文章编号:1003－6199(2024)01－0061－05 DOI:10.16339/j.cnki.jsjsyzdh.202401010

基于时序特征数据高效索引技术的物联网感知设备安全自动监测技术

王伟†,尚东方,韩雪

(交通运输部天津水运工程科学研究所,天津 300456)

摘 要:物联网设备已经被广泛应用于各个领域,为保证物联网的安全,排除内部隐患,基于时序特征数据高效索引技术设计物联网感知设备安全自动监测方法。结合时序特征数据高效索引技术提取物联网信息特征,在报文传输过程的基础上,区分不同流量数据之间的差异、恶意攻击软件与感知设备的系统特征,计算样本数据的表征值,得到物联网感知设备的原始信息特征。对数据特征进行分类,计算其数据内的缺失值和错误值,得到特征向量的筛选优化结果,计算训练损失函数,调整实际操作的阈值,保证数据特征分类的准确性。搭建物联网感知设备监测模型,训练判别器,进行物联网的自动监测。分别对数据包、字节以及数据流量进行识别,该监测技术可以准确地区分良性数据与攻击数据,从而保证物联网感知设备的安全。

关键词:时序特征数据;高效索引技术;物联网;感知设备安全;自动监测技术

中图分类号:TN06 **文献标识码**:A

Automatic Security Monitoring Technology of IoT Sensing Devices Based on Efficient Indexing Technology of Time Series Characteristic Data

WANG Wei†, SHANG Dongfang, HAN Xue

(Tianjin Research Institute For Water Transport Engineer M. O. T, Tianjin 300456, China)

Abstract: IoT devices have been widely used in various fields. In order to ensure the security of the Internet of Things and eliminate internal hidden dangers, an automatic security monitoring method for IoT sensing devices is designed based on efficient indexing technology of time series feature data. Combined with the efficient indexing technology of time series feature data to extract the networking information features, on the basis of the message transmission process, distinguish the differences between different traffic data, the characteristics of the system between malicious attack software and sensing devices, calculate the representative values of sample data, and obtain the original information features of the sensing devices of the Internet of Things. Classify the data features, calculate the missing and wrong values in the data, get the screening and optimization results of feature vectors, calculate the training loss function, adjust the threshold value of actual operation, and ensure the accuracy of data feature classification. Build the monitoring model of IoT sensing devices, train discriminators, and conduct automatic monitoring of the Internet of Things. Data packets, bytes and data traffic are identified respectively. This monitoring technology can accurately distinguish benign data and attack data, so as to ensure the security of IoT sensing devices.

Key words: time series characteristic data; efficient indexing technology; internet of things; perceive device security; automatic monitoring technology

收稿日期:2022－12－13

作者简介:王伟(1989—),男,天津人,本科,工程师,研究方向:交通信息化。

† **通信联系人**,E-mail: shaoliaohanp@163.com

物联网已经渗透进我们衣食住行的各个领域，频发的智能设备使个人的隐私安全受到严重的威胁，关键基础设施在实现数字化联网转型时也面临巨大风险。物联网安全不应局限在技术方面提高智能设备的安全性能，在处理随之产生的大量数据时，也需要合理的法规和完善的管理方案，确保风险的及时发现、准确定位和高效恢复。

当下政府以及其他科研单位都开始将物联网的安全问题作为重点研究对象，国内外一些研究尖端科技的公司已经开始着手于物联网感知设备安全监测产品的研发。如文献[1]为防止物联网被大规模入侵，设计了一种针对可疑设备的接入控制系统，同时建立了一整台识别机制。通过模型训练，构建流量特征指纹，并在实验验证时得到了 96% 以上的识别准确率。文献[2]在计算能力、存储空间等设备受限的基础上，设计了一种针对物联网感知设备异常点的检测机制。该检测算法能够在云服务器中感知模型内的数据，同时可以对异常数据进行识别，在网络流量与能耗方面具备较好的性能。文献[3]则是设计了一整套针对物联网安全管控的系统，该系统的技术架构包括表现层、应用层、服务层和数据层，具备较好的安全组织能力，可以对未知危险进行识别与定位，监测精度较高。本文结合上述文献，设计了一种基于时序特征数据高效索引技术的物联网感知设备安全自动监测技术。

