doi :10.3969/j.issn.1671-1122.2015.08.004

基于平衡二叉决策树 SVM 算法的物联网安全研究

张晓惠[1]，林柏钢[2]

（1. 福州大学至诚学院计算机工程系，福建福州 350002；2. 网络系统信息安全福建省高校重点实验室，福建福州 350108）

摘　要：物联网是继计算机、互联网和移动通信之后的又一次信息产业革命。目前，物联网已经被正式列为国家重点发展的战略性新兴产业之一，其应用范围几乎覆盖了各行各业。物联网中存在的网络入侵等安全问题日趋突出，在大数据背景下，文章提出一种适用于物联网环境的入侵检测模型。该模型把物联网中的入侵检测分为数据预处理、特征提取和数据分类 3 部分。数据预处理主要解决数据的归一化和冗余数据等问题；特征提取的主要目标是降维，以减少数据分类的时间；数据分类中引入平衡二叉决策树支持向量机（SVM）多分类算法，选用 BDT-SVM 算法对网络入侵数据进行训练和检测。实验表明，选用 BDT-SVM 多分类算法可以提高入侵检测系统的精度；通过特征提取，在保证精度的前提下，减少了检测时间。

关键词：入侵检测；平衡二叉决策树；支持向量机；物联网安全

中图分类号：TP309　　**文献标识码：**A　　**文章编号：**1671-1122（2015）08-0020-06

中文引用格式：张晓惠，林柏钢 . 基于平衡二叉决策树 SVM 算法的物联网安全研究 [J]. 信息网络安全，2015，（8）:20-25.

英文引用格式：ZHANG X H, LIN B G. Research on Internet of Things Security Based on Support Vector Machines with Balanced Binary Decision Tree[J]. Netinfo Security, 2015, (8):20-25.

Research on Internet of Things Security Based on Support Vector Machines with Balanced Binary Decision Tree

ZHANG Xiao-hui, LIN Bo-gang

(1. *Department of Computer Engineering, Zhicheng College of Fuzhou University, Fuzhou Fujian* 350002, *China*;
2. *Key Lab of Information Security of Network System in Fujian Province, Fuzhou Fujian* 350108, *China*)

Abstract: The Internet of Things (IoT) is another information industry revolution after the computer, the Internet and the mobile communications. At present, IoT has been officially listed as one of the national strategic emerging industries, and its application range covers almost all areas. Secure problems such as network intrusion in the IoT art prominent increasingly. In the big data context, this paper proposes an intrusion detection model that is suitable for IoT which divides the intrusion detection procedure into three parts, which are data preprocessing, features extraction and data classification. Data normalization and data redundancy reduction are solved in the data preprocessing. The main goal of features extraction is to reduce the dimension and thus to reduce the time of data classification. Support vector machine with balanced binary decision tree algorithm that is named BDT-SVM is introduced in the data classification for training and testing the network intrusion data. Experimental results show that it can improve the accuracy of intrusion detection system by using the BDT-SVM algorithm and reduce the detection time with features extraction in the premise of ensuring accuracy.

Key words: intrusion detection; balanced binary decision tree; support vector machines; IoT security

收稿日期：2015-07-03

基金项目：国家自然科学基金 [61075022]；福建省教育厅科技项目 [2014JB14224]

作者简介：张晓惠（1984-），女，福建，讲师，硕士，主要研究方向：信息安全、智能算法；林柏钢（1953-），男，福建，博士生导师，教授，主要研究方向：网络与信息安全、编码与密码。

通讯作者：张晓惠 hanyuzxh@126.com

0 引言

物联网的概念自1999年由麻省理工学院（MIT）首次提出后，在世界范围内得到了快速发展。它引领了信息产业革命的第三次浪潮，成为未来社会发展的基础设施。但物联网的研究与实际应用目前仍处于初级阶段，很多理论和技术有待于进一步突破，其中较为突出的是物联网中核心网络的信息传输与信息安全问题。物联网中节点数量庞大，会导致在数据传播时，因大量机器的数据发送而使得网络拥塞，产生拒绝服务攻击和网络入侵等安全问题[1]。物联网是全球商品互联的网络，一旦出现商业信息泄露，将会造成巨大的经济损失，危及国家经济安全。

