# 面向物联网系统的机器学习架构

## 摘要

物联网安全越来越受到学术界和工业界的关注。事实上，物联网设备容易出现各种安全攻击，从拒绝服务(DoS)到网络入侵和数据泄漏。本文提出了一种新的基于机器学习的安全框架，该框架自动应对与物联网领域相关的扩展安全方面。该框架利用软件定义网络(SDN)和网络功能虚拟化(NFV)使能器来缓解不同的威胁。该 AI 框架结合了监控代理和基于 AI 的反应代理，该代理使用 ML 模型分为网络模式分析，以及物联网系统中的基于异常的入侵检测。该框架利用监督学习、分布式数据挖掘系统和神经网络来实现其目标。实验结果证明了所提出方案的效率。特别是，使用数据挖掘方法的攻击分布在以高性能和低成本检测攻击方面非常成功。关于我们针对物联网的基于异常的入侵检测系统 (IDS)，我们使用单类 SVM 在真实智能建筑场景中评估了实验。异常的检测精度达到 99.71%。进行了一项可行性研究，以确定要采用的当前潜在解决方案，并促进对开放挑战的研究。

## 1.概述

物联网 (IoT) 的破坏性加速正在彻底修改当前的 ICT 景观，这些设备预计未来几年将部署大量蜂窝物联网设备。物联网设备正在接管我们当前生活的各个方面，如医疗保健、运输和家庭环境[1]。由于分析和云计算技术的巨大增长，它们有望能够在不进行人工交互的情况下使用彼此的自主通信来提供相关的上下文数据。所有这些设想的好处正在迅速推动这项技术的采用。在频谱的另一端，物联网节点可以由恶意攻击者利用资源约束和相关漏洞组成。考虑到物联网安全威胁的广泛应用，会造成严重的隐私问题和经济损害。由于它们正在成为我们日常生活中必不可少的元素，维护隐私、安全和业务操作/机会是非常高优先级的。例如，物联网设备可用于各种目的，可以部署在家庭、医疗保健和工业环境等不同的地方。因此，它们可以携带敏感的个人数据，例如用户信息和日常活动。对这些物联网设备的攻击可能导致敏感信息泄漏，并可能导致工作流中断，从而影响产品的质量。为了适应物联网系统的约束和异构性，软战网络似乎是最引人注目的解决方案。网络软战是最近一种很有前途的趋势，旨在通过在网络服务[2]中采用云计算技术和软件模型来从根本上推进电信行业。

这次革命背后的主要支柱是SDN和NFV。一方面，SDN通过解耦控制和数据平面，引入了一种新的网络可编程性水平。逻辑上集中的控制器负责监督网络状态，并为网络元素提供规则，以便适当地管理交通流。另一方面，NFV利用虚拟化技术将网络元素部署为软件实例，从而在服务供应方面允许更高水平的灵活性和弹性。此外，NFV 通过将专用昂贵的硬件替换为能够托管基于软件的网络设备的商用服务器，可以显着减少 CAPEX/OPEX 成本。虽然SDN和NFV是两个独立的范式，但它们的联合使用可以进一步提高网络提供的潜在安全服务，满足新的物联网应用带来的广泛日益增长的需求。预期物联网设备的爆炸式增长，基于位置的移动游戏应用的广泛扩散，触觉互联网应用都是苛刻场景的重要代表，暴露了广泛的新漏洞和安全问题。利用SDN和NFV集成提供的灵活性和可扩展性，telco运营商将成功能够加强物联网领域[3]的相关安全策略。在这个发明的背景下，一些工作已经研究了实现安全即服务 (SECaaS) [4]、[5] 的模型。

通过利用SDN和NFV特性，工业界和研究界正在推动在物联网网络领域实现类似模型的巨大努力。另一方面，快速增长的物联网攻击需要自适应框架，该框架可以使用不同的监控输入来处理未知类型的攻击。物联网系统中引入的新服务和特征暴露了新的和看不见的漏洞类型。在这种情况下，机器学习非常令人信服。最先进的AI算法利用机器学习来识别攻击，并通过根据攻击的威胁级别对攻击进行分类来适应和响应新的潜在网络安全风险。此外，当将深度学习原则纳入系统时，它们实际上可以随着时间的推移而适应，为网络管理员在网络犯罪分子[6]上提供边缘。与传统的基础设施不同，物联网中的入侵检测不仅要考虑网络系统指标，还要考虑从物理环境中处理和测量。

本文提供了一个完整的框架，该框架利用机器学习(ML)技术和5G使能技术SDN、NFV和物联网控制器，高效、快速地检测和防止网络安全攻击。本文的贡献有很多折叠：

•通过闭环自动化、自主和协调的方式监控、检测和防止网络安全威胁，与ETSI ZSM[7]视觉一致的统一AI安全框架;

•实现和验证物联网AI安全框架，该框架利用机器学习技术来处理，不仅通过网络模式/签名识别进行基于知识的入侵检测，而且基于设备正常行为的偏差进行基于异常的入侵检测，其报告的数据由框架的监控能力观察;

•提出了三种方法，利用ML技术检测基于网络模式的网络安全攻击;

•统一的AI安全框架具有识别物联网中新型网络攻击(0天攻击)的能力，无法通过网络模式识别检测;

•利用基于SDN/NFV的安全管理特征，根据框架推断的基于人工智能的上下文决策动态有效地缓解检测到的网络攻击;

