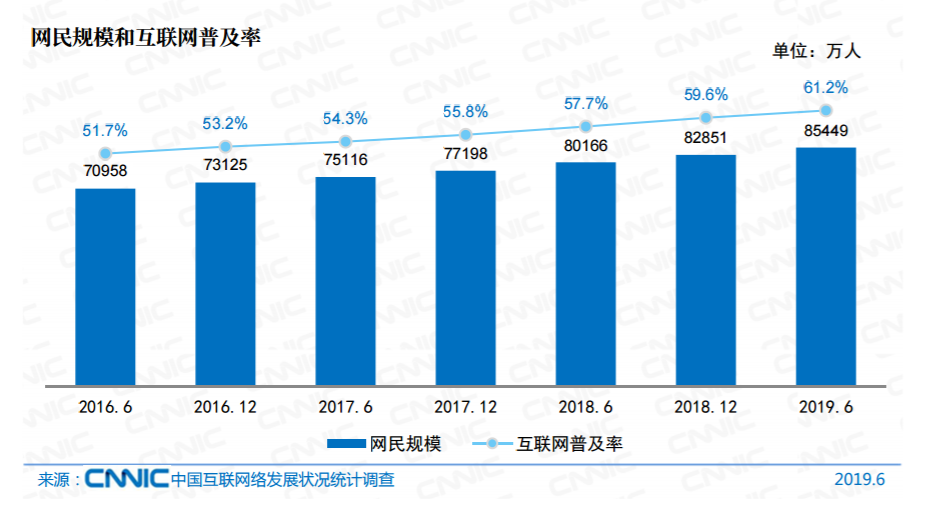
**1.1研究背景及意义**

近年来，Web服务由于其高度可用性和便利性，已经成为越来越多信息服务的主要提供方式。根据第六届世界互联网大会发布的《中国互联网发展报告2019》和[第44次《中国互联网络发展状况统计报告》](http://www.cac.gov.cn/pdf/20190829/44.pdf)显示，截至2019年6月，中国网民规模为8.54亿人，互联网普及率达61.2%，网站数量518万个。



Web应用规模的快速增长带来便利的同时，日益扩增的网民规模吸引了网络攻击者们通过互联网非法牟利。尽管这些不法活动的手段各有不同，但是为了达到攻击目的，它们都需要用户在不知情的情况下访问攻击者提供的网页地址，这些网页也因此被称为恶意网页。这些恶意网页通过用户对网页的信任而发起一系利恶意攻击，如网页挂马，跨站请求伪造等,对用户的数据安全以及财产安全产生了巨大的威胁。

根据国家信息中心联合瑞星公司日前发布《2019年中国网络安全报告》显示，2019年病毒样本总量上亿，勒索软件和挖矿病毒呈现爆发态势，瑞星“云安全”系统去年共截获病毒样本1.03亿个，病毒感染次数4.38亿次，病毒数量比2018年同期上涨32.69%，报告期内，新增木马病毒6557万个。

根据卡巴斯基实验室2019年年度报告《卡巴斯基安全报告2019》，19.8%的用户计算机在过去一年中至少遭受过一次恶意软件的网络攻击，2.73782113亿个URL被Web防病毒组件识别为恶意URL。此外，根据2019年4月百度安全发布的《互联网安全报告：态势分析与生态治理》报告数据显示，2018年全网日均新增检出恶意网页1456.6万，同比增长89.96%。

然而，传统的检测方法大多都基于恶意URL规则库进行匹配，检测方式固定，检测性能不佳并缺少灵活性，很容易被黑客所绕开。因此，研究如何对恶意网页进行主动性检测已经成为亟待解决的网络安全领域的重点。

**1.2国内外研究现状**

针对恶意网页识别问题，研究人员提出很多识别技术和解决方案：主要有客户端蜜罐技术、静态检测技术和机器学习技术三类。

传统的蜜罐技术是通过主动暴露系统漏洞，被动地等待恶意攻击，是一种对攻击方进行欺骗的技术，蜜罐实际上是经过周密布置设下的“黑匣子”，看似漏洞百出却尽在掌握之中，因此，蜜罐是一个安全资源，它用于被探测、攻击和损害。传统的蜜罐技术面向服务器，主要有低交互式服务器蜜罐和高交互式服务器蜜罐，而相对与传统蜜罐技术的客户端蜜罐技术顾名思义是面向客户端的。然而不同于面向服务器时的被动等待，客户端攻击只有主动与服务器交互时才能被检测出来。2004年一种新型的蜜罐技术——客户端蜜罐技术应运而生，客户端蜜罐技术又分为低交互式客户端蜜罐和高交互式客户端蜜罐，结合这两种客户端，又设计出了混合式客户端蜜罐。

