软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(7):1863−1879 [doi: 10.13328/j.cnki.jos.005356]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向工业物联网环境下后门隐私泄露感知方法*

沙乐天[1,2], 肖 甫[1,2], 陈 伟[1,2], 孙 晶[3], 王汝传[2]

[1](南京邮电大学 计算机学院,江苏 南京 210023)

[2](江苏省无线传感网高技术重点实验室,江苏 南京 210023)

[3](南京通信技术研究所,江苏 南京 210007)

通讯作者: 肖甫, E-mail: xiaof@njupt.edu.cn

摘 要: 伴随着工业物联网相关技术的高速发展,后门隐私信息的泄露成为一个重大的挑战,严重威胁着工业控制系统及物联网环境的安全性及稳定性.基于工业物联网环境下后门隐私的数据特征定义若干基本属性,根据静态及动态数据流安全威胁抽取上层语义,并基于多属性决策方法聚合生成静态与动态泄露度,最终结合灰色关联分析计算安全级与安全阈值,以此实现后门隐私信息在静态二进制结构及动态数据流向中的泄露场景感知.实验选择目标环境中27种后门隐私信息进行测试,依次计算并分析基本定义、上层语义及判决语义,通过安全级与安全阈值的比较成功感知多种后门泄露场景.实验还将所做工作与其他相关模型或系统进行对比,验证了所提方法的有效性.

关键词: 工业物联网;后门隐私;多属性决策;泄露感知

中图法分类号: TP309

中文引用格式: 沙乐天,肖甫,陈伟,孙晶,王汝传.面向工业物联网环境下后门隐私泄露感知方法.软件学报,2018,29(7):1863−1879. http://www.jos.org.cn/1000-9825/5356.htm

英文引用格式: Sha LT, Xiao F, Chen W, Sun J, Wang RC. Leakage perception method for backdoor privacy in industry Internet of Things environment. Ruan Jian Xue Bao/Journal of Software, 2018,29(7):1863−1879 (in Chinese). http://www.jos.org.cn/1000-9825/5356.htm

Leakage Perception Method for Backdoor Privacy in Industry Internet of Things Environment

SHA Le-Tian[1,2], XIAO Fu[1,2], CHEN Wei[1,2], SUN Jing[3], WANG Ru-Chuan[2]

[1](School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

[2](Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210023, China)

[3](Nanjing Telecommunication Technology Institute, Nanjing 210007, China)

Abstract: Leakage of backdoor privacy has become a major challenge with rapid development of industry Internet of Things (IIoT), causing serious threat to security and stability of industrial control system and internet of things. In this paper, some basic attributes are defined based on data feature of backdoor privacy in IIoT, upper semantics are extracted based on security threat in static and dynamic data flow, static and dynamic leakage degrees are generated based on multi-attribute decision-making, and finally security level and threshold are computed with grey correlation analysis. As a result, perception for leakage scenarios of backdoor privacy can be

* 基金项目: 国家重点研发计划(2018YFB0803403); 国家自然科学基金(61373137, 61572260, 61702283); 江苏省高校自然科学研究计划重大项目(14KJA520002); 江苏省杰出青年基金(BK20170039)

Foundation item: National Key Research and Development Program (2018YFB0803403); National Natural Science Foundation of China (61373137, 61572260, 61702283); Major Program of Jiangsu Higher Education Institutions (14KJA520002); Science Foundation for Outstanding Young Scholars of Jiangsu Province (BK20170039)

本文由“面向隐私保护的新型技术与密码算法”专题特约编辑黄欣沂教授推荐.

收稿时间: 2017-05-28; 修改时间: 2017-07-13; 采用时间: 2017-08-22; jos 在线出版时间: 2017-10-17

CNKI 网络优先出版: 2017-10-17 13:37:58, http://kns.cnki.net/kcms/detail/11.2560.TP.20171017.1337.003.html

accomplished in static binary structure and dynamic data flow. Twenty seven types of backdoor privacy are chosen for testing in target environment to compute and analyze basic definitions, upper semantics and judgment semantics, and successful perception for leakage scenarios is performed via comparison between security level and threshold. In addition, effectiveness of this work is validated through comparison with other models and prototypes.

Key words: IIoT (Internet of Thing); backdoor privacy; multi-attribute decision-making; perception of leakage

2005 年 11 月国际电信联盟正式提出了物联网(Internet of Things,简称 IoT)一词[1],引起世界各国广泛关注,被认为是继计算机、互联网之后,世界信息产业的第三次浪潮.目前,我国物联网发展与全球同处于起步阶段,呈现出良好的发展态势.据工业和信息化部统计数据显示,2016 年我国整个物联网产业的销售收入达到 9 000 亿元以上,充分体现了其强劲的发展势头.而“工业 4.0”概念提出要以智能制造为主导发起第四次工业革命,该战略旨在通过充分利用信息通信技术和网络空间虚拟系统相结合的手段,将制造业向智能化转型.“工业物联网”的概念正是在这样的背景下孕育生成,其核心理念是物联网技术与工业生产、加工、运输过程的高度融合,从而大幅提高生产制造效率,将传统工业提升到智能工业[2].

1 研究框架

工业物联网环境是由多种异构网络接入而成,根据实际环境中数据流及控制流的特征可划分为设备层、中间层、应用层,其中涉及到的后门隐私信息主要存在于设备层的终端节点及中间层的组态软件中,主要用于工业管理员或超级用户建立远程会话以实现控制、调试等目的.如图 1 所示.

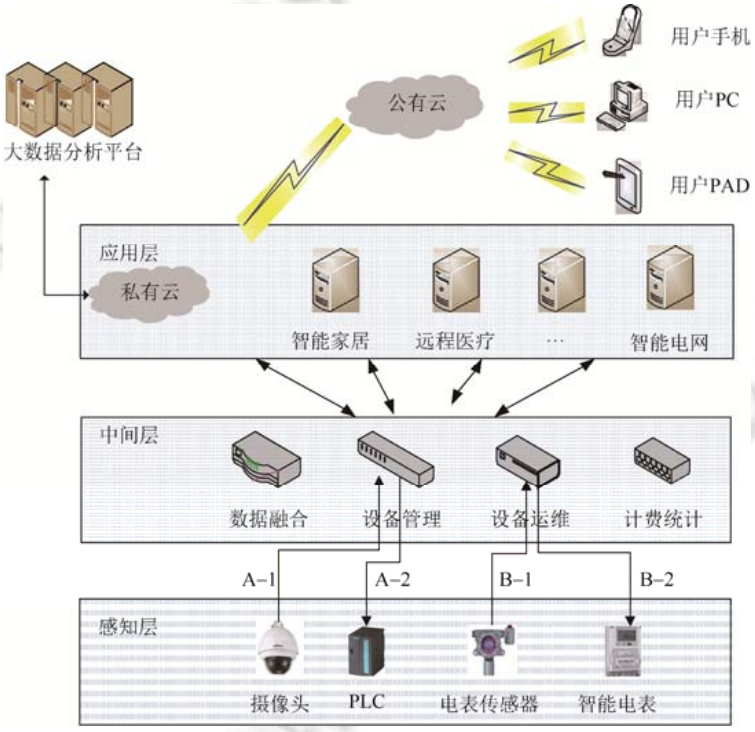

Fig.1 Industry IoT architecture based on sensed data

图 1 基于感知数据的工业物联网架构

工业物联网环境下设备层的终端节点主要包括摄像头、PLC、传感器、智能电表等等,负责工业控制设备与感知数据采集,中间层则实现生产设备管控、生产数据传输、生产现场统计等功能,而应用层则在此基础上为各种应用场景提供上层服务,并与公有云端对接,完成融合网络模式下的应用拓展.工业物联网与传统物联网

的主要区别在于:基于感知数据的自动化控制[3].如图 1 中基于 A-1 的设备监控完成 A-2 过程的 PLC 操作,具体过程可理解为:通过生产过程的视频监控,可了解生产流水线运行过程中的设备状态,基于感知数据分析给出目标设备的动作决策.

近年来,面向工业控制系统及物联网的恶意攻击事件频频爆发,尤其是基于后门及漏洞利用的攻击场景对工业生产及国家利益形成巨大威胁.大量爆发的变种病毒,如 Duqu、Flame 等大都通过社交网络攻击获得目标系统的外围信息,而后通过针对性的后门利用提升权限,从而在目标系统中潜伏下来实现持续性的敏感信息窃取.2014 年~2016 年间,相继爆发 Havex 及 Blackenergy 等针对工业控制系统的恶意攻击样本,同样通过针对性的后门利用达到了大量传播及持久潜伏的攻击目的.而 2016 年 10 月美国爆发的互联网瘫痪事件则源于面向物联网设备的僵尸网络攻击 Mirai,该恶意软件通过后门扫描及利用感染大量网络摄像头、DVR 及路由器,发起大规模的 DDos 攻击.由此可见,后门信息的泄露及利用对工业控制系统及物联网环境造成巨大威胁,严重损害了国家利益.作为一种重要的公共用户隐私,后门泄露及利用的恶意场景需要被安全防护系统所感知并捕获,从而为拒绝实时攻击提供支持与参考.面向未来将广泛应用的工业物联网环境,此类研究工作迫在眉睫.