此外，根据框架推断的基于人工智能的上下文决策，该框架基于SDN/NFV的安全管理特性允许对检测到的网络攻击进行动态和有效的缓解。

本文的其余部分安排如下。在第 II 节中，我们总结了文献中的相关工作。框架架构和相关技术在第三节中描述。第四节介绍了 AI 代理在这两种方法中的性能评估结果。最后，V 节包含这项工作并突出了开放的挑战。

## 2.相关工作

物联网安全是一个发明的研究领域，引起了研究界越来越多的关注。有很多作品涵盖了这一重要方面。例如，[8] 中的作者提出了智能基础设施的物联网安全框架，例如智能家居和智能建筑。它采用连续监控来捕获传感器的操作数据，以检测物联网域中的异常行为。该数据用于识别传感器并将其行为与“正常”行为进行比较。如果检测到攻击，它根据异常的类型对其进行分类，并采取相关的恢复动作，例如传感器重新认证、丢弃传感器数据或更改网络配置。尽管结果表明该系统在检测攻击方面能够提供高水平的准确性，但可能的缓解措施非常有限，经常导致服务中断。此外，该平台不提供 E2E（端到端）安全性，即必须因为攻击可以针对物联网框架的任何层。

通过定义基于SDN的安全框架，在[9]、[10]的工作中利用了SDN的灵活性。SDN技术提供的额外功能能够集成新的安全工具，如细粒度路由操作、流量过滤和安全网络通道使用来传输敏感数据。在NFV范围内，一些研究论文集中于使用容器[11]、[12](如入侵检测系统(IDS)和防火墙)评估在边缘上运行虚拟安全设备的性能和可行性。尽管这种较轻的虚拟化技术显示出巨大的效率，但考虑到资源受限的物联网设备来说，它被证明是具有挑战性的。事实上，大量的流量可以产生高能量和CPU消耗，从而影响设备的可用性。安全物联网系统的另一种方法是使用机器学习技术。[13]中提出了利用SDN技术和ML技术实现网络入侵检测系统的不同解决方案。这项工作还描述了与网络入侵检测系统实现相关的实现挑战。

[14] 中的作者提出了一种使用深度学习方法预测城市公交车位置的解决方案。在提出的解决方案中，基于长短期记忆 (LSTM) 的神经网络已被用于预测位置和数据速率。[15] 中的作者提出了一种解决方案，该解决方案利用块链来管理可扩展的物联网系统。[16] 中的作者提出了一种解决方案，以确保物联网设备和 MEC 之间的通信。所提出的解决方案采用学习方法来识别服务组合和交付的候选者。[17]的作者研究了人工神经网络的使用，以检测从网关到边缘设备[18]的异常网络流量。在他们的方法中，他们使用温度传感器作为边缘设备，树莓派作为物联网网关。该系统从边缘设备收集多个数据样本，并将它们存储在网关的数据库中。然后，他们将这些输入拆分为训练和测试数据。一旦使用训练数据训练神经网络，测试数据用于评估模型的准确性。虽然结果显示异常检测的安全性水平有所提高，但该系统的能力受到物联网网关上有限的资源阻碍，影响用户体验，设备的寿命在此过程中为负。[19]中提出了一种运行在联网车辆之上的入侵检测系统。所提出的框架采用深度信念网络和决策树机器学习机制来检测不同的攻击。

AI可以利用入侵检测系统(IDS)进行物联网，从而基于物联网设备报告的网络系统和物理测量指标来检测异常行为。Mehta等人[20]为物联网提供了一种基于人工智能的IDS方法，该方法利用一组给定的传感器数据时间序列之间的关系来检测异常。尽管如此，我们的 AI 框架旨在通过检查先前已知漏洞和攻击的连续签名和模式来应对不仅基于异常的 IDS [21]，还旨在应对基于知识的 IDS [22]。在这方面，迄今为止所做的大部分研究工作都集中在事件检测阶段。我们的框架旨在涵盖反应阶段，一旦确定了攻击。

我们强烈认为，由于SDN控制器的全局网络视觉，理想的解决方案将保证端到端安全，并使用AI进行适当的安全策略定义和细化。由于托管在云上的虚拟网络安全设备提供的高级功能，将强制执行这种相关的安全策略。因此，我们为物联网系统引入了我们新颖的基于人工智能的安全框架。

## 3.系统框架

### A.技术背景

#### （1）SDN架构

SDN是一种相对较新的范式，旨在将控制平面与数据平面解耦，以提高网络的灵活性和可编程性，允许外部应用程序以一种简单高效的方式控制网络的行为。SDN 提供了根据动态应用程序请求动态调整网络流的新功能。支持SDN的网络三个主要组成部分是：交换机、控制器和通信接口，其中SDN控制器是强制交换机中的认知决策的集中式实体，维护整个系统的状态，例如它通过更新交换机上的相关流规则来决定流量路由。

在物联网(启用SDN的物联网系统中)中采用SDN被认为是未来物联网系统成功和可行性的重要组成部分。利用SDN在路由流量和优化网络利用率方面的智能是管理物联网网络中大量数据流和消除瓶颈的关键使能函数[23]。这种集成可以在物联网网络的不同级别实现，例如访问（生成数据）、核心和云网络（处理和服务数据），这使得物联网流量从端到端管理成为可能。此外，SDN 还可用于为物联网系统提供先进的安全机制。例如，不同租户之间的流量隔离，利用网络的全局视觉和边缘的流量下降进行集中安全监控，使恶意流量通过网络传播。

