·热点与综述· 文章编号：1000-3428(2020)07-0036-07 文献标志码：A 中图分类号：TP393

基于工控系统功能码特征的同源攻击分析

王建华，陈永乐，张壮壮，连晓伟，陈俊杰

（太原理工大学 信息与计算机学院，太原 030024）

摘 要： IP 溯源是追踪攻击者源头的主要方法，工业控制系统（ICS）需要精确的 IP 溯源以提高其防护能力。现有 IP 溯源方法存在开销大、恶意 IP 所属组织识别效率低的问题。为此，通过采集和分析 ICS 蜜罐数据，提出一种基于工控协议功能码特征的同源攻击分析方法，以识别攻击行为相似的组织并提高 IP 溯源的效率和准确性。用工控协议功能码的粗粒度统计特征和细粒度序列特征来量化攻击行为，采用粗糙集和聚类模型分别对 2 类特征进行建模，在此基础上分析蜜罐数据中的同源攻击。实验结果表明，该方法具有较高的准确率和召回率，结合威胁情报后能够在蜜罐数据中发现包括 shodan 在内的 10 个恶意组织。

关键词： IP 溯源；工业控制系统；功能码序列；同源攻击分析；恶意组织

开放科学（资源服务）标志码（OSID）：

中文引用格式： 王建华，陈永乐，张壮壮，等. 基于工控系统功能码特征的同源攻击分析[J]. 计算机工程，2020，46(7)：36-42.

英文引用格式： WANG Jianhua，CHEN Yongle，ZHANG Zhuangzhuang，et al. Same origin attack analysis based on features of industrial control system function code[J]. Computer Engineering，2020，46(7)：36-42.

Same Origin Attack Analysis Based on Features of Industrial Control System Function Code

WANG Jianhua，CHEN Yongle，ZHANG Zhuangzhuang，LIAN Xiaowei，CHEN Junjie

（College of Information and Computer，Taiyuan University of Technology，Taiyuan 030024，China）

【Abstract】 IP traceback is one of the main methods of attack group identification. Industrial Control System（ICS）need accurate IP traceback to improve their self-protection. However，existing IP traceback methods are costly and inefficient in identification of the group a malicious IP belongs to. To address the problem，by collecting and analyzing the honeypot data of ICS，this paper proposes a same origin attack analysis method based on ICS function code features，so as to find out the attack group with similar attack behavior and improve the efficiency and accuracy of IP traceback. This method uses coarse-grained statistical features and fine-grained sequence features of industrial control function codes to quantify the attack behavior. Then the two kinds of features are modeled by using coarse set and clustering model. On this basis，the same origin attacks in honeypot data are analyzed. Experimental results show that the proposed method can use threat intelligence to discover more than 10 malicious groups including shodan in honeypot data with a high accuracy and recall rate.

【Key words】 IP traceback；Industrial Control System（ICS）；function code sequence；same origin attack analysis；malicious group

DOI： 10.19678/j.issn.1000-3428.0057547

0 概述

随着工业互联网技术的推广普及，工业控制系统（Industrial Control System，ICS）安全为国家安全和社会经济发展提供重要保障。然而，网络环境下的不安全因素对 ICS 构成极大威胁，2010 年的“震网病毒”事件以及 2015 年的“波兰航空公司黑客攻击”事件等，充分显示出 ICS 遭遇攻击后的严重危害

基金项目： 山西省自然科学基金（201701D111002，201601D021074）；山西省重点研发项目（201903D121121）。

作者简介： 王建华（1995—），男，硕士研究生，主研方向为物联网安全；陈永乐，副教授、博士；张壮壮、连晓伟，硕士研究生；陈俊杰，教授、博士。

收稿日期： 2020-03-02 **修回日期：** 2020-04-14 **E-mail：** chenyongle@tyut.edu.cn

性。因此,如何保障 ICS 安全是当今网络安全领域广泛关注的问题。攻击者溯源能够为 ICS 安全提供主动防护,蜜罐技术是实现攻击者溯源追踪的手段之一[1-2],通过蜜罐技术收集攻击者信息,分析攻击者的攻击行为、攻击特征等,能够对工控网络的组织和安全态势进行更准确的分析。目前,IP 溯源技术是攻击者溯源的主要手段,但是,IP 溯源需要对路由设备进行更改,存在资源开销大、识别精度低以及有效验证难等问题。此外,现有的攻击手段多数采用攻击代理来实施攻击行为[3],提高了 IP 溯源的难度。事实上,目前多数工控系统攻击是有组织、有规模的攻击行为,攻击者并非单一个体,对恶意组织进行全面溯源更有实际价值。然而,IP 溯源往往针对单个 IP 进行溯源,对于恶意 IP 所属组织的识别效率较低。

