组会汇报

汇报人：侯富豪

2025年3月18日

汇报提纲

- 文献一: SeIoT: Detecting Anomalous Semantics in Smart Homes via Knowledge Graph

- 文献二: Unveiling Context-Related Anomalies: Knowledge Graph Empowered Decoupling of Scene and Action for Human-Related Video Anomaly Detection

文献一： SeIoT: Detecting Anomalous Semantics in Smart Homes via Knowledge Graph

ISSN	期刊名	综合评分	期刊指标	中国科学院分区	学科领域	SCI收录	是否OA	录用比例	审稿周期	近期文章	查看数
1556-6013	IEEE Transactions on Information Forensics and Security IEEE T INF FOREN SEC	8.7 ★★★★☆	h-index:95 CiteScore:14.40	1区	大类：计算机科学 小类：计算机：理论方法	SCI SCIE	No	较易	约6.8个月	文章	275061

收稿日期：2023 年 11 月 30 日；接受日期：2024 年 6 月 24 日。出版日期：2024 年 7 月 15 日；

通讯作者： Qing Li

Ruoyu Li, Yucheng Huang, Qingsong Zou, Zhengxin Zhang, and Yong Jiang ： 鹏程实验室战略与前沿交叉研究部，深圳；清华大学深圳国际研究生院，深圳

Qing Li and Dan Zhao ： 鹏程实验室战略与前沿交叉学科研究部，中国深圳

Fa Zhu：南京林业大学信息科学与技术学院和人工智能学院，南京

thanasios V. Vasilakos：现供职于沙特阿拉伯达曼伊玛目阿卜杜拉赫曼-本-费萨尔大学计算机科学与信息技术学院网络与通信系，以及挪威克里斯蒂安桑阿格德尔大学人工智能研究中心（CAIR）。

论文研究主题及意义

研究主题

1. 提出SeIoT，一个基于知识图谱的智能家居双模态异常检测框架。
2. 利用智能家居中的语义信息，包括流量周期性和设备/环境交互，来提高异常检测的准确性。

研究意义

1. **提高检测准确性：** 通过考虑设备间的交互和环境属性，SeIoT能够更准确地识别异常行为，包括那些难以通过传统方法检测到的攻击。
2. **增强智能家居安全性：** 通过实时监测和异常检测，SeIoT有助于保护智能家居免受各种网络攻击，增强用户的安全感。
3. **适应性与泛化能力：** SeIoT不依赖于特定制造商或平台，具有很好的适应性和泛化能力，能够广泛应用于不同的智能家居环境。

该研究主题面临的不足与挑战

1. **平台依赖性强，获取常态困难**[1]

   需厂商配置文件或平台源代码，无法适配封闭平台，限制实际部署

2. **交互与环境语义缺失**

   现有方法忽略设备间交互及环境动态关系，难以检测平台攻击（如指令注入/拦截）。

3. **多状态设备建模困难（常态的异质性）**[2]

   多功能设备（如摄像头）在不同状态（空闲/活跃）流量差异显著，统一建模复杂度高。

4. **特征表示局限性**[3]

   依赖欧几里得空间特征（如流量统计），难以捕捉智能家居中设备之间的复杂交互关系(非欧关系)。

本文解决思路

活动流量[1]和**空闲流量**[2]具有两种语义信息:

与**交互**相关的语义[3]：设备的触发动作与环境属性和住宅中其他设备的状态有关。

与**时间**相关的语义[4]：一些设备具有某些周期性的始终在线的行为。

1.动作指纹模块

1.仅依赖网络流量（包长、时序）区分空闲/活跃流量，实现跨品牌、跨平台兼容。

2.知识图谱表征语义

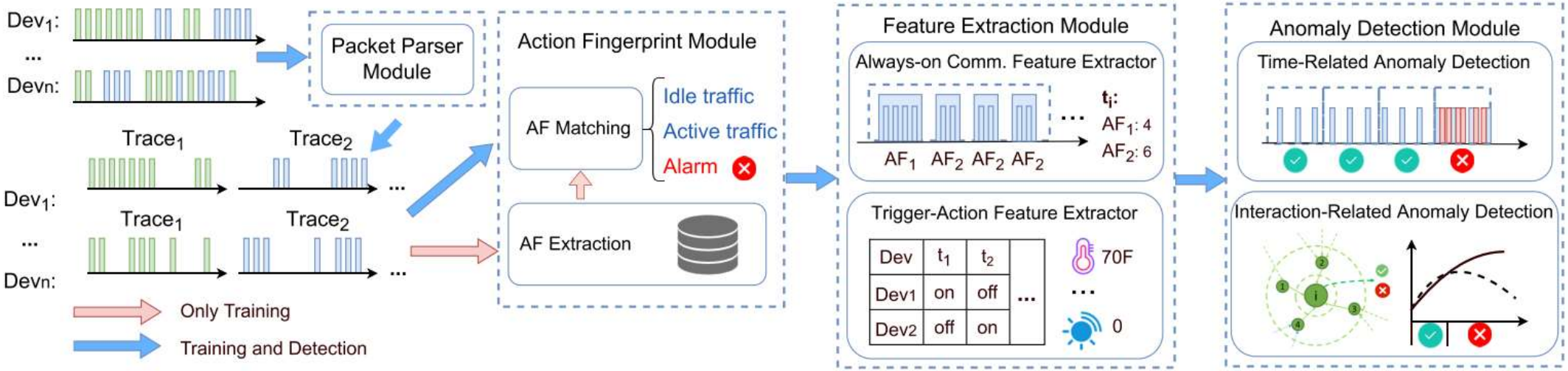
1.构建异构图（设备、环境、云节点），动态表征智能家居全局状态，解决非欧关系建模难题。