2 相关研究

工业物联网的理论框架及产业推广目前尚处于初级阶段,主要通过整合工业控制系统及物联网环境的结构特征及技术特点而成.而后门信息作为国家级公共用户的重要隐私之一,其数据安全的保护需要基于典型的隐私保护技术来实现.因此本文相关的研究工作主要包括以下几方面:第一,工业控制系统中的数据安全;第二,物联网环境中的数据安全;第三,典型的隐私保护方法及技术;第四,工业物联网环境中的数据安全.具体地,

第一,工业控制系统中的数据安全.文献[4]以工控环境中典型系统 SCADA 为对象展开综述,总结阐述了工控环境下的 24 个风险评估方法,根据各自的评估目标、应用场景、风险管理、风险对象论证了以上风险评估的必要性,并提出了若干研究热点,其中尤其强调 SCADA 系统中信息泄露所导致的控制流劫持风险评估.从入侵检测的角度分析,文献[5]实现了以 PLC 为代表的工业控制环境入侵建模及实时检测,通过对传输私有协议中关键信息的提取,可定位重要的入侵行为特征.同样实现的入侵检测方法如文献[6],针对医疗工业控制系统,该文献提出行为规则的特征定义及状态机背景下的行为演化模型,从而将攻击者的恶意行为检测等价于该环境中目标信息在行为规则及状态机中的非法转移度量.文献[7]从攻击效果评估的角度对工业控制系统进行分析,以通用的 ModBus 协议为例,设计阈值度量方法,成功检测隐藏的攻击行为.而针对工控环境中底层的设备固件,有大量工作围绕控制流及数据流中关键信息的度量与挖掘而展开研究.C-FLAT[8]针对设备层的嵌入式系统进行控制流的远程安全证明,在黑盒情况下面向基本的控制流运行块设计控制流安全验证方法,并最终基于 ARM 的可信基设计原型系统.而文献[9]则是针对固件镜像所提供的 Web 层服务展开后门与漏洞挖掘,通过自动化的固件解包方法来抽取固件中 Web 端相关的文件结构,并基于自定义的 Web 漏洞类型特征分析其中若干个关键数据的采样点,最终给出漏洞模式判决结果,并自动化生成漏洞利用程序.其实验分析中获得了大量未知后门与漏洞,但目前其自动化固件解包分析方法的适用性不强,且关键数据的抽取局限于 Web 类型中.文献[10]则主要针对设备层固件中的二进制数据部署函数级的相似度识别,在屏蔽 CPU 架构的情况下成功地对 x86、arm、mips 实现兼容,通过该方法对已知后门在大量平台上进行了验证.Avatar[11]工具直接针对固件设备展开控制流及数据流离线分析,基于动态符号执行及 qemu 虚拟机执行来完成静态及动态相结合的分析方法,可实现漏洞及后门的挖掘,但执行效率较低.

第二,物联网环境中的数据安全.从物联网技术发展的全局角度及物联网系统结构的框架设计出发,目前的研究工作具体从以下角度展开:文献[12]同样针对密文搜索方案进行分类研究,并拓展描述了基于访问控制、基于隐藏访问模式的隐私保护技术,最终提出面向搜索数据保护的隐私保护研究方向.文献[13,14]针对物联网的应用场景提出安全分析及实践性的安全数据保护方法,在实际的应用场景中给出安全威胁评估方法,定义目标数据集合,并提出低开销的保护框架.另外,从物联网的位置隐私数据保护角度出发,文献[15]介绍了相关背景知识,包括位置服务应用场景、体系框架等等,随后讨论 LBS 中的攻击者模型和隐私保护度量指标,最终给出未来

LBS 隐私保护技术潜在的研究方向.文献[16]依托移动设备上的传感器探索用户位置及路径的推导方法,将传感器中数据聚合后设计用户位置及用户行为规则的语义聚合算法,从攻击的角度向移动设备上的位置隐私保护提出挑战.同时,两层传感网环境下的隐私保护也是研究热点之一,文献[17,18]分别给出面向两层传感网的隐私保护方法,分别基于 *k*-NN 和 Top-*k* 方法设计查询处理算法,前者针对节点俘获攻击和共谋攻击实现较好的防御效果,后者则给出安全证明过程及能耗分析过程.

第三,典型的隐私保护方法及技术.以口令数据为例,文献[19]从口令生成的脆弱行为入手,介绍了中英文用户口令的特征、分布和重用程度;总结近年来提出的主流口令猜测算法,并根据具体的攻击对象信息差异进行分类,最终对比了当前主流的口令强度评价器.同样从隐私信息本身的特点出发,文献[20]提出一种基于博弈论的隐私保护模型,在允许访问者对隐私相关信息进行访问的同时,有效地阻止访问者获取被访问者隐私信息的行为.该模型主要分析访问者与被访问者之间不同的博弈策略所对应的收益,最终完成对访问请求的判决.同时,结合场景特征的隐私保护方法发展较快,如文献[21]针对移动智能终端中参与式感知进行隐私保护综述,介绍了参与式感知的基本应用和攻击模型,对比分析了各种隐私保护技术的性能并总结主要优缺点.文献[22]综述了云计算环境、云数据、云计算应用中的各种关键安全问题,其中就包括云环境中的隐私计算、密文存储、密钥管理等问题.另外,作为引申方法论,安全协议的形式化分析技术同样适用于网络层的隐私数据保护[23].

第四,工业物联网环境中的数据安全.文献[24]针对工业物联网环境中的 SCADA 系统进行综述,具体到智能交通、智能电网、智能医疗等应用场景,提出系统环境安全、数据隐私安全及网络通信安全的缺陷及加固方法.文献[25]则从传感隐私数据保护的角度出发,设计并实现面向工业物联网的一种零知识协议,用于保护目标隐私数据在链路传输过程中的安全.文献[26]则针对工业物联网中智能电网的典型环境提出一种基于正弦波电压的新型时钟同步方法,用以抵抗延迟攻击.文献[27]从工业控制系统中典型的攻击事件出发,分析震网、夜龙等典型 APT 的影响与危害,提出面向工业物联网的安全数据访问控制模型.

针对以上情况,本文提出一种面向工业物联网环境下后门隐私的泄露感知方法,基于后门隐私在特定场景下的数据特征定义安全相关的基本属性,针对静态文件泄露及动态数据流泄露的场景特征抽取上层语义,结合多属性决策及灰色关联分析生成动态安全级及安全阈值,最终通过两者的比对完成泄露感知.本文设计并实现了系统架构及模块化功能,并通过大量实验及数据分析验证了该方法的有效性.

3 面向工业物联网下后门隐私的泄露感知方法

本文主要针对工业物联网环境设计一种后门隐私的泄露感知方法,本节首先给出后门隐私的基本定义,即何为后门隐私信息以及从后门隐私数据流中抽取的若干特征属性.之后基于多属性决策方法对后门泄露的危险数据流建立感知模型,实时检测后门泄露场景.

3.1 相关定义

为准确描述后门隐私在该系统架构中的相关特征,本文定义后门隐私信息为 BP(backdoor privacy),相关定义为后门隐私的属性集合:*BP.Attribute*(简记为 *BP.A*),主要可分为三大类:基本定义,用于描述后门隐私信息与泄露场景相关的基本属性,取值可从数据流场景中实时抽取并计算;上层语义,用于描述后门隐私信息在静态控制流及动态数据流中与泄露场景相关的特征化语义,数值经实时场景数据采集后需通过算法及模型处理后才能获取;判决语义,用于判决目标后门信息最终的动态安全级,取值经多种基本定义及上层语义汇聚所得.由于基本定义从实时后门隐私泄露及静态隐私信息数据特征中抽取并分类、聚类,因此可保证相关定义覆盖后门隐私安全相关的所有特征属性.以下给出各项的具体内容.

(1) 网络层级(network layer),简记为 *BP.nl*,属于基本定义.主要指后门隐私所处的位置,通常有两种可能:第一是设备层(device layer),简记为 dl,第二是中间层(middle layer),简记为 ml.

(2) 地址属性(address attribute),简记为 *BP.aa*,属于基本定义.根据后门隐私信息所处的地址属性,可判决目标后门信息的威胁程度.具体包括如下几种可能:第一,地址不可读不可写(nonread-nonwrite),简记为 nr-nw;第二,地址可读且不可写(read-nonwrite),简记为 r-nw;第三,地址可读且可写(read-write),简记为 rw.

(3) 数据格式(data sheet),简记为 *BP.ds*,属于基本定义.后门隐私的数据格式具体可分为两种:明文(plain-text),简记为 pt,代表后门信息未被加密,格式为可见明文;密文(cipher-text),简记为 ct,代表后门信息已被某种或几种加密算法加密,格式为不可见密文;混淆(obfuscated-text),简记为 ot,代表后门信息通过混淆处理变成部分不可见格式.

(4) 数据内容(data content),简记为 *BP.dc*,属于基本定义.后门隐私信息的数据内容可根据其完整程度分成两种:第一,完整可见的数据内容(integrated-visible,简记为 iv);第二,不完整可见的数据内容(non-integrated-visible,简记为 niv).

(5) 操作主体(data operator),简记为 *BP.do*,属于基本定义.主要描述利用后门开启远程控制的主体信息,具体可分为:管理员(administrator,简记为 ad)、超级用户(super user,简记为 su)以及攻击者(attacker,简记为 at).

(6) 数据功能(data function),简记为 *BP.df*,属于基本定义.后门隐私信息被利用后可开启一定的控制功能,根据功能的大小可分为两种:第一,开启远程设备控制功能(remote control,简记为 rc),可导致目标设备的行为控制或启停,此类权限可定义为最高权限;第二,开启部分远程控制功能(partial remote control,简记为 prc),可导致目标设备开启端口或开启特征函数接受特征数据包,此类权限可定义为部分权限.