入侵检测系统（intrusion detection system, IDS）的作用是：根据开发者事先设定的安全策略，系统采取相应的响应行为，以阻止外来入侵。传统IDS存在对未知网络攻击检测能力差、误报率高等缺点。为了提高入侵检测的效率，降低漏报率和误报率，把机器学习的方法引入到IDS中已经成为IDS的重要发展方向[2]。物联网数据融合概念是针对多传感器系统而提出的。在多传感器系统中，信息表现形式的多样性、数据量的巨大性、数据关系的复杂性以及数据处理要求的实时性、准确性和可靠性都已大大超出人脑的信息综合处理能力，在这种情况下，多传感器数据融合技术应运而生。多传感器数据融合由美国国防部于20世纪70年代最先提出，之后在英、法、日等国做了大量研究。近40年来数据融合技术得到了巨大的发展。

英国的Leusse提出一种面向服务的物联网安全架构[3]，它利用SOA Security Autonomics、Usage&Access Management等模块组建了一个具有自组织能力的安全物联网模型；瑞士Zurich大学的Mattern F等人指出从传统互联网过渡到物联网可能会面临许多安全问题[4]。本文借鉴现有分布式入侵检测系统的结构特征，提出了一种面向物联网环境的基于平衡二叉决策树SVM算法的物联网入侵检测模型，把入侵检测分为数据融合和入侵分类两个阶段，其中数据融合又细分为系统数据预处理和特征提取等过程。实验表明，该模型相比于SVM二分类算法能够取得更好的检测精度，降低了误报率和漏报率，对于一些未知的入侵检测类型也能做出准确的判断。

1 平衡二叉决策树SVM算法

1.1 SVM算法简介

支持向量机（support vector machine，SVM）[5]算法是一种机器学习方法，它是在统计学理论的基础上发展而来的，其原理基于结构风险最小化和VC维理论，具有良好的推广性和较好的分类精确性。SVM有二分类和多分类两种分类模型，其中多分类是在二分类的基础上发展而来的。

1.2 SVM二分类算法

设训练样本集 $S=((x_1,y_1),(x_2,y_2),\ldots,(x_l,y_l))\subseteq(X,Y)^l$，其中 $X=R^n$，$Y\in\{+1,-1\}$。n 维空间中线性判别函数的一般形式为 $g(x)=w\cdot x+b$，分类面方程为

$$w\cdot x+b \quad \cdots\cdots(1)$$

当样本线性可分时，可以假设

$$\begin{cases} x_i\cdot w+b\geq+1 & \text{if} \quad y_i=+1 \\ x_i\cdot w+b\leq-1 & \text{if} \quad y_i=-1 \end{cases}$$

我们将判别函数进行归一化，合并得

$$y_i(x_i\cdot w+b)-1\geq 0\ (i=1,2,\ldots,l)\cdots\cdots(2)$$

公式（2）使得离分类面最近的样本的 $|g(x)|=1$，这样分类间隔就等于 $2/||w||$。因此要使两类数据的间隔最大，等价于使得 $||w||$ 最小。使得公式（2）等号成立的那些样本叫做支持向量。

SVM二分类算法发展到现在已经比较成熟，也出现了很多经典的算法，如HUSH的GDA算法[6]和OSUNA的分解算法[7]等。

1.3 SVM多分类算法

实际应用中，如人脸识别、入侵检测等，都是把数据分为多个类别，因此SVM多分类算法在近几年成为研究的热点。目前SVM多分类算法有基本算法、一对一算法、层分类算法等，其中比较常用的是基本算法和一对一算法。

1）基本算法。对于 k 类数据集，构造 k 个二分类的判决函数，第 i 个判决函数的作用就是把第 i 类和其他类分开。当一个样本数据输入时，分别使用这 k 个判决函数做出决策。如果只有第 i 个判决函数输出该样本属于第 i 类，其他 $k-1$ 个判决函数输出该样本都属于其他类，那么这个样本数据就属于第 i 类；否则不进行任何操作。这种算法容易产生一个样本属于多类别或一个样本不属于任何一个类