3.双模态检测框架

1. **时间相关检测**：分析空闲流量周期性，识别设备异常（如DDoS）。
2. **交互相关检测**：利用 FS-HAN（特征分离异构图注意力网络）建模设备和环境间的复杂交互，以检测平台攻击和设备间异常联动。

设备状态

所提出的模型/方案： SeIoT



数据包解析器模块:

- 对物联网流量进行解析，将混合设备流量分成流量级跟踪。

动作指纹模块:

- 提取动作指纹,推断物联网设备状态
- 对异常情况进行初步过滤

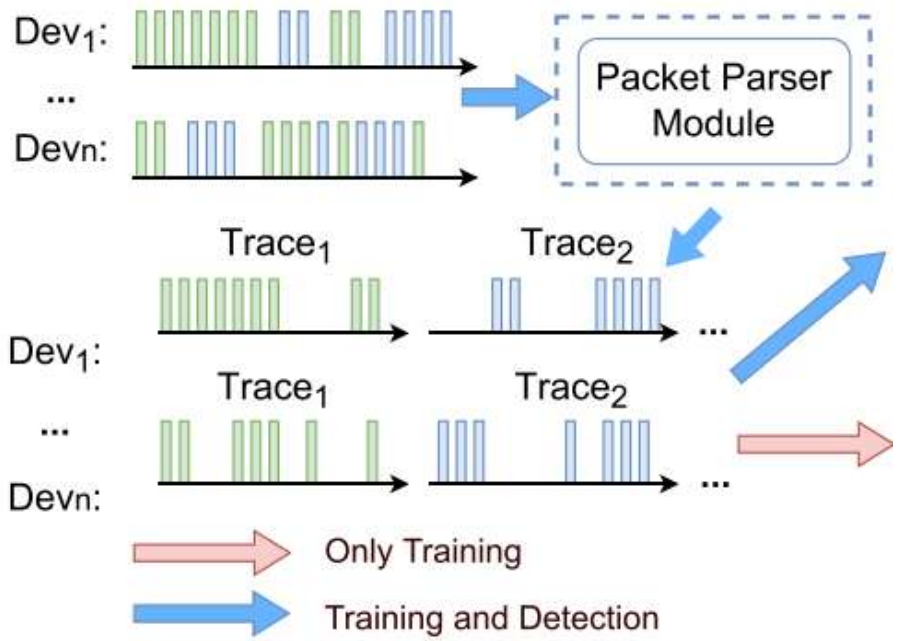
特征提取模块:

- 输入 AF 模块的匹配结果，获得始终在线的通信特征和触发行动特征，构建知识图谱

异常检测模块:

- 提出了一种双模异常检测机制，以解决物联网行为的异质性问题

数据包解析器模块



核心功能

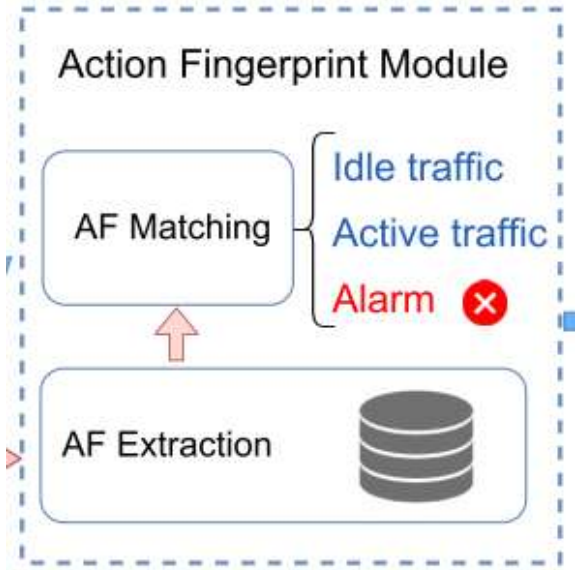
1. **流量分离**：基于 tcpdump 和 Scapy 等外部流量嗅探库构建，解析五元组信息或广播地址，将混合设备流量分成流量级跟踪。
2. **跨协议支持**：适配TCP/IP（五元组）和BLE（广播地址）。
3. **技术实现**（Trace Key定义）

   TCP/IP设备：双向五元组 <源地址, 目的地址, 源端口, 目的端口, 协议>。

   BLE设备：广播地址作为唯一标识符。
4. **作用于后续动作指纹模块**：

   通过流量级别的粗粒度分离提供数据输入，以进一步提取设备行为模式。

动作指纹模块



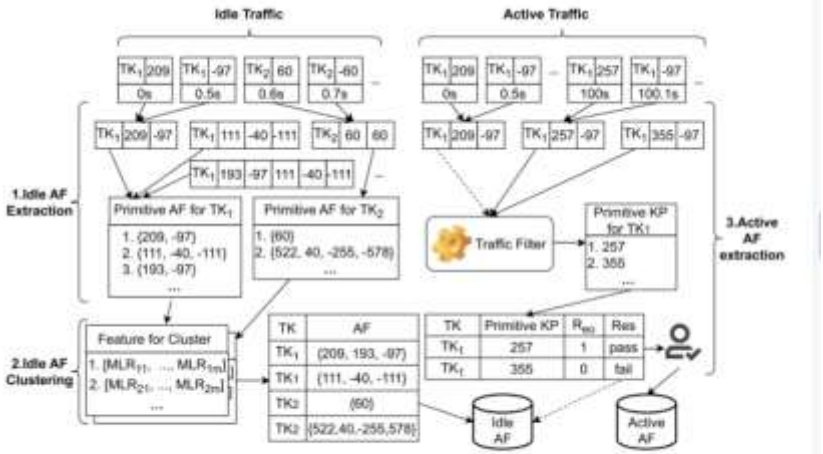


Fig. 3. An example of AF extraction.

