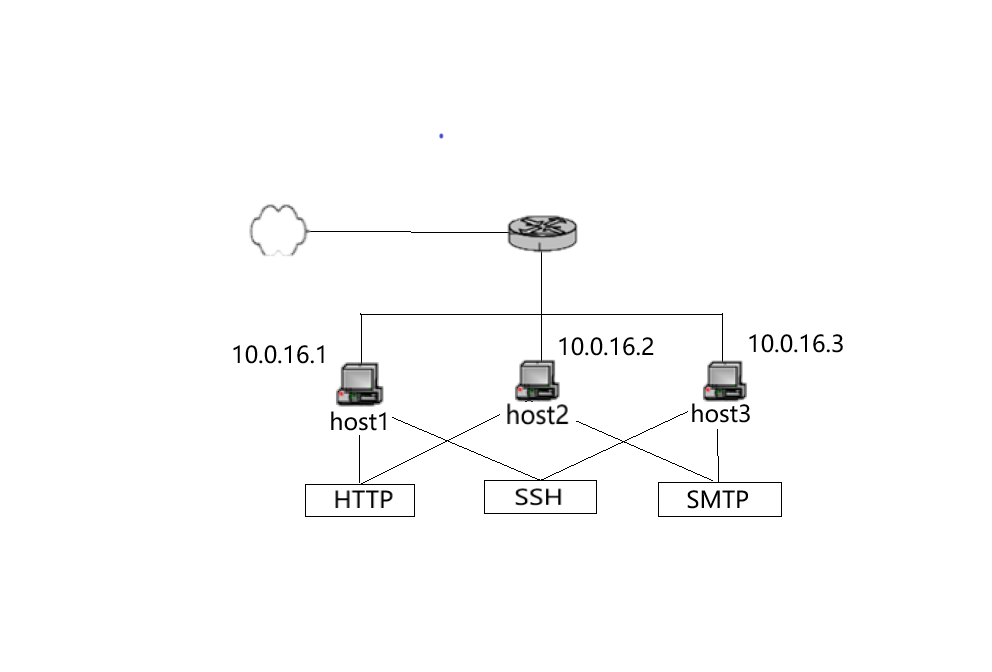


图3.2 局域网拓扑图

其中，suricata是sort的升级版，除了多一点功能，其他的全部一样，其内置了filebeat负责将日志文件传输到logstash。logstash负责收集输日志，进行处理，最后将数据传输至elastic。elastic负责日志的搜索和一些高级操作，elastic处理完成之后，最后传输给kibana进行数据可视化。具体操作过程见附件。

在虚拟环境中模拟正常用户访问各服务器。同时在特定时间模拟网络攻击对系统中的各主机和服务器进行攻击，攻击包括Dos攻击、字典攻击和端口扫描。利用Snort对其进行记录。

## 3.2 网络安全态势评估模型

层次化网络安全态势评估模型（如图3.3所示）是将整个网络分成从高到低的3个层次，即系统级、主机级和服务级，并根据后两者在系统中的重要性对其赋权值。同时确定各服务所受到的各种安全事件的威胁严重指数，以此计算安全态势值。

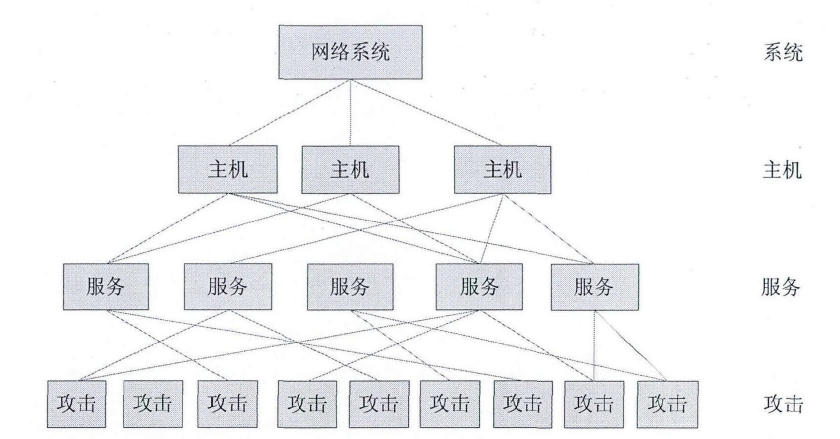
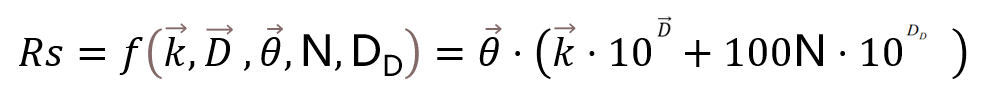


图3.3 层次化网络安全态势评估方法

定义威胁严重指数D：威胁严重指数D是针对网络攻击对指定目标可能带来的后果的一个量化，根据第二章介绍的网络安全评估标准，可以把网络攻击的严重程度分为高、中、低三个级别，量化为严重指数分别对应3，2，1。

### 3.2.1 服务级

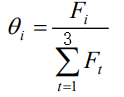
以Δt为计算时间间隔，对于在任意Δt内某服务受到的安全态势值Rs计算，给出以下公式：

 （1）

其中：

1. =（k1,k2,k3）为在Δt内该服务受到的不同严重指数的网络攻击的次数。例如，k1即为Δt内严重指数为1的网络攻击发生的次数。

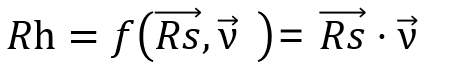
1. =（1，2，3）为威胁严重指数向量，与相对应。为提高评估的精准度，在各种严重指数的网络攻击的等效关系上，直接借用由陈秀真给出的标准[1]，即严重指数每提高一级，反映在安全态势值上提高一个数量级。特别说明DD为Dos攻击的严重指数。
2. =（θ1，θ2，θ3）为正常访问量系数，本文将一天分为三个魇时间段，t1(0:00-8:00), t2（8：00-18：00）和t3（18：00-24：00）。θi即为系统各时间段的访问总量归一化后值。如下，Ft为ti段内的总访问量。