本文借鉴恶意代码常用的家族同源判断分析方法[4-5],将具有相同攻击者信息或相似攻击特征的攻击定义为同源攻击。使用基于粗、细粒度功能码特征结合的特征提取和建模方法来判定具有攻击特征的攻击者是否属于同一恶意组织,从而实现攻击者溯源。此外,针对蜜罐数据同一攻击源判定验证较难的问题,使用开源威胁情报库验证恶意 IP 所属组织[6-7]。

1　相关工作

攻击者溯源的研究主要依赖于 IP 溯源技术,IP 溯源技术作为网络安全主动防护的关键手段,主要包括概率包标记溯源法和日志溯源法 2 类。概率包分组标记技术[8]将标识信息(如 IP 地址)写入转发分组的头域,也即标记域中,然后受害者从收到的分组中找回标记信息,最终确定攻击路径。日志溯源法[9]是路由器在转发分组前记录下分组相关的信息并进行重构。然而,设备的修改以及日志格式的不统一,使得传统 IP 溯源技术成本开销高、误报率高、可操作性低,在工控领域难以实现。

工控蜜罐技术通过搜集信息来分析工控系统攻击者的行为,包括攻击方式、攻击手段等。Glastopf 项目发布了第一个开源工控蜜罐框架 Conpot[10],其实现了协议栈上的请求-应答交互。文献[11]提出了一种基于物联网设备的蜜网系统,提升了蜜罐的交互能力和仿真度。对蜜罐数据中攻击者行为进行分析依赖于攻击特征的提取,Honeycomb[12]作为 Honeyd 蜜罐的扩展模块,其提出了利用蜜罐捕获数据进行攻击特征提取的基础方法。Nemean 系统[13]改进了 Honeycomb,其将原始数据包转换为会话树,生成基于语义特征最后转换为入侵检测系统的特征规则格式。但是,上述特征提取技术多数没有基于工控协议数据特性的攻击特征提取方法。

攻击行为的相似性分析有助于更加全面地对攻击者进行溯源。文献[14]将一个攻击源发送的有效载荷转换为字符串并进行连接生成攻击指纹,然后通过简单的字符串距离度量比较指纹,从而分析攻击源。文献[15]使用 Micro-Honeypot 框架内浏览器指纹来追踪攻击者,并提出一种指纹关联算法,将浏览器中 cookie、IP 信息和指纹进行关联,生成字符串并作比较,从而确定同源攻击者。然而,上述同源攻击者判断方法仍无法直接应用于工控网络。文献[16]部署分布式蜜罐系统来收集威胁数据库,根据 3 种不同的工控协议蜜罐数据对攻击方法、攻击模式和攻击源进行分析,提出一种针对攻击组织的聚类算法。然而,人工划分攻击模式无法分析大规模数据,且该算法识别的攻击源组织并未进行结果验证。网络攻击同源性是指不同的网络攻击事件具有内在相似性,源自同一个人或同一组织等[4]。现有同源攻击的分析主要集中在恶意代码家族判定方面,此外,也有部分针对蜜罐数据进行组织判定的研究[16-17]。针对工控协议进行同源判定及组织分析的研究较少,且结果验证不足,为此,本文依据工控蜜罐数据进行同源攻击分析,并重点研究同一组织的判定问题。

2　工控攻击行为特征提取

2.1　工控功能码

根据工控协议规约,功能码用于标明一个信息帧的用途,当主设备向从设备发送信息时,功能码将通知从设备需要执行哪些行为。文献[18]使用 Modbus 功能码来界定某一攻击者对于设备的请求,根据请求报文中的意图进行分类并总结出攻击方法。然而,在真实情况下,单一功能码无法较好地体现整个攻击序列的企图。文献[18]将连续的 Modbus TCP 数据包抽象简化为 Modbus 功能码序列,利用序列顺序进行系统异常检测。wireshark 为常用的网络封包分析软件,图 1 所示为 wireshark 解析器下工控协议 Modbus 蜜罐数据流 PCAP 文件的各字段解析示例。其中,加框字段为功能码字段。基于 Modbus 协议报文格式,利用 python 脚本对蜜罐数据流 PCAP 文件进行处理,提取出 Modbus TCP 数据包中的初始数据包特征以便进行后续处理。对 PCAP 数据流进行处理后的工控信息库示例如表 1 所示。