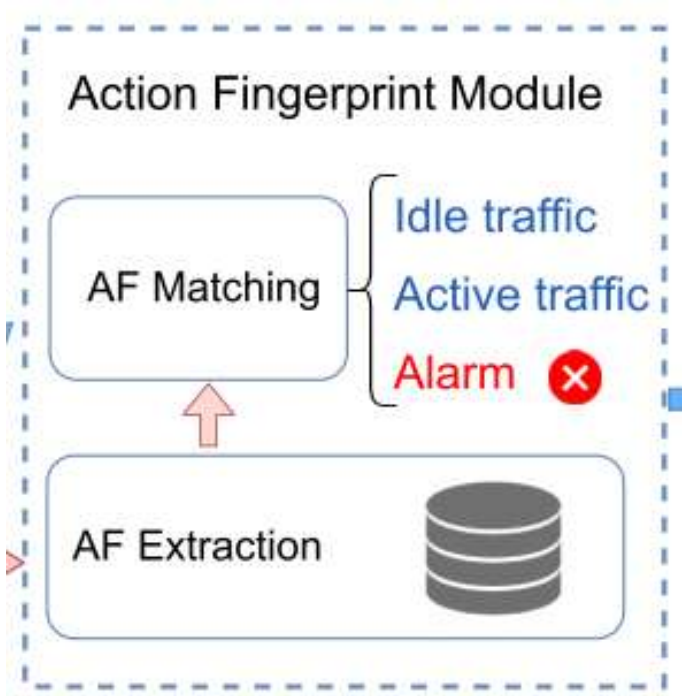
核心功能：

识别正在运行的设备的空闲流量和活动流量，并作为一个“白名单”式的检测机制，对异常情况进行初步过滤，为后续模块提供知识。

行动指纹提取：

1. 空闲AF提取：从无自动化规则的空闲流量中提取突发，统计每个突发的数据包长度集作为原始AF。
2. 空闲AF聚类：利用改进的Levenshtein距离（MLD）来计算不同AF之间的相似性，并通过DBSCAN算法合并相似长度集，生成最终的空闲AF。
3. 活动AF提取：从含自动化规则的流量中提取突发，过滤掉与空闲AF匹配的数据包长度，获得活动 AF（需结合应用日志进行验证，5秒内）。