 （2）

1. N为当前Δt内平均宽带占用率，取值（0<N≤100）,故乘100将其转整数。
2. Rs为服务在Δt内的安全态势值，Rs越大，代表在此Δt内服务受到的威胁越大，当引起重视。计算Rs的意义在于通过计算出一段时间内的Rs，经过合理分析可预测其变化趋势。同时也是后续计算主机级或更高级安全态势值的必要条件。

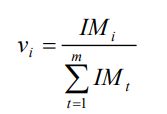
### 3.2.2 主机级

在Δt内主机受到的安全态势值有：

 （3）

其中：

1. 为服务安全态势值向量，是主机中不同服务在Δt的安全态势值。
2. 为主机内服务的权重向量，与一一对应。中νi的取值由该服务在该主机重要性IMi决定。服务重要性分低、中、高，量化为1，2，3。νi即为各服务重要性量化归一后的值，具体计算公式为:

 （4）

1. 服务重要性IM的判断本文主要是根据客观统计信息和陈秀真给出的主观经验知识[1]——主流服务的用户数越多、访问频率越大，服务重要性越高。并制定了服务重要性断定标准如表3.1所示。

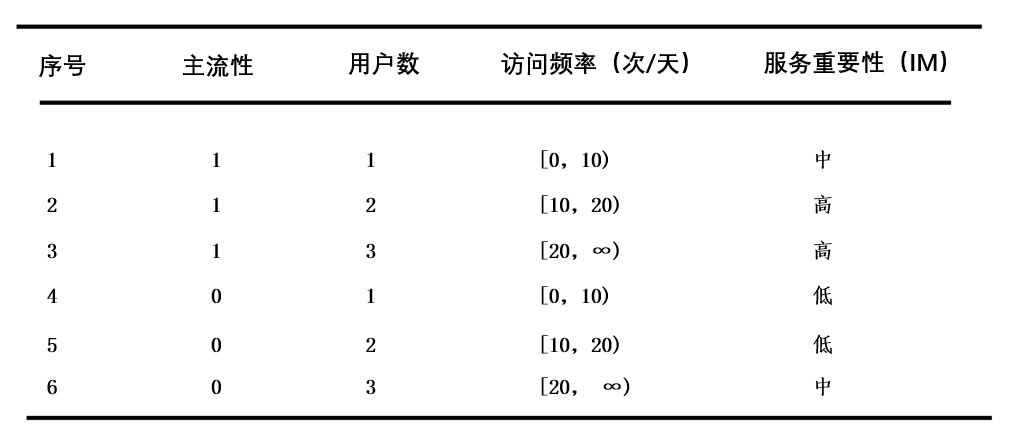
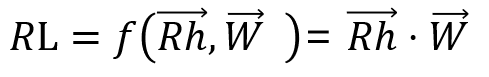


表3.1 服务重要性断定标准

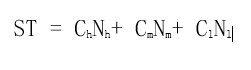
### 3.2.3 系统级

在Δt内主机受到的安全态势值有：

 （5）

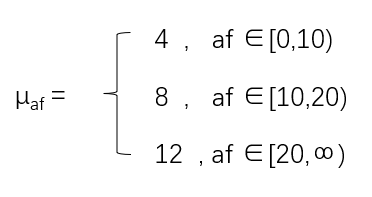
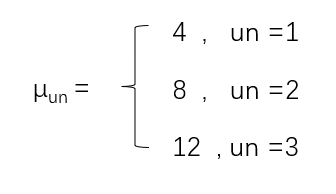
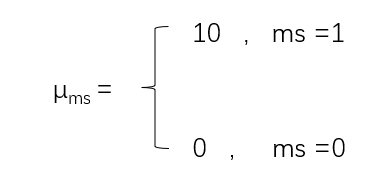
其中：

1. 为主机安全态势值向量，是系统中不同主机在Δt的安全态势值。
2. 为系统内主机的权重向量。中ωi取值由主机在系统中的地位ST决定。
3. 定义主机在系统中的地位ST由各个重要级别的服务数量确定，计算公式为：

 （6）

其中，Ch、Nh分别为重要性为高的服务的量化分值和数量；同理，Cm、Nm和Cl、Nl分别对应重要性为中和低的服务的量化分值和数量。C的取值基于表3.1的判断标准，定义如下：

1. 服务主流性ms、用户数un、访问频率af的量化分值为

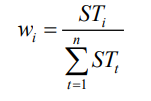


1. 对于μms、μun、μaf三者，根据安全管理人员的经验知识，给予三者如下系数关系

Iμ = μms + μun + μaf （7）

1. K的取值即为Iμ 的中值。

（4）计算出ST后，对其进行归一化处理即得到主机的 相对权重：

 （8）

## 3.3 评估模型的实现

### 3.3.1 各级态势值计算

（1）服务级

根据3.2节中网络攻击威胁严重指数定义，对于3.1.3节中数据采集时模拟的攻击类型，可给出表3.2如下：

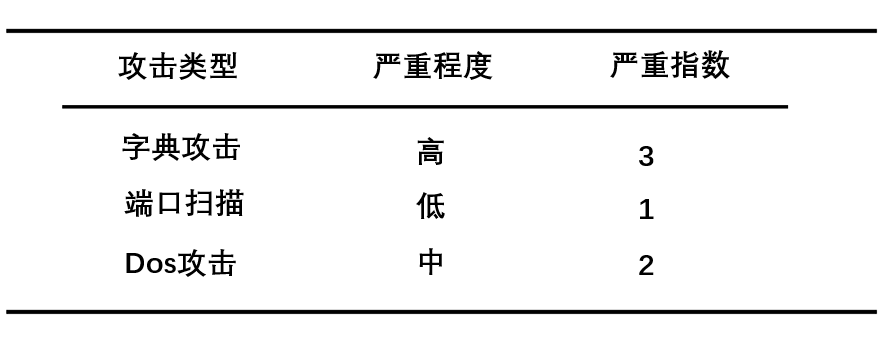


表3.2 攻击类型与威胁严重程度和严重指数

将Snort安全日志解析为xsl格式，利用Excel进行简单处理并根据3.2节里的要求进行赋值，再利用Python结合表3.2实现公式（1）计算得到服务级的安全态势值，以主机host1中的http服务为例，其态势值如下：