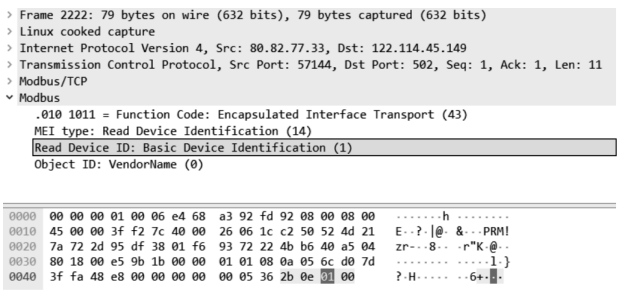

图 1　Modbus 蜜罐数据流 PCAP 文件

Fig. 1　Modbus honeypot data flow PCAP file

表 1 工控信息库示例

Table 1 Example of industrial control information base

序号	源 IP	功能码序列	交互时间/s	TTL 值
1	110.184.29.104	'2b' '5a' '5a' '5a' '5a'	3.997 460	50
2	85.25.79.124	'11' '2b'	3.668 198	38

2.2 基于统计信息的功能码粗粒度特征

根据文献[19]提出的 249 个基于流量的特征，结合所搜集的工控蜜罐中 Modbus TCP 数据包信息特征，本文提出一种基于统计信息的功能码粗粒度特征，其中包括功能码类型占比、攻击频率和稀有评级占比。攻击频率属于数据流特征，其余属于数据包内功能码信息特征。图 2 所示为 Modbus 协议的报文格式以及功能码在整个报文中的位置。

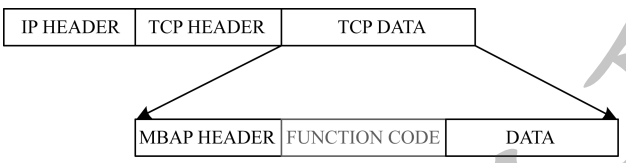

图 2 Modbus 协议报文格式

Fig. 2 Modbus protocol message format

功能码类型占比是功能码序列的功能码类型占整体功能码的比例。功能码类型占比表示为 $T_i = T_f/T$，其中，T_f 为功能码序列中功能码类型个数，T 为蜜罐数据中功能码类型个数。

攻击频率体现功能码序列中的时间特征，其值为功能码数量与交互总时长的比例，即 $F_i = \text{Num}/\text{Time}$，其中，Num 代表序列中功能码数量，Time 代表序列交互总时长。由于存在自动化攻击单位时间内发包数大于 1 的情况，因此本文对攻击频率进行 0-1 标准归一化处理。

稀有评级可以反映功能码序列在蜜罐数据中的特殊程度。以 Modbus 协议数据为例，在超过 1.7 万条 Modbus 数据包中，0x2b、0x11 出现频率最高，将其分别划为 1、2 级，其余按表 2 所示进行评级。稀有评级占比为 $R_i = D_i/D_{\max}$，其中，R_i 表示稀有评级占比，D_i 为功能码序列 i 中的最高稀有评级，$D_{\max}$ 表示整体数据中的最高评级 5。表 3 所示为基于 wireshark 解析器的报文解析样例，其中包括 2 个 Modbus TCP 粗粒度特征提取过程。从表 2 可以看出，蜜罐中的功能码类型数为 5，因此，计算出这一组 Modbus TCP 功能码类型占比为 0.4。出现功能码的稀有评级最高为 2，即其稀有评级占比为 0.4。攻击频率则根据时间戳进行计算，值为 3.876 8，后续再依据所有蜜罐数据整合进行归一化处理。

表 2 功能码稀有评级

Table 2 Function code rare rating

类型	次数	频率	评级
0x01	46	0.002 0	4
0x07	18	0.000 7	5
0x11	1 226	0.054 3	2
0x2b	21 128	0.935 1	1
0x5a	177	0.007 8	3

表 3 基于 wireshark 解析器的报文解析样例

Table 3 Example of message parsing based on wireshark parser

序号	时间戳	IP	事务处理标志	协议标志	长度	单元标志	功能码
1	2018-03-21 01:07:59.484 913	188.138.125.155	00 00	00 00	00 05	00	11
2	2018-03-21 01:08:00.000 809	188.138.125.155	00 00	00 00	00 02	00	2b

2.3 基于功能码序列的细粒度特征

细粒度特征即功能码序列特征，十六进制 Modbus 功能码序列为 07、01、01、11、2b、11、2b。由于 Modbus 功能码序列的有序性，本文使用窗口滑动的方法，通过一个大小为 2 的窗口对整个功能码序列进行截取，通过计算该功能码组合在整个序列各项组合中出现的概率，对 vector 内的相同功能码组合进行概率值填充，进而得到所有蜜罐数据中攻击 IP 的细粒度特征向量。

图 3 所示为细粒度特征提取方法示意图，其中，左半部分为样例功能码序列的处理过程，其得到对应攻击者 IP 和所有攻击者 IP 的向量值。功能码序列特征划分如表 4 所示，其中，$n = 2$，n 代表维度。