#### （2）网络功能虚拟化NFV

网络功能虚拟化(NFV)是指在网络环境中采用虚拟化技术。与传统的网络设备不同，NFV将软件与硬件解耦，带来增值特征和显著的资本和运营支出收益。ETSI（欧洲电信标准研究所）已经领导了这种方法的标准化，定义了能够实现上述优势的新颖架构。

ETSI NFV 架构确定了三个主要构建块：

1)虚拟化基础设施:这一层包括为虚拟化网络功能(VNFs)部署提供所需资源抽象所需的所有硬件和虚拟化技术。这包括存储、计算和网络资源，这些资源通常由云平台管理。

2)虚拟网络功能:NFV的核心思想是用基于软件的网络功能实例(即VNFs)替换专用硬件设备。它们可以在多个环境中部署和管理，提供可扩展且具有成本效益的网络功能。

3)管理和编排:NFV管理和编排(MANO)块与ETSI NFV体系结构中的基础设施和VNF层交互。它负责管理全局资源分配，包括:实例化、配置和监控VNFs。

将虚拟化网络资源引入物联网生态系统中带来了多个增值特征，考虑了它们的异质性和快速增长。当与SDN相结合时，NFV不仅可以提供先进的虚拟监控工具，如入侵检测系统(IDSs)和深度数据包检查器(DPIs)，还可以提供和配置按需和可扩展的网络安全设备，如防火墙和认证系统，以应对监控代理[24]、[25]检测到的攻击。此外，将这些资源受限的物联网设备的安全引起的额外处理卸载到虚拟实例[26]节省了能量，提高了效率，使更多的空间留给其他有用的应用程序。上述NFV的灵活性和先进的安全特性在当前现成的物联网安全硬件中缺乏。尽管NFV不是为了完全取代当前的物联网解决方案，但其互补值添加的特征被证明是物联网安全格局中非常引人注目的和革命性的。

#### （3）机器学习技术

机器学习 (ML) 是人工智能领域，它集成了一组技术和算法，为计算机和智能设备提供智能。ML技术，如监督学习、无监督学习和强化学习，已广泛应用于网络安全领域。它采用它来准确检测和定义特定的安全策略来加强数据平面。挑战在于微调相关安全协议的不同参数，以减轻某种类型的攻击，或者通过标记网络流量或定义访问控制策略。事实上，不同的ML技术可以解决各种物联网攻击。例如，神经网络可用于检测恶意软件检测中的网络入侵[27]和DoS攻击和K-NN[28]。

1)监督学习:在监督算法中，数据的内部关系可能是未知的，但模型的输出是。通常，该模型的训练需要一组数据来“学习”，另一个用于测试和评估驱动模型。安全景观中的一个常见例子是将攻击模式与一组已知的攻击相匹配。

2）无监督学习：与监督学习方法不同，在无监督学习技术中，模型是未知的，这意味着数据不必被标记。相关类型的模型试图找到数据之间的相关性并将其分类为不同的组。

3)强化学习：强化学习侧重于研究试图改进其模型的问题和技术。它有一个独特的模型训练方法，它使用试错和奖励函数。它监控其输出的结果，并使用奖励计算称为“价值函数”的值。根据这个值，模型知道其决策的准确性并相应地适应自身。

### B.架构概述

为了解决与物联网系统相关的不同安全问题，我们提出了一个结合SDN、NFV和ML的安全框架，如图1-1所示。图1-1(a)显示了所提出的安全框架中的组件及其交互，图1-1(b)显示了本文提出的闭环自动化，从监测和检测到攻击缓解。该系统通过整合前面小节中讨论的对策和使能器来提供全面的安全性。该框架允许执行安全策略，从设计到它们的部署和维护。

如图1-1(a)所示，该框架由两个主要层组成:

i)安全编排平面;