别的情况。基本算法常见的有 one-against-one 方法（OAO-SVM）、one-against-all 方法（OAA-SVM）、有向无环图方法和平衡二叉决策树方法等 [8]。

2）一对一算法。这种方式就是在 k 类数据集中，构造所有可能的二分类判决函数。通过组合可以得知 k 类数据需要构造 $k(k-1)/2$ 个判决函数。换句话说，需要构造 $k(k-1)$ 个分类器。当一个样本数据输入时，用所有的分类器进行分类，将所判决次数最多的类作为样本的最终结果。一对一算法由于目标函数过于复杂，所求的变量过多，导致求解困难，因此很少采用。

1.4 平衡二叉决策树SVM算法

平衡二叉决策树 SVM 算法借助二叉树的思想，并在此基础上结合 SVM 二分类算法。即在训练样本集时，首先建立一棵有效的平衡二叉决策树，再通过 SVM 二分类算法训练二叉树的所有节点。文献 [8] 以 7 类样本为例，其平衡二叉决策树如图 1 所示，叶子节点为样本的类别。构造二叉树的过程是自顶向下从根节点到叶子节点逐层进行。从图 1 可知，该二叉树需构造出 6 个判决函数（即 6 个分类器）。

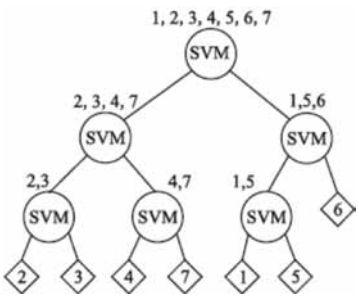

图1 平衡二叉决策树结构

由上述对 7 类样本的分析，可以把这个问题推广到一般状态。对于 k 类问题，该算法只需要构造出 $k-1$ 个决策面。文献 [9,10] 中的实验表明，这种方法保证了较高的识别准确度。算法经过 $\log_2^k$ 个节点最终到达叶子节点，以此可以判断数据样本的类别。这个过程是自顶向下进行的。

1.5 BDT-SVM算法

BDT-SVM 是一种平衡二叉决策树 SVM 算法 [11]。它的基本思想是：首先，从训练样本数据集中确定类间距离最大的两个类；然后，依次判断剩下的类别，采用最近原则分别向这两个类靠拢；最后，形成两个类别数相等的类簇。具体过程如下：

1）在多维空间下，分别计算 k 个类的类中心。

2）根据最大欧式距离，找出距离最大的两个类，将这两个类分别标记为类簇 c1 和类簇 c2。

3）在除去 c1 和 c2 的剩余类中，选出与 c1 的欧式距离最小的类 c3，将 c3 和 c1 合并，标记为类簇 c1，并重新计算该类簇的中心。

4）在除去 c1 和 c2 的剩余类中，选出与 c2 的欧式距离最小的类 c4，将 c4 和 c2 合并，标记为类簇 c2，并重新计算该类簇的中心。

5）循环执行步骤 3）和步骤 4），直到所有类分配完毕。

6）此时生成两大类簇 c1 和 c2，在 c1 和 c2 里，分别递归执行步骤 2）~5），直到所有的类完全分开。

通过上述 6 个步骤，就产生了一棵平衡二叉决策树，将树的每个节点用 SVM 二分类算法进行训练，由此实现了 BDT-SVM 多分类算法。

2 基于平衡二叉决策树 SVM 算法的入侵检测模型

入侵检测系统必须从各个子网和主机收集网络入侵信息，根据各种决策，做出相应响应。由于当前网络结构复杂，收集到的数据量比较大，当数据没有经过预处理时，可能存在严重冗余。本文通过研究各种复杂的网络拓扑结果，提出基于平衡二叉决策树 SVM 算法的入侵检测模型的总体框架，如图 2 所示。该模型主要包括数据预处理、特征提取和入侵分类 3 大步骤。