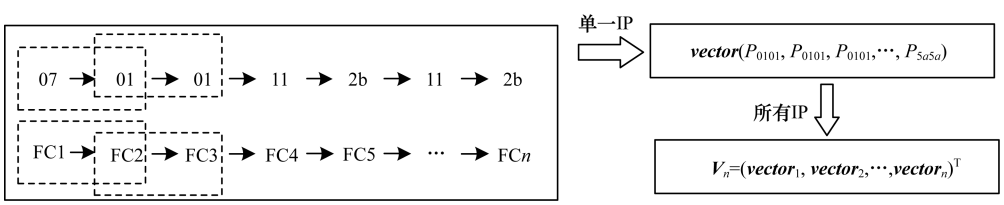

图 3 细粒度特征提取方法示意图

Fig. 3 Schematic diagram of fine-grained feature extraction method

表 4　功能码序列特征划分样例

Table 4　Example of feature division of function code sequence

功能码组合	次数	概率
0701	1	0.167
0101	1	0.167
0111	1	0.167
112b	2	0.333
2b11	1	0.167

在表 4 中,功能码组合表示当窗口长度 $n=2$ 时得到的十六进制值,次数表示功能码组合出现的次数,概率表示某二元功能码出现的几率。

将功能码组合进行功能码排序,得到有关功能码序列的多维向量。根据蜜罐数据以及功能码序列的有序性,向量种类表示为(0101,0107,0111,…,5a5a),特征向量为 $\boldsymbol{vector}=(P_{0101},P_{0107},P_{0111},\cdots,P_{5a5a})$。

3　工控攻击行为特征建模

3.1　基于粗糙集思想的粗粒度特征聚类

2.2 节介绍了功能码类型占比、攻击频率和稀有评级占比 3 个粗粒度行为特征。基于粗糙集思想的粗粒度特征建模方法,使用 Canopy 聚类算法将这 3 个行为特征转换为 (T_i,F_i,R_i) 三维向量。文献[20]实现了传统的 Canopy 算法,其具有快捷、简单等特点,为后续聚类算法消除了 K 值的不确定性,以便合理地使用初始聚类中心。

3.2　基于功能码序列的聚类模型

在机器学习中有一些基于标识数据的有监督机器学习算法,如 SVM、KNN 和决策树等。对于组织溯源领域,除一部分公开的设备扫描安全厂商外,其余攻击者的 IP 均没有样例标识。因此,需要通过无监督或半监督的机器学习算法来进行聚类分析。本文提出基于 K-Medoids 的改进围绕中心点划分(Partitioning Around Medoid,PAM)聚类算法,建立基于攻击行为的聚类模型,并使用基于密度的离群点检测方法对离群点进行检测和处理。

引入轮廓系数(Silhouette coefficient,S)来体现簇内数据的紧凑程度和簇间距离的分离程度。设数据集中的每个对象为 O,对象 O 与 O 所属的簇内其他对象之间的平均距离为 $A(O)$,$B(O)$ 指对象 O 到不包含 O 的所有簇的最小平均距离。所有点的轮廓系数平均值越接近 1,则显示聚类模型内聚度和外分离度越好,模型性能越优。轮廓系数的计算如式(1)所示:

$$S(O)=\frac{B(O)-A(O)}{\max\{A(O),B(O)\}} \tag{1}$$

本文使用改进的 PAM 聚类算法对攻击行为向量进行聚类并对攻击行为实现建模,流程如图 4 所示。

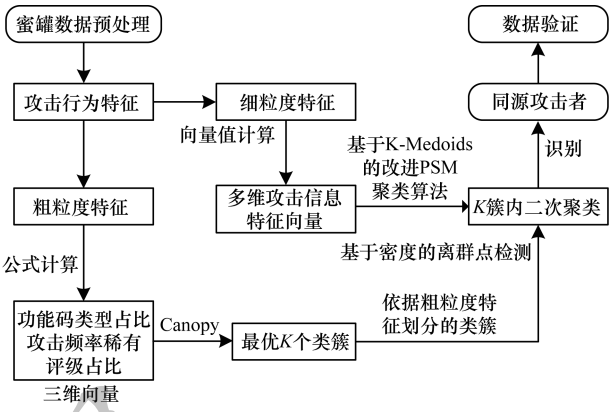

图 4　改进的 PAM 聚类算法流程

Fig. 4　Procedure of improved PAM clustering algorithm

本文算法具体步骤如下:

1)根据粗粒度、细粒度特征对数据集数据进行预处理。

2)使用基于粗糙集思想的粗粒度特征处理方法将粗粒度特征转换为三维向量,使用 Canopy 聚类算法计算最优类簇数量并进行第一次聚类。

3)使用基于功能码序列的细粒度特征处理方法将细粒度特征转换为攻击行为特征向量,并通过改进的 PAM 聚类算法对 K 簇进行聚类。基于 K-Medoids 改进的 PAM 聚类算法的基本思想为:

(1)任选 $k<k_{\max}$ 个初试点作为 Medoids。

(2)迭代使用 Medoids 外的其余非代表对象点来代替初始点进行聚类,找出更好的中心点,并根据部分已知组织信息,用半监督的方式取得更好的轮廓系数和聚类性能。

(3)由于相同 IP 可能有几种不同的扫描方式,因此引入惩罚因子 ξ 用于减小相同 IP 的不同向量间欧式距离的差异值。惩罚因子 ξ 计算公式如式(2)所示:

$$\xi=\frac{1}{N}\sum_{i\subset\{\mathrm{IPs}\}}\left|ED_{\mathrm{vector}}(i,i+1)\right| \tag{2}$$

其中,N 为集合{IPs}中相同 IP 攻击信息向量的个数,$ED_{\mathrm{vector}}(i,i+1)$ 表示 2 个向量的欧式距离差异值。

(4)对于学习样本(X,y)和模型$\hat{y}=f(X,w)$,半监督聚类方法的损失函数计算如式(3)所示:

$$L(y,\hat{y})=\left|ED[f(X,w),p(y|X)]-g(\xi)\right| \tag{3}$$

4)使用基于密度的离群点检测方法对离群点进行检测和处理,并根据半监督标识实现迭代学习,迭代轮廓系数 $S(O)$,直到得到更好的聚类精度。

5)使用本文提出的验证方法并应用第三方恶意 IP 库进行数据验证,计算精度和召回率,得到聚类结果。

3.3 基于密度的离群点检测方法

局部离群因子(Local Outlier Factor,LOF)[21]是基于密度的离群点检测方法中一个较有代表性的算法。以 q 为中心,与任意对象 O 距离为半径的所有对象集合记为 $N_k(q)$。将对象 q 和 O 的所有可达距离表示为 $\sum_{O\in N_k(q)} rd(q,O)$,其中,$rd(q,O)$ 代表对象 q 和 O的距离。局部可达密度 LRD 如式(4)所示:

$$\mathrm{LRD}_k(q)=\frac{|N_k(q)|}{\sum_{O\in N_k(q)} rd(q,O)} \tag{4}$$

则局部离群因子 LOF 计算如式(5)所示:

$$\mathrm{LOF}_k(q)=\frac{\sum_{O\in N_k(q)}\frac{\mathrm{LRD}_k(O)}{\mathrm{LRD}_k(q)}}{|N_k(q)|} \tag{5}$$

通过已有方法找出聚类模型中的离群点因子,为了得到更好的轮廓系数、聚类性能和结果,本文提出收缩因子 $\alpha(0\leqslant\alpha\leqslant1)$进行离群点处理,$\alpha$ 的迭代取值遵循 $\alpha=\arctan x+1$,以降低生成粗糙集时随机质心对离群点的判断误差并迭代出更优的轮廓系数 $S(O)$。聚类精确度不再提高时收缩停止。

4 实验结果与分析

4.1 数据集

本文数据来源于文献[16]收集到的分布式工控协议蜜罐数据,时间跨度超过 12 个月。实验及数据验证部分使用 Modbus 协议蜜罐数据,该协议是工控系统中常用的具有公开协议规约的通信协议。图 5 记录了 Modbus 蜜罐在一段时间期间的数据收集情况,其中,在 2018 年 8 月收集到了 2 987 条数据包,总数据包数量超过 1.7 万条。