ii)安全强制平面。

在下文中，我们将描述这两个平面，以及它们的内部和内部通信，以确保闭环自动化检测和减轻不同的威胁。

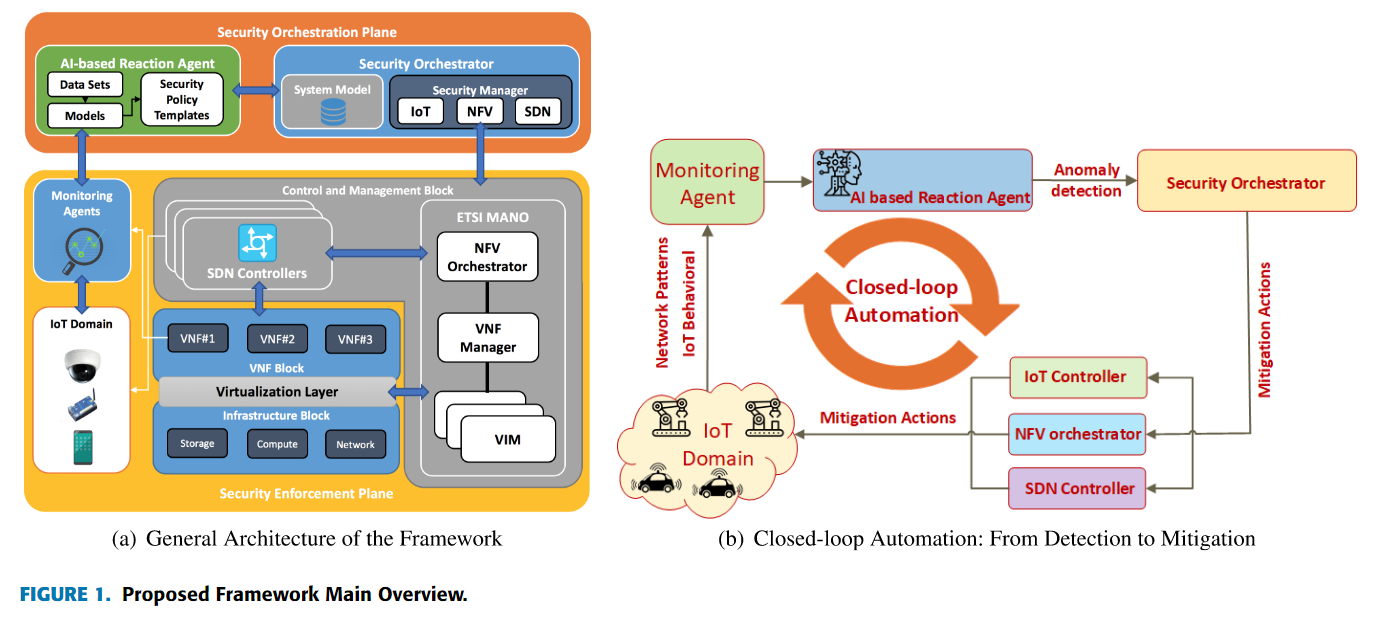


图1-1 集成框架

#### 安全加密方案

由于部署在不同云和边缘和物理网络功能(PNFs)上的不同VNFs，物联网设备和最终用户之间的通信就会发生。这些网络功能(即VNFs和PNF)之间的通信，物联网设备和最终用户通过遗留网络或基于SDN的网络发生。在物联网领域，我们区分了两种类型的攻击，它们是内部和外部攻击。虽然后者是在终端用户(即外部)网络向物联网域(即内部)网络启动的，但前者是由于恶意和入侵者物联网设备造成的。后者要么针对其他合法的物联网设备和/或外部网络生成攻击。主要是，攻击将在以下级别得到缓解:

i)利用物联网控制器的物联网设备;

ii)利用SDN控制器的网络级别;

iii)利用NFV协调器云/MEC级别。

该框架定义的安全属性应该通过部署安全VNF并通过SDN网络配置连接来适当地在物联网域中强制执行。安全执行平面被设计为完全符合SDN/NFV标准，分别由ETSI NFV和ONF(开放网络基金会)SDN规范指定。设想的安全执行对策将涉及三个逻辑块，如图1-1(a)所示。

A:VNF域

解释了部署在虚拟化基础设施上的VNFs，以使用不同的网络服务强制执行安全性。具体的注意力将得到解决，以提供高级安全VNFs(如虚拟防火墙、IDS/IPS等)，这些VNF应该能够提供安全策略请求的保护和威胁对策。

B:控制和管理域

考虑管理SDN和NFV环境所需的组件。为此，它包括 ETSI MANO 堆栈模块和 SDN 控制器。由于NFV通常与SDN相结合，根据资源和策略对网络进行编程调整，NFV协调器和SDN控制器之间的紧密交互，以实现部署适当的安全功能。

C：基础设施域

包括所有能够通过利用适当的虚拟化技术提供计算、存储和网络能力来构建基础设施即服务 (IaaS) 层的物理机器。该平面还包括负责流量转发的网络元素，遵循SDN控制器的规则，以及用于数据收集的分布式安全探针集，以支持监控服务。

D:监控代理

主要负责报告网络流量和物联网行为来检测不同的攻击。在所提出的框架中，检测机制既可以使用网络模式，也可以使用物联网不当行为。由于通过SDN完成的流量镜像，它们将知道通过网络流的所有流量。每个监控代理向托管在安全编排平面中的基于人工智能的反应代理发送包含相关可疑活动的日志。

E：物联网域

代表支持SDN的物理设备网络，从安全摄像头、温度传感器、家用电器到任何其他交换数据的智能设备不等。考虑到这些设备的高脆弱性，我们的框架旨在加强该领域的安全策略，以确保数据隐私和完整性。

#### （2）安全组织方案

该平面负责基于最新监测数据的安全策略的运行时配置及其上下文感知细化。它是我们的架构的创新层，负责通过对安全强制平面进行相关请求，将相关的安全策略强制到物联网领域。这包括实例化、配置和监控不同的虚拟安全使能器，以应对当前的攻击。

主要交互可以在图 1-2 中描述的图表中看到，该图总结了我们框架的组件之间的不同交互。如图1-1(b)和图1-2所示，本文从监测代理、基于人工智能的反应代理到安全协调器开始，提出了一种闭环自动化机制。后者分别通过物联网控制器、SDN控制器和NFV编排器来缓解威胁。