(7) 触发方式(trigger method),简记为 *BP.tm*,属于基本定义.该定义主要描述后门隐私信息被利用开启控制的触发方法,主要包括两种:第一,主动触发(active trigger,简记为 at),是指管理员或攻击者可主动向目标设备建立远程连接,而后输入后门信息开启权限;第二,被动触发(passive trigger,简记为 pt),是指管理员或攻击者发送特征数据包到目标设备,目标设备被动接受数据包后开启权限.

(8) 逆向痕迹(reverse trace),简记为 *BP.rt*,属于上层语义.该定义用于描述目标设备被攻击者逆向分析的痕迹检查结果,具体赋值主要包括:第一,完全逆向破解(full reverse-engineering),简记为 fre;第二,半逆向破解(semi reverse-engineering),简记为 sre;第三,未逆向破解(non reverse-engineering),简记为 nre.具体赋值过程见第 3.2 节.

(9) 流量检测(flow monitoring),简记为 *BP.fm*,属于上层语义.该定义用于描述攻击者利用后门进行远程设备控制时的动态流量检测结果.普通的流量检测至多只能做到对数据包内容的解析,而本文的流量检测主要实现针对后门利用场景的深度数据包检测,检测结果主要包括:第一,存在后门利用(existing backdoor leakage),简记为 ebl;第二,不存在后门利用(non-existing backdoor leakage),简记为 nebl;第三,存在后门探测利用(existing test backdoor leakage),简记为 etbl.具体赋值过程见第 3.2 节.

(10) 静态泄露度(static leakage degree),简记为 *BP.sld*,属于判决语义.该定义抽取并聚合网络层级、地址属性、数据格式、数据内容、操作主体、逆向痕迹等,用于描述后门隐私在静态逆向分析中对攻击者的泄露程度.具体赋值过程见第 3.2 节.

(11) 动态泄露度(dynamic leakage degree),简记为 *BP.dld*,属于判决语义.该定义抽取并聚合网络层级、地址属性、数据格式、数据内容、操作主体、流量检测等,用于描述后门隐私在动态数据流向中对攻击者的泄露程度.具体赋值过程见第 3.2 节.

(12) 安全级(security level),简记为 *BP.sl*,属于判决语义.该定义用于描述后门隐私信息的动态安全级别,根据以上基本定义及上层语义通过多属性决策方法聚合生成,具体生成规则见第 3.3 节.

(13) 泄露阈值(leakage threshold),简记为 *BP.lt*,属于判决语义.该定义用于描述后门隐私信息泄露场景下的动态安全级别,可用于度量后门隐私动态安全变化与泄露场景之间的关联性,具体生成规则见第 3.3 节.

3.2 上层语义生成规则

上层语义主要从后门隐私的泄露场景中抽取若干特征聚合生成,包括 4 种:逆向痕迹(*BP.rt*)、流量检测(*BP.fm*)、静态泄露度(*BP.sld*)、动态泄露度(*BP.dld*).逆向痕迹和流量检测的赋值用于评估后门隐私在离线与在线环境中被破解与利用的可能性,而静态与动态泄露度则聚合若干基本定义及以上两种上层语义而成,综合评价后门信息在静态二进制结构及动态数据交互中的泄露程度.具体赋值过程如下.

3.2.1 逆向痕迹生成规则

离线的逆向分析是后门隐私泄露的主要途径之一,但目前尚未出现设备离线逆向分析的形式化痕迹检查方法,主要由于攻击者进行逆向分析时只要针对生产环境中同类设备或软件展开,不需要专门针对生产环境中在线设备进行逆向,因此痕迹检验的范围较广,难度较大.本文的逆向痕迹主要依据工业物联网中设备层固件及中间层软件的结构化特点生成,在目标生产环境中随机选取目标软件或固件中的二进制代码块,依次进行 x86 指令集下及其他指令集下的跳转指令扩展,而后插入特征化污点标记,最后完成代码空间平衡后结束重编程过程.通过生产运行中定时化的标记抽取来校验目标系统的一致性及完整性,达到痕迹检查的目的,具体如算法 1 所示.

算法 1. 逆向痕迹生成算法.

Input: code_block in bin file of software and firmware //软件或固件中的二进制文件结构

Output: BP.rt //逆向痕迹检测结果

1. if BP.nl == ml //面向中间层的组态软件
2. for random code_block *k* in reversed bin file //随机选择二进制中代码块
3. {
4. if useful space is not enough, goto code_block *k*+1; //判定代码空间是否足够可用
5. code_block k^J=Extended_JMP(*k*); //对目标代码块进行 JMP 扩展
6. code_block k^T=Marked_Taint(k^J,$Taint^k$); //对扩展后代码块进行污点标记
7. reach the balance of remained code block space; //完成代码空间平衡
8. }
9. for code_block *k*+1, goto 4; //对下一代码块进行处理
10. endif
11. if BP.nl == dl //面向设备层的固件
12. for random code_block *i* in reversed bin file //随机选择二进制中代码块
13. {
14. if useful space is not enough, goto code_block *i*+1; //判定代码空间是否足够可用
15. code_block i^J=Extended_Bxx(*i*); //对目标代码块进行 Bxx 跳转扩展
16. if Bxx.Distance>Distance(code_block), goto code_block *i*+1; //判决跳转距离是否在合法范围内
17. code_block i^T=Marked_Taint(i^J,$Taint^i$); //对扩展后代码块进行污点标记
18. reach the balance of remained code block space; //完成代码空间平衡
19. }
20. for code_block *i*+1, goto 4; //对下一代码块进行处理
21. endif
22. SetTime(Checked_Taint(*k*,$Taint^k$,srand(k))) //定时地对随机的代码段中的污点标记进行检查
23. if all taints are original, BP.rt=nre //若所有污点均保持不变,表示目标二进制未被逆向破解
24. if some taints are modified, BP.rt=sre //若部分污点被修改,表示目标二进制部分被逆向破解
25. if all taints are modified, BP.rt=fre //若所有污点被修改,表示目标二进制全部被逆向破解

3.2.2 流量检测生成规则

流量检测通常用于审计设备或软件所面临的外部安全威胁,常用的检测方法是根据数据包的格式、数量、内容等特征进行匹配分析,最终捕获恶意攻击行为并溯源恶意攻击者.本文的流量检测重点针对工业物联网中设备层的节点之间通信过程及中间层软件与设备层节点的通信过程部署深度数据包检测,首先判决是否存在主机端操作系统渗透测试数据,而后检测是否存在默认共享的渗透测试数据,还需检测是否存在反弹连接,满足

以上 3 个条件后,针对攻击者释放的文件类型进行分类处理,若释放的是固件相关文件,则基于改进后的 Bloom 算法匹配检查固件通信的数据包中是否包含后门数据内容,根据检查结果对 *BP.fm* 赋值.具体如算法 2 所示.若释放的是 PE 格式文件,同样基于改进的 Bloom 算法匹配检查固件与软件之间的数据包,完成 *BP.fm* 赋值.具体过程也如算法 2 所示.

算法 2. 流量检测生成算法.

Input: Pack(x,x) between different nodes in device layer, Pack(x,y) between middle layer and device layer
//输入:固件层节点之间、中间层与固件层之间的数据包;

Output: *BP.fm* //输出:流量检测结果.

```
1. if OS_Judgement(Pack(x,x),Pack(x,y))>0   //是否存在操作系统渗透测试
2.   if default_Share(Pack(x,x),Pack(x,y))>0   //是否存在默认共享渗透测试
3.     if rebound_Link(Pack(x,x))>0   //是否存在反弹连接
4.       if released_bin(Pack(x,x)) is firmware file   //若释放文件为固件相关文件
5.         ImprovedBloomMatch(BP.dc,Pack(x,x))==1, BP.fm=ebl
6.         ImprovedBloomMatch(BP.dc,Pack(x,x))==0, BP.fm=nebl
7.         ImprovedBloomMatch(BP.dc,Pack(x,x))∈(0,1), BP.fm=etbl
8.       endif
9.       if released_bin(Pack(x,y)) is PE file   //若释放文件为 PE 文件
10.        ImprovedBloomMatch(BP.dc,Pack(x,y))==1, BP.fm=ebl
11.        ImprovedBloomMatch(BP.dc,Pack(x,y))==0, BP.fm=nebl
12.        ImprovedBloomMatch(BP.dc,Pack(x,y))∈(0,1), BP.fm=etbl
13.      endif
14.    endif
15. endif
```

3.3 判决语义生成规则

判决语义主要包括静态与动态泄露度、动态安全级、安全阈值 4 种,此类语义主要通过基本定义与上层语义聚合而成,从多角度评估后门隐私在静态二进制结构及动态数据流向中的泄露程度.因此,本文引入多属性决策进行判决,该决策是在具有相互冲突、不可判决的多个属性情况下取得最优解决方案的基本方法,具体来看,根据对象属性特征的区别分成定量及定性信息的多属性决策评估,本方案中各属性均有量化标准,因而采用定量信息的多属性决策模型.同时,本方案需要基于定量多属性决策生成综合指标用以判决实时敏感信息动态流向中的敏感度,并与敏感度阈值比较,从而衡量其泄露程度与风险.因此选择了理想优基点法.其原始算法以理想解和反理想解为参照基准,其中每一属性中极大值聚合生成理想优基点,而极小值聚合生成反理想优基点,采用欧几里德距离计算决策方法与两者之间的差值,以此评价决策的优劣.但针对本文中有限的属性集合,且各属性值之间不确定的关联关系,使用传统算法难以保证决策结果的正确性.灰色关联分析是挖掘数据内部不确定关联关系的有效方法之一,其基本思想是对数据序列曲线几何形状的相似度进行比较分析,以曲线间相似程度大小作为关联程度的衡量尺度.灰色关联分析对关联关系不清晰或者根本缺乏事物关联原型的灰关系序列化、模式化,使关系量化、显化,能为复杂系统的建模提供重要的技术分析手段.因此本方案引入一种基于灰色关联度的理想优基点决策方法,以此生成 4 种判决语义.