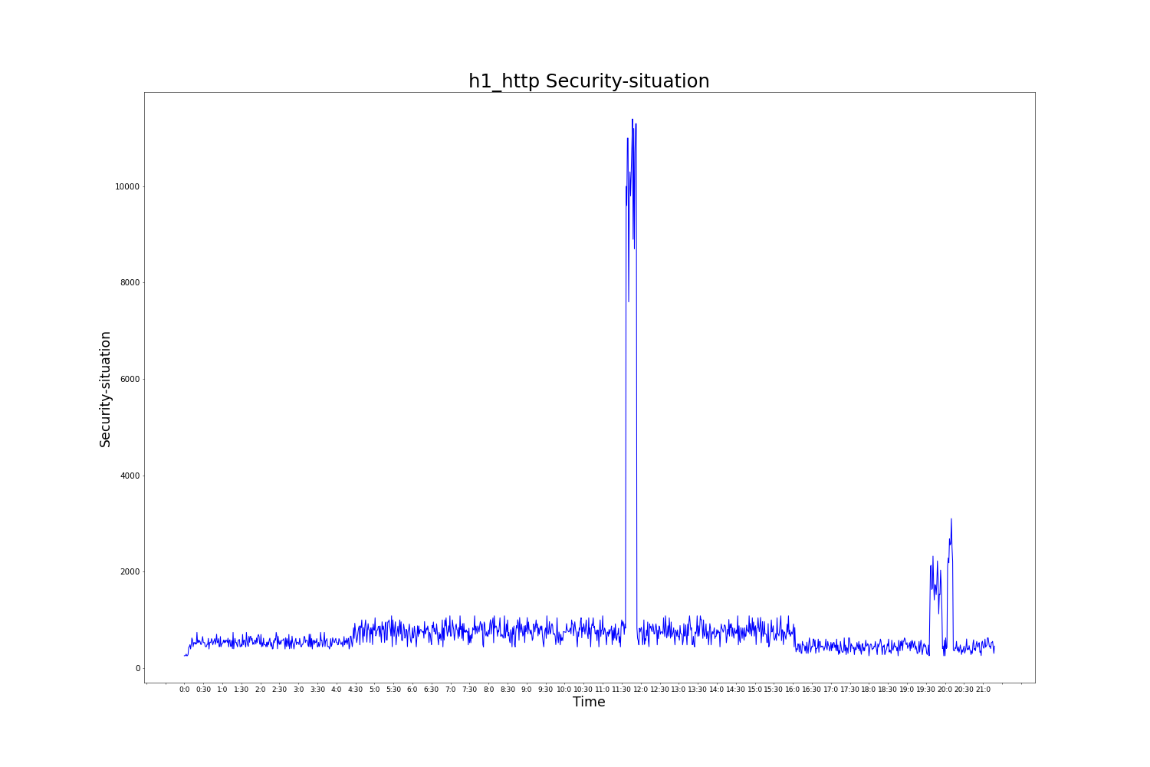


图3.4 主机host1中http服务安全态势值

（2）主机级

根据3.2.2节中主机内服务权重的定义，给出表3.3如下：



表3.3 主机内服务的重要性权重

结合表3.3实现公式（3）计算，结果如下：

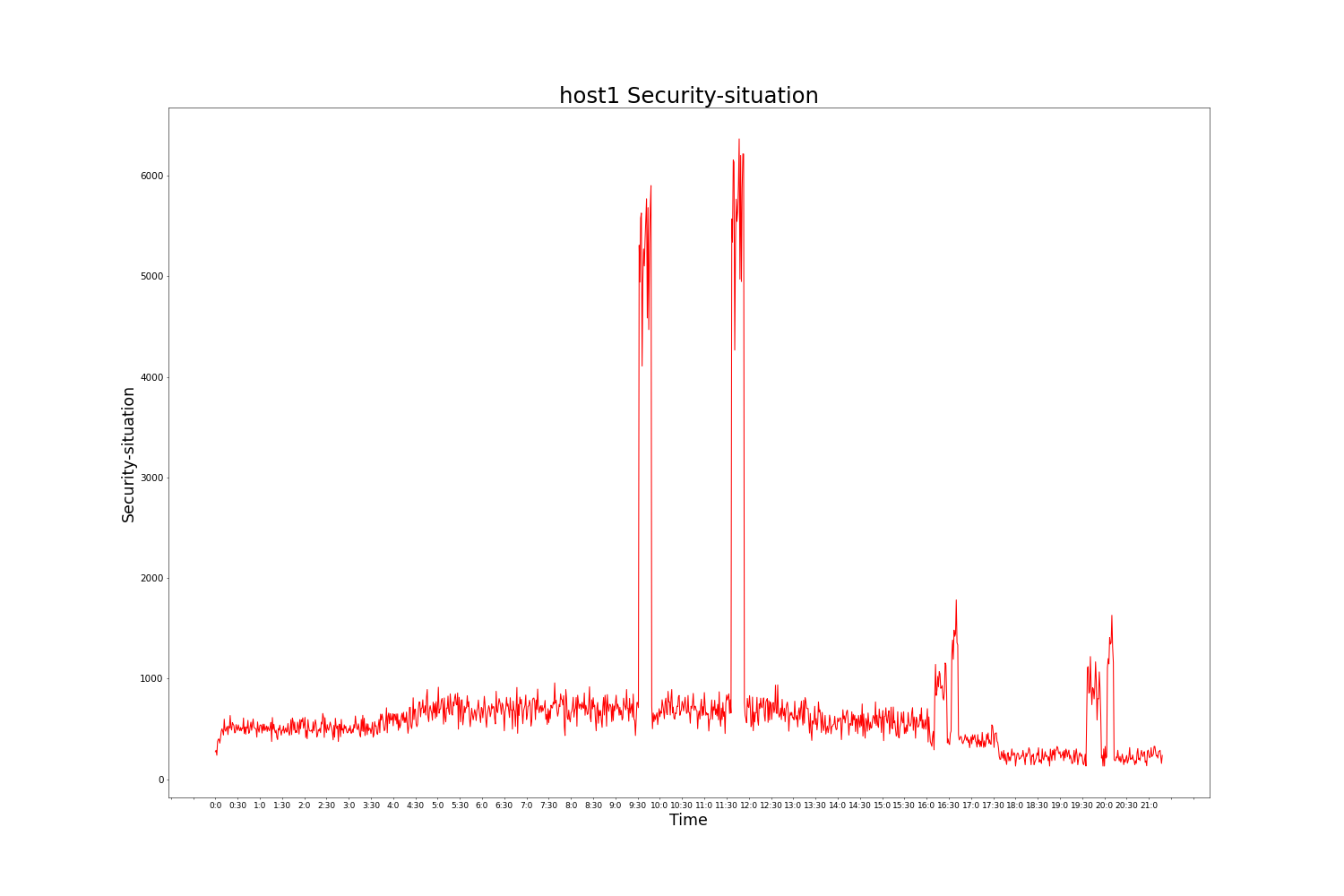


图3.5 系统内主机host1的安全态势值