动作指纹模块



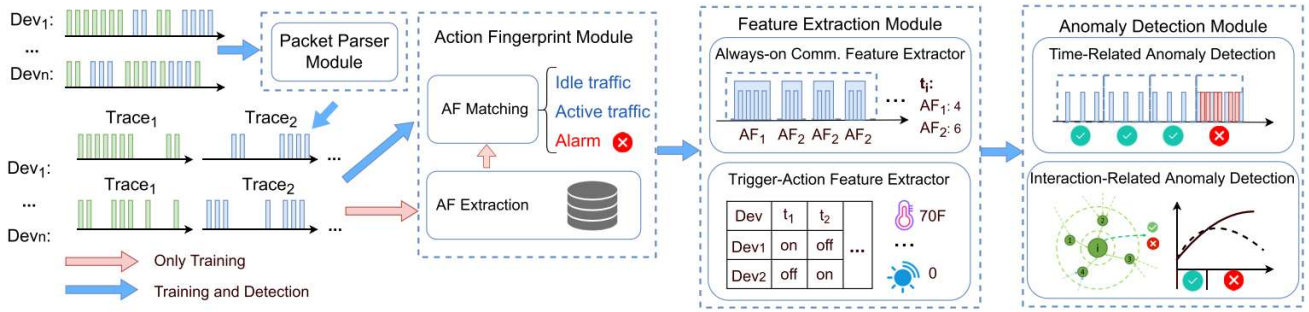
实时匹配与异常检测：

在运行期间，SeloT 会对比传输流量，

1）若匹配 Active AF，则标记为活动流量，

2）若匹配 Idle AF，则标记为空闲流量。

3）未匹配任何 AF 的数据包会被视为**可疑流量**。



特征提取模块

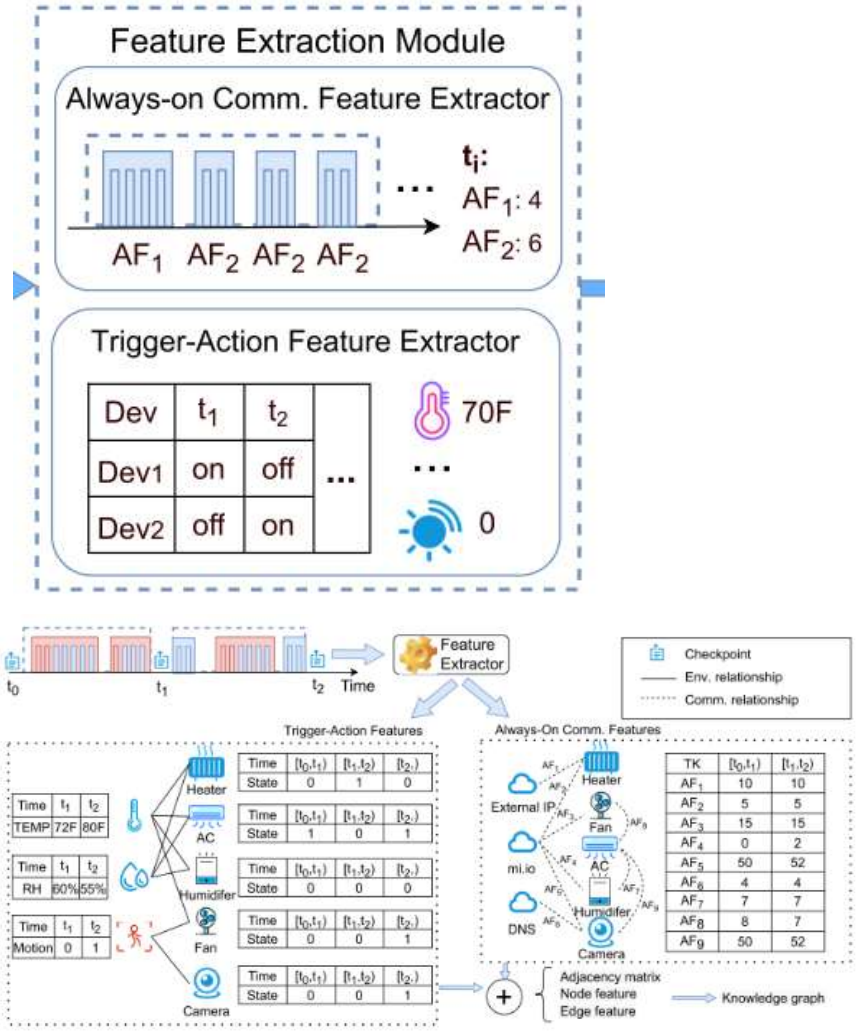


Fig. 4. Feature extraction module and knowledge graph.

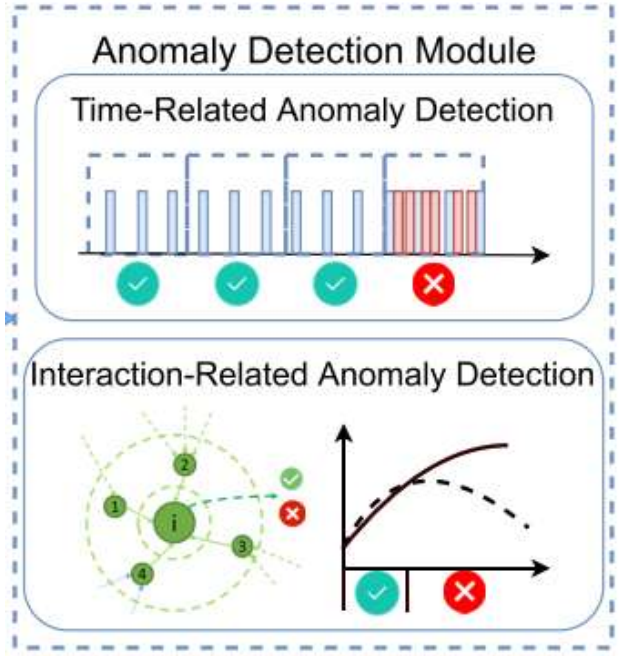
目标：

输入AF模块的匹配结果，获得两类语义特征，构建知识图谱，作为智能家居的整体状态表示，用于描述交互语义。

知识图谱构造

- **节点类型：**
  - IoT设备、环境属性、云服务
- **边类型：**
  - **通信关系：** 设备-设备、设备-云服务器
  - **环境关系：** 设备-其影响的环境变量
- **持续通信特征**
  - 描述设备在**空闲状态**下的**周期性流量**
  - 统计每个检查点（每分钟）匹配Idle AF的数据包数量
- **触发-动作特征**
  - 捕捉设备与环境的交互语义
  - 设备节点特征： 设备ID、先前状态、当前状态
  - 环境节点特征包括： 传感器的先前数值、传感器的当前数值

异常检测模块-时间相关异常检测



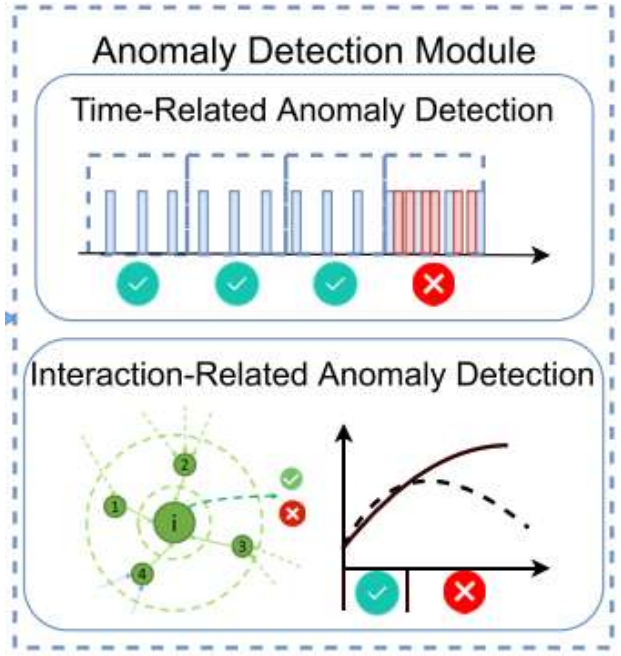
模块目标

通过双模态检测机制（时间相关+交互相关），识别智能家居中的异常行为。

时间相关异常检测

- 适用于周期性流量异常检测
  - **无传感器设备**（如摄像头）： 设定数据包数量阈值，超出阈值则判断为异常。
  - **传感器设备**（具有常态异质性）：设定高变速阈值和低变速阈值用于判断数据上传频率是否异常。
- **运行时流程:**
  - 检查 AF 流量是否超出设定阈值，超出则触发异常报警。

