



基于载荷特征与统计特征的Shodan流量识别

连晓伟,马 垚,陈永乐,张壮壮,王建华

(太原理工大学 信息与计算机学院,太原 030024)

摘 要:针对Shodan扫描流量对工业控制系统产生的不安全问题,结合载荷特征与统计特征,构建一种将确定性有限自动机(DFA)与支持向量机(SVM)相结合的流量识别DFA-SVM模型。通过分析应用层的流量特征,以提取协议功能码序列作为载荷特征,并结合传统的流量统计特征对流量进行识别。采用VPS部署6个分布式蜜罐系统对处理后的32 522个样本进行Shodan流量识别。实验结果表明,相比仅使用单一特征的模型,该模型可有效识别出27个Shodan扫描器IP,识别精度达到99.38%。

关键词:载荷特征;统计特征;确定性有限自动机;支持向量机;Shodan流量

开放科学(资源服务)标志码(OSID):



中文引用格式:连晓伟,马垚,陈永乐,等.基于载荷特征与统计特征的Shodan流量识别[J].计算机工程,2021,47(1):117-122.

英文引用格式:LIAN Xiaowei, MA Yao, CHEN Yongle, et al. Shodan traffic identification based on load characteristics and statistical characteristics[J]. Computer Engineering, 2021, 47(1): 117-122.

Shodan Traffic Identification Based on Load Characteristics and Statistical Characteristics

LIAN Xiaowei, MA Yao, CHEN Yongle, ZHANG Zhuangzhuang, WANG Jianhua
(College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China)

[Abstract] To address the security risk caused by Shodan scanning traffic in industrial control systems, this paper proposes a traffic recognition DFA-SVM model combining Deterministic Finite Automata (DFA) and Support Vector Machine (SVM) based on the load characteristics and statistical characteristics. By analyzing the traffic characteristics of the application layer, the model extracts the protocol function code sequence as the load feature, and combines the traditional statistical characteristics of traffic to identify it. Six distributed honeypot systems are deployed by VPS to identify the Shodan traffic in 32 522 samples. The experimental results show that compared with the model that only uses a single feature, the proposed model can effectively identify 27 Shodan scanner IPs from Shodan scanning traffic, with a recognition accuracy of 99.38%.

[Key words] load characteristics; statistical characteristics; Deterministic Finite Automation (DFA); Support Vector Machine (SVM); Shodan traffic

DOI: 10.19678/j.issn.1000-3428.0056888

0 概述

工业控制系统广泛应用于油气管道、供水系统、电网与核电站等关键基础设施^[1]。随着工业信息化的快速发展,工业控制系统与外部互联网的连接更加频繁,使得大量工控设备接入到互联网中,因此在工控领域出现愈来愈多的网络攻击,对工控系统造成严重威胁^[2-3]。由于Shodan搜索引擎可对网络攻击引起的威

胁进行有效识别以及索引面向互联网的工控系统组件,因此受到研究人员的广泛关注^[4]。

2009年,程序员约翰·马瑟利提出Shodan搜索引擎,它是全球第一个对全网设备进行扫描的搜索引擎,且带有图形用户界面,可有效识别面向互联网的设备。与传统搜索引擎不同,Shodan可以识别具有可路由IP地址的设备,包括计算机、网络打印机、网络摄像头以及工业控制设备等^[5]。Shodan每周

基金项目:山西省自然科学基金(201701D111002, 201601D021074)。

作者简介:连晓伟(1994—),男,硕士研究生,主研方向为物联网安全;马 垚,讲师、博士;陈永乐,副教授、博士;张壮壮、王建华,硕士研究生。

收稿日期:2019-12-12 **修回日期:**2020-01-15 **E-mail:**chenyongle@tyut.edu.cn

7天、每天24小时都在运行,且每月可收集大约5亿台联网设备的信息^[6],它将收集到的设备信息存储于一个可搜索的数据库中,该数据库可通过Web接口或Shodan API进行访问。用户可以使用一系列过滤器查询Shodan数据库,这些过滤器主要包括国家名、主机名、网络信息、操作系统与端口等。

