机器学习在网络空间安全研究中的应用

宇翔

2020 年 7 月 12 日

1 研究背景

传统的安全问题解决方案在面对海量数据变得效率低下，机器学习以其强大的自适应性、自学习能力为安全领域提供了一系列有效的分析决策工具为解决传统方法难以建模的网络空间安全问题提供了可能性。

2 本文工作

本文阐述了机器学习技术在网络空间安全研究中的应用流程，然后从系统安全、网络安全和应用安全三个层面，着重介绍了机器学习在芯片及系统硬件安全、系统软件安全、网络基础设施安全、网络安全检测、应用软件安全、社会网络安全等网络空间安全领域中的解决方案，重点分析、归纳了这些解决方案中的安全特征及常用机器学习算法。最后总结了现有解决方案中存在的问题，以及机器学习技术在网络空间安全研究中未来的发展方向和面临的挑战。

3 内容摘要

机器学习在网络空间安全的应用及流程

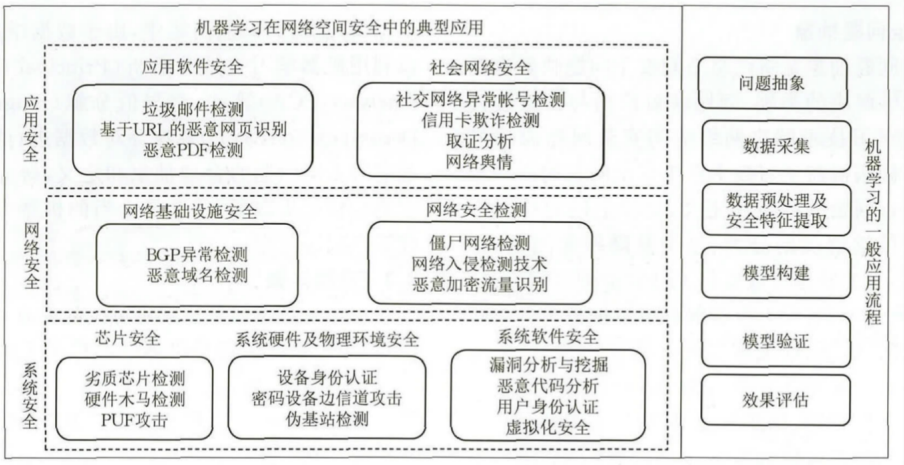

图 1 机器学习在网络空间安全研究中的应用及流程

图 1: 机器学习在网络空间安全研究中的应用及流程

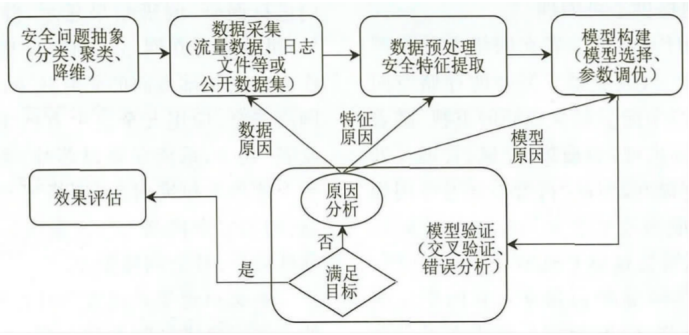

图 2 机器学习在网络空间安全研究中的应用流程

图 2: 机器学习在网络空间安全研究中的应用流程

3.1 安全问题抽象

安全问题抽象是将网络空间安全问题映射为机器学习能够解决的类别。

3.2 数据采集

自行采集数据和公开数据集

3.3 数据预处理及特征提取

3.3.1 数据预处理

分析统计数据，对缺失值、异常值、重复值、噪音数据等进行清洗，清洗之后对数据进行归一化操

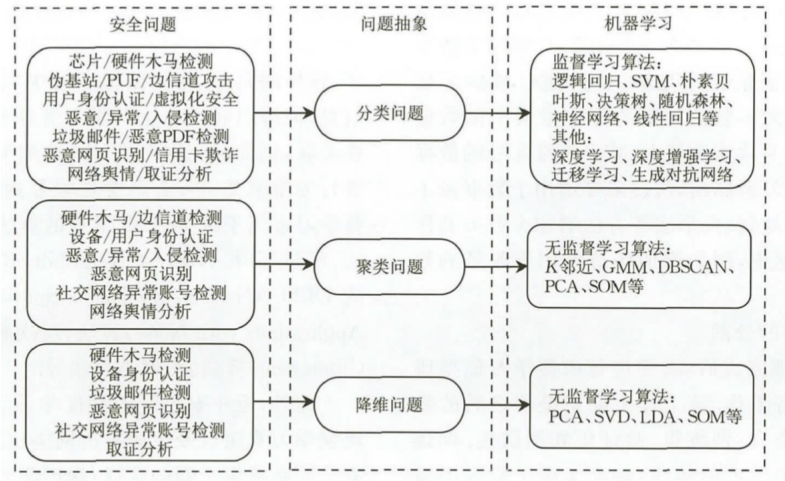

图 3 安全问题抽象

图 3: 安全问题抽象

序号	数据集名称	说明
1	DARPA Intrusion Detection Data Sets[13-15]	网络入侵检测数据集(包含1998、1999、2000 三个数据集)
2	KDD[16]	网络入侵检测数据集(包含1998、1999 两个数据集)
3	UCI's Spambase [17]	垃圾电子邮件数据集
4	Honeynet Project Challenges[18]	网络攻击行为数据集
5	Internet Traffic Archive[19]	网络包数据集，包含路由信息
6	Alexa 网站域名[20]	Alexa. com 收集的知名网站域名
7	DMOZ Open Directory Project[21]	URL 地址集
8	RIPE RIS 和 Route Views[22]	域间路由数据集
9	PhishTank[23]	钓鱼网站 URL 地址集

图 4: 数据集

作。

3.3.2 数据缺失处理及异常值的处理

采集数据集中某个特征缺失值较多时，通常会将该特征舍弃，否则可能会产生较大的噪声，影响机器学习模型的效果.当某个特征的缺失值较少时，可用采用固定值填充、均值填充、中位数填充、上下数据填充、插值法填充或者随机数填充等方法。

3.3.3 非平衡数据的处理