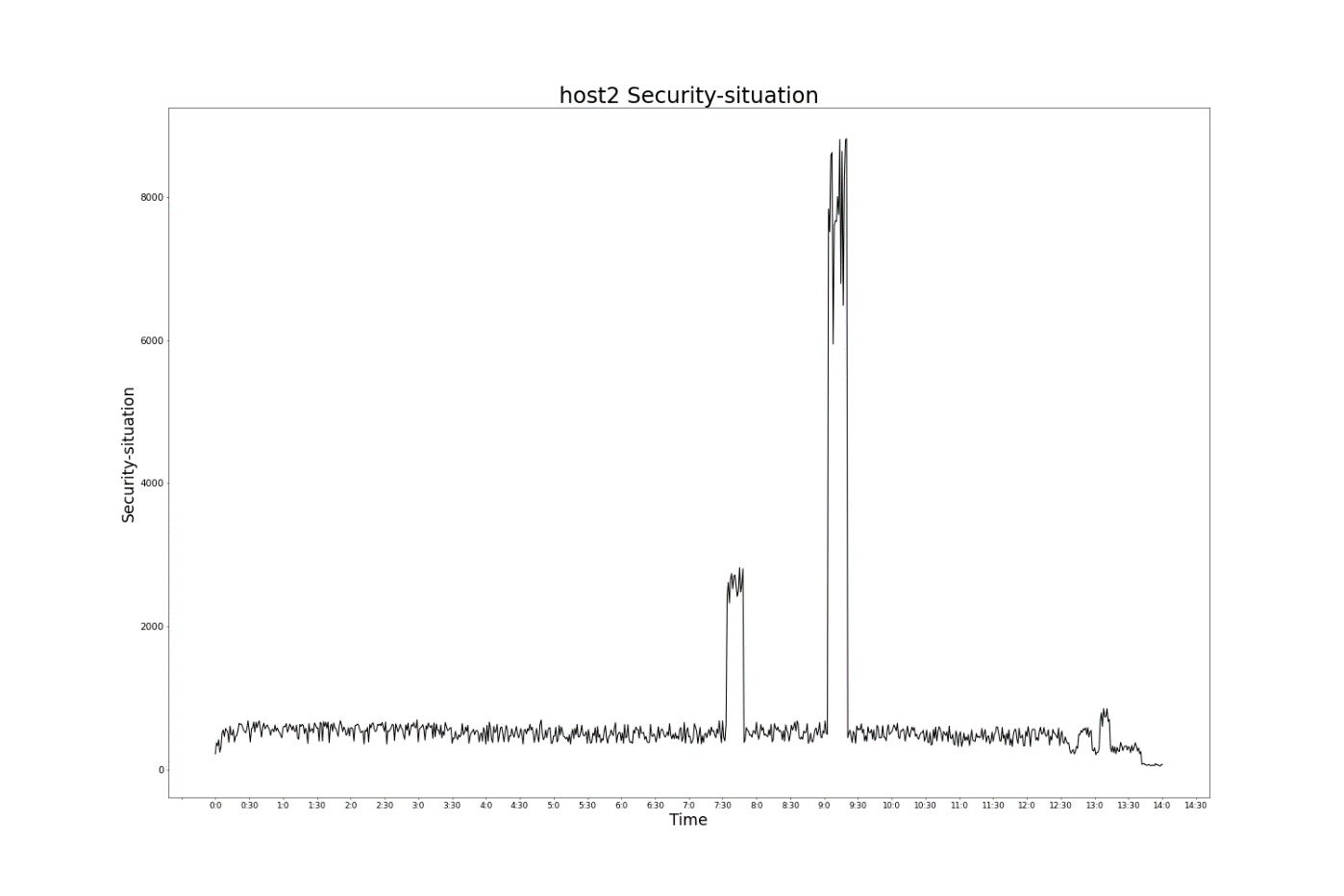


图3.6 系统内主机host2的安全态势值

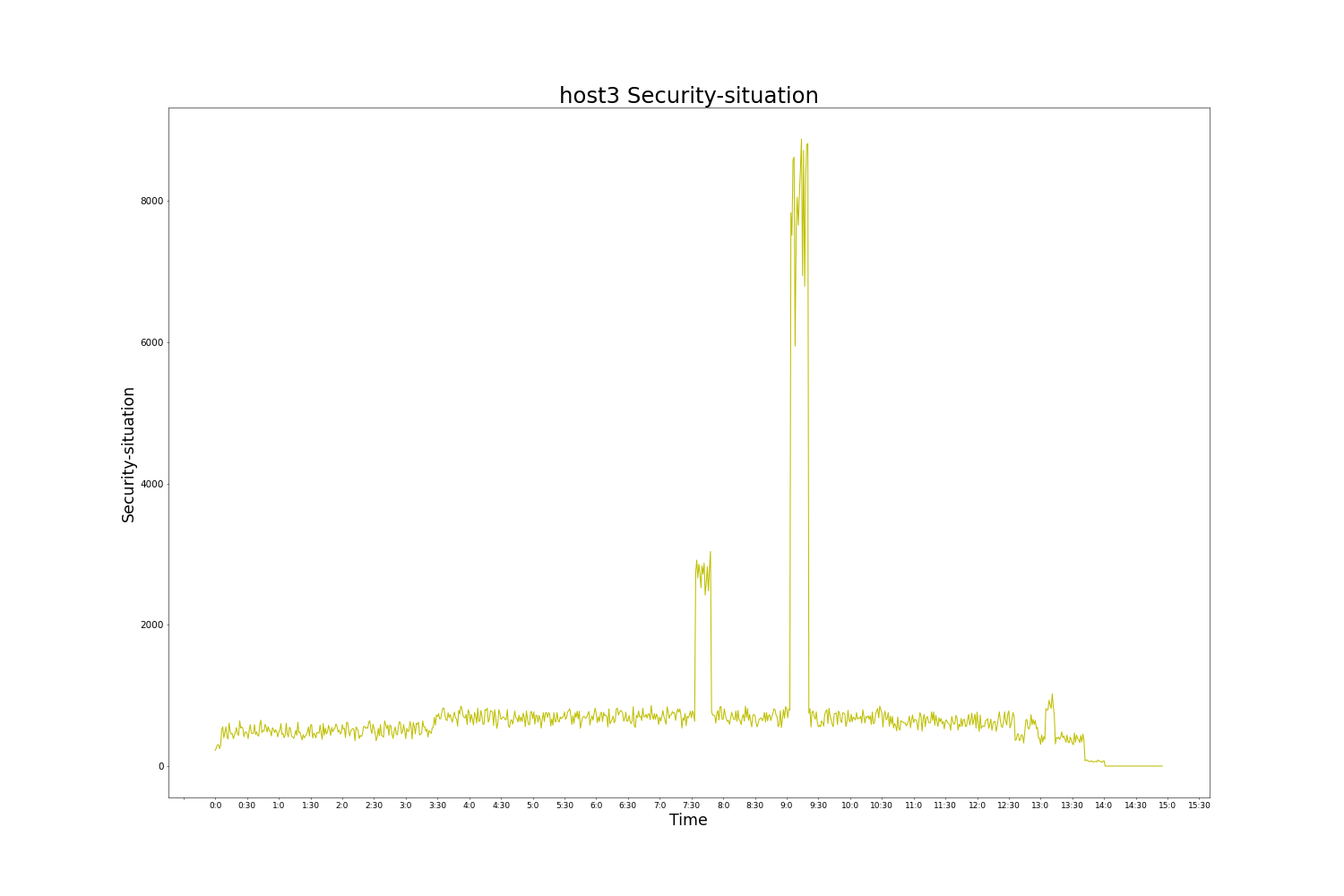


图3.7 系统内主机host3的安全态势值

1. 系统级

根据3.2.3节中对主机重要性权重的定义，结合公式（6）、（7）、（8）计算出=（0.431，0.296，0.273）分别为主机host1、host2、host3在系统中的权重，结合公式（5）计算出系统的安全态势值如下：