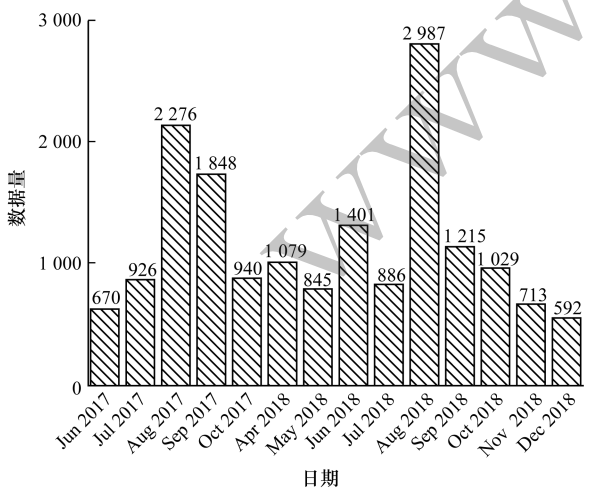

图 5 Modbus 协议数据量

Fig. 5 Modbus protocol data scale

4.2 数据验证

文献[16]指出,由于没有相似性实验,工控蜜罐数据攻击源分析后很难进行结果对比与验证。本文通过反向解析攻击者 IP 发现,只有很少一部分公开的设备扫描网站信息。在寻找公开数据集时,得知 AbuseIPDB[22]和 ipvoid[23]可以对滥用 IP(通常包括与可疑主机公司、僵尸网络、被黑客入侵的服务器或其他由黑客控制的机器相关的 IP 地址)进行记录、将网络中自动化软件扫描和攻击事件进行存储,而 IBM X-Force Exchange[24]公开的情报分析库甚至可以找到近 5 年内某一 IP 的活动情况,包括其何时被识别为扫描 IP、何时自动运行木马以及是否运行恶意软件。本文的数据验证方式是将同源判定的结果与公开网站数据进行匹配验证。当无法在开源滥用 IP 数据库中找到其明显组织后,使用物理位置判定方法,在数据验证时将同一网段的攻击者默认归于同一攻击源。

4.3 结果分析

文献[25]在观察大规模蜜罐攻击数据时发现,可追溯同源的类似 DDOS 攻击源的恶意服务器 TTL 值经常在某 2 个连续值之间波动。本次实验将对比本文同源攻击判定方法与 TTL 方法[22]、位置信息方法[3]的聚类性能。从图 6 可以看出,在 Modbus 协议数据集中,使用 Canopy 方法和 PAM 方法后,通过粗粒度特征向量生成 k_1 个类簇,簇中的细粒度聚类结果随着 k_2 值的变化而变化,其中,k_2 是对粗糙集中细粒度特征向量使用改进的 PAM 算法后的结果。当 $k_1=1$ 时,F1 值最高的是 $k_2=4$,即 $k_{21}=4$,F1 值为 0.60,表示在类簇 1 中,继续细分为 4 个簇时精确度和召回率的综合表现最佳。在其余 3 组中,F1 值最高的分别是类簇 $2k_2=3$,F1 值为 0.81;类簇 $3k_2=2$,F1 值为 0.90;类簇 $4k_2=1$,F1 值为 1.00。因此,类簇组织的最优数量为 $k_{21}+k_{22}+k_{23}+k_{24}=10$,平均 F1 值为 0.827 5。

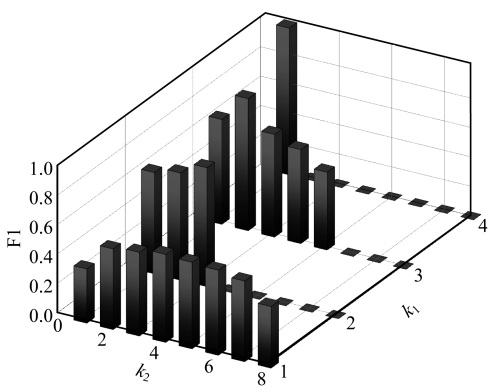

图 6 不同 k_1 和 k_2 值时算法聚类性能比较

Fig. 6 Comparison of clustering performance of algorithms with different k_1 and k_2 values

不同的同源攻击分析方法性能对比如图 7 所示。图 7(a)显示了不同初始簇数下方法的精确度,图 7(b)为召回率,图 7(c)使用 F1 值计算分析方法的精确率和召回率,图 7(d)显示聚类模型随类簇数改变时 SSE 值的变化情况。从中可以看出,本文方法较其他 2 种方法具有更高的 F1 值,数据的综合精确度和查全率具有更好的表现。理论上来说,随着类簇数 k 的增大,样本划分会更精细,每个簇的聚合程度会逐渐提高,则 SSE 值会逐渐变小。实验结果也证明,当 k 小于某一类簇数时,由于 k 的增大会大幅增加每个簇的聚合程度,而当 k 到达某一数值时,再增加 k 时所得到的聚合程度反馈会迅速变小,所以 SSE 的下降幅度会随着 k 值的继续增大而趋于平缓,这意味着当 k 值为 12 时,聚类模型具有更合理的 SSE 值,约为 7.376。然而,观察其 F1 值的结果可知,当 k 为 10 或 11 时,达到最高 F1 值 0.792。而且,在实验过程中,当 k 值大于 10 时,结果出现了测试集为空的情况,这显示出实验中可能存在过拟合现象。综合图 6、图 7,本文选取最优的类簇数为 10。