Shodan搜索引擎的设计目的是搜索互联网,并试图识别与索引与之相连的设备,且其已识别出数万个与工业控制系统相关的面向互联网的设备。然而,识别工控相关设备的能力引起了重大的安全问题,美国国土安全部发布一份关于Shodan的报告,该报告详细说明了工业控制设备暴露在互联网中存在的风险^[7]。文献[8]认为Shodan是互联网中最强劲的搜索引擎。事实上,Shodan为攻击者提供一个强大的侦察工具,攻击者通过Shodan可以便捷地发现暴露在互联网上的工业控制设备以及与该设备相关的IP地址,以及开放的服务与存在的漏洞等信息,进而通过这些信息发动攻击,从而对工控系统造成严重破坏^[9]。

针对Shodan的不安全性,本文采用蜜罐技术对Shodan扫描流量进行深入研究。利用蜜罐模拟工控设备将蜜罐部署到互联网中,并通过开放相关端口吸引攻击者的攻击,从而捕获所有的攻击数据。本文针对这些攻击数据,构建一种将确定有限自动机(Deterministic Finite Automata, DFA)与支持向量机(Support Vector Machine, SVM)相结合的Shodan扫描流量识别模型。该模型利用状态机模型对扫描序列进行过滤,排除不具有Shodan扫描序列特征的流量,再通过SVM模型对接收的流量进行识别,从而得到最终识别结果。

1 相关工作

互联网流量识别方法主要分为3种:基于端口的识别方法,深度包检测(DPI)识别方法,基于机器学习的识别方法^[10]。其中,基于端口的识别方法根据端口与网络应用的映射关系进行流量识别。例如,FTP服务使用21端口,SSH远程登录服务使用22端口,基于HTTP协议的Web服务使用80端口。深度包检测识别方法通过分析目标流量协议特征,提取数据包载荷中的特征码并对流量进行识别。基于机器学习的流量识别方法通过网络流量中提取一系列独立于荷载的统计特征,采用机器学习方法对统计特征进行流量识别。

基于端口号的流量识别方法并不适用于Shodan流量的识别。在深度包检测识别中,文献[11]基于

DPI技术设计一个分级分类器,将流量正确分类为20多个细粒度的类,并建立阶层式自学习的分类模型,该集成识别模型将传统DPI技术的准确性与其他技术相结合,有效改善了DPI技术的不足。文献[12]结合DPI技术提出一种RocketTC的流量分类架构,该架构对网络流量的识别准确率达到97%。文献[13]提出一种基于DPI技术的流量识别方法,该方法结合软硬件优点,采用正则匹配法实现流量的识别。在基于机器学习的流量识别中,文献[14]提出以249个统计特征作为流量识别的分类依据,后续研究在上述统计特征的基础上,采用不同的机器学习算法对流量进行识别。文献[15]提出一种新的互联网流量识别方案,该方案基于熵的算法对每个流的前4个包的大小进行离散化,利用KNN、SVM和朴素贝叶斯3种分类器确定未知流的标签。接下来,使用4种组合器方案对3种分类器的输出进行组合,从而对未知流的标签进行最后决策。文献[16]提出可变特征空间的SVM集成方法,为每个两分类SVM构建具有最优区分能力的独立特征空间并集成成为多分类器,以有效提高流量分类器的精度与召回率。文献[17]提出一种基于距离的最近邻优化算法,该算法能够改善非平衡流量的分类性能。文献[18]通过提取恶意软件C&C的通信特征,并对多条加密流量进行合并,使用卷积神经网络对加密C&C数据流进行识别。

仅基于流量统计特征的机器学习流量识别方法由于缺少应用层流量的有效特征,而导致识别效果较差。为此,本文通过对Shodan流量进行分析,将深度包检测技术应用到基于机器学习的流量识别方法中,并构建一种基于载荷特征与统计特征相结合的DFA-SVM识别模型。该模型通过提取应用层协议功能码序列,将其与流量统计特征相结合,以提高模型识别准确率,有效识别Shodan流量。

2 基于统计特征和载荷特征的识别模型

本节主要描述基于统计特征与载荷特征相结合的DFA-SVM识别模型的构建过程,DFA-SVM识别模型如图1所示。首先,对原始流量进行预处理并完成载荷特征和统计特征的提取;其次,通过IP反查域名检查关联的PTR记录是否属于Shodan的子域,标记已知样本,并生成完整的数据集;接下来,对载荷特征中的功能码序列进行DFA匹配,未被接收的数据将视为非Shodan流量;最后,对前一步实验接收的数据进行基于统计特征的SVM识别,完成对Shodan流量和非Shodan流量的分类。