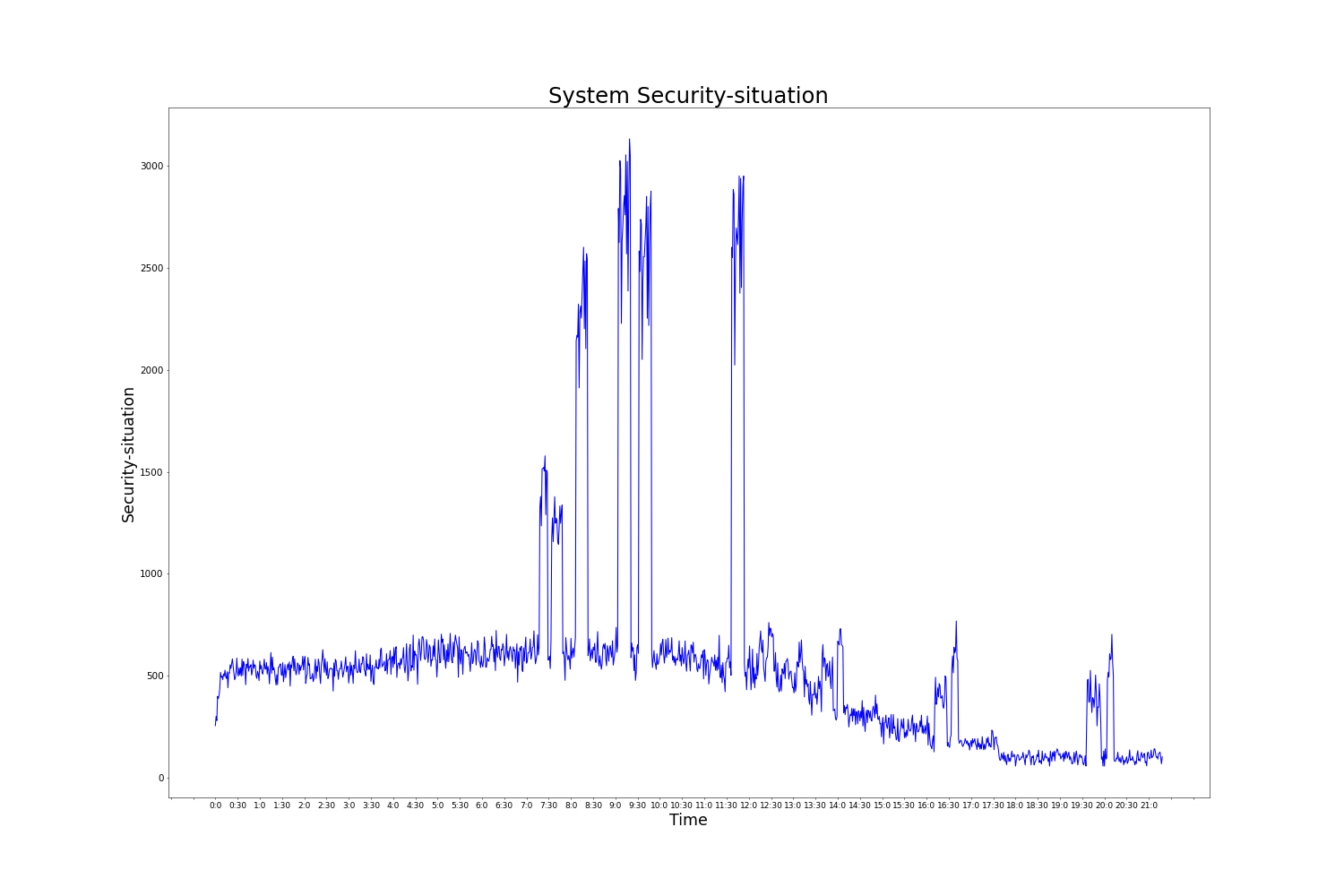


图3.8 系统的安全态势值

### 3.3.2 结果分析

本文在模拟局域网上采集数据时，在9点30分左右和11点30分左右分别针对ssh服务和http服务发起了字典攻击和Dos攻击，从主机级的安全态势值和系统级的安全态势值中可明显反映出来。同时在攻击前也曾进行过端口扫描，通过各自的安全态势曲线可分析出来。