高交互式客户端蜜罐在与网站进行交互时利用真实的系统真实地模拟人与网页的交互，主要是基于行为检测的方法，攻击者很难发现这是一个蜜罐，并且这种蜜罐会详细地记录攻击日志，因此正常的网页很难被高交互式蜜罐误判为恶意网页，即它的误判率非常低，可以说检测准确率为100%，然而由于识别过程需要等待攻击行为出现而且需要对系统的状态的变化进行详细地分析，高交互式客户端蜜罐的检测速度非常缓慢，并不适合对大规模网页的检测。

不同于高交互式客户端蜜罐，低交互式客户端蜜罐在与网站进行交互时利用模拟的系统与攻击者进行交互程度很低的交互或者根本不进行交互，因此它非常容易暴露。低交互式客户端蜜罐借鉴了传统的恶意代码检测机制，利用静态页面检测算法分析检测网页代码，因此会存在误判率较高的问题，虽然不适合用来对恶意网页的攻击行为做深入的检测分析，但是它的检测速度很快、系统结构简单配置快速且占用系统资源少，可以用于大规模网页的初步检测筛选。

低交互式客户端蜜罐可以模拟真实系统的一般功能虚拟出大量的IP地址但是它的交互能力差，误判率高，很难进一步收集攻击者的信息，高交互式客户端蜜罐虽然判断精准且能收集大量详细的攻击者信息和动向，但是它结构复杂，维护成本高；混合式客户端蜜罐技术则吸取了低交互式客户端蜜罐和高交互式客户端蜜罐的优点，在保证可拓展性和准确率的同时，又降低了成本。

静态检测算法主要包括模式匹配、静态代码分析和启发式规则这三种。

1.2.1.2.1模式匹配

模式匹配是检测恶意代码最基本的一种方法。它利用能唯一识别一种恶意代码的字符串，也就是特征码来检测计算机中的恶意软件。特征码通常需要足够的长，才能确保此特征码只识别特定的恶意代码，要不然就有可能会将正常的代码判为恶意代码。相比之下，使用含有通配符的特征码是一种更好的方法，因为这种方法可以检测出更多变种恶意代码。模式匹配不仅使用在恶意代码检测领域中，入侵检测领域也使用这种方法来检测入侵攻击。著名的网络入侵检测系统Snort就把模式匹配特征码封装在Snort特征库里。模式匹配也是检测恶意网页代码最基本的一种静态检测算法，它利用特征库里的特征码来检测网页代码，也可以采用更加灵活的正则表达式来检测网页代码中特定的字符序列。为了利用开源的特征库，低交互式客户端蜜罐HoneyC采取Snort特征格式来封装这些模式特征。

模式匹配检测算法虽然简单有效，但是它有以下几个缺点：第一，模式匹配算法不能检测未知攻击。第二，模式匹配算法不能从语义上分析代码。第三，模式匹配算法不能检测出利用了代码混淆技术的网页恶意代码。

1.2.1.2.2静态代码分析

静态代码分析是编译器理论的基础，它根据上下文无关文法对代码进行词法标

记和语法分析。静态代码分析可以检测出代码中危险的命令序列和安全漏洞，

比如Java开发的应用程序和Web服务器脚本语言的安全漏洞。静态代码分析也可以用来检测包含在网页脚本中的恶意代码。首先，通过词法标记，定义和标记网页代码中的字符序列为有意义的标记。然后利用解析算法形成抽象语法树，丢掉没有用的语法信息，最后得到网页代码结构的模板。这个模版表示了一个网页代码的语义特征，我们可以利用这个模版来查询最匹配的模板，通过计算它与已知模板的距离，用与它距离最近的模板的类型来决定这个网页的类型。文献[26]中，作者的实验包括以下四个步骤：(1)采集并且分类恶意网页代码样本；(2)利用静态代码分析技术创建模板；(3)获取正常的网页代码样本；(4)随机采样和评估。他们把恶意网页代码样本划分为不同的类别，每一类含有一种恶意网页代码以及它们的变形。接着，从每一类里面随机选取一个样本并利用静态代码分析技术创建一个模板，就把它当作这一类别的模板。从作者的实验结果可以看出，利用基于语义的静态代码分析方法来检测网页代码，可以到高达99.29%的检测率。虽然静态代码分析可以从语义上分析代码，但是它和模式匹配检测方法缺点一样，静态代码分析方法的模板需要事先定义，所以它也不能有效地检测未知的攻击。