异常检测模块-交互相关异常检测



交互相关异常检测

- 适用于复杂设备联动检测，利用FS-HAN分析知识图谱中设备与环境的动态交互关系

检测流程：

- 输入：知识图谱（设备状态、环境属性）。
- 模型推理：FS-HAN生成预测结果。
- 异常判定：
  - 设备状态（分类）：预测与当前状态不符则为异常。
  - 环境属性（回归）：预测值与实际值偏差超过阈值则为异常

实验设置与结果分析--Experimental Datasets

测试平台搭建

1. **设备与协议**：19个现成物联网设备，覆盖TCP/IP和BLE协议。
2. **环境模拟**：真实公寓环境（430平方英尺），包含客厅、卧室等区域，定义20条自动化规则。
3. **数据收集**：
    1. **正常流量**：2个月采集，总计6.4GB。
    2. **攻击流量**：部署感染Mirai/Bashlite的树莓派，重放公开僵尸网络攻击流量（C&C、DDoS、扫描等），总计1.89GB。

实验设置与结果分析--Evaluation Metrics

评估指标

1. **动作指纹模块**：活跃流量精确度（Precision）、活跃/空闲流量召回率（Recall）。
2. **攻击检测**：
    1. 真阳率（TPR）
    2. 真阴率（TNR）
    3. 检测延迟（LAT）
    4. 攻击容量（C）
    5. 异常定位精度 (Prel)：被正确识别的恶意环境属性数 / 被检测为恶意的所有环境属性数。
    6. 异常定位召回率 (Recl)：被正确识别的恶意环境属性数 / 所有真实恶意环境属性数。
3. 运行时性能：CPU/内存占用、推理时间。

实验设置与结果分析--Competing Methods

对比基线方法

1. **动作指纹模块**：
    1. 空闲流量，对比
        1. **最先进**的使用空闲流量的物联网指纹识别方法
        2. 学习空闲突发模式的**自动编码器模型（AE）**
        3. 训练集（一周）和测试集（三周）
    2. 活动流量，对比
        1. PingPong：**最先进**的事件签名方法
        2. 学习活动突发模式的**自动编码器（AE）**
        3. 训练集（两天）和测试集（两天）
2. **设备定向攻击检测**：对比Kitsune（集成自动编码器）和IoTEnsemble（多模型融合）。训练集（一周）和测试集（一周）
3. **平台攻击检测**：对比HoMonit（DFA规则匹配）、HAWatcher（假设检验）和自动编码器（AE）。训练集（一周）和测试集（一周）。

实验设置与结果分析--Experimental results

1.动作指纹模块

1.空闲流量召回率（Rec_i）达99.6%，活动流量精确度（Pre_a）和召回率（Rec_a）均为100%。

2.匹配时间比SP和AE显著减少，意味着动作指纹模块在处理网络流量时的速度显著提高。

TABLE IV

STATISTICS OF IDLE AFS AND TIME CONSUMPTION OF AF MATCHING (L_{Max} IS THE MAXIMUM LENGTH OF AFS)

Device	ϵ	$\#AF$ (ours)	L_{Max}	t_{AF}	t_{SP}	t_{AE}
Sensor-equipped	1	30	18	**1.32us**	9.42us	22.88us
	5	7	40	**1.10us**		
	10	6	44	**0.87us**		
Sensorless	1	48	20	**2.63us**	8.52us	22.88us
	5	15	32	**1.21us**		
	10	7	57	**1.27us**		

TABLE V

COMPARISON OF ACTION FINGERPRINTING METHODS

Metric	AF (ours)	PingPong	SP	AE
Pre_a	**100%**	**100%**	\	61.6%
Rec_a	**100%**	98.4%	\	99.4%
Rec_i	**99.9%**	\	99.9%	99.6%

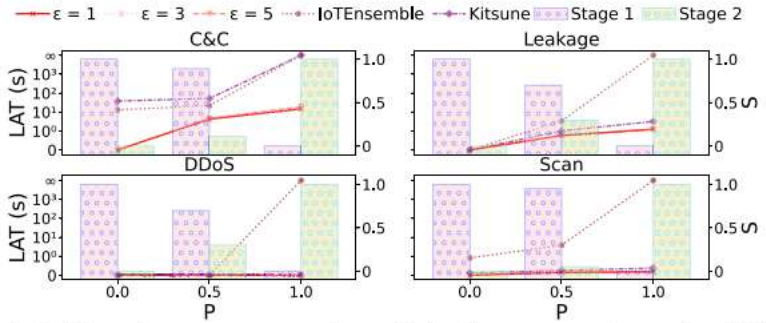
其中，ϵ 为DBSCAN参数；#AF (ours)为空闲动作指纹（AFs）的数量；