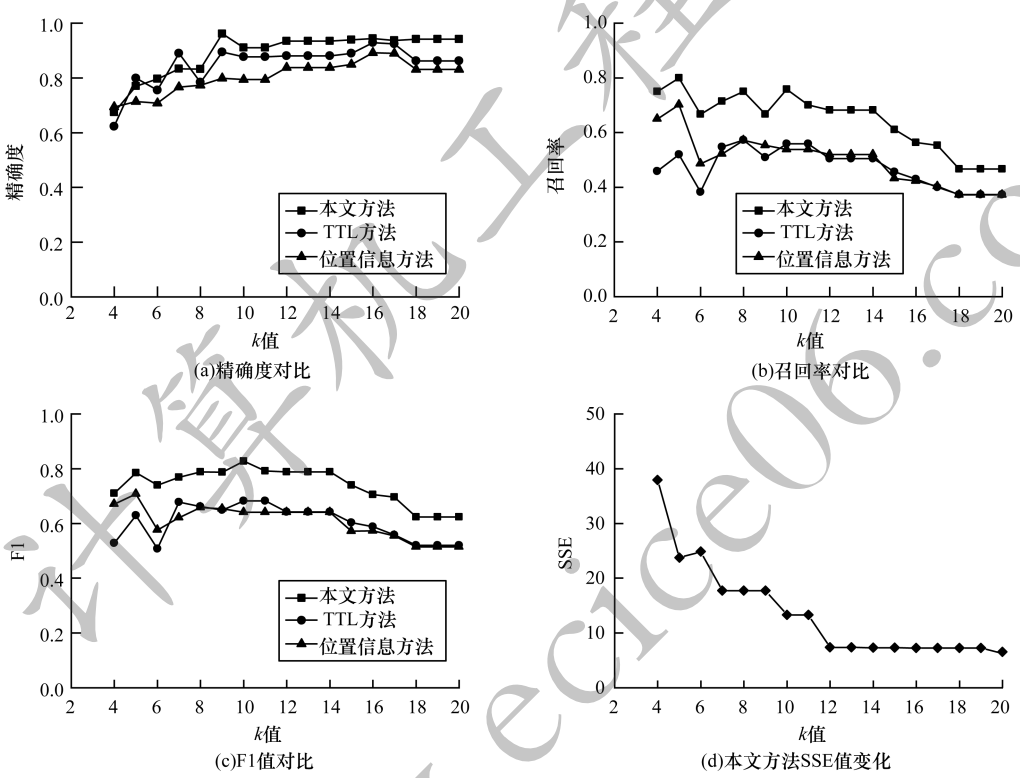

图 7　同源攻击分析方法性能对比结果

Fig. 7　Performance comparison results of same origin attack analysis methods

4.4　实验结论

蜜罐数据分析的验证是一个难题,本文通过反查 DNS,调用开源滥用 IP 库进行同源攻击者结果验证,相比于 TTL 值判别方法,本文方法具有更高的精确度和 F1 值。在判别威胁情报库中识别的 shodan IP 时具有 100% 的精确度和召回率,而对于其余恶意 IP 所属组织,如 plcscan、censys[26]、University of Michigan 等的判别,精确度达到 0.91。4 种协议攻击 IP 数前 10 的组织溯源结果如表 5 所示。其中,包括物联网设备识别企业和安全服务提供商,如 shodan、plcscan 和 censys 等。

表 5　同源攻击分析结果

Table 5　Same origin attack analysis results

组织	IP 数量
China Unicom Liaoning	35
shodan	27
University of Michigan	24
Nagravision SA	21
plcscan	18
censys	16
logicweb. com	15
China Telecom Sichuan	10
Northeastern University	10
onyphe. io	8

5 结束语

本文提出一种基于工控功能码序列的同源攻击分析方法,根据攻击者特征找到其物理位置特征,依据细粒度特征和粗粒度特征生成攻击行为模型。分析结果表明,该方法针对开源滥用 IP 库识别时具有较高的精确率,能够识别出 shodan、censys 和 plcscan 等 10 个组织。下一步将扩展工控蜜罐的适用协议,包括 S7comm、BACnet 等使用率较高的工控协议。此外,将攻击行为处理方法加入到入侵检测系统中实现一定程度的安全防护也是今后的研究方向。

参考文献

[1] SPITZNER L. Honeypots:catching the insider threat[C]//Proceedings of the 19th Annual Computer Security Applications Conference. Washington D. C., USA: IEEE Press,2004:15-26.

[2] MCGREW R,VAUGHN R B. Experiences with honeypot systems: development, deployment, and analysis [C]//Proceedings of the 39th Ha-waii International International Conference on Systems Science. Washington D. C., USA: IEEE Press,2006:256-269.

[3] FRAUNHOLZ D, ZIMMERMANN M, HAFNER A, et al. Data mining in long-term honeypot data[C]//Proceedings of IEEE International Conference on Data Mining. Washington D. C., USA:IEEE Press,2017:588-596.