1.2.1.2.3启发式规则

启发式规则已经被用于检测恶意代码和垃圾邮件。通常情况下，启发式规则中每条规则拥有一种通用特征，并且每条规则对应一个分数。一个完整的检测算法有可能采用多种启发式规则。启发式规则算法在检测时，每当匹配上一条规则的时侯，算法就把这条规则的分数加到总和里。如果最后得到的分数高于预先设定的阈值，就说检测的对象是恶意的。对于恶意软件检测，启发式规则具有一定的通用性，只要有一条规则出现在检测的软件中，那么这个软件就有可能是恶意软件，因此启发式规则检测算法可以检测出未知的恶意软件，但是它的缺点是误判率比较高。网页检测中的静态启发式规则检测方法利用的是恶意网页的通用元素。恶意网页包含三个核心元素：攻击代码、引诱浏览者的传播机制和躲避检测的隐藏方法。

(1)攻击代码：攻击代码是恶意网页的中心部分和主要元素，它必然出现在恶意网页中。攻击代码的目标是浏览器、插件以及操作系统的漏洞。最常见的攻击目标是ActiveX组件。这方面的规则包括：插件的数量、脚本标签的数量和XML处理指令的数量。

(2)传播机制：攻击代码可能不直接包含在一个网页，而是从一个攻击服务器导入。导入方法有两种：直接导入资源和重定向。直接导入资源可以从本地和远程服务器导入资源，它存在于很多HTML标签上。重定向可以通过HTML或JavaScript命令浏览器打开一个新网页。这方面的规则包括：包含源信息的内嵌框架的数量、重定向的标志和包含源信息的脚本标签。

(3)隐藏方法：通过混淆技术隐藏攻击代码和传播机制是恶意网页常用的一种方法。恶意网页脚本代码给出的是混淆的形式，同时还会有一个自定义的反混淆函数，用来将代码转换成简洁的形式。一旦转换完成，这些代码就会被运行。这方面的规则包括：标志着脚本混淆的函数和元素，以及内嵌框架的可视性和大小。但文献[28]中作者的实验结果不是很理想，虽然误判率只有5.88%，但漏判率却高达46.15%。

1.2.1.3机器学习

检测恶意代码时很重要的一点是能尽早发现恶意代码。尽早发现，尽快处理，这样就能尽可能地减少损失。早期恶意代码检测的方法是采用基于签名的方法。这些签名由文件名、代码字符串或字节等许多不一样的元素组成，同时保护系统安全也要考虑如何弥补这些恶意代码所产生的安全漏洞。现在经常使用的恶意代码检测方法是通过与一个已经有的恶意代码库进行比较的方法，这种方法在很大程度上需要靠已知的代码特征描述来进行检测。然而，一个新的恶意代码也许不存在任何已有的特征库里的特征，所以一个新的恶意代码是无法通过传统的基于签名的方法检测出来的。所以可以检测出从来没有出现过的恶意代码是设计的恶意代码检测系统的一个非常重要的问题。为了解决这个问题，基于机器学习的检测方法就成为恶意代码检测领域中的一个焦点。利用机器学习方法来检测恶意代码就能在恶意代码运行前将其检测出来，这样能有效地防止计算机系统遭到恶意代码的恶意破坏。机器学习算法应用于恶意代码检测，主要是利用了其分类的作用，即有监督学习。分类包括以下两个步骤：第一步，通过训练样本训练出一个分类模型。每个训练样本都有一个类标签，表示它属于哪一类，然后通过有监督的学习算法，生成可以用来分类的数学模型。第二步，使用学习到的模型进行分类。先用测试样本评估分类模型的预测准确率。测试样本是与训练样本相互独立的。用分类模型去分类测试样本，再拿分类结果与每个测试样本的类标签进行比较。假如分类模型的准确率达到了想要的阈值，就能用这个分类模型进行分类。利用机器学习技术，可以从大量的恶意代码和正常代码数据中学习到正确有效的信息，生成能分类恶意代码与正常代码的模型，进而通过分类模型检测代码是恶意代码还是正常代码。基于机器学习的恶意代码检测方法可以检测出未知的恶意代码，最大程度上防止恶意代码破坏计算机系统，因此机器学习的检测方法成为了现在恶意代码检测研究领域的重点和热点。