1 基于时序特征数据高效索引技术提取物联网信息特征

物联网感知设备可以通过传感器接收到外界的静态信息，这些信息一般都会直接存储在数据报文中。想要设计一种物联网感知设备的安全监测技术，并判断感知设备所接收到的数据是否具备攻击性，就需要对其信息特征进行提取[1]。TCP 在传输的过程中共包括八个主要的步骤，如表 1 所示。

表 1 报文传输过程

序号	状态标量	符号	含义
1	0x01	FIN	数据传输结束
2	0x02	SYN	第一次握手
3	0x04	RST	重置连接
4	0x08	PUSH	信息推送
5	0x10	ACK	确认数据
6	0x20	URG	指针集合
7	0x40	ECN	数据拥塞
8	0x80	CWR	拥塞窗口减少

通过这些连接特性之间的差异性，可以表明不同恶意软件与普通感知信息之间的差别，进而对物联网的安全性进行监测。不同恶意软件与普通感知信息在数据传输时间片段之间存在较大的差异，有一些攻击型流量会集中在同一个时段内传输，而另一些攻击型流量则会周期性传输，这样的差异会导致不同的数据报文之间存在时间差，通过时间差的计算，可以区分流量之间的差异，其计算公式可以表示为：

$$T_{\text{sub}} = \frac{\sqrt{T_{\text{cur}} - T_{\text{beg}}}}{T_{\text{beg}}} \tag{1}$$

式中，T_{sub} 表示单位基础报文时间内网络数据流量的时间差；T_{cur} 表示数据流显示的当前时间；T_{beg} 表示第一个报文时间[2-3]。除时间差之外，通信端口之间也存在一定的差异性。恶意攻击软件在设计之初，为保证常用端口不会被轻易检测出来，会规定一个随机化的端口来传递信息，这样的端口是具备一定规则的，因此可以通过这类特征获取恶意攻击软件与感知设备本系统之间的差距。

$$P_{\text{sub}} = \frac{\sqrt{P_{\text{cur}} - P_{\text{beg}}}}{P_{\text{beg}}} \tag{2}$$

式中，P_{sub} 表示单位基础报文时间内，随机化传递信息的端口差；P_{cur} 和 P_{beg} 分别表示数据报文的当前端口编号以及第一个报文的基础端口号[4-5]。除此以外，数据包大小的差异性也可以作为体现恶意攻击软件与自身软件之间差别的检测工具。对于已经被捕捉到的数据样本，可以进一步进行密度分布的分析，其表征值的计算公式为：

$$d_k = \begin{cases} \left| \left\{ d_i : d_i < W_i, d_i \in \dfrac{1}{L_p} \right\} \right|, d_i > 1 \\ \left| d_i : T_{i-1} < d_k < T_{i+1} \{ d_i < W_i \} \right|, 1 \leqslant d_i \leqslant x \\ \left| \left\{ d_i : d_i > W_i, d_i \in \dfrac{1}{L_p} \right\} \right|, d_i > x \end{cases} \tag{3}$$

式中，d_k 表示密度分布的表征值，该表征值通过分段函数来表示。当其中的第 i 个表征值 d_i 大于 1 时，该表征值的时域功耗小于全部的功耗信息集合，将 W_i 作为功耗信息的下界[6]。当第 i 个表征值在 1 到随机数 x 之间时，其时域功耗在全部功耗集合的范围内，x 为表征值 d_1 与 d_{i-1} 之前的最大值。当第 i 个表征值大于随机数 x 时，其时域功耗大于全部的功耗信息集合。L_p 表示原始特征库集合，T_{i-1} 和 T_{i+1} 分别表示该密度分布表征值计算的前一个时域以及后一个时域。据此可以得到

该时段内物联网的原始信息特征。

2　数据特征分类

对所收集到的数据信息进行预处理，避免报文数据在接收之后，出现缺失、错误等情况，此时可以通过线性插值法计算其缺失以及错误值。

$$p(x)=p_0+\frac{\sqrt{(h_i-h_0)p_1}-\sqrt{(h_i-h_0)p_0}}{\sqrt{h_1-h_0}} \tag{4}$$