3.3.1 静态泄露度生成规则

静态泄露度的生成主要根据多种基本定义及逆向痕迹聚合所得,首先根据 10 点标度法定义各种基本定义及逆向痕迹对静态泄露度的影响因子 *Factor*(*BP.A*),而后根据极差变换法对影响因子进行去量纲化处理.通过整合静态泄露度取值各种相关属性建立决策矩阵,计算设备层及中间层的理想与反理想优基点,最终求得灰色欧几里德距离作为静态泄露度.具体过程如下.

(1) 影响因子计算

$$BP.A_{sld}=\left(BP.nl,BP.aa,BP.ds,BP.dc,BP.df,BP.rt\right),$$

$$Factor_{BP.nl}=\left(0.50_{BP.nl==ml},0.99_{BP.nl==dl}\right),$$

$$Factor_{BP.aa}=\left(0.33_{BP.aa==nr\text{-}nw},0.67_{BP.nl==r\text{-}nw},0.99_{BP.nl==r\text{-}w}\right),$$

$$Factor_{BP.ds}=\left(0.33_{BP.ds==ct},0.67_{BP.ds==ot},0.99_{BP.nl==pt}\right),$$

$$Factor_{BP.dc}=\left(0.50_{BP.dc==niv},0.99_{BP.dc==iv}\right),$$

$$Factor_{BP.df}=\left(0.50_{BP.df==prc},0.99_{BP.df==rc}\right),$$

$$Factor_{BP.rt}=\left(0.33_{BP.rt==nre},0.67_{BP.rt==sre},0.99_{BP.rt==fre}\right).$$

如上所示,静态泄露度相关的属性包括:网络层级、地址属性、数据格式、数据内容、数据功能以及逆向痕迹,根据 10 点标度规则及属性中各取值对静态泄露的安全影响程度给予赋值.

(2) 去量纲化处理

$$Factor^{s}(BP.A)=\frac{F(BP.A)-F^{\min}(BP.A)}{F^{\max}(BP.A)-F^{\min}(BP.A)}.$$

根据极差变换法对各属性的影响因子进行去量纲化处理.

(3) 计算理想与反理想优基点

$$\boldsymbol{De}^{i\Delta}=(0.99,0.99,0.99,0.5,0.5,0.67),$$

$$\boldsymbol{De}^{i\nabla}=(0.50,0.33,0.33,0.50,0.50,0.33).$$

根据理想与反理想优基点的定义,结合静态泄露最大及最小化的场景判决,为各影响因子赋值,取得理想及反理想优基点.

(4) 计算灰色欧几里德距离

$$\exists\ x^{i}=BP_{(i)},\ \ x_{j}^{i}=BP_{i}.A.$$

生成决策矩阵:

$$D_{j}^{i}=(x_{1}^{i},...,x_{6}^{i}),\ \ i\in[1,N],j\in[1,6].$$

生成一个决策属性对应的决策权重:

$$w_{j}^{i}=(w_{1}^{i},...,w_{6}^{i}),\ \ i\in[1,N],j\in[1,6].$$

计算后门出现点处与理想和反理想优基点的距离:

$$d_{i}^{\Delta}=\sqrt{\sum_{k=1}^{6}\left(w_{k}^{i}\times Factor^{s}(x_{k}^{i})-De_{k}^{i\Delta}\right)^{2}},\ \ d_{i}^{\nabla}=\sqrt{\sum_{k=1}^{6}\left(w_{k}^{i}\times Factor^{s}(x_{k}^{i})-De_{k}^{i\nabla}\right)^{2}}.$$

计算灰色关联系数矩阵:

$$r_{j}^{i\Delta}=\frac{\min\limits_{i}\min\limits_{j}\left|De_{j}^{\Delta}-D_{j}^{i}\right|+\varepsilon_{sld}\cdot\max\limits_{i}\max\limits_{j}\left|De_{j}^{\Delta}-D_{j}^{i}\right|}{\left|De_{j}^{\Delta}-D_{j}^{i}\right|+\varepsilon_{sld}\cdot\max\limits_{i}\max\limits_{j}\left|De_{j}^{\Delta}-D_{j}^{i}\right|},$$

$$r_{j}^{i\nabla}=\frac{\min\limits_{i}\min\limits_{j}\left|De_{j}^{\nabla}-D_{j}^{i}\right|+\varepsilon_{sld}\cdot\max\limits_{i}\max\limits_{j}\left|De_{j}^{\nabla}-D_{j}^{i}\right|}{\left|De_{j}^{\nabla}-D_{j}^{i}\right|+\varepsilon_{sld}\cdot\max\limits_{i}\max\limits_{j}\left|De_{j}^{\nabla}-D_{j}^{i}\right|},$$

其中,ε_{sld} 为计算静态泄露度的分辨系数,由 10 点标度及后门泄露场景分析可知:

$$\varepsilon_{sld}+\varepsilon_{dld}=1,\ \varepsilon_{sld}\times 2=\varepsilon_{dld},\ \varepsilon_{sld}=0.33,\ \varepsilon_{dld}=0.67.$$

而后计算灰色关联度:

$$r^{i\Delta}=\frac{1}{n}\sum_{j=1}^{n}r_{j}^{i\Delta},\ r^{i\nabla}=\frac{1}{n}\sum_{j=1}^{n}r_{j}^{i\nabla}.$$

随后计算灰色理想优基点及灰色反理想优基点:

$$s^{i\Delta}=\alpha\cdot d_i^{\nabla}+\beta\cdot r^{i\Delta},\ s^{i\nabla}=\alpha\cdot d_i^{\Delta}+\beta\cdot r^{i\nabla},\ \alpha+\beta=1.$$

最终由灰色欧几里德距离来生成静态泄露度:

$$BP_i.sld=\frac{s^{i\Delta}}{s^{i\Delta}+s^{i\nabla}}.$$

3.3.2 动态泄露度生成规则

动态泄露度则根据基本定义及流量检测聚合生成,同样包括影响因子计算、去量纲化处理、理想与反理想优基点计算、灰色欧几里德计算等过程,主要区别在于:第一,动态泄露度生成的相关属性包括:网络层级、数据内容、操作主体、数据功能、触发方式、流量检测,主要涉及后门信息在动态流向中的属性特征;第二,根据动态泄露度各相关属性取值影响因子、计算理想与反理想优基点;第三,引入动态泄露度的分辨系数计算灰色关联系数矩阵;第四,最终根据灰色理想与反理想优基点计算欧几里德距离生成动态泄露度 $BP_i.dld$.限于篇幅,不再描述重复部分的计算过程.

3.3.3 安全级生成规则

安全级的生成需要综合考虑后门隐私在静态及动态两种情况下的泄露度,且通过两种场景的对比分析可知,攻击者往往首先进行静态后门挖掘,而后再部署动态后门利用,因此后门泄露的感知需考虑静态泄露度与动态泄露度之间的关联性.同样基于灰色关联分析定义关联指数:

$$\delta=\alpha\cdot s_{sld}^{i\nabla}+\beta\cdot s_{dld}^{i\nabla},\quad \alpha+\beta=1.$$

由于静态泄露之后通常伴随动态泄露场景,因此,基于决策偏好程度对静态泄露度及动态泄露的反理想优基点进行加权计算,生成关联系数.最终的安全级在参考静态泄露及动态泄露的情景特征后,其生成过程如下所示:

$$BP_i.sl=\frac{BP_i.sld+BP_i.dld}{BP_i.sld+\delta\cdot BP_i.dld}.$$

3.3.4 安全阈值生成规则

安全阈值生成取决于后门隐私同时满足静态及动态泄露条件下安全级的取值,而此类场景中安全级的值以关联系数σ趋近于反理想优基点,因此可知敏感度阈值为

$$\sigma=\frac{r^{i\nabla}}{r^{i\Delta}+r^{i\nabla}},$$

$$d_T^{\Delta}=\sqrt{\sum_{k=1}^{6}\left(w_k^i\cdot\sigma\cdot De_k^{i\nabla}-De_k^{i\Delta}\right)^2},$$

$$d_T^{\nabla}=\sqrt{\sum_{k=1}^{6}\left(w_k^i\cdot\sigma\cdot De_k^{i\nabla}-De_k^{i\nabla}\right)^2},$$

$$BP_i.ST=\frac{d_T^{\nabla}}{d_T^{\Delta}+d_T^{\nabla}}.$$

4 实 验

4.1 原型系统设计

为验证该泄露感知方法的正确性并测试其有效性,拟针对工业物联网环境部署并实现后门泄露感知的原型系统 BPLeakDetection.该系统基于模拟的工业物联网环境实现,面向中间层及设备层挖掘并定位后门隐私信息,通过静态逆向分析及动态污点跟踪为各相关属性赋值,而后基于算法 1 和算法 2 生成逆向痕迹及流量检测结果,最终基于多属性决策方法计算目标后门隐私的安全级与安全阈值.随后通过在线流量监控与离线逆向监控综合分析目标后门各个出现点,通过动态安全级与安全阈值的比对综合感知后门泄露的程度,并结合实时场

景的具体分析验证泄露感知的有效性.具体流程如图 2 所示.