1.2.2解决方案

|  |  |  |
| --- | --- | --- |
| 题目 | 作者 | 解决方案 |
| 基于在线学习算法的恶意网页检测系统 | 王庆 | 提取网页URL特征，利用在线学习算法中的感知器算法、PA算法和CW算法，训练出高效的分类器，并用分类器实现了一个实时的恶意网页检测系统。 |
| 基于深度学习的恶意URL检测算法研究与设计 | 左雯 | 设计了基于关键字的用于恶意URL检测的神经网络模型。该模型利用从安全厂商处获取的数据集，结合词嵌入技术和URL中的恶意关键字，用卷积神经网络进行URL的特征抽取，并且使用GRU进行时间维度上的特征捕获。 |
| 基于多特征的恶意网页检测研究 | 岳涛 | 研究并分析了常用的网页特征提取方法，针对现有方法的不足并结合恶意网页检测的实际特点，提出了一种基于网页源代码和网址属性的恶意网页特征提取方法。在提出的特征提取方法的基础上，设计并实现了一个恶意网页检测系统，并且考虑到大规模网页分类的问题，运用了来优化近邻算法以减少时间开销。 |
| 代价敏感的多分类恶意网页识别系统研究与实现 | 陈博深 | 提出了全新的ＣＳＳ特征，并证明了这些特征在对恶意软件下载的恶意网页识别的有效性，推导出三分类误分类代价度量指标，设计实现了一个针对钓鱼网页和恶意软件下载网页的多分类恶意网页识别系统。 |
| 恶意网页智能检测技术研究与实现 | 庞玉敏 | 首先，使用基于特征融合的静态检测技术对网页进行检测，其次，若为恶意网页则利用 Capture-HPC 进行动态检测。静态检测技术和动态检测技术的综合运用使得该系统同时具有二者优点：高检测效率、高准确率。 |

**1.3研究内容和目标**

主要内容：

1.学习SVM算法，了解其原理，并掌握使用方法。

2.收集URL样本数据，并对数据进行分析以及预处理，并使用SVM算法建立高效的检测模型：1）对数据进行特征提取，作为SVM算法的输入；2）通过实验选取最优SVM核函数和参数并分析其对检测性能的影响；3）构建高性能的恶意网页检测模型；

3.设计一个基于SVM算法的恶意网页检测系统：系统可以分为输入模块、数据预处理模块、检测模块和结果展示模块；输入模块负责用户操作，进行数据的上传；数据预处理模块负责对原始数据进行预处理以及特征提取，包含对中间数据的存储等；检测模块包含SVM检测模型的构建，并使用该模型对输入数据进行分类；结果展示模块用于展示系统的检测结果。

4.编程实现基于SVM算法的恶意网页检测系统的程序设计，并进行调试；

5.分析系统检出率和误报率，总结其优缺点

目标：通过收集大量的URL样本，使用SVM算法对正常样本和恶意样本进行分类训练以及建模，并基于此实现一个高效的意网页检测系统。

**二、恶意网页**

**2.1Web安全概述**

**2.2恶意网页攻击技术的研究**

**2.3URL**

**三、支持向量机算法**

## 3.1 支持向量机原理

**定义3.1** 一个超平面，向量被超平面按下式分类：

 （3.1）

称之为–间隔分类超平面。

**定理3.1** 在统计学习理论中，设向量分布在一个半径为*R*的球中，则–间隔分类超平面集合的VC维满足：

 （3.2）

其中表示取整。

根据定理3.1，最小化VC维等价于最大化分类间隔，支持向量机正是通过这种方式实现了VC维的最小化，进而降低系统的期望风险。

支持向量机种类很多，包括线性支持向量机、广义线性支持向量机、非线性支持向量机、最小二乘支持向量机等[34-37]。本节以最基本的线性支持向量机和线性回归支持向量机为例介绍支持向量机的原理。

3.1.1 线性支持向量机

对于两类分类问题，取个样本，样本的的训练集是，，，训练集能被超平面H：无差错分开，并且离超平面最近向量与超平面距离最大，我们称超平面H就是最优超平面，如图2.1所示。