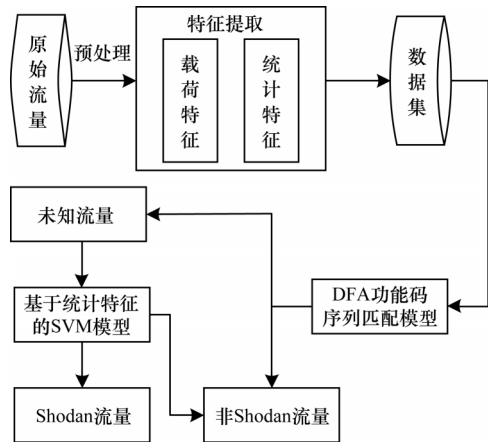


图 1 DFA-SVM 识别模型

Fig.1 DFA-SVM recognition model

2.1 基于状态机的载荷特征识别模型

2.1.1 确定性有限自动机

确定性有限自动机是一种能实现状态转移的自动机。对于给定的属于该自动机的状态和属于该自

动机字母表求和的字符,其都能根据事先给定的转移函数将它们转移到下一个状态,通常使用五元组 (Q,Σ,δ,s,F) 构成的数学模型表示。其中, Q 为状态的有限集, Σ 为字母表, δ 为转移函数, s 为开始状态, F 为一个接受状态集。

确定性有限自动机从起始状态开始,每一个输入都会使状态机的状态发生转移,如果能够从起始状态转移到接受状态,则识别输入序列。确定性有限自动机对于任何确定的输入都只有唯一确定的转移,且不存在空字符串的状态转移。

2.1.2 基于状态机的功能码序列匹配

基于状态机的应用层协议功能码序列匹配方法将一次完整的通信看作一个交互过程,对交互过程中每一个阶段的状态进行提取,分析并找出这些状态的特征,从而建立一个该协议的串行状态规则。下文以基于状态机的 S7comm 协议分析流程为例进行分析,其中,S7comm 协议的报文格式如图 2 所示。

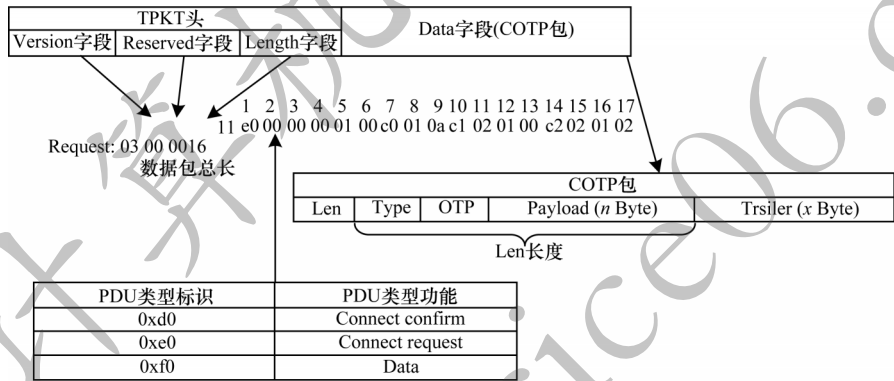


图 2 S7comm 协议报文格式

Fig.2 Message format of S7comm protocol

Shodan 对 S7comm 协议设备的扫描流程如表 1 所示。从表 1 可以看出,Shodan 扫描器与被扫描设备通过 3 次握手建立 TCP 连接,再建立 COTP 连接和 S7comm

连接,接着发送 2 条读系统状态列表数据包,用来分别请求 Module Identification 和 Component Identification,关闭连接后则扫描结束。

表 1 Shodan 对 S7comm 协议设备扫描流程

Table 1 Scanning flow of S7comm protocol device by Shodan

序号	PDU_Type	功能码	功能码组	子功能码	数据包功能
1					TCP 3 次握手
2	0xe0				建立 COTP 连接
3	0xf0	0xf0			建立 S7comm 连接
4	0xf0	0x00	0x44	0x01	读系统状态列表,请求 Module Identification
5	0xf0	0x00	0x44	0x01	读系统状态列表,请求 Component Identification
6					关闭连接,扫描结束