## 3.4 本章小结

结果表明,所提出的模型能够直观地得出对整个网络、主机器以及服务的三个级别安全态势,从而使得网络管理员能够快速掌握整个网络的安全动向。找出变化根源,修改安全策略,最大程度的提升网络安全性。另外,安全状态变化可分析态势的长期曲线来判断。此证明了该系统可用于安全态势预测，能够合理评估网络攻击对安全系统构成的威胁，并将管理员从分析警报数据的艰巨任务中解放出来。

# 第四章 网络安全态势预测

## 4.1 BP神经网络介绍

1986年科学家Rumelhart和McClelland开发的BP (Back Propagation)神经网络是一个多层神经网络，具有误差逆传播反馈能力，是最常用的神经网络模型之一。。

### 4.1.1 基本原理

BP神经网络中不需预先确定一个数学方程式,仅利用机器学习训练来映射输入与输出信号间的关联,从而获得与给定输入数据中的预期数据值最相似的结论。为了人工智能信息处理系统,用来完成其操作的人工神经网络的核心技术就是计算机。而BP神经网络则是一个经过误差反传技术训练的多层前馈网络体系,这个方法也叫做BP方法,其主要思路是采用渐进搜索技术以及逐渐减小缩放的方法,以达到在网络实际和预期的值中偏差的最小方差。

BP的基本计算过程包含了二种,用于传递前向信息和差错反传的步骤。也就是说,在传递到输入方向上计算误差输出值,并在输入输出方向上调节权重和阈值。在正向中,将输入消息穿过输入输出节点中的隐式层,从而完成了非线性转换过程,如实际输出值和预期输出信号不一致时,即生成输出信息,从而进入了错误的反传递过程。而反向传输误差则是将输出误差透过隐藏层反向传输到最上方的一个层次,将误差消息散布在各个层面上的每个模块中,使在各个层面接受到的错误信息,成为了调节每个块权重值的基础。通过设定入口节点与隐藏层节点的连线强度,以及调整隐藏层节点与输出节点之间的连线功率及其阈值并停止训练,可以使得误差向梯度下降,并通过重复练习和培训可以设定与最小偏差对应的网络参数(权重和阈值)。此时,经过训练后的神经网络将能够为此类样本输入信号,并以很小的输出误差独立处理经过线性变换后的信息。

### 4.1.2 BP神经网络的优缺点

（1）优点

①非线性映射能力。BP神经网络实质意义上完成了一个由输入信号到输出信号映射的功能，数学理论早已证明了三层的神经网络将可以以几秒任意的精度来逼近任意一个非线性连续函数。这使得其尤其地适用于那些解决内部机制复杂性极强的问题。换言之，BP神经网络具有强的非线性映射能力。

②自学习和自适应能力。BP神经网络的训练就是学习和自己主动提取出输入信号、输出信号之间的潜在规则，并将学习内容保存在隐藏层的权值中。总的来说就是BP神经网络的自学习和自适应能力很强。

③泛化能力。泛化能力是指在设计模式分类器时，要确保分在类正确，还要确保网络训练完成后，能对未知或有噪声污染的对象分类正确。也就是说，BP神经网络有能力将学习结果应用于新知识。

④容错能力。BP神经网络局部或者少数神经元对结果输出的影响极小，换句话说就是系统局部损伤时可以正常工作。即BP神经网络具有一定的容错能力。

其中BP神经网络的第一和第二个优点是研究的关键，也是选题的根本目的。由于网络安全态势预测的指标复杂，直接利用数学模型处理困难，所以建立从安全因素到安全态势的神经网络模型，能够很好将各种指标因素统一起来，更方便进行预测。通过分析预测结果，找出影响网络安全的根源，及其所造成危害的严重程度，对网络进行有针对性的主动防护和预防。

（2）缺点

BP神经网络无论在网络理论还是在性能方面已比较成熟。其突出优点就是具有很强的非线性映射能力和柔性的网络结构。网络的中间层数、各层的神经元个数可根据具体情况任意设定，并且随着结构的差异其性能也有所不同。但是BP神经网络也存在以下的一些主要缺陷：

①学习速度慢，即收敛慢，而且在处理简单问题时会把问题复杂化从而造成资源浪费。

②容易陷入局部极小值。

③网络层数、神经元个数的确定缺乏科学理论指导。

④网络推广能力有限。

针对这些问题，如今已有众多改进甚至解决措施，研究热点就是收敛速度和局部极小值的问题的改进。

## 4.2 预测模型设计及训练

实验利用第三章计算得到的系统网络安全态势值，选取前999个数据作为BP神经网络的训练集，其余数据作为测试集。

为减少因个别数据出错而导致的误差，同时也为了更好、更容易提取出态势数据间的强关联，使BP神经网络模型的预测达到高的可靠性。在此以编组的方法提升模型预测准确度,对序号为1-999的态势值，以每三个为一组作为模型的输入信号，下一个态势值作为模型的预测输出信号，如表4.1所示：

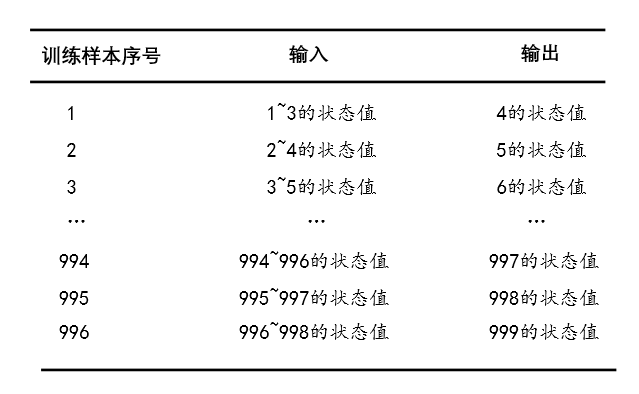
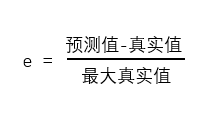


表4.1 样本的构建

考虑到实验的输入输出信号都是很简单的数据,因此我们的BP神经网络模型采用三层来结构,神经网络的输入层有三个信号(即前三个的网络安全态势值),输出层有一个信号(后一个的网络安全态势值),而隐层神经元数量则采用了经验公式并确定为12。并定义误差为：