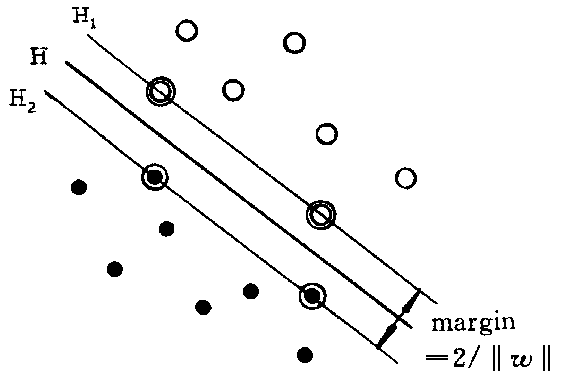


图3.1 支持向量机最优超平面示意图

定义两个标准超平面，H1：，H2：。H1和H2超平面平行于分类超平面H，并且这两个标准超平面是经过各类样本中离最优超平面最近的样本点。它们之间的距离为分类间隔。最大化分类间隔可以转化为最小化，前提是H1与H2之间无样本，即样本集中的所有样本点满足：

 （3.3）

因此，支持向量机就是采用下列凸二次规划问题对样本正确分类：



 （3.4）

这是一个凸二次规划问题，可以通过构造拉格朗日函数求解，即：

 （3.5）

式中为拉格朗日乘子（Lagrangian Multipliers）。

分别求的偏导数，并令偏导数等于零，有

 （3.6）

将式3.6代入式3.5，有



=

= （3.7）

上式称为拉格朗日函数对偶变换，该形式更容易数值求解，因此构建最优超平面问题转化为简单的对偶二次规划问题：



 （3.8）

式3.8有唯一解，设为式的最优解，则

 （3.9）

其中是最优分类超平面的权系数向量，从上式可以看出可以表示为训练样本的线性组合。

**定义3.2** 训练集中的输入为支持向量（Support Vector，SV），如果对应它的。

取值的样本中，当时，该样本位于标准超平面上，将这类样本称为支持向量，支持向量是最能提供分类间隔信息的样本，通常只占有样本集的一部分。由于数目不多，因此支持向量具有稀疏性，在决策函数构造时，只有这一部分样本对于决策有意义，这种性质对于降低模型复杂度有着重要意义[38-40]。

根据Karush–Kuhn–Tucker（KKT）条件，该问题必须满足

 （3.10）

将不为零的代入式可求解。为了提高决策函数的准确性，这里的可以求得多个解之后取平均值。

求解上述问题即可得到最优分类函数为

 （3.11）

式中为符号函数。

3.1.2 线性回归支持向量机

在机器学习问题中，回归问题是一类重要问题，可以对系统进行辨识建模，预测系统的输出。因此建立回归型支持向量机（Support Vector Regression，SVR）很有必要。本节简要讨论线性SVR的构建，相关详细证明见参考文献[40-43]。

对于线性SVR，采取线性回归函数估计样本集，在损失函数的选取上，这里引出线性不敏感损失函数，这里的可以理解为对损失函数的贡献度，可以表示为。

当时，，对损失无贡献。当或时，对损失有贡献，假设所有训练在精度下无误差的用线性函数拟合

 （3.12）

其中为样本点的输入，为在下样本的真实输出，为在输入下的预测输出值。

将回归函数代入式3.12有

 （3.13）

优化目标与线性支持向量机相同为



在更一般情况下，考虑到在线性不可分和发生噪声等情况下上述约束条件无法实现，这里引入松弛变量和，则约束条件变为

 （3.14）

优化目标变为



式中C（C>0）代表规则化参数。

上述问题是一个凸二次规划问题，可以构造其拉格朗日函数







 （3.15）

通过仿照线性支持向量机的求解方法，首先构建对偶二次规划，然后考虑KKT条件，最后得到的线性SVR回归函数为

 （3.16）

综上所述，采取上述方法就可以在经验风险误差较小的情况下，同时取最大分类间隔，这样可以使得最终的真实风险值最小[44-46]。支持向量机方法很好的解决了小样本建模问题，并且决策函数的推广性很好，具有很好的应用前景。这里需要注意的是，本章只简要介绍了线性可分支持向量机和线性回归支持向量机，其他支持向量机构造原理相似，只是引入了其他的构造方法。