式中，$p(x)$ 表示待补充的缺失值和错误值；p_0 表示前一条被记录的特征值，p_1 表示后一条被记录的特征值；h_i 和 h_0 分别表示前一条和后一条被记录的特征值所在位置[7-9]。假设网络攻击的报文存在一个系列化的集合 $K_{ij}=\{h_1,h_2,\cdots,h_i\}$，且每一个报文均遵循 $h_i=\{d_{i1},d_{i2},\cdots,d_{in},f_i\}$。此时 h_i 表示第 i 个网络攻击报文，d_{in} 表示第 i 个网络攻击报文中样本的第 n 个特征，且该报文中最多包含 n 个特征，f_i 表示该网络攻击报文的样本标签。结合该网络攻击报文的特征值，可以得到特征向量的筛选优化结果：

$$\arg\min\left\{\frac{\sum_{i=1}^{m}[h_i-(\boldsymbol{w}^{\mathrm{T}}x_i+b_i)^2]}{\sum_{i=1}^{n}(\boldsymbol{w}^{\mathrm{T}}x_i-2b_i)}+\lambda_f\|\boldsymbol{w}\|_1^2\right\} \tag{5}$$

式中，m 和 n 分别表示两个相近报文中的特征维度；h_i 表示该报文内的第 i 个样本标签，当 h_i 为正值时，被记作 1，当 h_i 为负值时，被记作 -1；$\boldsymbol{w}^{\mathrm{T}}$ 表示获选最优系数的特征维度向量；x_i 表示报文中的第 i 个特征样本；b_i 表示第 i 个特征的维数；λ_f 表示模型系数[10-11]。通过该公式，可以得到最优的特征系数。在整个训练过程中，需要实时计算训练的损失函数：

$$F(x,p(g(x)))=\sqrt{\frac{\sum_{i=1}^{n}(p(g(x))-x_i)}{N_m}} \tag{6}$$

式中，$g(x)$ 表示输入层向隐含层发送的编码格式；$p(g(x))$ 表示解码器接收到的数据解码；$F(x,p(g(x)))$ 则表示网络训练过程中的损失函数；x_i 表示第 i 个输出向量值；N_m 表示输出值的数量。在整个分类过程中，都需要保证 $F(x,p(g(x)))$ 大于等于 0，才能使数据特征完整地完成分类步骤[12-13]。为达到有效的监测目标，可以调整实际操作的阈值：

$$\sum_{i=1}^{n}\frac{\boldsymbol{k}_i^{(h)}}{\boldsymbol{d}_i^{(h)}}>\eta_g \tag{7}$$

式中，$\boldsymbol{k}_i^{(h)}$ 表示通过第 h 个样本维度获取的误差向量；$\boldsymbol{d}_i^{(h)}$ 表示第 h 个模型的输出维度；η_g 表示假定阈值。当式(7)中的不等式成立时，则标志着本文中的数据特征分类能够完整地实现。如果该不等式不成立，则需要通过调整阈值来令其实现。通过上文中的方法，可以对物联网感知设备中接收到的数据进行准确的分类，区分接收到的数据是不是攻击型数据。

3　搭建物联网感知设备监测模型

在物联网感知设备的安全自动监测系统内，主要起到分辨自身数据与恶意攻击数据的模型由两部分组成，分别是判别器和生成器。

如图 1 所示，在博弈的基本程序中，存在一个初始时报文样本，该报文样本在形成之后，会直接进入判别器中，用于信息的安全识别。而具备攻击性的样本数据则是通过生成器生成，此类具备攻击性的报文样本想要进入物联网的感知设备，也需要经过判别器。通过这样的方法，对数据进行监测与识别[14-15]。由于第一代的报文全部被判别器拦截，因此一部分生成器就会在训练下得到更加有效的第二代生成器，第二代生成器会形成第二代报文，这样一来，判别器也需要通过训练，形成更具备识别准确率的判别器。以此类推，判别器与生成器均会不断迭代。