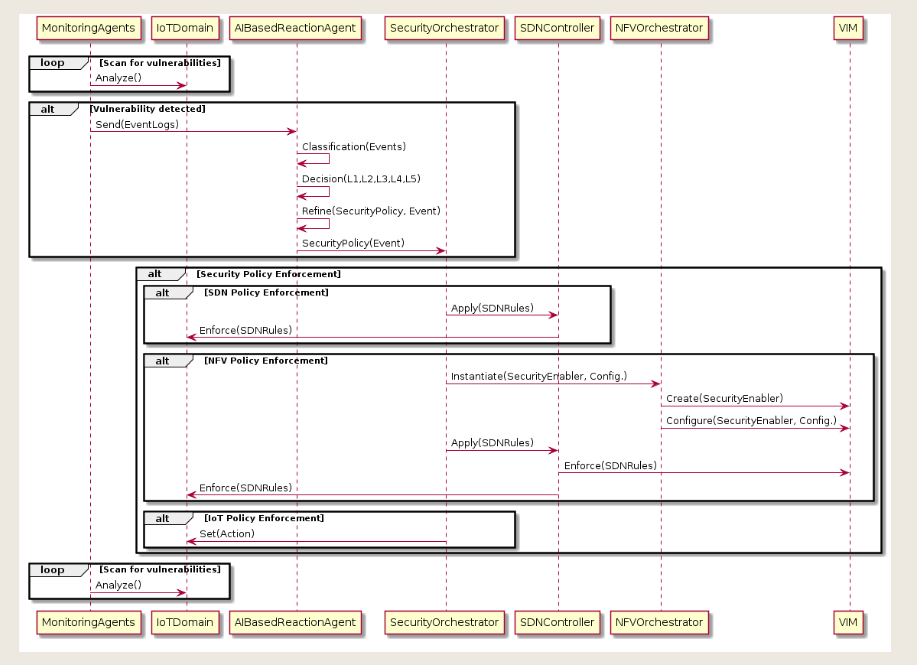


图1-2基于 AI 的物联网系统安全框架组件之间的交互概述

A:基于AI的回馈机制

该组件负责指示安全编排器采取的缓解措施。如图 1-1(b) 和图 1-2 中的第一个块所示，该组件使用由于监控代理，从网络和物联网域收集的数据。该组件使用基于网络模式和物联网行为训练的机器学习模型来检测威胁。这些机器学习模型将能够规定应该发送到安全协调器的适当安全策略模板。如图1-1(b)和图1-2中的第二个块所示，基于物联网行为和/或网络模式检测安全威胁。然后，将识别和发送到安全协调器的威胁级别（每个级别 - L1、L2、L3、L4、L5 对应于预定义的安全策略）。

如图1-1(b)所示，基于人工智能的反应代理使用不同的机器学习算法，包括J48、Byes Net、RandomForest、Hoeffding、支持向量机(SVM)和深度学习，用于检测不同的攻击相关的物联网行为和/或网络模式。有关该组件实现的更多信息将在第 IV 节中提供。

B:安全编排器

该组件是闭环自动化的一部分，可用于强制执行 AI 反应代理定义的安全策略。它与控制和管理块交互，以便在物联网域中使用SDN和NFV强制执行相关的安全策略。如图 1-2 中的第三个块所示，安全协调器要么通过实例化、配置然后监控虚拟安全设备或使用 SDN 操纵恶意流量，要么使用物联网设备本身采取直接操作，例如关闭受损设备。安全编排器还包含一个系统模型数据库，其中包含与数据平面和强制策略相关的所有信息，例如反应代理请求、SDN 控制器和交换机、当前运行 VNF 及其配置和物联网设备相关信息。

C：实施工具

在本小节中，我们对我们提出的解决方案的潜在实现进行了评估研究。为此，我们概述了用于启用建议框架的预期开源项目。

1）ONOS SDN控制器

ONOS（开放网络操作系统）是一个开源项目，旨在为通信和服务提供商创建一个 SDN 操作系统。众所周知，它的高性能、可扩展性和高可用性。它使用标准协议，例如 OpenFlow 和 NetConf，以便通过其应用程序公开高级流量操作功能。这些应用程序提供了高级抽象，同时提供有关网络的详细信息，例如现有节点、特定流量和现有链路的数据包数量，使得应用程序开发更加简单。

2）ETSI 开源管理与编排平台

OSM 是一个 NFV Orchestrator，于 2016 年在世界移动国会 (WMC) 上正式推出，由 Mirantis、Telefnica、BT、Canonical、Intel、RIFT.io、Telekom Austria Group 和 Telenor 成立。它符合 ETSI NFV MANO 参考架构，并为多云和 SDN 供应商支持（OpenStack、AWS、ONOS、Opendaylight.）提供支持。它由三个基本组件组成：

•服务编排器(SO):负责端到端服务编排和供应，它提供了一个web界面和一个目录，其中包含不同的NFV描述符。

•资源编排器(RO):用于在给定位置的特定IaaS提供商上提供服务。它直接与 VIM 交互以实例化虚拟资源

• VNF 配置和抽象 (VCA)：使用 Juju Charms LXD 容器执行初始 VNF 配置和恒定监控。

## 4.基于AI的反馈代理实现和性能评估

本节提供实验设置和评估基于 AI 的反应代理（详见第 III 小节）。基于 AI 的反应代理通过以下方式检测威胁：

i) 分析第 IVA 小节中介绍的网络模式。提出了一种基于知识的入侵检测框架来检测不同的网络攻击；

ii) 分析物联网系统中的异常行为，如第 IV-B 小节所述。在本小节中，基于物联网系统中异常行为的分析来检测网络攻击。

我们使用监督学习算法来准确分类攻击级别并正确选择合适的安全模板。使用监控代理的相关输入，基于人工智能的反应代理将利用多种机器学习技术来减轻给定的威胁。

### A：网络模式分析

入侵系统的评估是证明框架效率的首要步骤。有几个数据集广泛用于此目的，例如 DARPA [29]、KDD99 [30] 和 DEFCON [31]。我们基于包含 20 多个攻击的 NSL KDD 数据集构建 IDS，例如 Neptune-dos、pod-dos、smurfdos、缓冲区溢出、rootkit、satan、泪滴等。 NSL KDD 是原始数据集 Kdd99 的改进，它存在重大问题，可能导致 IDS 评估效率低下。基于[32]所做的工作，新的NSL KDD数据集解决了几个严重的问题，其中它消除了大约77个冗余记录。出于这个原因，为了设计我们基于 AI 的反应代理，我们使用了 NSL KDD 数据集。