L_{max}为每个动作指纹中包含的最大数据包数量，反映了设备在空闲状态下的行为模式的复杂性。

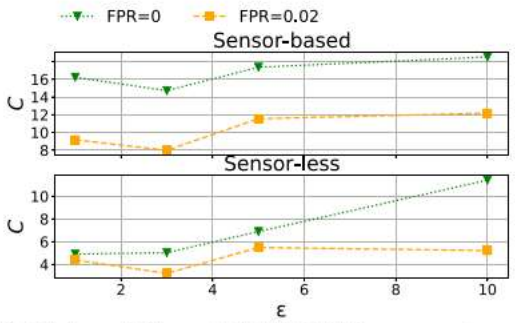
实验设置与结果分析--Experimental results

2.设备定向攻击检测

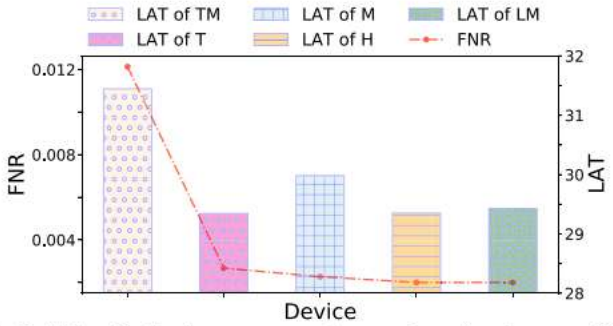
1. **检测效率**：平均检测延迟优于Kitsune和IoTEnsemble。
2. **准确性**：活跃流量的规避攻击测试中，表现出**较低的假阴性率**，表明其能够有效地识别攻击行为。



(a) LAT and stage proportion of device-targeted attacks (idle traffic).



(b) C for different DBSCAN parameter ϵ.



(c) LAT of device-targeted attacks (active traffic) and False Negative Rate (FNR).

(a)中，Stage1代表动作指纹模块， Stage2代表异常检测模块；

横轴（X轴）：表示规避攻击的比例（P），即攻击流量中模仿正常空闲流量的比例；

纵轴（右Y轴）：表示攻击被检测到的阶段（Stage1 / Stage2）比例（S）。

实验设置与结果分析--Experimental results

3.平台攻击检测

1. **命令注入/拦截**：SeIoT的真阳率、真阴率显著高于HoMonit和HAWatcher。
2. **精确度和召回率**：SeIoT系统在精确度和召回率上均优于基线方法。
3. **异常定位**：SeIoT在异常定位精度和异常定位召回率指标上均优于规则基方法。

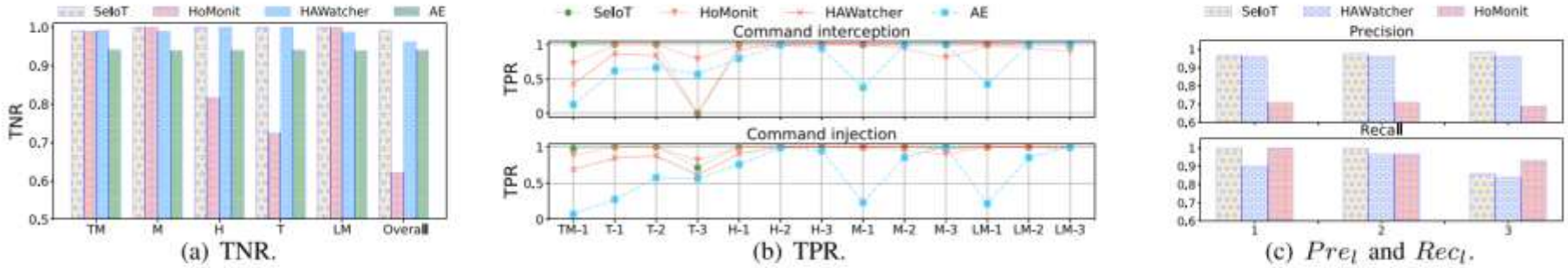


Fig. 8. Performance of detection against platform-based attacks.

横轴（X轴）：表示不同类型的设备或攻击场景（如光和运动相关设备、温度和湿度相关设备）

Command Interception：命令拦截；Command Injection：命令注入

实验设置与结果分析--Experimental results

4.运行时性能

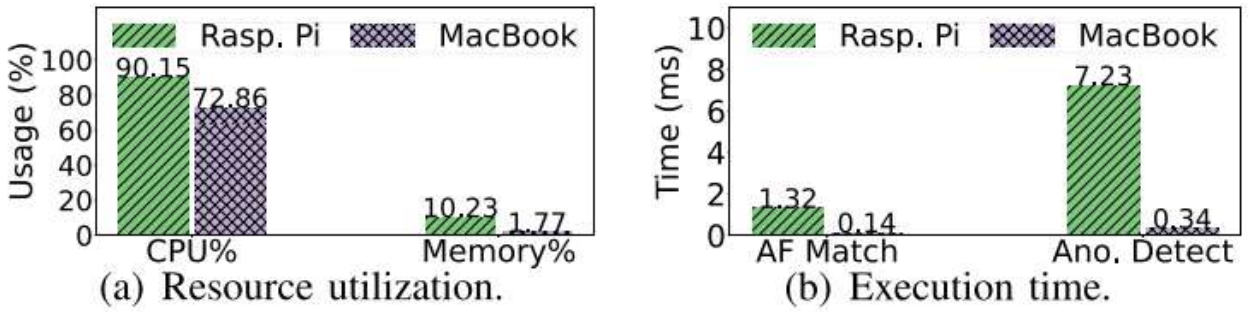
1. **资源消耗**:

   Raspberry Pi上内存占用仅10.23%，CPU利用率仍然可以接受，允许Chrome等其他应用程序正常运行

2. **实时性**:

   1. 单次推理时间仅毫秒级，与检查点时间窗口（1分钟）相比，决策延迟可以忽略不计，能够实时检测异常。

结论：在消费级机器上部署SeIoT具有可行性。



(a) Resource utilization. (b) Execution time.

Fig. 9. Runtime performance on consumer-grade machines.