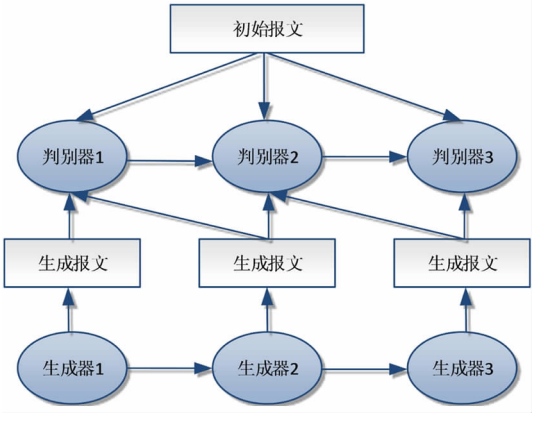


图 1　安全监测模型示意图

4　实验研究

4.1　系统运行环境与实验步骤

为测试上文中设计的基于时序特征数据高效索引技术的物联网感知设备安全自动检测技术是否具备有效性，设计如下实验。搭建物联网原形演

示系统，其中包含三个台式计算机、四个笔记本计算机、一个物联网网关、两个局域网作为系统硬件。该系统的整体运行环境如图 2 所示。

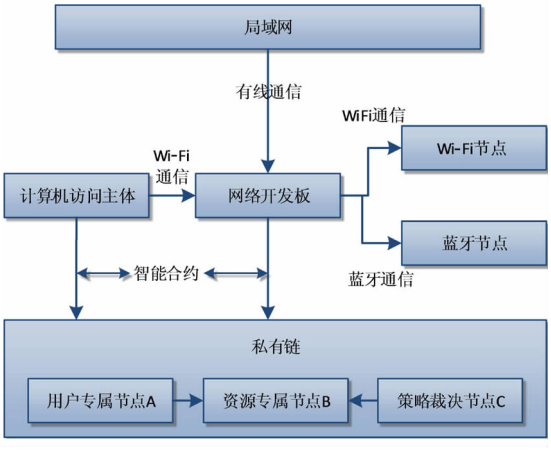


图 2 系统运行环境

图 2 所示的实验环境下，系统内的硬件配置均与 geth 以太客户端相连，同时通过智能合约在节点的配置下连接到私有链中，物联网网关的开发板则会通过 Remix 集成开发环境控制智能合约。选择某公共数据集中的物联网设备数据作为本实验的初始数据，并随机分配初始数据，构建一个良性流量数据集以及一个攻击流量数据集，二者作为安全环境与被攻击环境下的对照组。

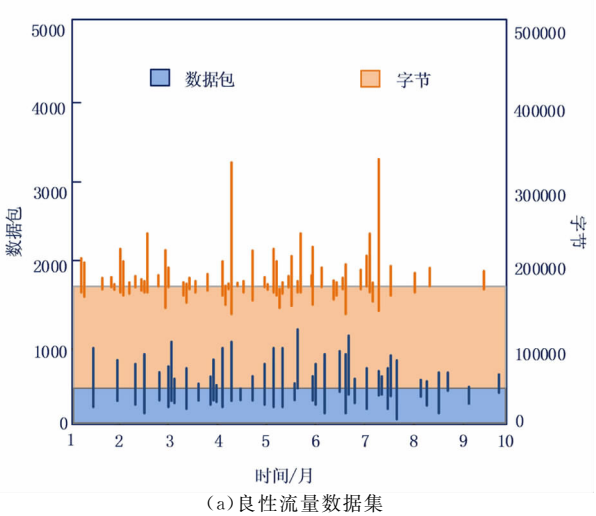
在实验中，首先需要读入数据集的 pcap 文件，由于本文针对物联网自动监测技术的测试不需要应用 pcap 文件的首部信息，所以可以直接略过最开始的 24 个字节。处理包头信息的同时，保存所有网络层以及感知层的协议类型，处理网络层和传输层的信息，同时记录协议的端口访问地址。针对安全环境与被攻击环境下，数据包、字节以及流量的差异，构建网路行为画像，同时判断本文设计的物联网感知设备安全自动监测技术是否可以正常运行。