为了对基于 NSL-KDD 数据集的 IDS 进行评估，我们使用了一个名为 Weka 的预处理和可视化数据挖掘工具。Weka 用于对训练样本进行分类。KDD数据集包含125943个连接和41个特征，其中每个样本属于以下攻击之一:拒绝服务攻击(DoS)、用户根攻击(U2R)、远程本地攻击(R2L)和探测攻击。

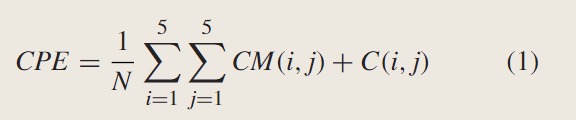
属性性质的多样性使得某些机器学习算法的学习是不可能的。当属性是连续的时，它使模型构建变得困难。因此，预处理步骤在构建分类模式之前是原始的，以最大化预测准确性 [33]。特别是，采用离散化方法来解决这个限制。离散化是一种数据挖掘技术，旨在通过将它们分组为间隔来减少连续变量的值的数量。在文献中，有两种离散化类型可以应用[34]:

•静态变量离散化:离散化是独立于其他变量执行一个变量。

•动态变量离散化:所有属性(变量)同时离散化。

除了离散化之外，我们还以仅具有主要攻击类别（DDoS、Probe、U2R、R2L）的方式对攻击进行分组。

1)性能比较测量:入侵检测系统的评估是一个基本问题，选择能够描述IDS[35]强度的指标是很重要的。IDS 的性能分别超出了分类率。我们根据模型准确度、检测率、精度和每个示例成本 (CPE) 评估我们的系统。在衡量性能时，一起使用的以下指标是必不可少的。



等式 1 表示每个示例 (CPE) 的成本，对于某些作品，它被称为成本敏感分类 (CSC) [37]。为了找到入侵检测系统错误分类的成本，这是一个重要的指标。其中 CM 是分类模型的混淆矩阵，C 对应于表 1 所示的成本矩阵，N 表示样本的总数。在下文中，我们提出了基于人工智能的不同系统。我们使用具有 16Go RAM 的 i5-8350U 计算机基于 10 折交叉验证来评估我们的系统。

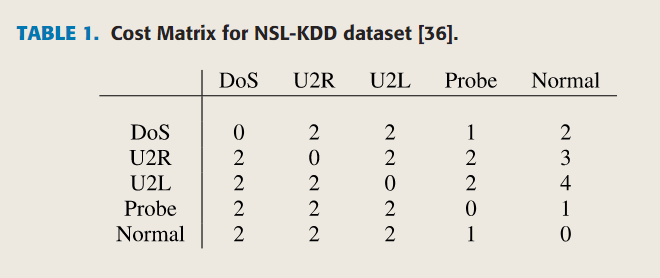


表1 NSL-KDD数据集的开销矩阵

2) 预处理、特征选择和分类：最初，我们提出了一种第一种方法，该方法包括预处理，然后使用不同的算法（J48、贝叶斯网络、随机森林和 Hoe ding Tree）对整个数据集进行分类。然后，我们选择了给出更好表现的最佳算法。

3)反向传播技术:在下面，我们探索了一种基于多层神经网络的技术，使用反向传播算法。多层神经网络由三层组成。第一层是输入层包含 41 个输入（数据集特征）。最后一层提供用于学习过程的分类答案（Dos、Probe、U2R、R2L、Normal）。在这种技术中，我们考虑一个隐藏层和 100 个神经元。这些参数是通过经验获得的，因为隐藏层的数量和神经元的数量的其他值，似乎在均方误差 (MSE) 方面没有显示出任何显着的改进。

4)分布式分类系统:在下面，我们提出了一个分布式分类系统，其中每个攻击类别(DDoD、Probe、R2L和U2R)分配给JRip算法。然后，采用AdaBoost算法合并得到的模型。

5) 结果讨论：表 2 中显示的结果表明，随机森林算法在整体精度和模型精度方面表现良好。但是，对于 U2R 和 R2L 攻击，它显示出非常低的精度。J48 以非常好的准确性和低误分类率（或 CPE）检测攻击。然而，J48 在 U2R 攻击的精度方面效率不高。Hoeffding 树算法表现出稳定的性能，但它也受到 U2R 攻击的低精度的影响。特别是，Bayes Net 算法显示出最差的结果，因为它无法很好地识别大部分 U2R 攻击。