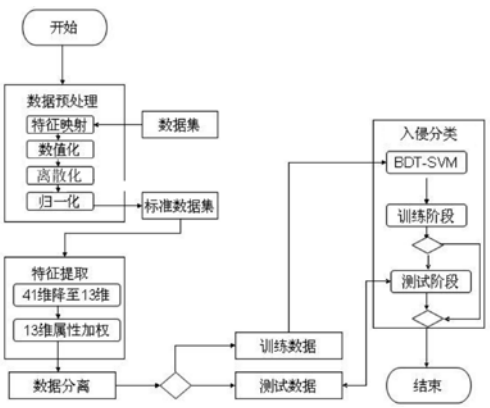

图2 基于平衡二叉决策树SVM算法的入侵检测模型

2.1 数据预处理

从网络连接事件提取模块得到的每个连接事件都是由 41 个属性（本文采用 KDD99 提供的一个标准数据集，每个数据具有 41 个属性）构成的。其中有连续性属性，也有不连续性属性；有些属性是字符串型的，也有的是整型和

浮点型的。各个属性的取值范围大多不一样。鉴于上述原因，需要对原始数据进行预处理，把数据数值化、离散化和归一化，为后续的特征选择和 SVM 预测做好准备。

1）数值化

数据的数值化很好处理，手工统计属性的个数，然后把每个属性对应为一个数值。

2）离散化

本文采用 Naïve Scaler 算法对原始数据集进行离散化。对每一属性 $a\in C$，进行如下过程：

（1）根据 $a(x)$ 的值，从小到大排列各个记录 $x\in U$。

（2）设 x_i 和 x_j 为两个相邻的记录，从上到下扫描，若 $a(x_i)=a(x_j)$，则继续扫描。否则，若 $d(x_i)=d(x_j)$，也就是说决策值相同，继续扫描；否则，得到一个断点 c，$c=(a(x_i)+a(x_j))/2$。也就是说在属性值和决策值都不相同的情况下得到一个断点 c。

3）归一化

数据的离散化已经处理完毕。然而，由于数据的值域不同，导致在进行 SVM 预测时，各个属性对结果的影响也不一样。为了均衡各个属性对结果的影响，需要对数据进行归一化。

本文采用 Randall Wilson D[12] 等人在异构数据集中定义的奇异距离函数对数据进行归一化。设异构数据集 上有两条记录 x 和 y，x_i 和 y_i 分别是 x 和 y 的第 i 个属性。于是，x_i 和 y_i 在第 i 个属性上的距离函数定义为

$$d(x_i,y_i)=\sqrt{\sum_{c=1}^{C}\left|\frac{N_{i,x,c}}{N_{i,x}}-\frac{N_{i,y,c}}{N_{i,y}}\right|^2} \quad \cdots\cdots(3)$$

其中，C 为数据的类别总数；$N_{i,x}$ 为所有样本数据中第 i 个属性取值为 x_i 的样本个数；$N_{i,x,c}$ 为所有样本数据中第 i 个属性取值为 x_i 且输出类别为 c 的样本个数。若 x_i 和 y_i 中有一个取值未知，则它们之间的距离定义为 1。

通过上述距离函数可以对数据进行归一化，结果为

$$x_i=\frac{d(x_i,x_{i.\min})}{d(x_{i.\max},x_{i.\min})}\times scale+low \quad \cdots\cdots(4)$$

其中，$x_{i.\max}$ 为数据集中第 i 个属性的最大值，$x_{i.\min}$ 为数据集中第 i 个属性的最小值，$scale$ 为 1，low 为 0。

2.2 特征提取

文献 [13] 利用 KDD99 的异常检测标准数据进行实验，分别使用粗糙集（RS）、SVDF、LGP 和 MARS 算法对数据集中的 41 维数据进行特征提取，每个算法选择 6 个属性。实验证明，在兼顾时间和准确度的情况下，把这 4 个算法提取出来的属性组成一个集合，利用这个集合中的属性判断入侵能取得较好的效果。图 3 给出了4 种特征提取算法提取出来的属性构成的集合（共有属性 13 个）。在保证时间效率的前提下，也保证了检测的精度。