[4] QIAO Yanchen, YUN Xiaochun, ZHANG Yongzheng. How to automatically identify the homology of different malware[C]//Proceedings of 2016 IEEE Trust-com/BigDataSE/ISPA. Washington D. C., USA:IEEE Press, 2016:12-36.

[5] WANG Liyan, XUE Jingfeng, CUI Yan, et al. Homology analysis method of worms based on attack and propagation features[M]//AVATANGELOU E, DOMMARCO R F, KLEIN M, et al. Communications in computer and information science. Berlin, Germany:Springer,2017:1-15.

[6] Shodan. Shodan is the world's first search engine for Internet-connected devices [EB/OL]. [2020-02-15]. https://www.shodan.io/.

[7] ICS security workspace [EB/OL]. [2020-02-15]. http://plcscan.org/blog/.

[8] SAVAGE S, WETHERALL D, KARLIN A, et al. Practical network support for IP traceback[J]. IEEE/ACM Transactions on Networking, 2002, 9(3): 226-237.

[9] SNOEREN A C, PARTRIDGE C, SANCHEZ L A, et al. Singlepacket IP traceback[J]. IEEE/ACM Transactions on Networking,2002,10(6):721-734.

[10] TAO Yaodong, LI Ning, ZENG Guangsheng. Overview of industrial control system security [J]. Computer Engineering and Applications, 2016, 52(13): 8-18. (in Chinese)
陶耀东,李宁,曾广圣. 工业控制系统安全综述[J]. 计算机工程与应用,2016,52(13):8-18.

[11] GUARNIZO J D, TAMBE A, BUNIA S S, et al. SIPHON: to-wards scalable high-interaction physical honeypots[C]//Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security. New York, USA: ACM Press,2017:26-39.

[12] KREIBICH C, CROWCROFT J. Honeycomb-creating intrusion detection signatures using honeypots [J]. Computer Communication Review,2004,34(1):51-56.

[13] YEGNESWARAN V,GIFFIN J T, BARFORD P, et al. An architecture for generating semantics-aware signatures[C]//Proceedings of Conference on Usenix Security Symposium. [S.l.]:USENIX Association,2005:97-112.

[14] POUGET F, DACIER M. Honeypot-based forensics [EB/OL]. [2020-02-15]. http://pdfs. semanticscholar. org/935e/d80c40367c8ccf2155fe66e3e52bc0fcdaad. pdf.

[15] JIA Z,CUI X,LIU Q,et al. Micro-honeypot:using browser fingerprinting to track attackers[C]//Proceedings of 2018 IEEE International Conference on Data Science in Cyberspace. Washington D. C., USA: IEEE Press, 2018: 197-204.

[16] LI K, YOU J, WEN H, et al. Collaborative intelligence analysis for industrial control systems threat profiling[C]//Proceedings of Future Technologies Conference. Berlin, Germany:Springer,2018:152-189.

[17] DACIER M, PHAM V H, THONNARD O. The wombat attack attribution method:some results[C]//Proceedings of International Conference on Information Systems Security. Berlin,Germany:Springer,2009:12-26.

[18] SHANG Wenli, ZENG Peng, WAN Ming, et al. Intrusion detection algorithm based on OCSVM in industrial control system[J]. Security and Communication Networks, 2016, 9(10):1040-1049.

[19] MOORE A W, ZUEV D. Internet traffic classification using Bayesian analysis techniques [J]. ACM SIGMETRICS Performance Evaluation Review, 2005, 33(1):50-56.

[20] ZHAO Yonghan,CHEN Bin,LI Mengyu. Parallel K-Medoids improved algorithm based on MapReduce[C]//Proceedings of 2018 International Conference on Advanced Cloud and Big Data. Washington D. C., USA:IEEE Press,2018:156-168.

[21] TAO Jing. Clustering-based and density outlier detection method [D]. Guangzhou: South China University of Technology,2014. (in Chinese)
陶晶. 基于聚类和密度的离群点检测方法[D]. 广州:华南理工大学,2014.

[22] AbuseIPDB. Making the Internet safer,one IP at a time[EB/OL]. [2020-02-15]. https://www. abuseipdb. com/.

[23] IPVOID. IP address tools online[EB/OL]. [2020-02-15]. https://www. ipvoid. com/ip-blacklist-check/.

[24] IBMX-Force Exchange. IBM security[EB/OL]. [2020-02-15]. https://exchange. xforce. ibmcloud. com.

[25] LUKAS K, KRUPP J, MAKITA D, et al. AmpPot: monitoring and defending against amplification DDoS attacks[EB/OL]. [2020-02-15]. https://christian-rossow.de/publications/amppot-raid2015. pdf.

[26] Censys. Actionable security insights about your attack surface[EB/OL]. [2020-02-15]. https://censys. io/.

编辑 吴云芳