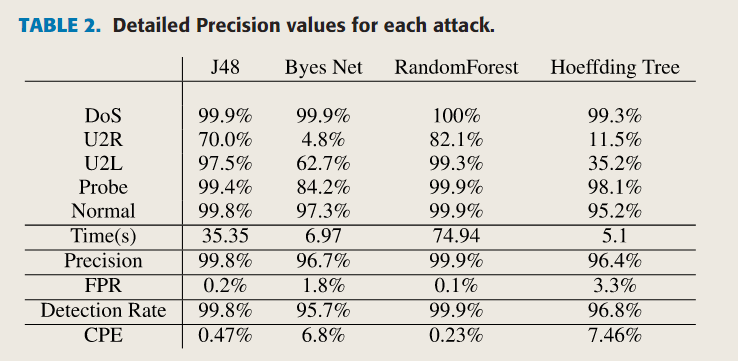


表2 每次攻击的精度值

与以前的方法在准确性、精度方面相比，反向传播算法显示出轻微的改进（表 3）。然而，错误分类的成本与处理时间相差一点。

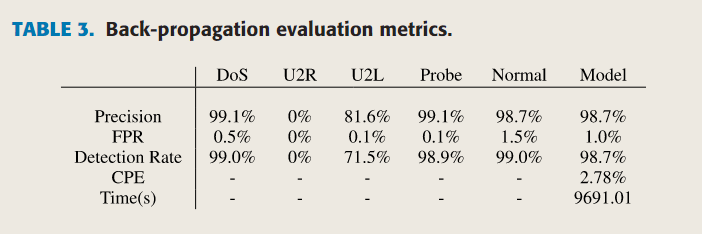


表3 反向传播算法评估指标

AdaBoost（表 4）在检测精度、检测率和每个示例成本（CPE）方面获得了增强模型。

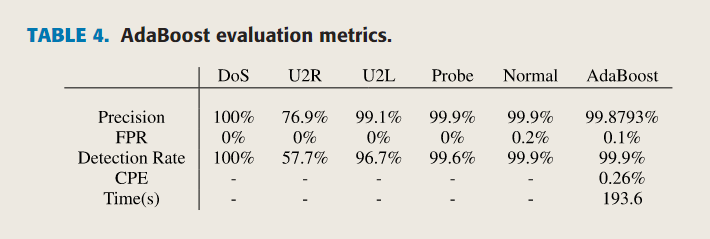


表4 AdaBoost评估指标

6) 比较研究：表 5 显示了性能结果。与之前的系统相比，该系统在检测精度、精度、检测率和每个示例成本 (CPE) 方面获得了增强模型。

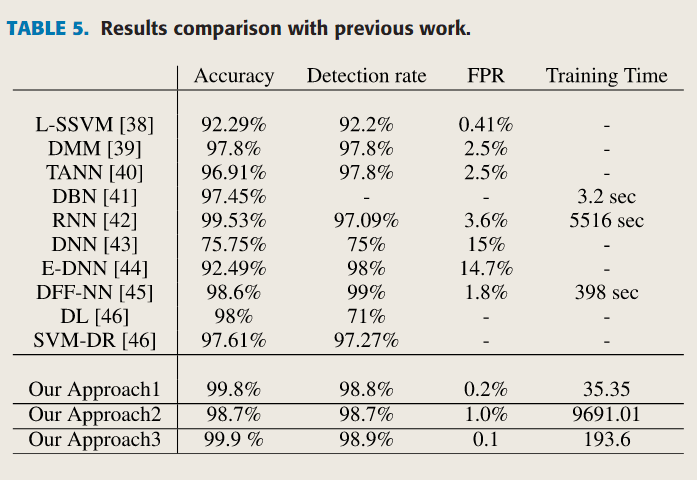


表5 实验对比

我们与最近的基于准确率、检测率、误报率和 CPE 的工作进行了比较。最近的工作总结在表 5 中。比较结果表明，我们基于系统的分布式 JRip 算法和集成方法是最好的，而其他系统的结果也很有希望。这些系统，即基于滤波器的支持向量机(F-SVM)[38]、Dirichlet混合模型(DMM)[39]、三角形区域最近邻(TANN)[40]、深度信念网络(DBN)[41]、递归神经网络(RNN)[42]、深度神经网络(DNN)[43]、[45]、[46]、Ensemble-DNN[44]、支持向量机降维[47]。

### B.异常检测

这一部分描述了实施和评估，以证明我们的AI框架在基于物联网系统中异常行为(非常见传感器数据值)分析来检测网络攻击的可行性和准确性。所提出的 AI 框架利用不同传感器数据之间的时间空间相关性来检测威胁。常见的传感器值表明，报告值的物联网设备可能会受到攻击，例如被一些恶意软件感染，或通过中间人冒充。具体来说，我们基于AI的框架检测物联网设备故障，并相应地实施反应对策。虽然超出了本文的范围，但为了完整起见，值得一提的是，我们的框架部署在智能建筑测试平台场景中，强制执行一个缓解计划：

1）重新配置 vAAA（虚拟身份验证代理），

2）启用 vChannelProtection 来建立安全的 DTL 通信，

3) 使用 SDN 强制执行新的流量过滤规则以丢弃恶意流量，以及

4) 可选地关闭和/或闪烁物联网设备。这些反应对策是在阿斯塔西亚欧盟项目[26]、[48]、[49]的范围内实施和评估的，超出了本文的范围，重点是评估机器学习机制来检测物联网系统中的网络攻击。

1.数据收集：我们的工作中采用的数据集是从我们智能建筑试验台的四个不同房间的真实传感器数据中获得的。我们在一个月的时间内每 2 分钟观察每个房间的温度和 CO2 的测量。该数据集用属性 (ID, Room, SensorValueCO2, SensorValue 温度, Class (Optional)) 描述，它包含 6876 个样本的测量，认为是正常值。我们为每个传感器构建了一个模型，其中包括 co2 和温度。图 1-3 描述了每个房间传感器数据的分布。我们注意到，另一方面，每个房间的 co2 值是不同的，温度在所有房间中的相同间隔，因此相同的模型可以处理所有这些。我们还可以使用第一个房间进行训练，而其他房间进行测试。