异常数据样本或恶意数据样本远远少于正常样本时，直接使用机器学习算法构建检测模型效果不佳。为了解决非平衡数据问题，通常使用过采样或欠采样方法构造平衡数据集。

3.3.4 数据集分割

将整理之后的数据集分为三个集合：训练集、验证集和测试集。训练集用于机器学习模型的构建，验证集用于验证模型及参数调优，测试集用于评估模型在实际使用中的泛化能力。

3.3.5 特征提取

根据数据预处理后的数据集及目标问题类型，在本阶段选择合适的学习算法，构建求解问题模型，包括算法选择和参数调优。深度学习凭借强大的自动提取特征的能力，被用于解决异常协议检测、恶意软件检测、网络入侵检测、BGP异常路由检测以及差分隐私保护等安全问题。（此部分知识欠缺，需继续学习各类算法及适合解决的问题类型）

3.3.6 模型验证

评估训练的模型是否足够有效，常用k倍交叉验证法。

3.3.7 效果评估

* 分类问题评估指标：正确率、查准率（精度）和查全率（又称召回率）
* 异常检测、入侵检测评估指标：误报率、漏报率
* 认证领域评估指标：误识率、拒识率
* 聚类问题评估指标：一类是将聚类结果与某个参考模型进行比较;另一类是直接考察聚类结果而不利用任何参考模型

4 机器学习在系统安全研究中的应用

对用户身份认证详细记录一下：在基于机器学习的用户身份认证研究方面，主要有利用机器学习攻击传统用户身份认证方法和利用机器学习设计新的用户身份认证机制两个研究点。

表 4 机器学习在系统安全中的应用

系统安全	安全问题	问题抽象	主要特征	机器学习方法	参考文献
芯片安全	劣质芯片检测	分类	芯片外形、边信道信号	OC-SVM、PCA、SOA、ANN	[57-59]
	硬件木马检测	分类/聚类/降维	边信道信号、芯片原理图	OC-SVM、PCA、非线性回归、K 近邻、SVM	[39,49,62-63]
	PUF 攻击	分类	PUF 激励-响应对	逻辑回归、SVM、进化策略、ANN	[65-66]
系统硬件及物理环境安全	设备身份认证	聚类/降维	暂态信号、调制信号、频谱响应、内部传感器响应	SOA、PNN、SVM、K 近邻、PCA、随机森林	[46,67-69]
	物理层边信道攻击	分类/聚类	边信道信号分布特征、差分能耗	LS-SVM、学习排序法、PCA、自组织映射、mRMR 算法、SOA、SVM、随机森林	[70-72]
	伪基站检测	分类	2G 和 3G 之间的模式转变、真正基站检测到的手机信号消失的时间、加密的禁用、基站位置数据库、WiFi 位置数据库和短信日志库	神经网络、SVM	[10,73]
系统软件安全	漏洞分析与挖掘	分类/聚类	源码 API 用途、代码语法	PCA、RNN、SVM、集成学习、N-Gram 模型	[75-81]
	恶意代码分析	分类/聚类	二进制文本、运行行为，信息、内容、时间和连接，动态行为、请求许可、请求时间序列、敏感程序接口	SVM、AdaBoost、贝叶斯网络、随机森林、K 近邻	[27,50,82-85]
	用户身份认证	分类/聚类	触屏特征、击键行为、验证码颜色与文本、计算机视觉、传感器边信道信号	SVM、朴素贝叶斯、马氏距离算法、K 近邻、DNN、LSTM 和 GRU 网络、*K*-means	[86-93]
	虚拟化安全	分类	虚拟机边信道信息	SVM	[98-100]