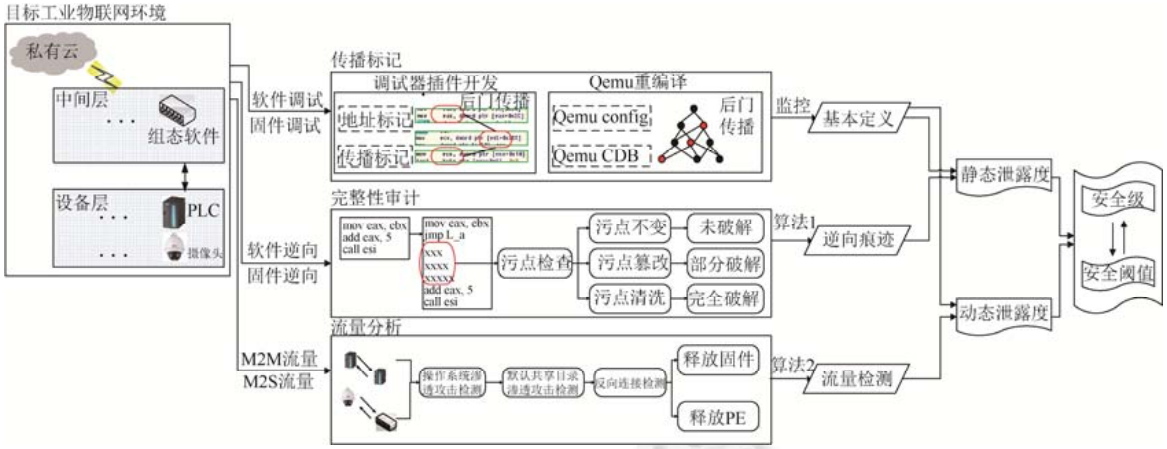

Fig.2 Control flow of prototype in BPLeakDetection

图 2 BPLeakDetection 原型系统流程图

4.2 实验环境搭建

鉴于工业物联网目前尚处于产品研发及行业推广阶段,标准化模型及设备配置还未统一,本文为达到精确仿真生产环境的目标,主要选取工业控制系统及物联网环境中通用化较强的一线硬件与软件设备进行实验,并结合部分已知后门泄露的生产设备搭建环境,方便验证原型系统的泄露感知效果.具体设备型号见表 1.

Table 1 Experiment environment

表 1 实验环境

设备层级	设备种类	设备厂商	设备型号
私有云层	ECS	阿里云	windows 1.4.x
	RDS	阿里云	windows 3.7.x
	VPC	阿里云	windows 2.2.x
	ESS	阿里云	windows 3.512.x
设备层	摄像头	TP-Link	IPC213/323/533
		Hikvision	DS2CD3T-10/25/35
		Le-orange	TP1/TP1S/TP1+/Q
	智能电表	Schneider	ION8600/PM810/EN4
		Siemens	PAC3100/3200/4200
		ABB	EMplus/EM/EM-M
	PLC	Schneider	NEZA/Quantum/Micro
		Siemens	S7 200/300/1200
		Omron	CPM1A/CP1L/CP1H
		Rockwell	1756/1769/1794
	工业路由器	TP-Link	TL WR/WVR/WAR
		DLink	DI 7002/7200/8003
		Tenda	AC6/AC18/i12
中间层	摄像头组态软件	TP-Link	TP-LINKsecurityV2.x
		Hikvision	SADPTool V2.3.x
		Le-orange	LeChange V1.10.x
	智能电表组态软件	Schneider	IONsetup V3.x
		Siemens	Powerconfig V2.x
		ABB	ABB V3.7.x
	PLC 组态软件	Schneider	Unity Pro V8.x
		Siemens	winCC V7.x
		Omron	CXProgrammer V9.x
		Rockwell	RSLogix V5000.x
	工业路由器组态软件	TP-Link	Web development
		DLink	DI-100.x
		Tenda	Tenda V15.03.4.x

如表 1 所示,实验围绕工业物联网的典型 3 层结构构建仿真环境,私有云端主要基于工业阿里云环境搭建,因此选择阿里云中 4 个主要部件构造私有云通信端.考虑到工业物联网应用场景繁多且设备型号多样,实验选择工业控制设备 PLC、工业视频采集设备摄像头、工业电量计量设备智能电表及工业路由器完成设备层环境搭建,并根据工业运行环境中物联网设备的使用率统计情况选定约 30 余种设备厂商的具体产品型号.中间层则根据设备层具体设备型号确定相关组态软件及操控软件,实现全局化良好兼容的目标.

4.3 实验数据分析

实验针对搭建的目标环境进行后门隐私挖掘与定位,对第 3.1 节中基本定义的各属性进行计算并赋值,进而根据第 3.2 节与第 3.3 节中的理论建模对上层语义和判决语义进行计算并赋值,最终在实验环境中根据安全级与安全阈值的比对完成泄露感知测试.实验共分析了包含设备层及软件层的设备固件及软件约 63 种版本,针对 27 种挖掘或已知的后门隐私进行数据统计与特征分析.

4.3.1 后门隐私信息定位

本文中目标后门隐私的来源主要包括:第一,自动化挖掘方法;第二,已知后门.自动化挖掘主要通过离线逆向分析与在线污点跟踪实现,基于字符串扫描、特殊权限函数扫描、函数调用关系收集等在离线状态下定位疑似后门,而后通过在线数据注入的方式在特殊权限函数处收集污点约束信息,确定完整的后门信息.本文通过自动化挖掘共获得不同设备或软件中的 11 种后门隐私信息,结合相同设备中已知的后门信息,扩展目标集合至 27 种.表 2 给出目标集合中部分后门的对比信息,针对相同设备或软件中的自挖掘与已知后门,分别比较其数据内容、开启权限、数据格式、触发方法等特征,为基本定义计算提供数据支持.鉴于挖掘所得后门尚未曝光,表 2 中隐去相关后门数据内容的核心部分(以 x 代替).

Table 2 Part of elements in target backdoor privacy set

表 2 部分目标后门隐私集合

来源	位置	内容	权限	格式	触发方式
挖掘	IPC 系列摄像头固件	Qxxxxk	远程调试用户	明文	主动
已知	IPC 系列摄像头固件	IPC7185/IPC	管理员用户	明文	主动
挖掘	ION8600 智能电表固件	akcmibxxxxxxxxxxx_qs	开启特殊端口	密文	被动
已知	ION8600 智能电表固件	8600kt/8600kcmti	开启特殊端口	明文	主动
挖掘	NOE771 PLC 固件	sysdxxg/kvxxxic	远程控制	明文	被动
已知	NOE771 PLC 固件	bbddRdkm9/bbddrdkm9	远程控制	混淆	被动
挖掘	TP-Link TL 路由器软件	axxxxxp/axxxxe	开启特殊端口	明文	主动
已知	TP-Link TL 路由器软件	root/5up	管理员用户	明文	主动
挖掘	Schneider IONsetup 电表软件	fwupxxxde/upxxxdefw	开启特殊端口	混淆	被动
已知	Schneider IONsetup 电表软件	qcm77101/fikdfcsd	管理员用户	明文	主动

如表 2 所示,选取目标集合中 10 种后门信息进行对比分析,分别取自摄像头固件、智能电表固件、PLC 固件、路由器软件及智能电表软件.从数据内容上看,已知后门的威胁较大,通常是以明文格式存在,数据内容较为简单,且可通过主动连接方式触发,权限涉及管理员较多.挖掘所得后门权限不一,包括管理员、远程调试、开启特殊端口等等,与已知后门相比权限略低,且数据格式存在混淆或密文,隐藏较深.同时,对比设备层固件与中间层软件可知,软件后门内容较为简单,主动触发方式较多,容易利用,且在软件中开启特殊端口或管理员权限后可进一步对下控固件进行恶意操作,威胁程度较大.

4.3.2 基本定义取值分析

原型系统对收集所得后门信息进行分类、聚类以及特征分析,对应后门隐私在全生命期的不同出现点处的各种特征计算基本定义并完成赋值,最终对目标后门集合中的全部元素进行基本定义取值的统计分析.如表 3 所示,共统计 27 种后门信息在全局生命期中 137 个出现点的 7 种基本定义取值,并分析取值分布特征,总结分布原因.