基于状态机的功能码序列匹配方法是通过串行顺序逻辑判断来实现的,因此需要得到一次完整通信中各阶段的数据状态,而工控协议中的功能码能够表示每条数据包的功能状态。根据工控协议规

约,功能码用于标明一个信息帧的用途,即指明数据包的功能,通常在协议数据包的某个固定字段指明。因为 Shodan 扫描为机器扫描,其扫描流量序列相对固定,所以可以从流量数据中提取功能码,并将其表

示为基于状态机的协议交互状态^[19]。对4种工控协议分别建立与其对应的自动状态机,通过对不同协议的数据包进行基于状态机的功能码序列匹配,即可区分出所捕获的数据包是否为Shodan流量。

2.2 基于统计特征的SVM识别模型

2.2.1 流量统计特征提取

由于网络流量特征中存在很多冗余特征和无关特征,而这些特征不仅会降低分类精度,还会增加分类模型的计算成本。本文根据原始数据特点,从文献[5]提出的249个流量统计特征中提取30个特征作为样本属性特征,并利用Relief特征选择算法将无关或冗余特征删除。Relief算法是一种特征权重算法,该算法的目的是根据各个特征和类别的相关性赋予每个特征不同的权重,并将小于某个阈值的特征删除^[20]。在实验中,将特征权重小于0.01的特征删除,从而得到13个特征,这13个特征的详细网络流量特征描述如表2所示。

表2 网络流量特征描述

Table 2 Description of network traffic characteristics

序号	表示	描述
1	Src_IP	源IP地址
2	Dst_IP	目的IP地址
3	Dst_Port	目的端口
4	IP_Length	IP包头长度
5	Src_Pack	源主机到目标主机数据包总数量
6	Dst_Pack	目标主机到源主机数据包总数量
7	Src_Bytes	目标主机到源主机数据字节数
8	Dst_Bytes	源主机到目标主机数据字节数
9	Duration	连接持续时间
10	Min_IAT	数据包到达时间间隔最小值
11	Max_IAT	数据包到达时间间隔最大值
12	Mean_IAT	数据包到达时间间隔均值
13	Var_IAT	数据包到达时间间隔方差

2.2.2 基于SVM的Shodan流量识别建模

针对Shodan流量统计特征的识别模型,其目的是设计一种对具有交互行为特征的流量数据进行分析处理的分类方法,以识别出Shodan扫描流量。由于工控网络数据具有高维、非线性等特点,针对流量交互特性设计的Shodan流量识别算法需要适应工控流量的特殊性,以达到更好的识别效果。基于机器学习的SVM是一种监督学习算法,其被广泛应用于统计分类以及回归分析,且分类效果较好,适用于流量识别^[21]。

利用SVM算法对Shodan流量识别进行建模,根据提取的流量统计特征,建立识别模型的训练集和测试

集。设定模型的各项参数,并对数据集进行训练以获得Shodan流量识别模型的决策函数,具体步骤为:

步骤1 根据流量特征提取阶段提取的特征对数据集进行处理,构建实验训练集和测试集。

步骤2 选择合适的核函数,并设定核函数的相关参数和惩罚系数 C ,其中, C 是用来控制寻找最大超平面和保证数据点偏差量最小的权重,并引入拉格朗日函数, α 是拉格朗日乘子向量, x_i 和 y_i 是样本点,构造并求解式(1):

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \end{aligned} \quad (1)$$

步骤3 通过计算得出 α 的最优解 $\alpha^*=(\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$,并计算式(2):

$$b^* = y_j - \sum_{i=1}^N \alpha_i^* y_j K(x_i \times x_j) \quad (2)$$

步骤4 求解最优分类函数:

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i^* K(x \times x_i) + b^* \right) \quad (3)$$