利用Python实现BP神经网络模型，将训练集导入，经过训练得到结果如下：

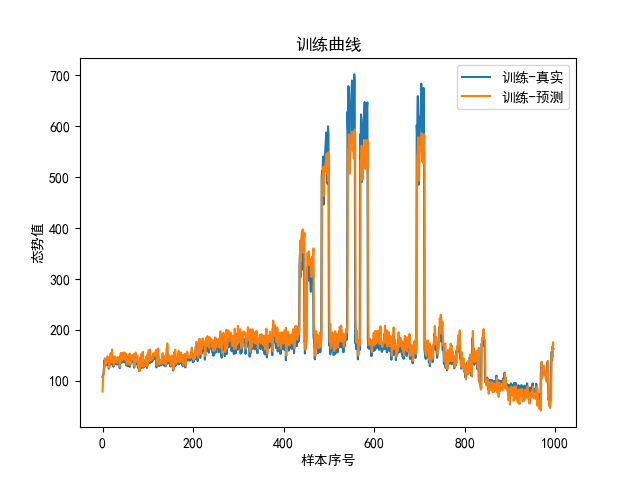


图4.1 预测模型的训练集曲线

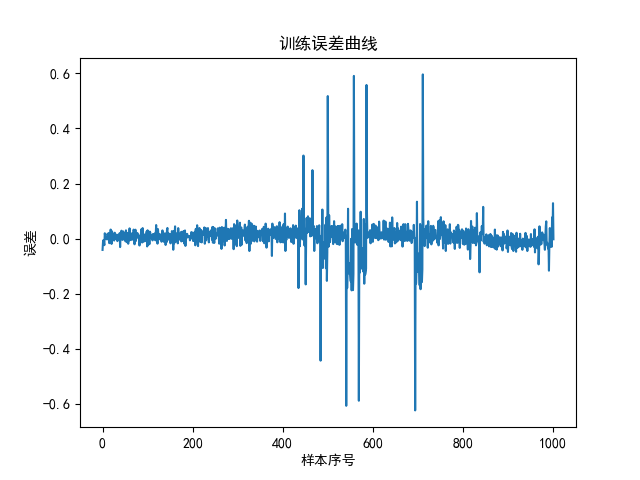


图4.2 预测模型训练时的误差曲线

## 4.3 预测实验

BP神经网络预测模型在经过4.2节中所述的训练并将误差减小到可接受范围后，利用系统网络安全态势值测试集对预测模型进行测试实验，结果如下：

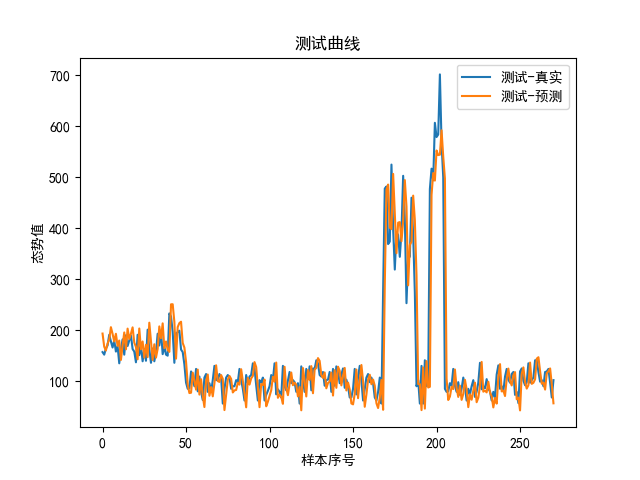


图4.3 预测模型经测试集测试结果

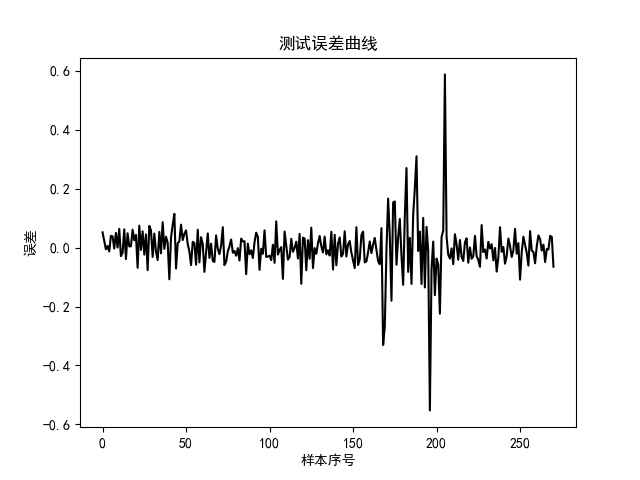
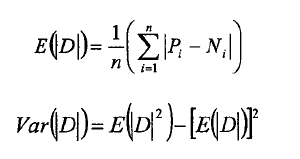


图4.4 预测模型测试时的误差曲线

测试结果分析：

在训练曲线和测试曲线图中，橙色曲线为由BP神经网络预测模型输出的态势预测值，蓝色曲线为训练集或测试集的真实态势值。通过图4.1和4.3可以看出橙色曲线和蓝色曲线基本重合，即BP神经网络预测模型输出的态势值走势大致正确，基本上能够较好的反映网络安全态势的发展趋势。

结合图4.2和图4.4（分别为训练和测试时的误差曲线），可以看出误差整体基本在[-0.1,0.1]范围内波动，但在态势值发生突变的地方还存在较大误差，即该模型在局部还存在些许缺陷待已改进。进一步对误差进行分析，利用概率统计相关知识计算绝对误差的期望和方差，公式如下：



其中，*E(|D|)*表示误差绝对值的期望，*V*ar*(|D|)*表示误差绝对值的方差，*P*i为某时刻态势的预测值，*N*i该某时刻态势的真实值。

计算得到*E(|D|)*= 0.0497612，*V*ar*(|D|)*=0.00419866。可以看出预测模型对实验数据的预测误差整体上是比较小且相对均衡稳定的，其预测结果可信度较高。