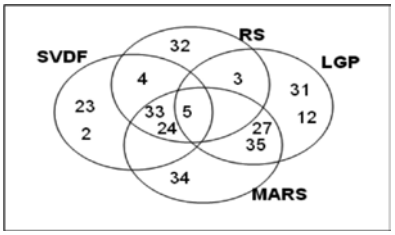

图3 4种算法特征提取结果

1）粗糙集理论 [14]（rough set theory，RST）由 Pawlak 于 1982 年提出，它是一种处理模糊和不确定性知识的数学理论和工具。粗糙集理论提出后得到了广泛的发展和运用。属性约简是粗糙集理论的核心问题之一，把它用于特征选择是理所当然的事情。

2）基于支持向量决策函数（support vector decision function，SVDF）的特征选择算法是将支持向量机（SVM）理论用于特征选择。SVM 算法诞生于数据分类，主要用于二分类。支持向量机也可以用于特征选择。

3）线性遗传规划（linear genetic programming，LGP）[15] 是遗传规划在线性基因组方面的应用。把线性遗传规划应用于特征选择基于它是处理机器代码的操作和它是基于评价的规划。LGP 算法从一组随机生成的初始可行解出发，通过复制、交叉和变异等操作，逐步迭代，逼近问题的最优解。LGP 算法中每一个被选择的特征子集都由一个适应度函数来衡量。适应度函数用于在约简空间和训练子集中衡量特征子集对入侵检测的效果。当迭代次数达到了算法要求，LGP 算法就停止。

4）多元自适应回归样条（multivariate adaptive regression splines，MARS）[16] 由 Friedman 于 1991 年提出，主要目的是从大量的独立变量 $x_1,x_2,\ldots,x_n$ 中预测一个连续的输出变量 $\hat{f}$，是一种用于解决多元数据问题的方法。MARS 将分段解释方程，并汇总组合出一个具有弹性的预测模型，然后自动建立准则模型，并利用这个准则模型来推测其连

续和间断的因变量。从多元自适应回归样条的名称来看，就是指多元逐步的回归程序。它最适合应用于高维问题中，也被视为广义的线性逐步回归，或者是通过修正分类与回归树模型方法而改善得到的多元自适应回归。目前 MARS 已被运用到很多不同的学术领域，如分类判别和特征选择等。

2.3 入侵分类

将从特征提取中降维得到的数据作为 BDT-SVM 的输入，进行攻击类型的识别。关于 BDT-SVM 算法的训练和测试流程在 1.5 节已做详细介绍。在此基础上，考虑不同特征信息对分类结果的影响，对样本数据的信息特征进行加权，并融合各个特征的属性值。

3 仿真与实验

3.1 数据集的选取

本文采用 KDD99 数据集进行仿真实验。该数据集中的异常数据总的可以分为四大攻击类别，分别为扫描与探测、拒绝服务攻击、获取根权限攻击和远程攻击。这四大攻击类别又可以细分为 38 种不同的攻击类型。该数据集中大约有 500 万条训练数据和 30 万条测试数据。由于数据量大，可以反映本算法的时间效率。

1）扫描与探测攻击（Probe）是扫描一个网段内的计算机，获取计算机的系统信息和脆弱点。这类攻击有 Ipsweep、Nmap、Portsweep 等。

2）拒绝服务攻击（DoS）是攻击者自己或者利用一些僵尸机占用大量网络资源，导致合法用户无法获得请求。这类攻击有 Back、Land、Neptune、Pod 等。

3）获取根权限攻击（U2R）是对本地超级用户的非法访问，也就是使得普通用户通过非法手段获得根用户权限，如 Buffer_overflow、Load_module、Perl、rootkit 攻击等。

4）远程攻击（R2L）就是非法用户通过发送网络数据包获得合法账号，如 Ftp_write、Guess_passwd、Imap、Multihop 攻击等。

本文采用的是 BDT-SVM 分类算法，存储核矩阵需要一个随着样本大小二次增长的内存空间，因此训练数据集规模大小的确定显得尤为重要。文献 [17] 指出，在入侵检测系统中，利用 KDDCUP99 数据集，在置信度为 0.95 且总的错误率小于 0.002 的情况下，SVM 训练数据集的样本数量要大于 74894 个。文章分析了 3 种规模的数据集，分别是接近 10000、接近 100000 和接近 500000。从规模 10000 到规模 100000，SVM 训练时间增长了 43.78 倍，预测错误率下降 69.51%。但是从规模 100000 到 500000，SVM 训练时间增长了 17.05 倍，可预测错误率却没有明显的下降。