其中, b 为分类超平面参数。

步骤5 利用建立的分类函数在测试集上进行训练测试,并不断优化参数的选择,直至达到满意的训练精度为止,从而建立高效的SVM识别模型。

3 实验结果与分析

3.1 数据集

为收集大量的流量数据,本文开发一个分布式蜜罐系统,该系统包含6个蜜罐,可以模拟4种可编程逻辑控制器(Modicon(BMX P34 2020)、s7-400、奥莱斯LGR25和ABB PM573-ETH)以及4种工业控制协议(Modbus、S7comm、IEC 60870-5-104和BACnet-APDU)。每个蜜罐都是在Conpot^[22]的基础上开发的,所有蜜罐都可在预先定义好的响应机制支持下响应请求,并捕获与攻击者的所有交互。每个蜜罐集成一个开源认证发布-订阅协议hpfeeds,并将捕获的数据传输到数据中心Mongodb数据库中。此外,为使蜜罐更具欺骗性,实验改变原蜜罐框架的硬编码特征,使得Shodan将本文蜜罐误识别为真实的工业控制系统。

在为期3个月的数据收集,实验总共收到来自145 720个IP的攻击。实验开始前,需要对数据进行预处理,仅保留有完整交互的流量,并对流量进行统计特征和功能码序列的提取,从而获得32 522个

样本。原始数据集中每个属性的取值范围不同,为使每个属性处于同一量纲上,本文采用线性变换将每个属性的取值范围映射到 $[-1,1]$ 。然后,对数据进行人工标注并将数据集随机分为训练集和测试集2个子集。实验数据集的具体信息如表3所示。

表3 实验数据集的具体信息

Table 3 Details of the experimental dataset

流量类型	数量	流量所占百分比/%
Shodan 流量	9 883	30.4
非 Shodan 流量	22 639	69.6

3.2 支持向量机参数选择与测试

在基于统计特征的SVM识别模型中,SVM中不同参数的选择对实验结果有显著影响。因此,本文在实验开始前,利用训练集进行测试以选择最优实验参数,并将最优参数用于后续实验。

实验利用径向基核函数对参数进行选择与优化,径向基函数是某种沿径向对称的标量函数,通常定义为样本到数据中心之间径向距离的单调函数。径向基核函数是一种比较常见的核函数,且常用的径向基核函数可表示为:

$$K(x, x') = \exp(\gamma \|x - x'\|_2^2) \quad (4)$$

实验研究了惩罚系数 C 对实验精度的影响,结果如图3所示。从图3可以看出,随着惩罚系数 C 的增大,实验精度呈现先增大后降低的趋势,当 $C=128$ 时,实验精度达到最大。因此,实验设置惩罚系数 C 为128。经过实验得出,当 $C=128, \gamma=0.008 2$ 时,实验结果较好。因此,将 $C=128, \gamma=0.008 2$ 作为后续实验的默认参数。

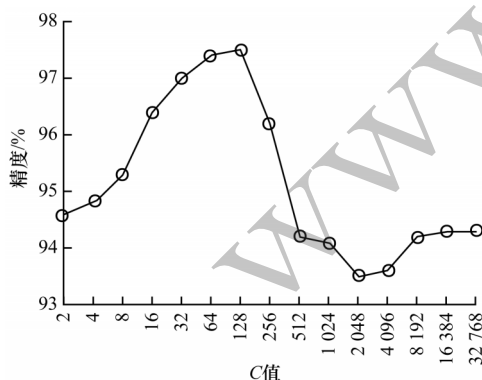


图3 惩罚系数对实验精度的影响

Fig.3 Influence of penalty coefficient on experimental accuracy

3.3 模型性能验证

本文从网络流量中分别提取统计特征和载荷特征,并建立相应的模型,将2种模型相结合完成对

Shodan流量的识别。实验对SVM模型、DFA模型与本文提出的SVM-DFA模型进行分类效果进行对比,分别进行10次实验,采取十折交叉验证的方式将实验数据集分为10份,取其中1份作为测试集,其余9份作为训练集,实验结果如图4所示。