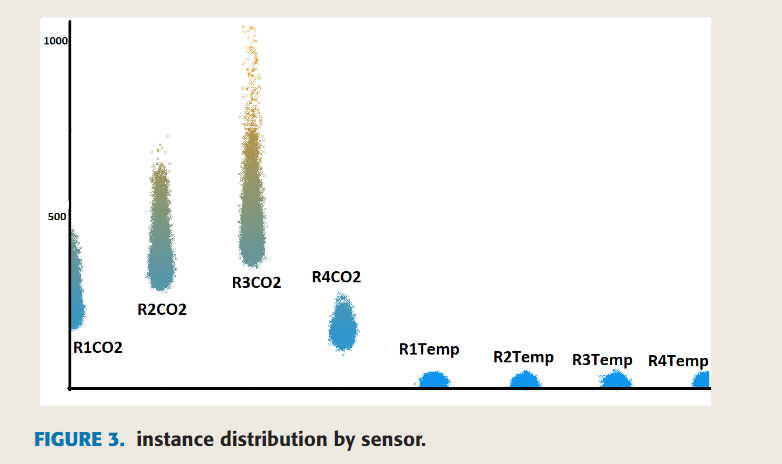


图1-3

2.数据集：

• 单值数据集 (SV)：生成值的简单数据集，它只将捕获的值和时间表示为特征。

• 前五个值 (P5V)：这种方法捕获了测量的传感器数据之间的时间相关性。由于温度是上下文，该数据集包括来自单个值数据集 [date, value] 的不同数据集中具有特征的先前值的上下文。为了保持事物清晰和限制标准，我们只使用了房间 1 数据集。该数据集包括每个值的 5 个先前值 [日期、值、值先例、值第二先例、第 5 个先例]。我们还注意到这些值之间存在很强的相关性。

• 之前不同的三个值（PD3V）：与之前的方法类似，这种方法利用了收集到的传感器数据之间的时间相关性。这种方法旨在通过每次只考虑最后三个不同的值来防止重复[日期、值、不同的先例、第二不同的先例、第三不同的先例]

•跨房间:由于所有房间的传感数据存在相关性，在这种方法中，我们通过结合房间值来检测异常来考虑这种相关性。通过利用这个数据集，我们结合了 mght 提高准确性的房间值，使 4 个房间最终达到以下数据集：[日期、房间 1、房间 2、房间 3、房间 4、标签]。

3. 一个类 SVM 模型：为了构建一个能够很好地识别数据集中的异常的模型，我们针对使用 python Scikit-learn 库实现和改编的单类支持向量机。我们提出的基于异常的IDS模型包括四个阶段。首先，对数据集进行预处理和清理。第二步是数据离散化，它包括将时间序列从连续值转换为离散区间。最新的阶段应用学习算法网格搜索步骤进行分类。对于温度数据集，我们将第一个房间值拆分为训练，第二个房间值用于测试。基于仅对温度数据存在空间相关性的观察，我们省略了用另一个房间测试 CO2 数据生成的模型。出于这个原因，我们根据训练数据集中的检测精度 33% 来评估学习模型。

4. 结果和比较：从温度值获得的结果表明，SV 和 P5V 在检测精度方面优于其他特征组合，其中达到了 98.86% 的检测精度。然而，在 CO2 案例中，p5V 数据集达到了 99.24%。

## 5. 结束语和开放研究挑战

在不久的将来，物联网系统有望彻底改变我们的日常生活。在潜在的增值特征中，按需安全措施的供应代表了面临网络安全攻击爆炸的突破。在本文中，我们研究了物联网系统最常见的威胁。然后，我们提供了一系列有前途的技术，并设计了一个安全框架，以全面的方式集成它们。事实上，我们强烈相信SDN、NFV和机器学习解决方案的联合使用可以使能够强制执行请求的安全策略的整体安全系统。我们还提供了一项研究，证明了我们基于 AI 的安全框架的可行性，该框架结合了基于知识的入侵检测和基于异常的入侵检测。一方面，对于基于知识的检测，三个不同的系统用于评估基于 NSL KDD 数据集的框架：

1)基于系统的分类算法，

2)基于分布式攻击规则关联的JRip算法，以及，

3)反向传播技术，其中我们执行了几种预处理技术，例如离散化。获得的结果非常有希望，其中评估指标使我们能够很好地评估框架并考虑错误分类攻击的效果。另一方面，我们的框架集成了一个用于传感器数据异常检测的 IDS，采用 One-Class SVM 在大多数提出的数据集组合中实现了超过 98% 的检测精度。

在下文中，我们描述了一些额外的研究挑战，这些挑战被我们的安全框架解决。首先，我们正在解决定义标准化接口以简化预期框架模块之间的交互的挑战，包括指定根据基于人工智能的决策做出反应所需的物联网安全策略的常见语言。其次，随着物联网格局不断演变，AI系统将需要自主重新配置，以处理额外的新兴(和潜在未知的)物联网网络攻击，这些攻击不遵循以前的网络/系统签名和模式。第三，另一个挑战是反应代理可以使用的机器学习方法和算法，以便动态规划最佳对策以根据不同的上下文强制执行。最后，我们还注意到，确保一定程度的安全性涉及额外的资源消耗和潜在的性能下降；因此，在反应模块中应深入研究安全要求和服务质量之间的权衡。