本文训练数据集是从 KDDCUP99 的 10% 的训练数据中随机抽取的，共 85355 条网络连接记录。其中涵盖了正常数据和异常数据，异常数据中包含了 DoS 攻击、U2R 攻击、R2L 攻击和 Probe 攻击。训练数据集中正常数据和异常数据大约各占 50%。

本文测试数据集是从 KDDCUP99 的 10% 的测试数据中随机抽取的，共 311029 条网络连接记录。其中包含了正常数据、DoS 攻击、U2R 攻击、R2L 攻击和 Probe 攻击。为了检测本文提出的异常检测代理应对未知攻击的检测能力，在测试数据集中不仅包含了上述训练数据集中的已有攻击类型，还包含了一些新的攻击类型。

这些新攻击类型所占各种攻击类别的比例如表 1 所示。

表1 新攻击类型所占各种攻击类别的比例

攻击类别	新攻击类型的样本数	总样本数	新攻击占百分比（%）
DoS	6147	229853	2.68
U2R	10196	16189	62.98
R2L	189	228	82.88
Probe	1789	4166	42.94

表 2 给出了训练数据集和测试数据集的数据组成。

表2 训练数据集和测试数据集的数据组成

数据类别		训练数据集	测试数据集
正常样本个数		44165	60593
异常样本个数	DoS 攻击样本个数	19630	229853
	U2R 攻击样本个数	2600	16189
	R2L 攻击样本个数	5630	228
	Probe 攻击样本个数	13330	4166

3.2 SVM多分类的结果

评价系统指标（检测精度、误报率、漏报率、训练时间和检测时间）的定义如下：

1）检测精度 = 正确分类的样本数 / 总样本数；

2）误报率 = 错误分类的样本总数 / 总样本数；

3）漏报率 = 异常样本被认为是正常样本的总数 / 攻击样本数；

4）训练时间 =SVM 训练得到支持向量的时间；

5）检测时间 =SVM 检测输入样本类别的时间。

本文把 SVM 二分类算法和多分类算法应用到入侵分类阶段的数据分析中。SVM 多分类算法（BDT-SVM）把数据分成正常、DoS 攻击、U2R 攻击、R2L 攻击和 Probe 攻击 5 类。通过表 3 可以得知多分类算法在训练时间和检测时间上比二分类算法长，这是由于算法本身的局限性，但却大大提高了系统的检测精度。在计算机硬件发展迅速的当前环境下，牺牲一些时间代价换取一定的检测精度是值得的。

表3 二分类算法和多分类算法的时间和检测精度对比

SVM 分类算法	训练时间（s）	检测时间（s）	检测精度（%）
二分类	4.6	2.5	86.72
BDT-SVM	8.3	3.9	92.62

本文提出的入侵检测系统的误报率较低，但漏报率相对而言比较高。各类数据的具体检测结果如表 4 所示。

表4 SVM多分类算法的检测精度

检测结果类型 / 数据包类型	Probe	U2R	DoS	R2L	正常
Probe	97.32%	0.04%	1.35%	0%	1.29%
U2R	0%	63.46%	0%	25.35%	11.19%
DoS	0.35%	0.18%	87.35%	1.95%	10.17%
R2L	0%	0.83%	0.36%	92.94%	5.87%
正常	0.68%	0.28%	0.16%	1.99%	96.89%

由表 4 可知，本文提出的基于平衡二叉决策树 SVM 算法的入侵检测模型的检测精度比较高，但系统的漏报率相对比较高，其中最高的是 U2R。

上述实验得到的结果是某些攻击的漏报率偏高，导致系统的检测精度也受到一定的限制。漏报率偏高对系统安全的影响是比较大的，甚至可能使系统遭受损失。而误报不会影响系统安全，降低甚至消除误报可以通过人机交互等方式。因此，系统急需解决漏报的问题。