实验设置与结果分析--Competing Methods

与基于 DFA 和基于自动编码器的方法相比， SeloT具有

1）高准确性：

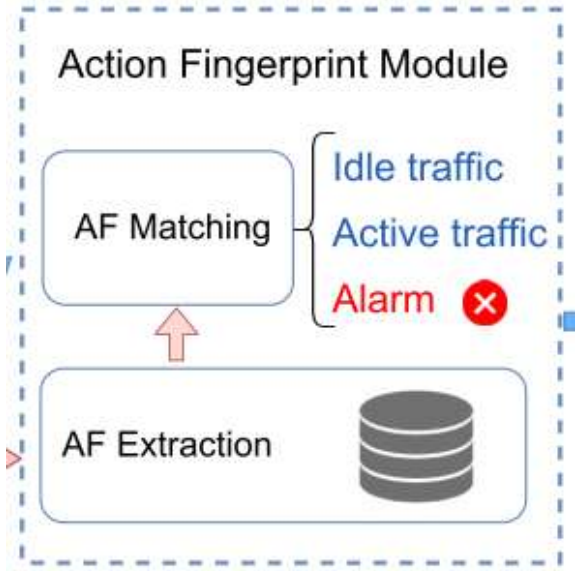
SeloT 在检测各种物联网攻击方面表现出更高的准确性，尤其是在识别新的攻击类型（如命令拦截）方面表现出色[1]。

2）可解释性：

SeloT构建了一个更结构化和可解释的数据表示，即一个描述智能家居整体状态的知识图；

由于FS-HAN可以精确定位导致异常的设备和环境属性，SeloT为用户提供了更好的可追溯性。

动作指纹模块补充



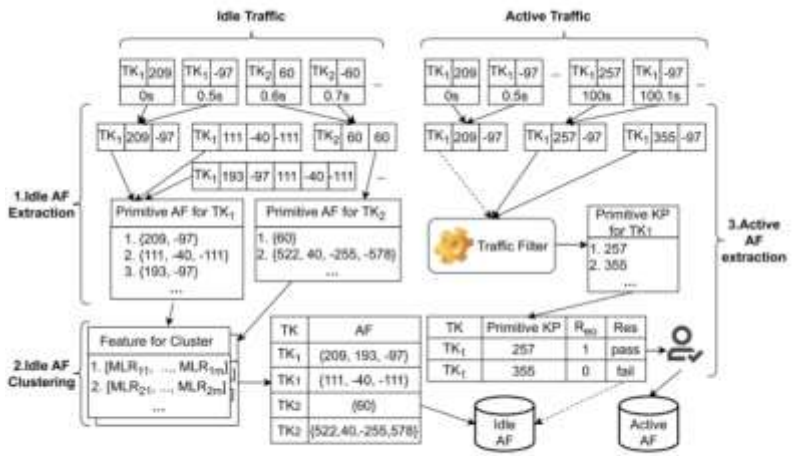


Fig. 3. An example of AF extraction.

核心功能：

识别正在运行的设备的空闲流量和活动流量，并作为一个“白名单”式的检测机制，对异常情况进行初步过滤，为后续模块提供知识。

行动指纹提取：

1. 空闲AF提取：从无自动化规则的空闲流量中提取突发，统计每个突发的数据包长度集作为原始AF。
2. 空闲AF聚类：利用改进的Levenshtein距离（MLD）来计算不同AF之间的相似性，并通过DBSCAN算法合并相似长度集，生成最终的空闲AF。
3. 活动AF提取：从含自动化规则的流量中提取突发，过滤掉与空闲AF匹配的数据包长度，获得活动 AF（需结合应用日志进行验证，5秒内）。

实时匹配与异常检测：

在运行期间，SeIoT 会对比传输流量，

1）若匹配 Active AF，则标记为活动流量，

2）若匹配 Idle AF，则标记为空闲流量。

3）未匹配任何 AF 的数据包会被视为**可疑流量**。

知识图谱的应用总结

对于SeIoT中知识图谱的应用方式:

训练阶段:

在特征提取模块利用正常状态（无异常）的数据来创建知识图谱（其内容会随着训练阶段实时数据的更新而更新）。之后知识图谱用于FS-HAN训练阶段的环境建模与训练。

检测阶段:

在特征提取模块利用设备或环境的实时数据（可能有异常数据）来动态更新知识图谱。之后知识图谱用于FS-HAN检测阶段用于预测和判断异常。

FS-HAN的工作流程及原理详解

1. 训练阶段：环境建模

- **任务**：学习环境属性与设备状态的正常关系。
- **输入**：知识图谱中的历史状态序列。
- **输出**：环境属性的预测模型（如温度变化模型）。
- **损失函数**：均方误差（MSE）或交叉熵，优化预测值与真实值的差距。

3. 检测阶段：异常判定

- **输入**：**实时**知识图谱状态（设备状态、环境属性值）。
- **预测环境属性值**：使用训练好的FS-HAN模型预测当前环境属性（如温度）。
- **异常判定**：

计算预测值 ypredypred 与实际值 yrealyreal 的差距（如绝对误差 |ypred−yreal||ypred−yreal|）。

若差距超过阈值（通过训练阶段统计正常数据确定），则判定为异常。

FS-HAN的工作流程及原理详解-示例场景

示例场景

假设知识图谱中存在元路径**空调→温度→窗户**：

- **正常情况**：温度超过阈值时，空调开启且窗户关闭。
- **攻击场景**：温度未超标，但窗户异常开启且空调未启动。
- **检测过程**：
  - FS-HAN通过元路径发现空调与窗户的逻辑关联。
  - 预测空调应处于“关闭”状态，但实际状态为“开启”（异常）。
  - 环境属性预测显示温度未达阈值，与实际值一致，排除传感器故障。
  - 最终判定为**平台型攻击**（如指令注入）。