Table 3 Statistic analysis for basic definition values

表 3 基本定义取值统计分析

基本定义	取值	数量(比例)	说明
BP.nl	dl ml	85(62.0%) 52(38.0%)	设备层中目标后门较多,主要由于设备层后门利用较中间层后门利用简单,且不易被检测
BP.aa	nr-nw r-nw r-w	21(15.3%) 103(75.2%) 13(9.5%)	地址属性大都为可读不可写状态,即可利用后门,但不可植入后门.最少情况为可读可写状态,即软件或固件为植入后门开放接口,为最差安全性
BP.ds	pt ct ot	64(46.7%) 27(19.7%) 46(33.6%)	数据格式半数为明文,即字符未加密,无需解密直接可利用.其次为混淆字符,部分内容置乱,需去混淆再利用.最少为加密字符,解密后可利用
BP.dc	iv niv	94(68.6%) 43(31.4%)	可见内容作为真实后门输入设备或软件中;不可见内容表示后门不能直接开启特殊权限,涉及多次、多点关联输入
BP.do	ad su at	61(44.5%) 54(39.4%) 22(16.1%)	数据主体主要为管理员,其通过检查输入源白名单可确定目标管理员通过后门实现远程控制;同理可检测目标远程用户利用后门实现控制.其他为后门泄露情况下的恶意攻击者,具体场景来自实验模拟过程
BP.df	rc prc	94(68.6%) 43(31.4%)	数据功能中 68%为开启远程控制权限,其他为开启部分远程控制权限,包括特殊端口开放、特殊函数开放等
BP.tm	at pt	79(57.7%) 58(42.3%)	触发方式中主动连接方式居多,易用性强.其他为被动连接方式,目标固件或软件接受特征数据包后可触发

4.3.3 上层语义取值分析

对上层语义包括逆向痕迹及流量检测两部分,首先收集在线及离线的场景信息,基于算法 1 与算法 2 生成两种语义的取值,并对比同一后门在不同出现点处、同一固件中不同后门、同一软件中不同后门的上层语义取值特征.具体如图 3 所示.

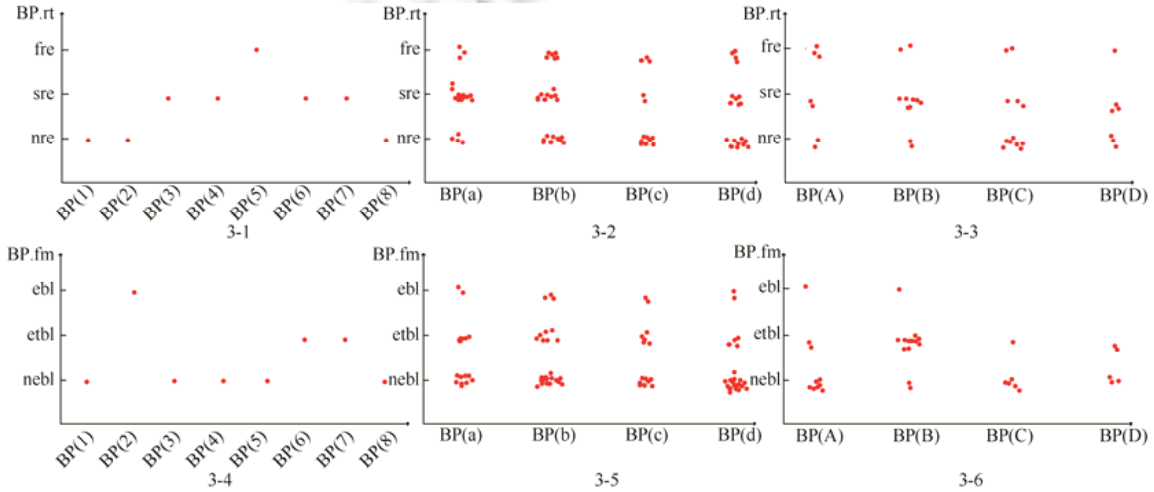

Fig.3 Statistic analysis for upper semantic values

图 3 上层语义取值统计分析

图 3 中选择同一后门隐私(IPC 系列摄像头中:IPC7185/IPC)进行逆向痕迹与流量检测的统计分析,如图 3-1 和图 3-4 所示;选择同一固件 NOE771PLC 中的 4 种不同后门:sysdxxg、kvxxxic、bbddRdkm9、bbddrdkm9(依次为 BP(a)、BP(b)、BP(c)、BP(d))作为分析对象对比分析逆向痕迹与流量检测取值分布,如图 3-2 与图 3-5 所示;选择同一软件 TP-Link TL 路由器中 4 种不同后门:axxxxxp、axxxxe、tl49xxxxk、root(依次为 BP(A)、BP(B)、BP(C)、BP(D))对比分析逆向痕迹与流量检测分布,如图 3-3 与图 3-6 所示.由图 3-1 可知,IPC 后门共 8 个出现点,只有第 5 个出现点处逆向痕迹为完全逆向破解,其他共 4 个点处为半逆向破解,其余为未逆向破解.可知在检测过程中半数以上加入的原始污点被破坏,而全部原始污点被破坏的情况下可认定目标固件已被攻击者逆向分析并去除安全保护的混淆过程.对比分析图 3-2 可知,在线流量检测结果中后门利用情况较少,只在第 2 个出现点处检测结果显示存在后门利用,在点 6、点 7 中显示存在后门探测利用,即不完全正确的后门利用过程,其

余为非后门利用过程.对比分析可知,对同一后门而言,离线逆向分析是在线利用攻击的基础,攻击者通常会在确保离线逆向破解成功的基础上谨慎部署在线后门利用,保证后门利用一次成功.对比同一固件中的不同后门,图 3-2 统计分析了 4 种后门在全局生命期中各个出现点处 *BP.rt* 的取值特征,由分布可知,取值显示完全逆向破解的情况总体偏少,而部分逆向破解的情况居多,未逆向破解的情况最多.统计中主要针对目标设备的常规运行状态完成数据采集,因此,表示常规运行状态下插入的原始污点未被修改或部分被修改,而完全被修改的情况较少,可认定被彻底破解.对比分析图 3-5 中数据可知,在线流量检测中可获得的后门攻击场景更少,绝对情况下可认定为后门利用的在线攻击在常规运行状态中基本没有,均来自实验构造的模拟过程.而中间层软件中的后门与设备层固件相比呈现更少的数据采集特征,由图 3-3 及图 3-6 的对比可知,在离线逆向痕迹及在线流量检测中的完全破解或在线攻击数明显减少,部分数据采集为 0.因此可证明软件内部的后门利用较固件来说频率较低,数据量不大,主要原因在于软件可实现较好的安全防护与较快的安全补丁,而固件层相关技术较匮乏.

4.3.4 判决语义取值分析

判决语义共包括 4 种:静态泄露度、动态泄露度、安全级与安全阈值.根据第 3.3 节中的判决语义生成规则可知,静态泄露度由网络层级、地址属性、数据格式、数据内容、数据功能以及逆向痕迹聚合生成,而动态泄露度由网络层级、数据内容、操作主体、数据功能、触发方式以及流量检测聚合而成,而安全级与安全阈值则基于静态与动态泄露度的反理想优基点聚合而成.如图 4 所示,统计分析目标后门集合中的各种后门隐私在各个出现点处静态泄露度与动态泄露度的取值,并通过安全级与安全阈值的比较实现后门泄露场景的感知.

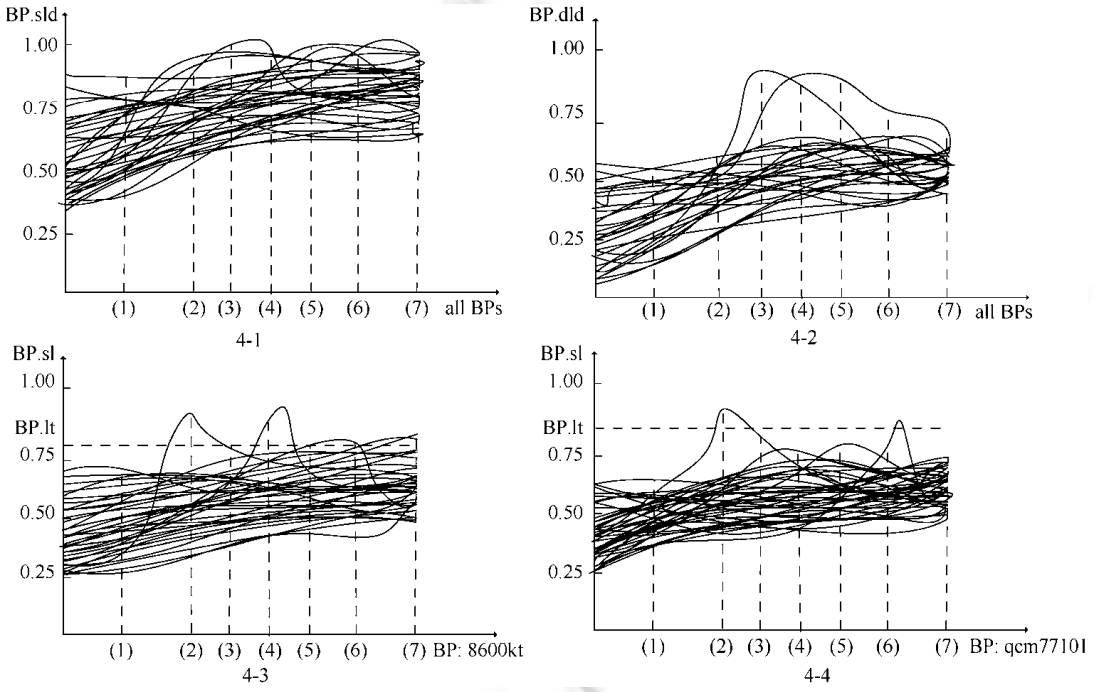

Fig.4 Analysis for judgement semantic

图 4 判决语义取值分析

图 4-1 及图 4-2 中选择目标后门隐私集合中全部 27 个元素进行静态泄露度及动态泄露度的取值统计分析,并对每种后门隐私选择 7 个出现点进行取值采样.静态泄露度的取值分布为[0.2743,0.9635],整体分布取值较大,主要集中在[0.5,0.85]的取值区间,尤其是在数据生命期末期,由于后门在静态文件结构回收或销毁操作的缺失导致静态泄露度增大,取值集中在[0.8,1]的区间.从取值变化上分析,取值曲线的斜率较大,会在某些出现点处呈现突然增大的情况,如 IPC7185 摄像头后门从第 2 出现点到第 3 出现点的过程中呈现静态泄露度取值从 0.7793~0.9446 的突变,而后又在第 4 出现点下降至 0.796 1,证明在出现点 3 处静态泄露情况严重,结合该位置处