4.2 安全监测结果

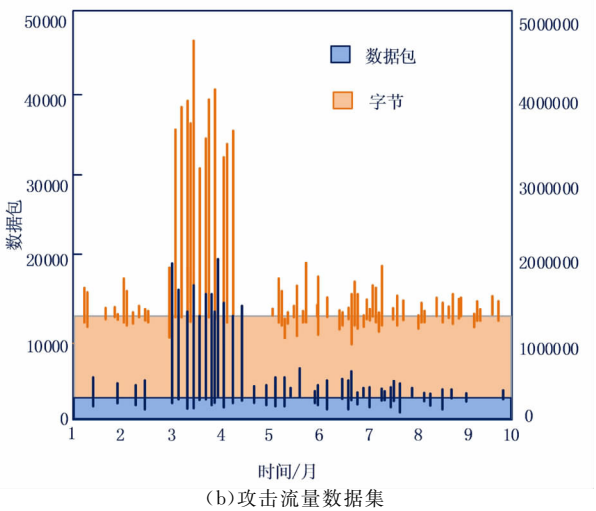
4.2.1 数据包与字节

在本实验中，分析的重点区域在于感知设备的数据包和字节。图 3 表示良性流量数据集以及攻击流量数据集下的感知设备数据包与字节的频率。

由图 3 可知，良性流量数据集和攻击流量数据集的数据包与字节频率均存在极大差异，很容易进行区分。这是因为在数据特征分类过程中，经过处理的初始数据形成了良性流量数据集，解决了其缺失及错误等问题，完善了良性流量数据，从而扩大了良性流量数据集与攻击流量数据集之间的差距。



(a)良性流量数据集



(b)攻击流量数据集

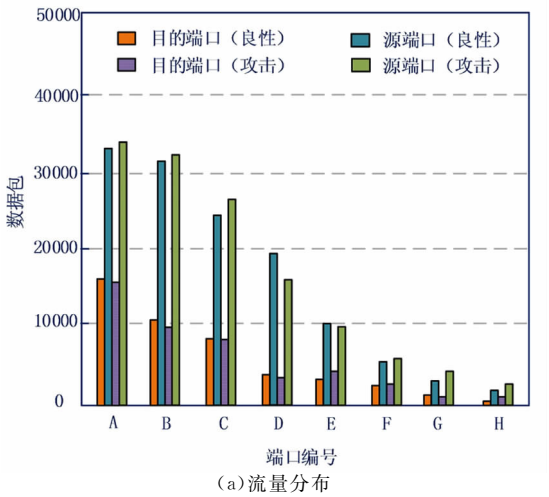
图 3 两种数据集下数据包与字节的频率

4.2.2 网络流量

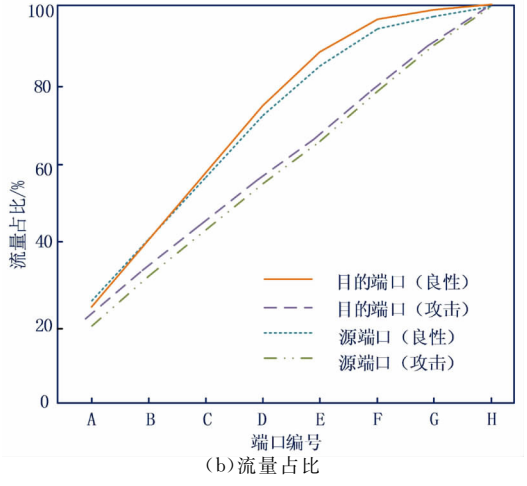
本实验中使用了 Python 脚本作为解析工具，在网络层与传输层之间，收集目的端口与原端口的网络流量分布以及占比。

如图 4 所示，目的端口和源端口在良性流量数据集以及攻击流量数据集内的流量分布以及所占比例。在 A 端口处，目的端口的数据包分布范围约为 15000，源端口处的数据包分布范围则为 33000 左右。在之后的端口编号中，数据包的分布范围逐渐减小，直至 H 端口下，其分布范围已经普遍降低至 3000 以下。在流量分布中，很难观察到良性流量数据集和攻击流量数据集的差异，但是在流量占比的图像中，两种数据集可以进行明显的区分。其中，良性流量数据集的端口流量占比曲线明显减缓，而攻击流量数据集的端口流量占比在该图像中则为一个正比例函数的趋势。这是因为判别器通过训练，形成更具备识别准确率的判别器，可