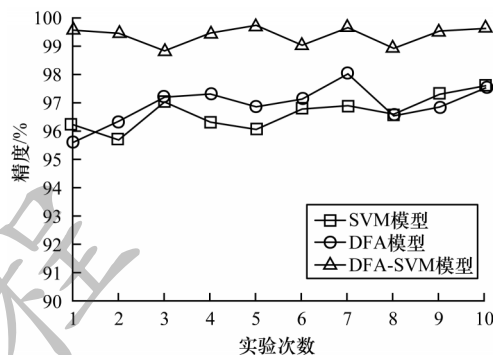


图4 3种模型的实验精度对比

Fig.4 Comparison of experimental accuracy of three models

从图4可以看出,基于统计特征与载荷特征相结合的DFA-SVM识别模型在识别准确率上优于其他2种基于单一特征的识别模型,这说明在加入功能码序列特征后,识别模型的精度提升了约3%,达到99.38%。识别模型的精度得到明显提升,这是由于Shodan扫描为机器扫描,且其扫描序列是相对固定的,因此本文综合扫描序列特征和统计特征,在使用统计特征前,利用状态机对扫描序列进行过滤,排除不具有Shodan扫描序列特征的流量。对于状态机模型接受的流量中,既包含真正的Shodan流量,也包含类似Shodan的流量,因此采用SVM模型对这部分流量进一步识别,得到最终识别结果。因此在加入功能码序列特征后,对模型识别精度有很大的提升作用。

实验进一步对KNN模型、C4.5模型、NB模型与本文DFA-SVM模型的准确率和召回率进行比较,结果如图5所示。从图5可以看出,本文DFA-SVM模型的准确率和召回率均优于其他3种模型,说明本文提出的识别模型具有更好的分类效果。

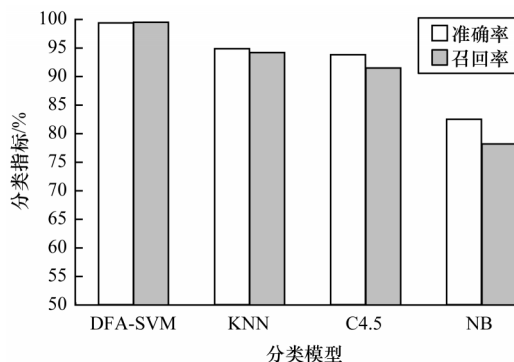


图5 4种模型的准确率与召回率对比

Fig.5 Comparison of accuracy and recall rate of four models

4 结束语

针对Shodan扫描流量识别问题,本文构建一种基于流量统计特征与载荷特征相结合的DFA-SVM识别模型。采用提取应用层中的协议功能码序列作为载荷特征,并将其与流量统计特征相结合对流量进行识别。实验结果表明,与DFA、SVM等模型相比,该模型可有效识别27个Shodan扫描器IP,显著提高流量识别精度。本文详细介绍了特征提取与模型构建部分,但对分类算法的优化还有待提高,下一步将采用梯度下降法对SVM的时间复杂度与空间复杂度进行优化,以提高模型识别效率。