的逆向痕迹检查结果可知,该后门在第 3 出现点处呈现被攻击者逆向破解的泄露场景.同理,分析 TP-Link 路由器后门 tl49xxxxk 可知,在出现点 5、点 6 的取值为 0.884 3 与 0.860 6,较之前与之后的各个出现点呈现出静态泄露度的陡增,结合逆向痕迹检测结果可以确定在出现点 5 与点 6 中被攻击者逆向破解.对比分析图 4-2 中的动态泄露度,取值分布为[0.1226,0.8420],整体分布取值较小,主要集中在[0.1,0.70]的取值区间,在全局生命期中的取值过程较平缓,斜率较小.从取值变化的角度来看,个别后门隐私的出现点会出现动态泄露度陡增,如 NOE771 PLC 固件后门 sysdxxg 与 siemens 电表组态软件 Powerconfig V2.x 后门 kvgxxxxxx27,分别在第 3、第 4 出现点呈现取值陡增,分别由 0.611 6 增大到 0.842 0,以及从 0.772 5 增大到 0.824 8.结合流量检测结果分析可知,这两种后门分别在这两个出现点被检测到有后门利用的恶意流量.

图 4-3 与图 4-4 选择 ION8600 智能电表固件中的后门 8600kt 及 Schneider IONsetup 电表软件中的后门 qcm77101 进行安全级与安全阈值的取值统计分析,随机选择 30 种系统运行状态,针对 7 个后门出现点进行取值采样.由图 4-3 可知,30 种系统运行状态中后门 8600kt 的安全级分布为[0.2606,0.8933],主要集中于[0.3,0.6]的取值区间,安全级在全局生命期中变化不大,且取值稳定,结合静态泄露度、动态泄露度及两种定义的反理想优基点取值分析可知,该后门相关的静态、动态泄露度的反理想优基点较大,导致关联系数取值较大,因此导致安全级取值较小,稳定在 0.5±0.2 左右.在第 7 运行状态与第 15 运行状态中,分别在第 2 与第 4 出现点呈现安全级突然增大,结合相关定义取值分析,第 7 运行态下的第 2 出现点处存在逆向痕迹,则静态泄露度出现突然增大的情况,在其他值取值基本保持不变的情况下导致安全级突然增大.而第 15 运行态下的第 4 出现点则存在动态泄露度增大的情况,该点处可判决为存在目标后门正相关的在线恶意流量,与之前的第 3 出现点的采样结果呈现出明显区别,因而在该点处表现为安全级突然增大.且两处的安全级取值均已超出安全阈值 0.815 1,可实现基于安全级比较的泄露感知.对比分析图 4-4 中的 Schneider IONsetup 电表软件中的后门,在第 12 运行状态的第 2 出现点及第 26 运行状态的第 6 出现点呈现安全级陡增,分别取值为 0.910 4 和 0.873 6,均超出安全阈值 0.869 6.进一步分析关联定义取值可知,第 12 运行状态的第 2 出现点处静态泄露度与动态泄露度取值均偏大,深入分析逆向痕迹与流量检测结果,逆向痕迹为半逆向破解,且存在后门探测利用.总结以上取值可以判定该场景中存在攻击者离线逆向破解与在线后门利用,但利用的准确度不高,证明目标后门并未被完全破解.而第 26 运行态的第 6 出现点处安全泄露度取值较大,具体分析流量检测结果发现,流量检测中存在外部主体的后门利用远程连接流量,可知在半逆向破解的情况下,后门已被攻击者成功利用实现远程控制.

4.4 对比分析

第一,安全性对比.为评估本方法的安全性,本文选择几种经典的访问控制模型进行对比分析[28],结合本文方法在基本定义、建模分析、威胁感知等方面的特点,共选择 3 种访问控制模型进行比较:强制访问控制模型 MAC、基于角色的访问控制模型 RBAC、基于属性的访问控制模型 ABAC.选择依据是:本文方法在本质上可理解为围绕后门隐私信息的数据泄漏感知方法,而经典的访问控制模型则是围绕目标信息构建以主体、客体的对象的访问策略集合与访问控制模型,因此有较大的可比性.同时,依据对目标信息进行属性定义的方法各有不同,选择以上 3 种典型的访问控制模型进行比对.

如表 4 所示,从模型本质上看,MAC 的特点在于主体到客体访问规则的强制性,BRAC 则定义了与角色关联的访问策略,ABAC 拓展到主体、客体、环境的属性范畴,而本文方法围绕工业物联网中的隐私信息建立了主体、客体、环境的多级属性定义,进而提出泄露感知方法.从属性定义上看,从 MAC 到 ABAC 的属性定义呈现出完整性优化的趋势,本文中的属性对象在 ABAC 进一步拓展,加入了面向场景的属性定义.从属性聚合的角度分析,前 3 种访问控制模型中只有 ABAC 涉及到部分的属性聚合过程,但并不包括全部的属性定义.本文则通过多属性决策方法将所有安全相关属性聚合为统一值,并利用灰色关联分析去除了属性间的不确定关系.从威胁关联和场景判决上看,由于访问控制策略重在事前防范,策略部署及实施的目标在于防范目标数据的恶意访问,因此均不涉及这两方面的功能.而本文方法针对泄露场景的实时特征抽取威胁关联关系,准确判决是否为泄露场景及泄露场景中泄露程度、泄露途径等具体泄露特征.

第二,可用性对比.鉴于目前面向工业物联网环境的数据泄露模型较少,而本文的原型系统从本质上看也可

等价于一个动态污点插桩系统,因此,本文选择 3 种典型的污点插桩系统进行性能比对,经预备实验结果分析可知,这 3 种典型插桩工具可在目标仿真环境中兼容运行,包括设备层、中间层及私有云层,因此具体同构对比的前提条件;且插桩平台主要用于污点跟踪系统部署前的目标程序修改,因此插桩过程的兼容则代表污点跟踪的可行性.实验主要从插桩方式、指令集支持种类、性能损耗几个方面对比系统的优劣性.

如表 5 所示,从插桩方式上看,Pin 针对镜像进行插桩时采用整体镜像缓存插桩,属于粗粒度插桩方法;而 DynamoRIO 在指令级控制插桩精度,更利用插桩分析实现准确、高效的插桩方法;DynInst 则是基于探针技术实现插桩,控制粒度较细,但开销较大.综合分析以上经典方法,BPLeakDetection 选择指令级缓存插桩,平衡插桩精度与系统开销.从指令集支持范围来看,Pin 和 DynamoRIO 主要支持 Windows 环境下的 x86 指令集,而 DynInst 则扩展至 PowerPC 指令集,本文的 BPLeakDetection 系统需兼容固件及 Windows 软件环境,因此支持 x86、arm、PowerPC 这 3 种指令集.最后对比性能损耗,由于本文基于指令集插桩实现,且扩展至多种指令集,因此损耗大于 Pin 和 DynamoRIO 平台,但优于 DynInst 平台,损耗在 210%左右.由于工业物联网环境对实时性要求较高,因而在部署插桩的过程中会对系统运行造成一定影响,但在部署完成后的污点检查及完整审计中对系统的性能影响基本可以忽略不计,对实时性影响较小.

Table 4 Comparison analysis for security

表 4 安全性对比分析

	MAC	RBAC	ABAC	本文方法
模型本质	强制主体服从客体访问策略的访问控制模型	角色与访问策略关联的访问控制模型	基于主体、客体与环境的属性设计的访问控制模型	包含主体、客体、环境等多级属性定义的数据泄露感知方法
属性定义	定义主体与客体的普通属性	围绕角色的不同定义主体与客体的属性	围绕主体、客体、环境 3 方面定义属性	围绕主体、客体、环境、场景 4 方面特征定义多级属性
属性聚合	不涉及	不涉及	部分涉及	利用多属性决策聚合多种属性,利用灰色关联分析去除属性之间的不确定性
威胁关联	不涉及	不涉及	不涉及	基于真实泄露场景的特征抽取实现感知方法与泄露威胁的实时关联
场景判决	不涉及	不涉及	不涉及	针对实时发生的泄露场景完成特征聚类与泄露判决

Table 5 Comparison analysis for availability

表 5 可用性对比分析

对比系统	插桩方式	指令集支持	性能损耗(%)
Pin	镜像级缓存插桩	x86	135.4
DynamoRIO	指令级缓存插桩	x86	168.1
DynInst[29]	探针插桩	x86/PowerPC	246.3
BPLeakDetection	指令级缓存插桩	x86/arm/PowerPC	211.8

当前隐私数据保护或隐私泄露防御的实现方法较多,对比同类研究成果,之前的研究工作针对某种典型环境中的典型隐私或敏感数据进行隐私保护方法或泄露防御方法研究,主要从泄露事前角度出发,大都基于加密算法或访问控制模型设计并实现数据保护或泄露防御.本文则主要从实时后门隐私泄露场景出发,从泄露场景中提取目标数据及数据环境的相关特征,并依据多属性决策方法生成判决语义,以此判定目标后门信息在静态结构或动态流向上的泄露程度,因此与同类方法相比,本文方法与泄露场景关联度较强,可用于泄露事前防御或泄露事中拒绝,并为泄露事后取证提供数据支持.