以区分良性流量数据集及攻击流量数据集。上述图像可以作为区分两类数据的依据。



(a)流量分布



(b)流量占比

图 4　两种数据集下不同端口的流量分布与占比

如上述两类监测结果所述，本文设计的基于时序特征数据高效索引技术的物联网感知设备安全自动监测技术可以成功地区分良性数据与攻击数据，从而达到保证物联网安全的目的。

5　结　论

设计了一种基于时序特征高效索引技术的物联网感知设备安全自动监测技术，该技术可以准确识别物联网内被恶意入侵的数据，并对其进行自动监测。实验结果表明：

(1)完整地实现了数据特征分类，突出了良性流量数据集与攻击流量数据集之间的差距。

(2)通过训练判别器，使判别器识别攻击流量数据集准确，增大了良性流量数据集和攻击流量数据集之间的区别。

因此，该监测方法可以有效地区分良性数据与攻击数据，完成物联网感知设备安全的自动监测，从而保证物联网的安全。在今后的发展中，物联网设备会更加广泛地应用于社会中，因此会更加迫切地需要此类监测技术，由于本文技术在通过流量分布区分良性数据与攻击数据的方面还存在不足，因此在今后的研究中，可以引入属性加密方法等对物联网感知设备安全进行自动监测，以确保物联网设备的安全性。

参考文献

[1] 齐波，冀茂，郑玉平，等. 电力物联网技术在输变电设备状态评估中的应用现状与发展展望[J]. 高电压技术，2022，48(8)：3012－3031.

[2] 李方平，吴楠，郭运华，等. 水电工程智能安全监测体系特征及发展趋势[J]. 人民长江，2021，52(S2)：259－264.

[3] 王俊淞，段斌，吴万波，等. 水电工程智能安全管控系统建设方案研究[J]. 中国安全科学学报，2021，31(S1)：96－102.

[4] 杜楚，杜新新，刁金. 面向物联网应用的压缩感知异常数据聚合机制研究[J]. 无线电工程，2021，51(11)：1335－1342.

[5] 张达，王济农，冀虎，等. 矿山微功耗安全监测物联网系统的研究与应用[J]. 通信学报，2020，41(2)：44－57.

[6] 任立民，薛晓，陈华敏. 基于物联网的煤矿井下机电设备安全监测系统设计[J]. 煤炭技术，2021，40(10)：166－168.

[7] 陈栋，张翔，陈能成. 智慧城市感知基站：未来智慧城市的综合感知基础设施[J]. 武汉大学学报(信息科学版)，2022，47(2)：159－180.

[8] 刘玉红，杨亮，朴春慧，等. 基于区块链的铁路工程施工安全监测数据共享关键技术研究[J]. 通信学报，2021，42(8)：206－216.

[9] 王永红，王诗瑶. 基于多协议的温室智能物联网系统研究[J]. 北方园艺，2021(5)：156－161.

[10] 李敏波，吴宇，卢晨耀. 面向情景感知的物联网设备智能控制系统[J]. 小型微型计算机系统，2021，42(12)：2637－2644.

[11] 何光辉，李晓伟，李鑫奎，等. 基于有限元与物联网的塔式起重机安全监测系统[J]. 中国安全科学学报，2020，30(11)：88－94.

[12] 范博，龚钢军，孙淑娴. 基于等保 2.0 的配电物联网动态安全体系研究[J]. 信息网络安全，2020，20(11)：10－14.

[13] 高祥斌，孔凡兴. 基于光纤复用的物联网节点安全控制域监测[J]. 激光杂志，2020，41(7)：133－136.

[14] 杨威超，郭渊博，李涛，等. 基于流量指纹的物联网设备识别方法和物联网安全模型[J]. 计算机科学，2020，47(7)：299－306.

[15] 于泊宁，任明，张志斌，等. 配电设备分布式局部放电感知技术的实现方法[J]. 高电压技术，2020，46(6)：1929－1938.