# 第五章 总结与展望

## 5.1 总结

针对当今复杂的网络安全问题，网络安全态势感知区别于对某一方面单个的预防措施，为网络安全问题的解决提供了新的研究方向和方法。如何利用安全设备产生的数量庞大的日志和的数据，从宏整个网络的角度评价网络环境再以此为依据提出策略修改意见，目前已成研究热点之一。网络安全态势感知从宏观角度去分析网络安全信息，分析不同安全事件之间的关联，科学地对网络安全状态给出评估，并作出有效、可靠的预测。BP神经网络作为一种已经发展得相对成熟的神经网络，其在机器学习和数据挖掘方面有较深和较广的应用。

本论文所做的主要工作主要包含以下几个方面：

（1）深入研究了态势感知的发展背景和国内外现状,包括安全态势预测理论与分析方法,并对比了不同预测方式。同时分析了国内外比较有名的信息安全评估标准，以此作为网络安全评估前提条件。此外，简单对网络安全态势的时间序列的可预测性的的方法进行了研究。

（2）借鉴陈秀真提出的层次化网络安全威胁量化标准，构建了网络安全态势评估模型。其中着重对模型的指标体系构建和指标提取进行了研究。本文进行实验的数据来自于在Docker搭建的局域网上利用Snort自行采集的数据，包括模拟的网络正常访问数据和网络攻击数据。因此，在理论研究方面还对Docker和Snort相关理论进行了研究。之后，对数据进行了处理，使其格式和形式满足网络安全态势评估模型所需。最后通过编写Python代码实现计算工作得出模拟局域网的网络安全态势评估值。

（3）在预测阶段，针对性地研究了BP神经网络基本理论，并阅读了一些针对BP神经网络缺陷提出的改进方法，初步构建起BP神经网络预测模型。利用得到的模拟局域网的网络安全态势评估值，进行训练直至模型的预测误差降低到可接受范围内，并重新模拟采集数据进行测试，结果仍在接受范围内。

## 5.2 研究展望

尽管本文最后使用BP神经网络预测模型得到的预测结果在可接受的范围，但是由于个人能力和实验环境、工具等限制，仍然有许多不足，需要再作改善。

（1）使用该BP神经网络预测模型可有效地预测网络安全态势值。但这个结果严重依赖网络安全态势评估方法自身的可靠性，只有评估够严谨，结果反映的现状才够真实和科学。本文在态势评估方法上直接借用了现有的态势评估方法，为进一步提出预测结果的真实性和科学性，对更加严谨和科学的网络安全态势评估方法研究势在必行。

（2）从结果上看，该BP神经网络预测模型的预测值在数据发生突变的地方虽然从整体上看能够明显反映出突变的趋势，但就结果本身而言其误差相对较大。对BP神经网络本身缺陷的改进也将是下一阶段工作的重点之一。

# 致谢

岁月如梭，时光荏苒，不知不觉四年的大学生活即将结束。在这充实且充满美好记忆的四年中，在干部、导师的殷勤教导和同学朋友的帮助之下，相较于四年前刚踏入这个校园时有了很大的进步，在此向他们表达真诚的感谢。

首先， 要感谢毕设导师孙佳佳教员。感谢您在毕业设计和论文整个过程中对我的指导，从查阅资料、选题、中期和最终答辩都提出了大量宝贵意见，帮助我顺利完成毕设。在这个过程中也使我对对专业知识技能有了进一步的强化，使我的个人动手能力和综合素质有了较大提升。

其次，要感谢这四年带过我的几位干部。有大一和大二时的蔡兆忠队长和贾斌、刘建、汪永飞三位教导员，感谢你们帮我扣好了这第一粒扣子，为我这四年的大学生活打下一个好的基础。还有大三之后的吴疆、冯为江两位队长和聂东发教导员，感谢你们对我的敦敦教导，在后两年中给予了我莫大的帮助和鼓励。

最后，感谢专家组的老师们在百忙之中为我的论文审稿，以及提出宝贵的意见和建议。

# 参考文献

1. [1]陈秀真,郑庆华,管晓宏,林晨光.层次化网络安全威胁态势量化评估方法[J].软件学报,2006(04):885-897. [↑](#endnote-ref-1)
2. [2]第48次《中国互联网络发展状况统计报告》发布 超十亿用户接入互联网 我国成全球最庞大数字社会[J].网络传播,2021(09):76-81. [↑](#endnote-ref-2)
3. [3]Endsley, Mica R . Measurement of Situation Awareness in Dynamic Systems[J]. Human Factors, 1995, 37(1):65--84. [↑](#endnote-ref-3)
4. [5]Zhang H , Shi J , Chen X . A Multi-Level Analysis Framework in Network Security Situation Awareness[J]. Procedia Computer Science, 2013, 17:530-536. [↑](#endnote-ref-4)
5. [7]胡红. 网络安全态势感知研究现状及分析[J]. 中国市场, 2011(49):2. [↑](#endnote-ref-5)
6. [8]王雪. 基于时间序列分析的网络安全态势预测模型研究[D].北京邮电大学,2015. [↑](#endnote-ref-6)
7. [9]文志诚, 陈志刚, & 唐军. (2016). 基于时间序列分析的网络安全态势预测. 华南理工大学学报：自然科学版, 44(5), 8. [↑](#endnote-ref-7)
8. [10]王文思. 基于贝叶斯网络和支持向量机的网络安全态势评估和预测方法研究[D].南京邮电大学,2020.DOI:10.27251/d.cnki.gnjdc.2020.000301. [↑](#endnote-ref-8)
9. [11]刘俊男,陈占芳,姜晓明,朱利莞.用于网络安全态势预测的SAGPSO-SVM模型研究[J].长春理工大学学报(自然科学版),2019,42(06):126-128. [↑](#endnote-ref-9)
10. [12]吴国强. 网络安全评估的研究[J]. 信息安全与技术, 2012, 03(1):10-12. [↑](#endnote-ref-10)
11. [13]http://www.itsec.gov.cn/webportal/portal.p. [↑](#endnote-ref-11)
12. [14]Dod -std. Department of Defense (DoD) Trusted Computer System Evaluation Criteria. 1999. [↑](#endnote-ref-12)
13. [15]张学工译. 统计学习理论的本质[M]. 北京：清华大学出版社. 2000年 [↑](#endnote-ref-13)
14. [16] VN Vapnik, AY Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities[J]. Theory of Probability and Its Applications, 1971, 16(2):264-280. [↑](#endnote-ref-14)