参考文献

- [1] LI Qiang, JIA Yuxuan, SONG Jinke, et al. Search of Internet of thing information in the cyberspace[J]. Journal of Cyber Security, 2018, 3(5): 38-53. (in Chinese)
李强,贾煜璇,宋金珂,等. 网络空间物联网信息搜索[J]. 信息安全学报, 2018, 3(5): 38-53.
- [2] WANG Xiaoshan, YANG An, SHI Zhiqiang, et al. New trends of information security in industrial control systems[J]. Netinfo Security, 2015, 15(1): 6-11. (in Chinese)
王小山,杨安,石志强,等. 工业控制系统信息安全新趋势[J]. 信息网络安全, 2015, 15(1): 6-11.
- [3] TAO Yaodong, LI Ning, ZENG Guangsheng. Review of industrial control systems 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.
- [4] BODENHEIM R, BUTTS J, DUNLAP S, et al. Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices[J]. International Journal of Critical Infrastructure Protection, 2014, 7(2): 114-123.
- [5] YU Bofei. Search engine safety and equipment protection security based on Internet of things technology[J]. Metal World, 2015(1): 47-50. (in Chinese)
于博菲. 基于物联网技术的搜索引擎与设备安全[J]. 金属世界, 2015(1): 47-50.
- [6] MA Cheng. The principles and security usages of cyberspace search engine[J]. Information Security and Technology, 2016, 7(5): 6-10. (in Chinese)
马程. 网络空间搜索引擎的原理研究及安全应用[J]. 网络空间安全, 2016, 7(5): 6-10.
- [7] KLIMBURG A. National cyber security framework manual [EB/OL]. [2019-11-10]. https://www.researchgate.net/profile/Eric_Luijff/publication/261984536_Organisational_Structures_Considerations/links/544f5c6e0cf2bca5ce90e65d.pdf.
- [8] GOLDMAN D. Shodan: the scariest search engine on the Internet[EB/OL]. [2019-11-10]. <https://money.cnn.com/2013/04/08/technology/security/shodan/>.
- [9] ALEXANDRU V S, OBERMEIER S, YU D Y. ICS threat analysis using a large-scale honeynet[C]// Proceedings of the 3rd International Symposium for ICS & SCADA Cyber Security Research. Washington D. C., USA: IEEE Press, 2015: 20-30.
- [10] PENG Lizhi. A survey of Internet traffic identification[J]. Journal of University of Jinan (Science and Technology), 2016, 30(2): 95-104. (in Chinese)
彭立志. 互联网流量识别研究综述[J]. 济南大学学报(自然科学版), 2016, 30(2): 95-104.
- [11] GRIMAUDO L, MELLIA M, BARALIS E. Hierarchical learning for fine grained Internet traffic classification[C]// Proceedings of the 8th International Wireless Communications and Mobile Computing Conference. Washington D. C., USA: IEEE Press, 2012: 1-6.
- [12] FU Wenliang, SONG Tian, ZHOU Zhou. RocketTC: a high throughput traffic classification architecture on FPGA[J]. Chinese Journal of Computers, 2014, 37(2): 158-166. (in Chinese)
付文亮,嵩天,周舟. RocketTC: 一个基于FPGA的高性能网络流量分类架构[J]. 计算机学报, 2014, 37(2): 158-166.
- [13] HUANG Jianwen. Development and design of traffic identification system based on DPI[J]. Electronic Design Engineering, 2017, 25(11): 14-18. (in Chinese)
黄健文. 基于DPI的流量识别系统的开发与设计[J]. 电子设计工程, 2017, 25(11): 14-18.
- [14] MOORE A W, ZUEV D. Internet traffic classification using Bayesian analysis techniques[J]. ACM SIGMETRICS Performance Evaluation Review, 2005, 33(1): 50-60.
- [15] GHOFrani F, KESHAVARZ-HADDAD A, JAMSHIDI A. Internet traffic classification using multiple classifiers[C]// Proceedings of the 7th Conference on Information and Knowledge Technology. Washington D. C., USA: IEEE Press, 2015: 1-8.
- [16] QIAN Yaguan, GUAN Xiaohui, YUN Bensheng, et al. Internet traffic classification using SVM with flexible feature space[J]. Telecommunications Science, 2016, 32(5): 105-113. (in Chinese)
钱亚冠,关晓惠,云本胜,等. 基于可变特征空间SVM的互联网流量分类[J]. 电信科学, 2016, 32(5): 105-113.
- [17] WU D, CHEN X, CHEN C, et al. On addressing the imbalance problem: a correlated KNN approach for network traffic classification[M]. Berlin, Germany: Springer, 2014: 138-151.
- [18] CHENG Hua, XIE Jinxin, CHEN Lihuang. CNN-based encrypted C&C communication traffic identification method[J]. Computer Engineering, 2019, 45(8): 31-34, 41. (in Chinese)
程华,谢鑫鑫,陈立皇. 基于CNN的加密C&C通信流量识别方法[J]. 计算机工程, 2019, 45(8): 31-34, 41.
- [19] WANG Shan, CHEN Jian, HUANG Zhigen. Protocol identification method and content analysis of application protocol based on finite automaton[J]. Electronic Measurement Technology, 2011, 34(12): 109-112. (in Chinese)
王珊,陈健,黄志根. 基于状态机的应用层协议识别和内容分析[J]. 电子测量技术, 2011, 34(12): 109-112.
- [20] HUANG Xiaojuan. Research on relief algorithm for feature selection[D]. Suzhou: Soochow University, 2018. (in Chinese)
黄晓娟. 面向特征选择的Relief算法研究[D]. 苏州: 苏州大学, 2018.
- [21] ESTE A, GRINGOLI F, SALGARELLI L. Support vector machines for TCP traffic classification[J]. Computer Networks, 2009, 53(14): 2476-2490.
- [22] Conpot[EB/OL]. [2019-11-10]. <https://plscan.org/blog/tools/conpot/>.