**四、设计与实现**

**五、测试与分析**

**六、结论**

**七、参考文献**

1. 郑黎明, 邹鹏, 贾焰, 等. 网络流量异常检测中分类器的提取与训练方法研究[J]. 计算机学报, 2012, 35(4): 719-729.
2. RENUKA D S, YOGESH P. A Hybrid approach to counter application layer DDoS attacks[J]. International Journal on Cryptography and Information Security, 2012, 2(2): 45-52.
3. BOLZONI D. ATLANTIDES: An architecture for alert verification in network intrusion detection system[C]. The 21st Large Installation System Administration Conference. Berkeley: USENIX. 2007: 141-152.
4. 穆祥昆, 王劲松, 薛羽丰, 等. 基于活跃熵的网络异常流量检测方法[J]. 通信学报, 2013, 34(Z2): 51-57.
5. 田志宏, 王佰玲, 张伟哲, 等. 基于上下文验证的网络入侵检测模型[J]. 计算机研究与发展, 2013, 50(3): 498-508.
6. LEUNG K, LECKIE C. Unsupervised anomaly detection in network intrusion detection using clusters[C]. in Proc. of Australasian Computer Science Conference, Newcastle, NSW, Australia, 2005, 333-342
7. RUBINSTEIN B, NELSON B, HUANG L, JOSEPH A, et al. Stealthy poisoning attacks on PCA-based anomaly detectors[A]. Proceeding of the ACM SIGMETRICS[C]. New York, USA, 2009.
8. 钱叶魁, 陈鸣, 叶立新. 基于多尺度主成分分析的全网络异常检测方法[J]. 软件学报, 2012. 23(2): 361-377.
9. BRAUCKHOFF D, SALAMATIAN K, MAY M. Applying PCA for traffic anomaly detection: problems and solutions[C]. The INFOCOM. New York: IEEE Press, 2009, 46-53.
10. HARVEY N J A, NELSON J, ONAK K. Sketching and streaming entropy via approximation theory[C]. The 49th Annual IEEE Symptom on Foundation of Computer Science. Piscataway, New York: IEEE, 2008: 489-498.
11. NYCHIS G, SEKAR V, ANDERSEN D G, et al. An empirical evaluation of entropy-based traffic anomaly detection[C]. The 8th ACM SIGCOMM Conference on Int Measurement, New York: ACM, 2008: 151-156.
12. 郑黎明, 邹鹏, 韩伟红, 等. 基于多维熵值分类的骨干网流量异常检测研究[J]. 计算机研究与发展, 2012. 49(9): 1972-1981.
13. 李锦玲, 汪斌强. 基于最大频繁序列模式挖掘的App-DDoS攻击的异常检测[J]. 电子与信息学报, 2013, 35(7): 1739-1745.
14. 高赟, 周薇, 韩翼中, 等. 一种基于文法压缩的日志异常检测算法[J]. 计算机学报, 2014, 37(1): 73-86.
15. 王秀利, 王永吉. 基于命令紧密度的用户伪装入侵检测方法[J]. 电子学报, 2014, 42(6): 1225-1229.
16. 周颖杰. 基于行为分析的通信网络流量异常检测与关联分析[D]. 成都: 电子科技大学, 9-14.
17. W. Eberle, L. Holder. Anomaly detection in data represented as graphs[J]. Intelligent Data Analysis, 2007, 11(6): 663-689.
18. M. Iliofotou, P. Pappu, M. Faloutsos, et al. Network monitoring using traffic dispersion graphs(TDGs)[C]. in Proc. of ACM SIGCOMM Internet Measurement Workshop (IMC), Kyoto, Japan, 2007, 315-320.
19. D. Q. Le, T. Jeong, H. E. Roman, et al. Traffic dispersion graph based anomaly detection[C]. In Proc. of International Symposium on Information and Communication Technology, Hanoi, Vietnam, 2011, 36-41.
20. Y. Jin, E. Sharafuddin, Z. Zhang. Unveiling core network-wide communication patterns through application traffic activity graph decomposition[C]. in Proc. of ACM SIGMETRICS, Seattle, WA, USA, 2009, 49-60.
21. 王风宇, 曹首峰, 肖军, 等. 一种基于Web群体外联行为的应用层DDoS检测方法[J]. 软件学报, 2013, 24(6): 1263-1273.