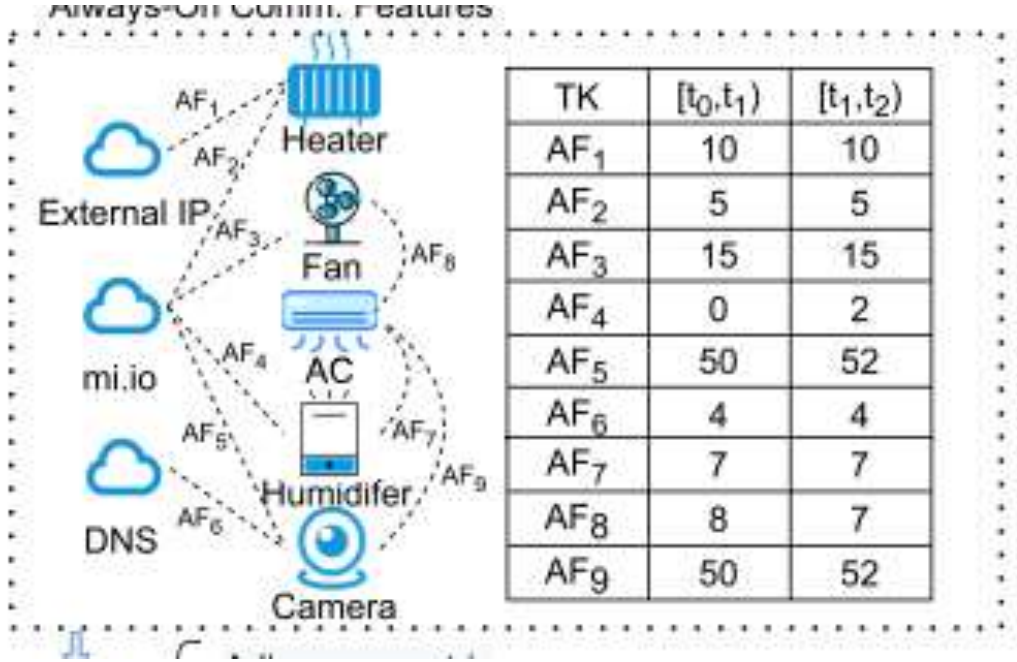
问题总结

问题1：

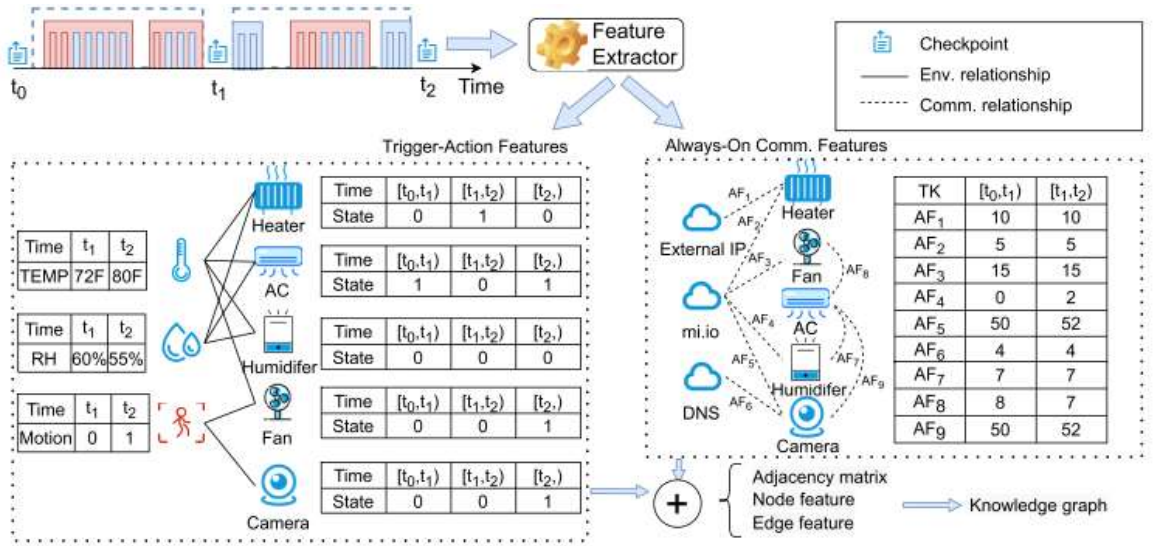
在**时间相关异常检测**中是否用到了知识图谱？

时间相关检测完全基于**统计分析方法**（如周期性特征的阈值判断），如果检查点中AF的数据包编号超出阈值，则认为是恶意的。

历史统计阈值等作为知识图谱的**静态**属性存在，但时间相关检测过程本身不依赖知识图谱的图结构或关系推理，仅使用这些存储的阈值进行直接比较。



TK	$[t_0,t_1)$	$[t_1,t_2)$
AF_1	10	10
AF_2	5	5
AF_3	15	15
AF_4	0	2
AF_5	50	52
AF_6	4	4
AF_7	7	7
AF_8	8	7
AF_9	50	52



知识图谱的应用总结

问题2:

为什么不直接用通常意义上的数据集来训练FS-HAN，还要建一个知识图谱，用知识图谱来训练FS-HAN。

文中提出:

1. 一种新颖的异常检测框架，可以简单地对网络流量进行嗅探，使其在物联网品牌和平台上具有很强的通用性。
2. 探索智能家居场景中的两种语义，并建议使用知识图来表示整个智能家居中的语义。
3. 一种双峰异常检测机制，用于检测各种攻击，特别是设计一种新的图形注意力网络，该网络考虑并有效地模拟设备和环境之间的复杂交互。

问题3

可以怎样推广?