5 结 论

本文面向工业物联网环境中的后门隐私实现了一种信息泄露的感知方法,通过定义后门的基本属性、上层语义及判决语义从后门的静态二进制结构及动态数据流向中检测泄露场景的特征,最终结合灰色关联分析与多属性决策生成动态安全级,通过与安全阈值的比对判决泄露场景的细粒度特征.最后,本文通过目标环境中的

多种后门信息验证该方法及原型系统的有效性,并测试系统性能开销.后继工作中希望能基于泄露感知进一步实现后门隐私的实时防护,或基于已知后门的泄露感知推演生成同类信息的泄露感知方法.

References:

[1] Ning HS, Xu QY. Research on global Internet of Things' developments and it's lonstruction in China. ACTA ELECTRONICA SINICA, 2010,38(11):2590–2599 (in Chinese with English abstract).

[2] Kang SL, Du ZY, Lei YM, Wang J. Overview of industrial Internet of Things. Internet of Things Technolegies, 2013,(6):80–82 (in Chinese with English abstract).

[3] Yang JC, Fang BX, Zhai LD, Zhang FJ. Reseach towards IoT-oriented universal control system security model. Journal on Communications, 2012,(11):49–56 (in Chinese with English abstract).

[4] Cherdantseva Y, Burnap P, Blyth A, Eden P, Jones K, Soulsby H, Stoddart K. A review of cyber security risk assessment methods for SCADA systems. Computers & Security, 2016,56(C):1–27.

[5] Ponomarev S, Atkison T. Industrial control system network intrusion detection by telemetry analysis. IEEE Trans. on Dependable & Secure Computing, 2016,13(2):1–13.

[6] Mitchell R, Chen IR. Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems. IEEE Trans. on Dependable & Secure Computing, 2015,12(1):16–30.

[7] Urbina DI, Giraldo JA, Cardenas AA. Limiting the impact of stealthy attacks on industrial control systems. In: Kruegel C, ed. Proc. of the ACM SIGSAC Conf. on Computer and Communications Security. New York: Academic Press, 2016. 1092–1105.

[8] Abera T, Asokan N, Davi L, *et al*. C-FLAT: Control-Flow attestation for embedded systems software. In: Jaeger T, ed. Proc. of the ACM SIGSAC Conf. New York: Academic Press, 2016. 743–754.

[9] Daming C, Manuel E, Maverick W, *et al*. Towards automated dynamic analysis for linux-based embedded firmware. In: Gunter CA, ed. Proc. of the Network and Distributed System Security Symp. University of California Press, 2016. 452–468.

[10] Eschweiler S, Yakdan K, Gerhards-Padilla E. discovRE: Efficient cross-architecture identification of bugs in binary code. In: Gunter CA, ed. Proc. of the Network and Distributed System Security Symp. University of California Press, 2016. 49–64.

[11] Ooi ST, Lorber B. Avatar: A framework to support dynamic security analysis of embedded systems' firmwares. In: Gunter CA, ed. Proc. of the Network and Distributed System Security Symp. University of California Press, 2014. 112–129.

[12] Wang JH, Liu CY, Fang BX. Survey on data preserving for the search of internet of things. Journal on Communications, 2016,37(9):142–153 (in Chinese with English abstract).

[13] Fernandes E, Jung J, Prakash A. Security analysis of emerging smart home applications. In: Peterson Z, ed. Proc. of the 37th IEEE Symp. on Security and Privacy. University of California Press, 2016. 312–328.

[14] Earlence F, Justin P, Amir R, *et al*. FlowFence: Practical data protection for emerging IoT application frameworks. In: Holz T, ed. Proc. of the 2016 USENIX Security Symp. Boca Raton: CRC Press, 2016. 207–225.

[15] Wan S, Li FH, Niu B, Sun Z, Li H. Research progress on location privacy-preserving techniques. Journal on Communications, 2016,37(12):124–141 (in Chinese with English abstract).

[16] Narain S, Vo-Huu TD, Block K, Noubir G. Inferring user routes and locations using zero-permission mobile sensors. In: Peterson Z, ed. Proc. of the 37th IEEE Symp. on Security and Privacy. University of California Press, 2016. 397–413.

[17] Peng H, Chen H, Zhang XY, Zeng JR, Wu YC, Wang S. Privacy-Preserving *k*-NN query protocol for two-tiered wireless sensor networks. Chinese Journal of Computers, 2016,39(5):872–892 (in Chinese with English abstract).

[18] Dai H, Yang G, Qin XL, Liu L. Privacy-Preserving top-*k* query processing in two-tiered wireless sensor networks. Journal of Computer Research and Development, 2013,50(6):1239–1252 (in Chinese with English abstract).

[19] Wang P, Wang D, Huang XY. Advances in password security. Journal of Computer Research and Development, 2016,53(10): 2173–2188 (in Chinese with English abstract).

[20] Zhang YX, He JS, Zhao B, Zhu NF. A privacy protection model base on game theory. Chinese Journal of Computers, 2016,39(3): 615–627 (in Chinese with English abstract).

[21] Zeng JR, Chen H, Peng H, Wu Y, Li CP, Wang S. Privacy preservation in mobile participatory sensing. Chinese Journal of Computers, 2016,39(3):595–614 (in Chinese with English abstract).

[22] Xue R, Ren K, Zhang YQ, Li H, Liu JQ, Zhao B, Zhu LH. Introduction for special issure of security research for cloud computing. Ruan Jian Xue Bao/Journal of Software, 2016,27(6):1325–1327 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5010.htm [doi: 10.13328/j.cnki.jos.005010]

[23] Xue R, Feng DG. The approaches and technologies for formal verification of security protocols. Chinese Journal of Computers, 2006,29(1):1–20 (in Chinese with English abstract).

[24] Sajid A, Abbas H, Saleem K. Cloud-Assisted IoT-based SCADA systems security: A review of the state of the art and future challenges. IEEE Access, 2016,4:1375–1384.

[25] Huberman BA. Ensuring trust and security in the industrial IoT—The Internet of Things (Ubiquity Symp.). Gastroenterology, 2002, 122(5):1235–1241.

[26] Viswanathan S, Tan R, Yau DKY. Exploiting power grid for accurate and secure clock synchronization in industrial IoT. In: Clack R, ed. Proc. of the Real-Time Systems Symp. Chicago: University of Chicago Press, 2017. 146–156.

[27] Hassanzadeh A, Modi S, Mulchandani S. Towards effective security control assignment in the industrial Internet of Things. In: Justin B, ed. Proc. of the Internet of Things. New York: Academic Press, 2015. 795–800.

[28] Li FH, Su M, Shi GZ, Ma JF. Research status and development trends of access control model. ACTA ELECTRONICA SINICA, 2012,40(4):805–813 (in Chinese with English abstract).

[29] Buck B, Hollingsworth JK. API for runtime code patching. Int'l Journal of High Performance Computing Applications, 2000,14(4): 317–329.

附中文参考文献:

[1] 宁焕生,徐群玉.全球物联网发展及中国物联网建设若干思考.电子学报,2010,38(11):2590–2599.

[2] 康世龙,杜中一,雷咏梅,张璟.工业物联网研究概述.物联网技术,2013,(6):80–82.

[3] 杨金翠,方滨兴,翟立东,张方娇.面向物联网的通用控制系统安全模型研究.通信学报,2012,(11):49–56.

[12] 王佳慧,刘川意,方滨兴.面向物联网搜索的数据隐私保护研究综述.通信学报,2016,37(9):142–153.

[15] 万盛,李凤华,牛犇,孙哲,李晖.位置隐私保护技术研究进展.通信学报,2016,37(12):124–141.

[17] 彭辉,陈红,张晓莹,曾菊儒,吴云乘,王珊.面向双层传感网的隐私保护 *k*-NN 查询处理协议.计算机学报,2016,39(5):872–892.

[18] 戴华,杨庚,秦小麟,刘亮.面向隐私保护的两层传感网 Top-*k* 查询处理方法.计算机研究与发展,2013,50(6):1239–1252.

[19] 王平,汪定,黄欣沂.口令安全研究进展.计算机研究与发展,2016,53(10):2173–2188.

[20] 张伊璇,何泾沙,赵斌,朱娜斐.一个基于博弈理论的隐私保护模型.计算机学报,2016,39(3):615–627.

[21] 曾菊儒,陈红,彭辉,吴垚,李翠平,王珊.参与式感知隐私保护技术.计算机学报,2016,39(3):595–614.

[22] 薛锐,任奎,张玉清,李晖,刘吉强,赵波,祝烈煌.云计算安全研究专刊前言.软件学报,2016,27(6):1325–1327. http://www.jos.org.cn/1000-9825/5010.htm [doi: 10.13328/j.cnki.jos.005010]

[23] 薛锐,冯登国.安全协议的形式化分析技术与方法.计算机学报,2006,29(1):1–20.

[28] 李凤华,苏铓,史国振,马建峰.访问控制模型研究进展及发展趋势.电子学报,2012,40(4):805–813.

沙乐天(1985－),男,江苏徐州人,博士,讲师,CCF 专业会员,主要研究领域为网络安全,物联网攻防.

孙晶(1985－),男,工程师,主要研究领域为通信网络技术,通信技术保障.

肖甫(1980－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为传感网,物联网.

王汝传(1943－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为物联网,网络安全.

陈伟(1979－),男,博士,教授,CCF 专业会员,主要研究领域为无线网络安全,移动互联网安全.