22. Jiang J, Papavassiliou S. Detecting Network Attack in the Internet via Statistical Network Traffic Normality Prediction [J]. Journal of Network and Systems Management, 2004,12(1):51~72.
23. 孙轶东,张德远,高鹏. 基于时间序列分析的分布式拒绝服务攻击检测. 计算机学报, 2005,28(5):767~773.
24. 唐成华, 刘鹏程, 汤申生, 等. 基于特征选择的模糊聚类异常入侵行为检测[J]. 计算机研究与发展, 2015, 52(3):718-728.
25. 张玲, 白中英, 罗守山, 等. 基于粗糙集和人工免疫的流量入侵检测模型[J]. 通信学报, 2013, 34(9): 167-176.
26. CHANDRASEKAR A, VASUDEVAN V, YOGESH P. Evolutionary approach for network anomaly detection using effective classification[J]. IJCSNS Int Journal of Computer Science and Network Security, 2009, 9(1): 296-302.
27. SHINGO M, CHEN C, LU N N. Intrusion-detection model based on fuzzy class-association-rule mining using genetic programming network[J]. IEEE Transaction on Systems, Man, and Cybernetics, 2011, 41(1): 130-139.
28. 杨雅辉, 黄海珍, 沈晴霓, 等. 基于增量式GHSOM神经网络模型的入侵检测研究[J]. 计算机学报, 2014, 37(5): 1216-1224.RUBINSTEIN B I, NELSON B, HUANG L, JOSEPH A D, et al. Stealthy poisoning attacks on PCA-based anomaly detectors[C]. The ACM SIGMETRICS. New York: ACM Press, 2009.
29. NOBLE C C, COOK D J. Graph-based anomaly detection[C]. in Proc. of ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Diego, CA, USA, 2003, 631-636.
30. 陆悠, 李伟, 罗军舟, 等. 一种基于选择性协同学习的网络用户异常行为检测方法[J]. 计算机学报, 2014, 37(1): 28-40.
31. DASH S K, REDDY K S. Adaptive naive Bayes method for masquerade detection[J]. Security and Communications Networks, 2011, 4(4): 410-417.
32. 刘大有, 陈慧灵, 齐红, 等. 时空数据挖掘研究进展[J]. 计算机研究与发展, 2013, 50(2): 225-239.
33. RINGBERG H, SOULE A, REXFORD J, et al. Sensitivity of PCA for traffic anomaly detection[C]. The ACM SIGMETRICS.New York: ACM Press, 2007, 78-89.
34. 白鹏，张喜斌，等．支持向量机理论及工程应用实例[M]．西安：西安电子科学技术大学出版社，2008．
35. 邓乃扬，田英杰．数据挖掘中的新方法——支持向量机[M]．北京：科学出版社，2004．
36. 范昕炜．支持向量机算法的研究及其应用[D]．杭州：浙江大学，2003．
37. Vapnik V N著，许建华，张学工译．Statistical Learning Theory[M]．北京：电子工业出版社，2004．
38. W. Eberle, J. Graves, L. Holder. Insider threat detection using a graph-based approach[J]. Journal of Applied Security Research, 2010, 6(1): 32-81.
39. YASAMI Y, MOZAFFARI S P. A novel unsupervised classification approach for network anomaly detection by k-Means clustering and ID3 decision tree learning methods[J]. Journal of Super computing, 2010, 53(1): 231-245.
40. 边肇祺，张学工．模式识别（第二版）[M]．北京：清华大学出版社，1999．
41. Suykens J A K，De Brabanter J，Lucas L，et al．Weighted least squares support vector machines：robustness and sparse approximation[J]．Neurocomputing，2002，48（1）：85-105．
42. 许建华，张学工，李衍达．支持向量机的新发展[J]．控制与决策，2004，19（5）：481-484．
43. 翟永杰，尚雪莲，韩璞等．SVR在传感器故障诊断中的仿真研究[J]．系统仿真学报，2004，16（6）：1257-1279．
44. 张浩然，韩正之，李昌刚．基于支持向量机的非线性系统辨识[J]．系统仿真学报，2003，15（1）：119-121．
45. 李涛. 数据挖掘的应用与实践[M]. 厦门: 厦门大学出版社, 2013, 65-68.
46. 武小年, 彭小金, 杨宇洋, 等. 入侵检测中基于SVM的两级特征选择方法[J]. 通信学报, 2015, 36(4):2015127.