图 5: 机器学习在系统安全中的应用

表 3 机器学习在用户身份认证技术中的应用

安全问题	安全特征	机器学习算法	相关文献
用户身份认证攻击	验证码的颜色和文本特征	SVM	[87]
	计算机视觉、触摸点	K-Means 算法	[88]
	边信道信号	随机森林、K 近邻、SVM、神经网络	[89]
用户身份认证设计	触屏特征(加速度、压力、大小和时间)	单类学习算法	[90]
	击键行为	SVM、朴素贝叶斯、K 近邻	[91]
	击键	DNN	[92]
	击键	LSTM	[93]
	手势	SVM	[97]

图 6: 机器学习在用户身份认证技术中的应用

5 机器学习在网络安全研究中的应用

对于恶意加密流量识别详细记录一下文中提要的已有研究：《Identifying encrypted malware trafic with contextual flow data》（Procedings of the ACM Workshop on Security and Artificial Intel ligence, 2016）在不解密网络流量的情况下，利用机器学习技术从加密的网络流量中识别出具有恶意行为的网络流量。首先采集了百万级的正常流量和恶意流量，然后分析了TLS流、DNS流和HTTP流的不同之处，具体包括未加密TLS握手信息、TLS流中与目的IP地址相关的DNS响应信息、相同源IP地址5min窗口内的HTTP流的头部信息；然后从上述信息中提取特征，将该特征采用零均值和单位方差进行归一化处理，随后利用L1逻辑回归分类器获得检测模型最优权值，并采用10倍交叉验证进行模型验证。《Analyzing android encrypted network trafic to identify user actions》（IEEE Transactions on Information Forensics and Security, 2015）利用机器学习技术分析网络加密流量，用于识别移动终端用户的行为。利用已知APP在移动终端生成网络流量，在网络侧截取网络流量，将网络流的时间序列进行标记，生成有标签的训练集；然后使用层次聚类算法将网络流进行聚类，相似的流被分组在同一个簇中代表有相似的用户行为，不同的流分组至不同的簇;利用整数形式表示每个簇的特征，再使用随机森林算法执行分类操作，将未知的流量分类至不同的流簇中。

表 7 机器学习在网络安全中的应用

网络安全	安全问题	问题抽象	安全特征	机器学习算法	参考文献
网络基础设施安全	BGP 异常检测	分类/聚类	BGP 更新消息特征、BGP 的时间序列特征等	层次聚类、决策树和 ELM、SVM 和隐马尔可夫模型、朴素贝叶斯、AdaBoost、LSTM	[22,35,102,104,107]
	恶意域名检测	分类/聚类	基于网络层的特征、基于区域的特征、基于时间的特征、基于 DNS 应答的特征、基于 TTL 的特征、基于域名信息等	决策树、X-Means	[47,110-111]
网络安全检测	僵尸网络检测	分类/聚类	快速收敛时间、P2P 节点的通信图、误用检测信息、域名中的语法特征、查询特征、分级域名的统计信息、n 元组的统计信息、加密特征、域名结构化的特征、Bot 意图查询、主机信息、图特征、NetFlow、C&C 通信中的关键特征、C&C 异常命令结合异常日志信息、流量的目的地址和端口聚合集合结合异常日志信息、网络流量数据关联 spam 信息以及 DNS 的日志记录	X-Means 聚类、图聚类、单链层次聚类算法、SVM、随机森林、隐马尔可夫模型、关联规则	[114-122]
	网络入侵检测	分类/聚类	协议标识符(ID)、源端口、目的端口、源地址、目的地址、ICMP 类型、ICMP 代码、原始数据长度、原始数据内容、TCP session 特征、攻击签名、telnet 会话、网络数据包	神经网络、关联规则、贝叶斯网络、决策树、隐马尔可夫模型、朴素贝叶斯、SVM、DBSCAN	[13-16,28,51-52,125-127,129-135,137,139-140]
	恶意加密流量识别	分类/聚类	未加密 TLS 握手信息、DNS 响应信息、HTTP 流的头部信息、网络流的时间序列等	逻辑回归、层次聚类、随机森林	[24,142]