4 结束语

本文针对物联网数据量大，网络安全隐患比较突出的问题，提出一个基于平衡二叉决策树 SVM 算法的入侵检测模型。该模型主要分为 3 个步骤：1）数据预处理。进行原始数据的归一化和消除冗余数据等操作；2）特征提取。利用特征提取，在保证检测精度的前提下，缩短训练时间和检测时间；3）入侵分类。把 SVM 多分类算法引入入侵检测中，提高检测精度，降低系统的误报率。实验表明，该算法模型具有实际的参考意义。

物联网有别于一般的信息网络，它有自身的特殊性，同时也面临着特殊的安全威胁，如碰撞攻击、信息篡改等。一些比较有效的抵制物联网攻击的方法有加密机制和密钥管理、数据融合安全等。下一步的工作目标是：1）在分析物联网传感层数据的基础上，利用数据融合、属性特征加权等技术，寻找预防物联网未知攻击的方法。2）寻找一种轻量级的身份鉴别机制，保证网络中用户身份的可信。例如，聚合签名引入了一种改进通信效率和计算效率的方法，以减少链路所占的带宽。轻量级的认知技术是物联网可信身份认证体系的核心，将成为以后研究的重点。（责编 马珂）

参考文献：

[1] 杨庚，许建，陈伟，等 . 物联网安全特征与关键技术 [J]. 南京邮电大学学报 ,2010,30(4):20-29.

[2]Tsai C F, Hsu Y F, Lin C Y, et al. Intrusion Detection by Machine Learning: A Review[J]. Expert Systems with Applications, 2009, 36(10): 11994-12000.

[3]De Leusse P, Periorellis P, Dimitrakos T, et al. Self managed security cell, a security model for the Internet of Things and Services[C]//2009 First Internatiional Conference on Future Internet. IEEE, 2009:47-52.

[4]Mattern F, Floerkemeier C. From the internet of computers to the internet of things[C]// From Active Data Management to Event-based Systems and More. Springer Berlin Heidelberg, 2010:242-259.

[5] 曹宏鑫 . 基于 SVM 的网络入侵检测研究 [D]. 南京：南京理工大学，2004.

[6]HUSH D, SCOVEL C. Polynomial-time decomposition algorithms for support vector machines[R]. LANL Technical Report LA-UR-00-3800, Los Alamos: Los Alamos National Laboratory, 2000.

[7]OSUNA E, FREUND R, GIROSI F. Training support vector machines: an application to face detection [C]//Proceedings of CVPR’97 Puerto Rico, 1997.

[8] 林志杰 , 余春艳 . 改进的基于平衡二叉决策树的 SVM 多分类算法 [J]. 小型微型计算机系统 , 2014,35(5):1128-1132.

[9]Gjorgji Madzarov, Dejan Gjorgjevikj, Iyzn Chorbev. A multiclass SVM classifier utilizing binary decision tree[J]. Informatica, 2009,33(2):233-241.

[10] 刁智华 , 赵春江 , 郭新宇 . 一种新的基于平衡决策树的 SVM 多类分类算法 [J]. 控制与决策 , 2011，26（1）：149-156.

[11] 张义荣 , 鲜明 , 肖顺平 , 等 . 一种基于粗糙集属性约简的支持向量异常入侵检测方法 [J]. 计算机科学 . 2006, 33(6): 64-68.

[12]Randall Wilson D, Tony R Martinez. Improved Heterogeneous Distance Functions[J].Journal of Artificial Intelligence Research,1997,6(1):1-34.

[13] 张晓惠 , 林柏钢 . 基于特征选择和多分类支持向量机的异常检测 [J]. 通信学报 ,2009,30（10A）:68-73.

[14] Pawlak Z. Rough Sets[J]. International Journal of Computer and Information Sciences, 1982, 11 (5):341-356.

[15] 王再见 . 模糊入侵检测规则的自动发现算法研究 [D]. 合肥：中国科学技术大学，2005.

[16] 张束 . 基于贝叶斯 MARS 的入侵检测算法研究 [D]. 哈尔滨：哈尔滨工程大学，2006.

[17] 边肇祺 , 张学工 . 模式识别 [M]. 第 2 版 . 北京：清华大学出版社 ,2000.