图 7: 机器学习在网络安全中的应用

6 机器学习在应用安全研究中的应用

6.1 应用软件安全

* 垃圾邮件检测（已有过滤技术已趋于稳定）

* 基于url的恶意网页识别（基于分类的恶意网页识别和基于聚类的恶网页识别）

* 恶意pdf检测（PDF文档的检测研究大多采用PDF文档内容或结构为特征，利用随机森林、SVM、决策树等分类器构建PDF检测器）

6.2 社会网络安全

表 9 机器学习在应用安全中的应用

应用安全	安全问题	问题抽象	安全特征	机器学习算法	参考文献
应用软件安全	垃圾邮件检测	分类/降维	邮件的发送者 IP 地址、邮件内容文本特征、域名特征	朴素贝叶斯、神经网络、SVM、决策树集成法	[17,25,44]
	基于 URL 的恶意网页识别	分类/聚类/降维	主机信息、URL 信息、网页信息、浏览器行为、URL 的重定向信息、网页跳转关系	决策树、贝叶斯网络、SVM、逻辑回归、DBSCAN	[12,21,23-29]
	恶意 PDF 检测	分类	PDF 文档内容、PDF 文件结构	随机森林、SVM、决策树、遗传编程	[48,150-153]
社会网络安全	社交网络帐号异常检测	分类/聚类/降维	用户的个人信息、用户行为、帐号创建时间、每天发布消息数量以及好友关系、消息文本中的 URL、消息内容等	随机森林、SVM、PCA、朴素贝叶斯、逻辑回归、K-Means、DBSCAN	[31,45,155-160]
	信用卡欺诈检测	分类	信用卡交易次数、行为等	神经网络、决策树、SVM、随机森林、隐马尔可夫模型	[161-164]
	取证分析	分类/降维	书写特征、笔迹内容特征、残差模式噪声特征、文件系统活动、网络流量日志	逻辑回归、ANN、SVM、SOM	[165-170]
	网络舆情	分类/聚类	用户生成内容的时间、空间、文本等特征	朴素贝叶斯、K 均值、SVM、隐马尔可夫模型	[174-179]

图 8: 机器学习在应用软件安全中的应用

表 8 机器学习在社交网络异常帐号检测中的应用

检测方法	主要特征	机器学习算法	社交网络	相关文献
基于帐号行为的检测方法	好友关系、消息内容、URL	随机森林	Facebook、Twitter	[45]
	用户点击行为	SVM	人人网、LinkedIn	[155]
	用户时间、空间行为	PCA	Facebook	[156]
	注册时间、点击历史等	朴素贝叶斯	LinkedIn	[157]
基于消息内容的检测方法	用户个人信息特征以及微博文本内容	DBSCAN、K-Means	Twitter	[31]
	时间、消息来源、消息主题、直接用户交互等	SVM	Twitter	[158]
	URL 的特征、URL 跳转的特征、HTML 内容、HTTP 头部、JavaScript 事件、DNS	逻辑回归	Twitter	[159]
	推文内容、推文来源、垃圾邮件目的域名	K-Means	Twitter	[160]

图 9: 机器学习在社交网路异常账号监测中的应用

7 研究展望与挑战

7.1 已有技术解决方案在模型的泛化能力、检测准确度、实时性上的能力提升

7.2 机器学习自身难点

* 如何能够使机器学习的解决方案具备可解释性、鲁棒性

* 如何能够提高实时监测效率以及解决基于机器学习技术的攻击

* 机器学习自身安全（用户隐私、输入样本攻击）

8 个人阅读小结

1 机器学习在网络安全方面能解决问题的能力有限，确定研究方向时需要深入调研，综合考虑研究意义、机器学习对具体问题的解决能力、数据获取等方面。

2 初步对恶意加密流量识别比较感兴趣，后续继续阅读相关文献进行总结